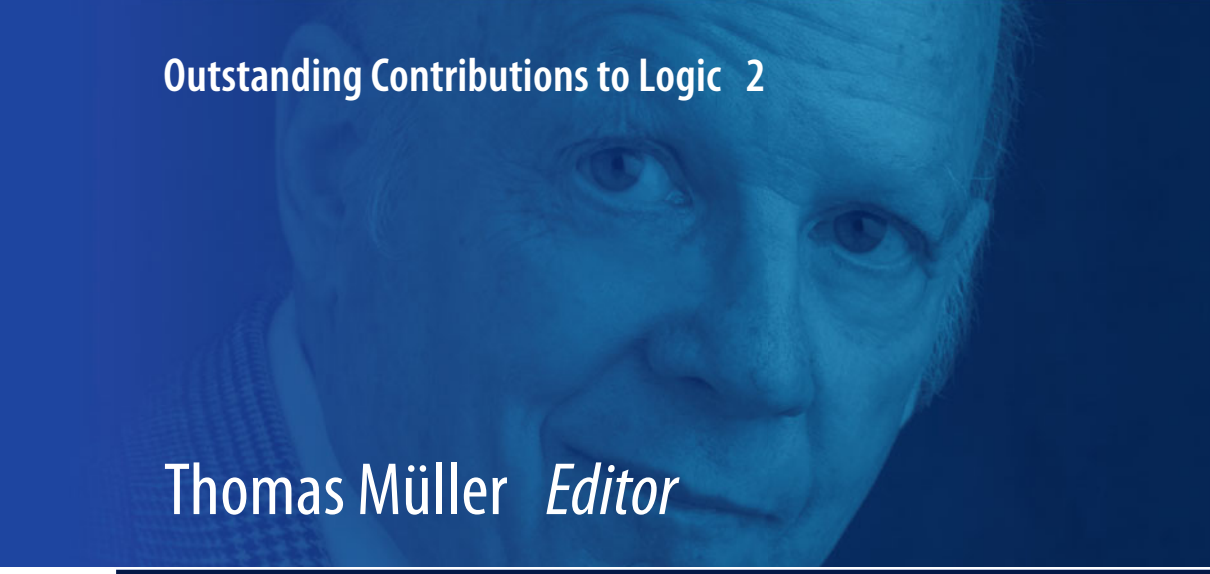
Outstanding Contributions to Logic 2



Thomas Müller *Editor*

Nuel Belnap on Indeterminism and Free Action

Springer Open

Outstanding Contributions to Logic

Volume 2

For further volumes:
http://www.springer.com/series/10033

Thomas Müller
Editor

Nuel Belnap on Indeterminism and Free Action

Springer

Editor
Thomas Müller
Department of Philosophy
Universiteit Utrecht
Utrecht
The Netherlands

ISSN 2211-2758 ISSN 2211-2766 (electronic)
ISBN 978-3-319-01753-2 ISBN 978-3-319-01754-9 (eBook)
DOI 10.1007/978-3-319-01754-9
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013957706

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Introduction: The Many Branches of Belnap's Logic

Thomas Müller

Abstract In this introduction to the *Outstanding contributions to logic* volume devoted to Nuel Belnap's work on indeterminism and free action, we provide a brief overview of some of the formal frameworks and methods involved in Belnap's work on these topics: theories of branching histories, specifically "branching time" and "branching space-times", the *stit* ("seeing to it that") logic of agency, and case-intensional first order logic. We also draw some connections to the contributions included in this volume. Abstracts of these contributions are included as an appendix.

Nuel Belnap's work in logic and in philosophy spans a period of over half a century. During this time, he has followed a number of different research lines, most of them over a period of many years or decades, and often in close collaboration with other researchers:[1] relevance logic, a long term project starting from a collaboration with Alan Anderson dating back to the late 1950s and continued with Robert Meyer and Michael Dunn into the 1990s; the logic of questions, developed with Thomas Steel in the 1960s and 1970s; display logic in the 1980s and 1990s; the revision theory of truth, with Anil Gupta, in the 1990s; and a long-term, continuing interest in indeterminism and free action. This book is devoted to Belnap's work on the latter two topics. In this introduction, we provide a brief overview of some of the formal frameworks and methods involved in that work, and we draw some connections to the contributions included in this volume. Abstracts of these contributions are presented in Appendix A.

[1] The biographical interview with Nuel Belnap provides some additional information on these research lines and on some of the collaborations.

T. Müller (✉)
Department of Philosophy, Utrecht University, Janskerkhof 13a,
3512 BL Utrecht, The Netherlands
e-mail: Thomas.Mueller@phil.uu.nl

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_1,

1 About this Book

This book contains essays devoted to Nuel Belnap's work on indeterminism and free action. Philosophically, these topics can seem far apart; they belong to different sub-disciplines, viz., metaphysics and action theory. This separation is visible in philosophical logic as well: The philosophical topic of indeterminism, or of the open future, has triggered research in modal, temporal and many-valued logic; the philosophical topic of agency, on the other hand, has led to research on logics of causation and action. In Belnap's logical work, however, indeterminism and free agency are intimately linked, testifying to their philosophical interconnectedness.

Starting in the 1980s, Belnap developed theories of indeterminism in terms of branching histories, most notably "branching time" and his own "branching space-times". At the same time, he pursued the project of a logic of (multi-)agency, under the heading of *stit*, or "seeing to it that". These two developments are linked both formally and genetically. The *stit* logic of agency is built upon a theory of branching histories—initially, on the Prior-Thomason theory of so-called branching time. The spatio-temporal refinement of that theory, branching space-times, in turn incorporates insights from the formal modeling of agency. Both research lines arise in one unified context and exert strong influences on each other.[2]

This volume appears in the series *Outstanding contributions to logic* and celebrates Nuel Belnap's work on the topics of indeterminism and free action. It consists of a selection of original research papers developing philosophical and technical issues connected with Belnap's work in these areas. Some contributions take the form of critical discussions of his published work, some develop points made in his publications in new directions, and some provide additional insights on the topics of indeterminism and free action. Nearly all of the papers were presented at an international workshop with Nuel Belnap in Utrecht, The Netherlands, in June 2012, which provided a forum for commentary and discussion. We hope that this volume will further the use of formal methods in clarifying one of the central problems of philosophy: that of our free human agency and its place in our indeterministic world.

2 State of the Art: BT, BST, *stit*, and CIFOL

In order to provide some background, we first give a brief and admittedly biased sketch of the current state of development of three formal frameworks that figure prominently in Nuel Belnap's work on indeterminism and free action: the simple branching histories framework known as "branching time" (BT; Sect. 2.1), its relativistic spatio-temporal extension, branching space-times (BST; Sect. 2.2), and the "seeing to it that" (*stit*) logic of agency (Sect. 2.3). In Sect. 2.4, we additionally introduce case-intensional first order logic (CIFOL), a general intensional logic offering

[2] Readers interested in the concrete history can find some details in Appendix B at the end of this introduction.

resources for a first-order extension of the mentioned frameworks. CIFOL is a recent research focus of Belnap's, as reflected in his own contribution to this volume.

2.1 Branching Time (BT)

It is a perennial question of philosophy whether the future is open, what that question means, and what a positive or a negative answer to it would signify for us. The question has arisen in many different contexts—in science, metaphysics, theology, philosophy of language, philosophy of science, and in logic. The logical issue is not so much to provide an answer to the question about the openness of the future, nor primarily about its meaning and significance, but about the proper formal modeling of an open future: How can time and possibility be represented in a unified way? Thus clarified, the logical question of the open future is first and foremost one of providing a useful formal framework within which the philosophical issue of multiple future possibilities can be discussed.

In the light of twentieth century developments in modal and temporal logic, that logical question is one about a specific kind of possibility arising out of the interaction of time and modality. That kind of possibility may be called historical possibility or, in the terminology that Belnap favors, *real possibility*. A formal framework for real possibility must combine in a unified way a representation of past and future, as in temporal logic (tense logic), and of possibility and necessity, as in modal logic. That combination is not just interesting from a logical point of view—it is also of broader philosophical significance. To mention one salient example, the interaction of time and modality reflects the loss of possibilities over time that seems central to our commonsense idea of agency.

Working on his project of tense logic, Arthur Prior devoted his first book-length study to the topic of *Time and modality* (Prior 1957). A leading idea was that temporal possibility should somehow be grounded in truth at some future time, where time is depicted as linearly ordered. In 1958, Saul Kripke suggested a different formal framework, making use of partial orderings of moments. His exchange with Prior is documented in Ploug and Øhrstrøm (2012). The leading idea, which Prior took up and developed in his later book, *Past, present and future* (Prior 1967), was that the openness of the future should be modeled via a tree of histories (or chronicles) branching into the future. In terms of the partial ordering of moments m, a history h is a maximal chain (a maximal linearly ordered subset) in the ordering—graphically, one complete branch of the tree, representing a complete possible course of events from the beginning till the end of time (see Fig. 1). If the future is not open, all possible moments are linearly ordered, and there is just one history; if the future is open, however, the possible moments form a partial ordering in which there are multiple histories. In that case, we can say that there are incompatible possibilities for the same clock time (or for the same instant, i), which lie on different histories. Tomorrow, as Aristotle's famous example goes, there could be a sea-battle, or there could be none, and nothing yet decides between these two future possibilities.

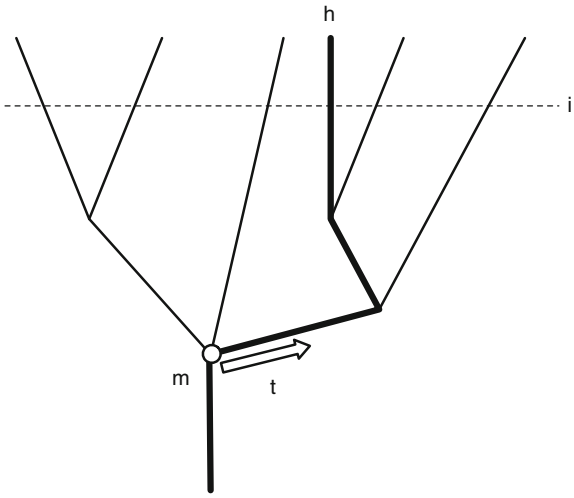


Fig. 1 BT structure. *m* is a moment, and *h*, indicated by the bold line, is one of the structure's six histories. *t* is one of the three distinct transitions originating from *m*. The dashed line, *i*, indicates an instant, a set of moments at the same clock time in different histories. The future direction is up

This approach to modeling indeterminism has come to be known as "branching time" (BT), even though Belnap rejects the label on the ground that time itself "never […] ever "branches" " (Belnap et al. 2001, 29). It is indeed better to speak of branching histories, since it is the histories that branch off from each other at moments. The label "branching time" is, however, well entrenched in the literature. Prior's own development of BT was not fully satisfactory, but Thomason (1970) clarified its formal aspects in a useful way, adding even more detail in his influential handbook article on "Combinations of tense on modality" (Thomason 1984). The most versatile semantic framework for BT, which goes under Prior's heading of "Ockhamism" due to an association with an idea of Ockham's, posits a formal language with temporal operators ("it was the case that", "it will be the case that") and a sentential operator representing real possibility. The semantics of these operators is given via BT structures. The distinctive mark of Ockhamism is that it takes the truth of a sentence about the future to rely on (minimally) two parameters of truth, a temporal moment and a history containing that moment.[3]

The Ockhamist set-up can be developed in various ways, and Belnap has explored many of these in detail. We mention a number of salient issues and give a few references. The contributions to this volume by Brown and by Garson both develop further foundational issues of BT: Brown relates, *inter alia*, to the notion of a possible world that can ground alethic modalities; Garson connects the issue of the open future to the question of what is expressed by the rules of propositional logic and argues for a natural open future semantics that allows one to rebut logical arguments for fatalism.

[3] See the article by Peter Øhrstrøm in this volume for discussion and historical details, including a hypothetical response to Belnap's employment of the BT framework on Ockham's behalf.

- BT, and also the earlier system of tense logic, brings out the dependence of the truth of a sentence on a suite of *parameters of truth*. For a simple temporal language, the truth of a sentence such as "Socrates is sitting" depends only on the moment with respect to which the sentence is evaluated. In Ockhamism, a sentence expressing a future contingent, such as "Socrates will be sitting at noon", or indeed "There will be a sea-battle tomorrow", is true or false relative to (minimally) two parameters, a moment and a history. Such a sentence, evaluated at some moment, can be true relative to one history and false relative to another. Relativity of truth to parameters of truth is nothing new or uncommon—it occurs already in standard predicate logic (see the next point). But in Ockhamism, one is forced to consider the issue of parameters of truth explicitly and in detail. A recognition of that issue has paved the way for a general semantics for indexical expressions (also known as "two-dimensional semantics"), as in the work of Kamp (1971) and Kaplan (1989). Belnap has pointed out the far-reaching analogy between "modal" parameters (such as m and h in Ockhamism) and an ordinary assignment of values to variables in predicate logic (Belnap et al. 2001, Chap. 6B).
- Working with this analogy, there is the interesting issue of how, given a context of utterance (or more generally, a context of use), parameters of truth receive a value that can be used in order to assign truth values to sentences. Belnap et al. (2001, 148f.) discuss this under the heading of "stand-alone sentences"; MacFarlane (2003) speaks of the issue of "postsemantics". In the case of the variables in predicate logic, it seems quite clear that unless some value has been assigned to x, the sentence "x is blue" cannot have a truth value. If all we have is "x is blue", the best we can do is prefix a quantifier, e.g., to read such a sentence as universally quantified, "for all x, x is blue". In Ockhamism, a sentence minimally needs *two* parameters, a moment m and a history h containing m, in order to be given a truth value. How do these parameters receive a value? It seems plausible to assume that a context of utterance provides a moment of the context that can be used as an initial value for m. But what about h? We make assertions about the future, but in an indeterministic partial ordering, there will normally be many different histories containing the moment m; there is no unique "history of the context" to give the parameter h its needed value. This problem is known as the *assertion problem*. It does not seem that quantification provides a way out. Universal quantification in the semantics (an option known under Prior's term "Peirceanism") seems out of the question—when we say that it is going to rain tomorrow, we are not saying that it will necessarily rain tomorrow, i.e., that it will rain on *all* histories containing the present moment. When it turns out to be raining on the next day, we are satisfied and say that our assertion was true when made; we do not retract it when we are informed that sunshine was really possible (even though it didn't manifest). These considerations also speak against the option of quantifying over the history parameter outside of the recursive semantics ("postsemantically"), as in supervaluationism (Thomason 1970). Similarly, one argues against existential quantifications over the relevant histories on the ground that when we say that it will be raining, we are claiming more than that it is possible that it will be raining. (On that option, we would have to say that both "It will be raining tomorrow" and

"It will not be raining tomorrow" are true, which sounds contradictory.) So, how do we understand assertions about the future?

- Together with Mitchell Green, Belnap has given a forceful statement of the problem of the uninitialized history parameter in Ockhamism and argued that it needs to be met head-on. According to Belnap and Green (1994), it will not do to posit a representation of "the real future" as a metaphorical "thin red line" singling out one future possibility above all others. They argue that marking any history as special, or real, would mean to deny indeterminism. (So, do not mistake the boldface line marking history h in Fig. 1 as indicating any special status for that history.) A number of solutions to the assertion problem have been discussed in the literature. Belnap (2002a) has argued that we can employ a second temporal reference point in order to assess future contingents later on. Before they can be assessed, a speech-act theoretic analysis can show their normative consequences. Here Belnap relies on the theory of word-giving developed by Thomson (1990). The current state of the debate appears to be that a "thin red line" theory is a consistent option from a logical point of view, but disagreements over the metaphysical pros and cons remain. In this volume, Øhrstrøm's contribution gives a well-argued update on this discussion and its historical predecessors, while Green holds that a "thin red line" comes at an unnecessarily high metaphysical cost and argues that a speech-act theoretic understanding of our assertion practices is also possible.[4]
- Belnap has pointed out the importance of the notion of immediate, "local" possibilities for the proper understanding of the interrelation of time and modality. He finds in von Wright (1963) the notion of a "transition", which is formally analyzed to be an initial paired with an immediately following outcome (Belnap 1999). Given an initial moment in a branching tree of histories, such a transition singles out a bundle of histories all of which remain undivided for at least some stretch of time. (Technically, one uses the fact that the relation of being undivided at a moment m is an equivalence relation on the set of histories containing m.) In Fig. 1, "t" indicates one of the three transitions (bundles of histories) branching off at m. Histories can then be viewed as maximal consistent sets of transitions. This allows for a generalization of the Ockhamist framework: instead of taking the parameters of truth to involve a moment/history pair m/h, one can employ a moment/set-of-transitions pair, m/T. Since sets of transitions are more fine-grained than whole histories, they can be used to represent the relative contingency of statements about the future, extending MacFarlane's notion of a "context of assessment". See Müller (2013a) and Rumberg and Müller (2013) for some preliminary results on this approach.
- Unlike theories developed in computer science, BT does not come with the assumption that the partial ordering of moments be discrete. While this assumption is certainly appropriate for many applications, it would trivialize some issues that can be usefully discussed in BT. An important case in point is the topology of branch-

[4] For a recent defense of the "thin red line", see also Malpass and Wawer (2012). MacFarlane (2014), on the other hand, defends assessment-relative truth of future contingents via his postsemantic approach.

ing. Assume that there are two continuous histories branching at some moment: Is there a last moment at which these two histories are undivided (a "choice point"), with the alternatives starting immediately afterwards, or should there be two alternative first moments of difference between these histories, so that there is no last moment of undividedness? McCall (1990) has illustrated these topologically different options. In BT, while assuming the existence of choice points is sometimes technically convenient, it makes no important difference which way one decides, as there is an immediate transformation of one representation into the other. This situation changes remarkably once we move to branching space-times.

2.2 Branching Space-Times (BST)

Branching space-times (BST) is a natural extension of the branching time framework, retaining the idea of branching histories for representing indeterminism but adding a formal representation of space in a way that is compatible with relativity theory. Belnap (2012) motivates the development of his theory of BST (Belnap 1992), which we will call BST1992, in the following way: Start with Newtonian space-time, which has an absolute (non-relativistic) time ordering and is deterministic. One way to modify this theory is to allow for indeterminism while sticking to absolute time. This corresponds to BT, in which the moments are momentary super-events stretching all of space. Another way to modify Newtonian space-time is to move to relativity theory, in which the notion of absolute simultaneity is abandoned in favor of a notion of simultaneity that is relative to a frame of reference. Combining the two moves, one arrives at a theory that is indeterministic (like BT) *and* relativistic (like relativistic space-time). Histories are no longer linear chains of moments ordered by absolute time, but whole space-times. Correspondingly, branching occurs not at space-spanning moments, but locally, at single possible point events.

The main technical innovation that makes BST1992 work, is the definition of a history not as a linear chain, but as an upward directed set in a partial ordering: a history contains, for any two of its members, a possible point event such that the two given members are in its causal past. In this way, one can work out branching history structures whose individual histories are all, e.g., Minkowski space-times (Müller 2002; Wroński and Placek 2009; Placek and Wroński 2009).

Historically, the origins of BST are somewhat different from the pedagogical set-up chosen by Belnap (2012). The story is interesting because it testifies to the mentioned intimate interrelation between indeterminism and agency. In the *stit* ("seeing to it that") approach to the logic of agency, the truth conditions for "agent α sees to it that ϕ" invoke the Ockhamist (BT-)parameters m/h. Briefly, for such a sentence to be true relative to moment m and history h, the agent α has to guarantee the outcome ϕ, which must not otherwise be guaranteed at m, by a choice determined by h. (See Sect. 2.3 for details.) Clearly, a single agent framework can only be the start; in fact, *stit* catered for multiple agents from the beginning. Now, intuitively speaking, what agents α and β choose to do at any given moment, should be independent: everybody

makes their own choices. It is reasonable to assume that this independence is guaranteed if agents α and β make their choices at different places at the same time, which implies that these choices are causally independent. But in a BT-based framework, there is no direct way to model that spatial separation. The solution in BT-based *stit* is, therefore, to introduce an additional axiom demanding independence. (See the contribution to this volume by Marek Sergot for a critical discussion of that axiom.) It would be much nicer if the agents' locations were modeled internally to the formalism, and the independence of their choices could accordingly be attributed to their spatial separation. An adequate notion of space-like relatedness is available in relativity theory, starting with Einstein's special theory of 1905. BST allows for a clear definition of space-like relatedness based on the underlying partial ordering: Two possible point events e and f are space-like related iff they are not order-related, but have a common upper bound (which guarantees that there is a history—a possible complete course of events—to which they both belong). Once agents are incorporated in BST (idealized as pointlike to begin with; see Belnap (2005a, 2011)), their choices can be taken to be events on their world-lines, and causal independence of such events can be directly expressed via space-like relatedness.

One can thus see two relevant motivations for constructing BST: as a relativistic extension of BT, and as a natural background theory for multi-agent logics of agency. The resulting quest for a reasonable framework for BST was mostly one of finding a useful definition of a history, and of fixing a number of topological issues, which become crucial in this development. Based on considerations of the causal attribution of indeterministic happenings, Belnap (1992) opts for the so-called "prior choice postulate", which guarantees the existence of choice-points: For anything that happens in one history rather than in another, there is some possible point event in the past that is shared among the two histories in question, and which is maximal in their intersection. This postulate, together with continuity requirements, fixes to a large extent the topological structure of BST 1992.[5] Figure 2 depicts a BST1992 structure with four histories, each of which is isomorphic to Minkowski space-time.

As in the case of BT, we mention a number of important issues and developments in BST to which Belnap has contributed. It will be obvious that he has been of central importance to all of them.

- To begin with topology, the original paper (Belnap 1992) mentions an approach to defining a topology for BST1992 that brings together different ideas from the theory of partial orders and from relativity theory. This topology, which Belnap attributes to Paul Bartha, has been researched in recent work by Placek and Belnap (2012); see also the contribution to this volume by Tomasz Placek. Naturally, the topological structure of a model of BST1992, which incorporates many incompatible histories, turns out to be non-Hausdorff (containing inseparable points);

[5] There are related frameworks for incorporating space-time and indeterminism. An early description occurs in Penrose (1979); see also the references in Müller (2011a). McCall (1994) gives an informal description of branching models incorporating a spatial aspect; Strobach (2007) discusses alternatives in space-time from the point of view of defining logical operators. See also the remarks on topology and on general relativity's challenges for BST in the main text below.

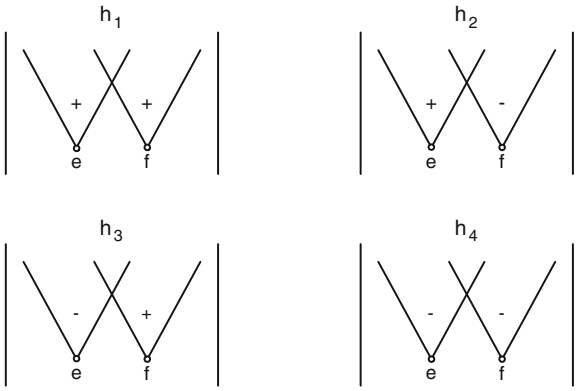


Fig. 2 Schematic diagram of a BST structure. e and f are choice points with two outcomes each, schematically denoted "+" and "−". The four histories $h_1, \ldots, h_4$ overlap outside the W-shaped forward lightcones of the choice points and in those parts of the light cones above e and f for which the labels coincide. The choice points e and f belong to all four histories. As in BT diagrams, the future direction is up

a single history is however typically Hausdorff. This makes good sense given indeterminism: If different possibilities exist for the same position in space-time, the corresponding possible point events may be topologically inseparable in the full indeterministic model.

- These topological observations are linked to the question whether BST can be viewed as a space-time theory. Earman (2008) has asked a pointed question about the tenability of BST as a space-time theory, sharply criticizing McCall's (1994) version of BST and raising doubts about Belnap's framework. His main challenge is to clarify the meaning of non-Hausdorffness that occurs in BST, since in space-time theories this is a highly unwelcome feature. Some recent literature, including Tomasz Placek's contribution to this volume, has clarified the situation considerably, highlighting the difference between branching *within* a space-time, which indeed has unwelcome effects well known to general relativists, and the BST notion of branching histories, in which the histories are individually non-branching space-times. The connection between BST and general relativity is only beginning to be made, and a revision of Belnap's prior choice principle may be in order to move the two theories closer to each other. (Technically, the issue is that the prior choice principle typically leads to a violation of local Euclidicity, which is, however, presupposed even for generalized, non-Hausdorff manifolds.) Apart from Placek's contribution, see also Sect. 6 of the contribution by Pleitz and Strobach, and Müller (2013b).
- Another area of physics that may be able to interact fruitfully with the BST framework is quantum mechanics. As BST incorporates both indeterminism and space-like separation, it seems to be especially well suited for clarifying the issue of space-like correlations in multi-particle quantum systems, pointed out in a famous

paper by Einstein et al. (1935). Following some pertinent remarks already in the initial paper by Belnap (1992), there have been some applications of the BST framework in this area, starting with Szabó and Belnap (1996), who target the three-particle, non-probabilistic case of Greenberger-Horne-Zeilinger (GHZ) states. These modeling efforts are connected with research on various types of common cause principles—see Hofer-Szabó et al. (2013). Placek (2010) brings into focus the epistemic nature of observed surface correlations vis-à-vis an underlying branching structure. For some remarks on a link between BT- or BST-like branching history structures and the quantum-mechanical formalism of so-called consistent histories, see Müller (2007).

- Even independently of applications to quantum physics, which may help to show the empirical relevance of the BST framework, there is the structural issue of how space-like correlations can be modeled in BST. Corresponding formal investigations were begun by Belnap (2002b) and continued in Belnap (2003), where the equivalence of four different definitions of modal correlations in BST1992 is proved. The basic observation is that it is possible to construct BST models in which the local possibilities at space-like separated choice points do not always combine to form global possibilities, i.e., histories. The simplest case corresponds exactly to the phenomenon pointed out in Einstein et al. (1935): Given a certain two-particle system, once its components are separated spatially, certain measurement outcomes for the components are perfectly correlated, meaning that it is impossible that a specific outcome on one side is paired with a specific outcome on the other side, even though no single outcome on either side is excluded. For an illustration, think of Fig. 2 with histories h_2 and h_3 missing: both choice points e and f then have two possible outcomes each, but the respective outcomes are perfectly modally correlated, admitting only joint outcomes $++$ and $--$. Müller et al. (2008) generalize Belnap's mentioned BST1992 results to incorporate cases of infinitely many correlated choice points. In this generalization, the notion of a transition, mentioned above in connection with BT, is crucial. For the use of sets of transitions to describe possibilities in BST, see also Müller (2010).
- The idea of (sets of) transitions as representatives of local possibilities is also the driving motor behind Belnap's highly original analysis of indeterministic causation (Belnap 2005b). In his approach, the relata of a singular causal statement "C caused E" are a transition (E) and a set of (basic) transitions (C). For a given effect E, described as "initial I followed by outcome O", it is possible in BST1992 to single out the relevant choice points (past causal loci) of that transition, and to describe the cause in terms of basic transitions in the past of O that lead from a choice point to one of its immediate local possibilities. These *causae causantes*, as Belnap calls them, are themselves basic causal constituents of our indeterministic world. Using various generalizations of the notion of an outcome, Belnap can prove that the *causae causantes* of an outcome constitute INUS conditions: insufficient but nonredundant parts of an unnecessary but sufficient condition for the occurrence of the outcome. (The notion of an INUS condition is famously from Mackie (1980).) Belnap's analysis provides a strong ontological reading of "causation as difference-

making" that appears to be well suited to modeling the kind of causation involved in human agency.

- Another useful employment of transitions in BST is in defining probabilities. Groundbreaking work was done by Weiner and Belnap (2006); a generalization to sets of transitions is given in Müller (2005), published earlier but written later. Paralleling earlier but independent work by Weiner, Müller (2005) shows that considerations of probability spaces lead to topological observations about BST as well. A general overview of probability theory in branching structures is given by Müller (2011b).
  The basic idea of defining probability spaces in BST is to start with local probability spaces, defined on the algebra of outcomes of a single choice point. The interesting issue is how to combine such local probability spaces to form larger ones. Here it becomes crucial to consider consistent sets of transitions and to exclude pseudo-events whose probabilities make no sense. Müller (2005) offers the notion of a "causal probability space" in an analysis of which probability spaces can be sensibly defined in BST.
- The formal structure of BST is rich and multiply interpretable. This volume's contributions by Strobach and by Pleitz and Strobach testify to the versatility of the BST framework by providing a biological interpretation. Further developments are to be expected in the interaction between BST and the *stit* logic of agency.

2.3 Seeing to it That (stit)

We already remarked on some aspects of the *stit* framework that show its relation to branching histories frameworks and specifically to the development of BST. *Stit* logic is based on BT structures and uses the Ockhamist parameters of truth m and h, as introduced in Sect. 2.1. In order to represent agents and agency, BT structures are augmented via a set A of agents and an agent-indexed family of choices at moments, $Choice_m^{\alpha}$, which represent each agent's alternatives at each moment as a partition of the histories passing through that moment. These choices must be compatible with the local granularity of branching (the transition structure) resulting from the underlying BT structure: Agents cannot choose between histories before they divide in the structure ("no choice before its time").

The semantic clause for "α sees to it that ϕ", evaluated at m/h, has two parts: a positive condition, demanding that α must settle the truth of ϕ through her choice, and a negative condition, which excludes as agentive those ϕ whose truth is settled anyway. More specifically, there are two different developments of *stit*, which Belnap et al. (2001) call the "deliberative *stit*" (*dstit*) and the "achievement *stit*" (*astit*), respectively. The difference between them is one of perspective on what it is that the agent sees to. Both are built upon the mentioned BT structures with agents and their choices, but *astit* uses an additional resource, viz., a partitioning of the set of moments into so-called *instants* that mark the same clock time across different histories (Fig. 1 depicts one such instant, i). The book by Belnap, Perloff and

Xu, *Facing the future*, gives a comprehensive overview of a large number of developments in the *stit* framework, and is highly recommended as a general reference (Belnap et al. 2001). We leave many of the topics treated in that book, such as normative issues, strategies, word giving, or details of the resulting logics, to the side and describe just the basic frameworks, *astit* and *dstit*. Even though *astit* is historically earlier (Belnap and Perloff 1988), we start with a description of the simpler deliberative *stit*.—It is important to stress that while mentalistic notions such as deliberation are mentioned in the *stit* literature, the basic frameworks do not go beyond modeling the indeterministic background structure of agency; agents' beliefs and epistemic states do not play a role in the formal theory. This keeps the framework simple and general. Specific applications, however, can call for extra resources. The contributions to this volume by Bartha, Van Benthem and Pacuit, Broersen, Sergot, Vanderveken and Xu all testify to this: each discusses specific and useful additional details. Bartha adds utilities and probabilities in order to ground normative notions; Broersen also treats normative issues, via an Andersonian "violation" constant; Sergot models normativity via flagged ("red" or "green") states. Van Benthem and Pacuit draw a comparison between *stit* and dynamic action logics, discussing a number of extensions that suggest themselves, including a dynamification of *stit*. Broersen adds probabilities for bringing about as well as subjective probabilities in order to anchor epistemic notions. Sergot employs a slightly different formal framework based on labeled transition structures, drops the independence of agents axiom, and emphasizes the importance of granularity of description for normative verdicts. Vanderveken adds a rich logic of propositional attitudes in order to analyze the logical form of proper intentional actions, extending the *stit* approach such as to give a logic of practical reason. Xu, in contrast, stays close to the austere *stit* framework; he explores in formal detail the extension of *stit* by group choices and group strategies. Further extensions of the basic *stit* approach are certainly possible.

Dstit was defined in Horty and Belnap (1995). The perspective is on securing a future happening due to a present choice, or deliberation. The positive clause for *dstit* demands that every history in the agent's current choice set satisfy the (future) outcome. The negative condition demands a corresponding witness for the violation of that outcome, which must belong to one of the other choices available to the agent. See Fig. 3 for an illustration; history h' fulfills the negative condition for α *dstit* : p, which is true at m/h. Large parts of *stit* can be developed without the negative condition, which greatly simplifies the logic; the corresponding *stit* operator is called *cstit*, after Chellas's employment of a similar idea in his analysis of imperatives (Chellas 1969). (A further simplification is possible if one assumes discrete time, see below.) Apart from the mentioned book by Belnap et al. (2001), see also Horty (2001).[6]

Belnap's historically first *stit* framework (Belnap and Perloff 1988) is based on the achievement *stit*, *astit*. As mentioned, instants are needed to define the *astit* operator. The perspective is different from that of *dstit*. For *astit*, a current result, or achievement, is attributed to an agent if there is a past witnessing moment at

[6] For an independent, similar development, see also von Kutschera (1986).

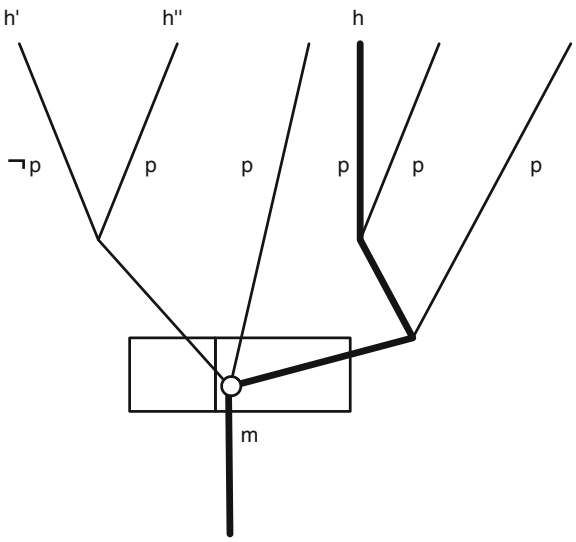


Fig. 3 Illustration of *dstit*. The BT structure is that of Fig. 1. At moment m, the agent α has two possible choices, marked by the two boxes. (For the other moments, the choices are not indicated to avoid visual clutter.) On history h, but not on history h' nor on history h'', α sees to it that p

which the agent's choice (as determined by the given history parameter) guaranteed the current result: All histories in that former moment's respective choice set must guarantee the result at the given instant (positive condition), and there must be another history passing the witnessing moment that does not lead to the result at that instant (negative condition). The logic of *astit* is interesting and quite complex; see Belnap et al. (2001, Chaps. 15–17).

In the recent literature, *dstit* plays the larger role. This may be due to its simpler logic, but perhaps also reflects the fact that the *dstit* operator is helpful for a formal representation of one of the main positions in the current free will debate, so-called *libertarianism*. According to the libertarian, free agency presupposes indeterminism. An influential argument given in favor of this assumption, Van Inwagen's so-called consequence argument (Van Inwagen 1983), proclaims that an action cannot be properly attributed to an agent if its outcome is already settled by events outside of the agent's control, and that would invariably be so under determinism. See the contribution to this volume by Robert Kane for a defense of libertarianism that points out the virtues of *stit* as a logical foundation for an intelligible account of free will based on indeterminism.

A helpful result in the logic of *dstit* is that refraining can itself be seen to be agentive, and that refraining from refraining amounts to doing. This result should be useful for clarifying the status of the assumption of alternative possibilities that is widely discussed in the free will debate and on whose merits or demerits much ink has been spilt. In *dstit*, if α sees to it that ϕ relative to the (Ockhamist) parameters m/h, this implies that there is a history h' containing m on which ϕ turns out false—that is the gist of the negative condition. As this history must lie in one of the agent's choices other than the one corresponding to h (this is due to the choices forming a

partition of the histories through m, and the positive condition demanding the truth of ϕ on all histories choice-equivalent to h), on that alternative, the agent *sees to it that she is not seeing to it that* ϕ. After all, making the choice corresponding to h', α is not seeing to it that ϕ (since on h', ϕ turns out false), but there is a history, viz., h, on which she *does* see to it that ϕ. So, "α sees to it that she is not seeing to it that ϕ" is the *stit* analysis of refraining. You can check that in Fig. 3, at m on history h', the agent refrains from future p (Fp) in exactly that sense. It is clear that the alternative of refraining from ϕ does *not* have to amount, on that analysis, to the agent's possibly seeing to it that non-ϕ, even though this is often taken to be implied by the assumption of alternative possibilities. (In Fig. 3, there is no history on which α sees to it that $\neg Fp$.) In our view, *stit* provides some desperately needed clarity here.[7] There is certainly much work to be done to integrate formal work on *stit* into the free will debate. See Kane's contribution to this volume for a discussion of a number of additional steps towards a fuller account of indeterminism-based free will.

Outside of philosophy proper, *stit* has had, and continues to have, a significant influence on the modeling of agency in computer science and artificial intelligence. Many of the contributions to this volume testify to *stit*'s usefulness in this area. Usually, such applications of the framework give up the initial generality of BT models (which allow for continuous structures) in favor of discrete orderings. While this means a limitation of scope, it makes the framework much more tractable and thus, useful from an engineering point of view. The availability of a "next time" operator suggests that one can read a *dstit*- or *cstit*-like operator as "an agent secures an outcome at all choice-equivalent possible next moments", thus doing away with a layer of complexity introduced by the usual handling of the future tense (which quantifies over all future moments on a given history, including moments that are far removed), and by the need for considering whole histories. In this volume, the contribution by Broersen explicitly builds upon discrete structures, and the transition system framework employed by Sergot is also typically discrete. Van Benthem and Pacuit in their contribution leave the basic *stit* framework unconstrained, but go on to employ the discrete view of stepwise execution that is basic for dynamic logic. With various refinements and extensions of *stit*, it seems fair to say that the computer science community currently provides the richest environment for the development of that framework. Interaction with the philosophical community can certainly prove to be beneficial for both sides, and we hope that this volume can be helpful in that respect.

It should also be stressed that while the *stit* framework has found many applications, it is by no means the only approach to the formal modeling of agency on the market. Two of the contributions to this volume draw explicit connections to other important existing frameworks. Sergot remains close to the *stit* framework,

[7] We refrain from entering a lengthier discussion of the free will debate, which has turned into a maze of arguments, counterarguments and, not too infrequently, confusion and talking past each other. From among the recent original and helpful contributions to the debate, we mention Helen Steward's plea for the libertarian position of "agency incompatibilism" (Steward 2012). She indicates connections to the *stit* framework as well (Steward 2012, 31).

but draws upon the formalism of Pörn (1977). Van Benthem and Pacuit provide a detailed comparison between the *stit* approach and the paradigm of dynamic logic that was developed in the formal study of computer programs. These comparisons are highly valuable, since they promise to help to bring related research lines operating in relative isolation closer together.

Since *stit* is so rich and multi-faceted, we do not attempt here to give an overview of recent developments akin to what we did for BT and BST above. We refer again to the book, *Facing the future* (Belnap et al. 2001), for the groundwork and a clear presentation of logical issues. For contemporary developments, we refer to the contributions in this volume.

2.4 Case-Intensional First Order Logic (CIFOL)

The development of all the three mentioned frameworks—BT, BST, and *stit*—is based on semantical considerations, though not necessarily with a view toward providing a semantics for an extant formal language. The common, semantically driven idea is to define structures that represent aspects of reality such that the truth or falsity of sentences can be discussed against the background of such a structure.

When one looks at applications that do relate to a formal language (such as the language of tense logic for BT), it turns out that most often, models based on the respective structures are thought of as providing the semantics for a *propositional* language, which does not use variables or quantifiers. This is probably mostly due to the fact that many actual applications arise in a computer science context, and propositional logic is computationally much more tractable than predicate logic. There may also still be a lingering worry about the tenability of quantified modal logic, even though Quine's influence is waning.[8] But perhaps the main reason for the fact that there is not a lot of BT-based predicate logic (let alone a predicate logic based on BST, or on *stit*) is that it is hard to get it right. For philosophical purposes, it is, however, clear that we need to take individual things seriously—after all, we, the biological creatures populating this planet, are agents, and it is not always fruitful to reduce the representation of one of us to a mere label on a modal operator. Thus, one of the areas in which much further logical development is to be expected, is an adequate representation of things, their properties and their possibilities in an indeterministic setting.

Quantified modal logic (QML) has long been an area of interaction between logic and metaphysics, not always to the benefit of logic. One of the most interesting recent developments in Belnap's work on indeterminism and free action is connected with the attempt of developing a metaphysically neutral quantified modal logic, which would be driven by applicability rather than by underlying metaphysical assumptions. Consider the handling of variables. Most systems of QML assume that modal logic

[8] For Quine's arguments against quantifying into modal contexts, see, e.g., Quine (1980, Chap. VIII). See Fine (2005, Chaps. 2 and 3) for extensive analysis and critique.

should be built on a modal parameter of truth that specifies a "possible world". Also, typically, a variable functions as a rigid designator: Each possible world comes with its domain of individuals (the world's "inhabitants"), and a variable designates the same individual in any world. Alternatively, a counterpart relation between the domains and a corresponding handling of variables is discussed.[9] Both moves make a certain view of the metaphysical status of individuals part of the quantificational machinery of QML. Accordingly, such logics cannot be used to represent dissenting metaphysical views about individuals. It would seem, however, that one of the main virtues of using a logical formalism is that it provides an arena in which different views can be formulated and arguments in their favor or against them can be checked. What good is a quantified modal logic if it does not allow one to discuss different theories and arguments about the metaphysical status of individuals?

Belnap argues for a broader, more general approach to QML that is based on a neglected but useful framework for quantified modal logic developed in the interest of clarifying arguments arising in the empirical sciences. Aldo Bressan (1972) developed his case-intensional approach to QML out of his interest in the role of modality in physics. His system is higher order and includes a logicist construal of the mathematics necessary for applications in physics; this makes it highly complex and may have stood in the way of its wider recognition or application. Belnap (2006) provides a useful overview of the general system. For many purposes it is, however, sufficient to look at the first-order fragment of Bressan's system, and to develop that as a stand-alone logical framework. One guiding idea is generality: instead of developing a modal logic based on the idea of a "possible world", or a temporal logic that is geared towards truth at a time, it is better to work with a general notion of a modal parameter of truth that we may call a *case*. This accords with ordinary English usage, and justifies S5 modalities built upon cases: necessary is what is true in any case; something is possible if there is at least one case in which it is true. Another guiding idea is uniformity. Rather than following standard systems of QML, which treat variables, individual constants and definite descriptions in widely different ways, one can use the most general idea of a term with an extension in each case, and an individual intension that represents the pattern of variation of the extension across cases. (Technically, the intension is the function from cases to extensions, and the extension at a case is the intension-function applied to that case. This recipe is followed uniformly for all parts of speech, generalizing Carnap's (1947) method of extension and intension.) Correspondingly, the most general option is used for predication as well: predication is not forced to be extensional, but is generally intensional, such that a one-place predicate for each case provides a function that maps intensions to truth values. This rich and uniform background provides for a simple yet powerful definition of sortal properties as allowing for the tracing of individuals from case to case. See Belnap and Müller (2013a) for a detailed description of the resulting framework of case-intensional first order logic (CIFOL). The framework has recently been extended to cases in a branching histories framework (Belnap and Müller 2013b). This application of CIFOL helps to disspell worries that have been raised against

[9] For an in-depth overview, see Kracht and Kutz (2007).

the idea of individuals in branching histories, such as famously in Lewis's argument against branching (Lewis 1986, 206ff): Using the resources of CIFOL, it is possible to model individuals and sortal properties successfully in a branching histories framework. Good news, surely, for those of us who believe that we are just that: individual agents facing an open future of possibilities.

In line with the development of BT, BST and *stit*, CIFOL is developed from a semantical point of view. The interface with a formal logical language is, however, much more pronounced in the case of CIFOL—the fact that we are considering a predicate logic necessitates close attention to the syntax as well. (For example, as the framework is required to remain first-order, while lambda-abstraction is unfettered, lambda-predicates may only occur in predicate position.) Naturally, it is to be expected that there can be fruitful discussions of CIFOL's proof theory and metatheory. Nuel Belnap, in his contribution to this volume, gives a highly interesting overview of a truth theory that can be developed within CIFOL+, a minimal extension of CIFOL. Given the framework's intensionality, it is possible to define terms representing the cases, and based on those, one can develop the theory of the mixed nector "that Φ is true at case *x*". You will, we hope, not go wrong in expecting further striking results about CIFOL and its connection to indeterminism and free action in the near, albeit open future.

Acknowledgments Research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement nr 263227, and from the Dutch Organization for Scientific Research, grant nr NWO VIDI 276-20-013. I would like to thank Nuel Belnap for continuing inspiration, and both Nuel Belnap and Antje Rumberg for helpful comments on a previous draft.

Appendix A: Abstracts of the Papers in this Volume

Paul Bartha (University of British Columbia): Decisions in Branching Time

This paper extends the deontic logic of Horty (*Agency and deontic logic*, 2001) in the direction of decision theory. Horty's deontic operator, the *dominance ought*, incorporates many concepts central to decision theory: acts, causal independence, utilities and dominance reasoning. The decision theory associated with dominance reasoning, however, is relatively weak. This paper suggests that deontic logic can usefully be viewed as *proto-decision theory*: it provides clear foundations and a logical framework for developing norms of decision of varying strength. Within Horty's framework, deontic operators stronger than the dominance ought are defined for decisions under ignorance, decisions under risk, and two-person zero-sum games.

Nuel Belnap (University of Pittsburgh): Internalizing case-relative truth in CIFOL+

CIFOL is defined in Belnap and Müller (*J Phil Logic* 2013) as the first-order fragment of Aldo Bressan's higher-order modal typed calculus MC^{ν} Bressan based his calculus on Carnap's "method of extension and intension": In CIFOL, truth is relative to "cases," where cases play the formal role of "worlds" (but with less pretension). CIFOL+ results by following Bressan in adding term-constants **t** for the true and **f** for the false, and a single predicate constant, P_0, which together with a couple of simple axioms enable the representation of "sentence Φ is true in case x" by means of a defined expression, $T(\Phi, x)$, where Φ is the sentence of CIFOL+ in question and where x ranges over a defined family of "elementary cases." (Whereas being a *case* is defined in the semantic metalanguage, *elementary cases* are squarely in the (first order) domain of CIFOL+.) A suitable suite of axioms guarantees that one can prove (in CIFOL+) that there is exactly one elementary case, x, such that x happens (i.e., such that $x = \mathbf{t}$), a fact that underlies the equivalence of $\Box(x = \mathbf{t} \rightarrow \Phi)$ and $\Diamond(x = \mathbf{t} \wedge \Phi)$. (Proofs are surprisingly intricate for first order modal logic). One can then go on to show that $T(\Phi, x)$ is well-behaved in terms of its relation to the connectives of CIFOL+, a result required for ensuring that $T(\Phi, x)$ is properly read as "Φ is true in elementary case x."

Jan Broersen (Utrecht University): A stit Logic Analysis of Morally Lucky and Legally Lucky Action Outcomes

Moral luck is the phenomenon that agents are not always held accountable for performance of a choice that under normal circumstances is likely to result in a state that is considered bad, but where due to some unexpected interaction the bad outcome does not obtain. We can also speak of moral misfortune in the mirror situation where an agent chooses the good thing but the outcome is bad. This paper studies formalizations of moral and legal luck (and moral and legal misfortune). The three ingredients essential to modelling luck of these two different kinds are (1) indeterminacy of action effects, (2) determination on the part of the acting agent, (3) the possibility of evaluation of acts and/or their outcomes relative to a normative moral or legal code. The first, indeterminacy of action, is modelled by extending stit logic by allowing choices to have a probabilistic effect. The second, deliberateness of action, is modelled by (a) endowing stit operators with the possibility to specify a lower bound on the change of success, and (b) by introducing the notion of attempt as a maximisation of the probability of success. The third, evaluation relative to a moral or legal code, is modelled using Andersons reduction of normative truth to logical truth. The conclusion will be that the problems embodied by the phenomenon of moral luck may be introduced by confusing it with legal luck. Formalizations of both forms are given.

Mark A. Brown (Syracuse University): Worlds Enough, and Time—Musings on Foundations

Belnap's work on stit theory employs an Ockhamist theory of branching time, in which the fundamental possibilia within models are commonly taken to be moments of time, connected into a tree-like branching structure. In the semantics for alethic

modal logic, necessity is characterized by quantification over relevant possible worlds within a model, yet Belnap refers to an entire model of branching time as our world, seemingly leaving no room for non-trivial quantification over worlds within a single model.

This paper explores the question how the notion of possible worlds should be understood in relation to an Ockhamist framework, in order to be able to combine an account of alethic modalities with an account of branching time and stit theory. The advantages and drawbacks of several alternative approaches are examined.

James W. Garson (University of Houston): Open Futures in the Foundations of Propositional Logic

This paper weaves together two themes in the work of Nuel Belnap. The earlier theme was to propose conditions (such as conservativity and uniqueness) under which logical rules determine the meanings of the connectives they regulate. The later theme was the employment of semantics for the open future in the foundations of logics of agency. This paper shows that on the reasonable criterion for fixing meaning of a connective by its rule governed deductive behavior, the natural deduction rules for classical propositional logic do not fix the interpretation embodied in the standard truth tables, but instead express an open future semantics related to Kripke's possible worlds semantics for intuitionistic logic, called natural semantics. The basis for this connection has already been published, but this paper reports new results on disjunction, and explores the relationships between natural semantics and supervaluations. A possible complaint against natural semantics is that its models may disobey the requirement that there be no branching in the past. It is shown, however, that the condition may be met by using a plausible reindividuation of temporal moments. The paper also explains how natural semantics may be used to locate what is wrong with fatalistic arguments that purport to close the door on a open future. The upshot is that the open future is not just essential to our idea of agency, it is already built right into the foundations of classical logic.

Mitchell Green (University of Virginia): On Saying What Will Be

In the face of ontic (as opposed to epistemic) openness of the future, must there be exactly one continuation of the present that is what *will* happen? This essay argues that an affirmative answer, known as the doctrine of the Thin Red Line, is likely coherent but ontologically profligate in contrast to an Open Future doctrine that does not privilege any one future over others that are ontologically possible. In support of this claim I show how thought and talk about "the future" can be shown intelligible from an Open Future perspective. In so doing I elaborate on the relation of speech act theory and the "scorekeeping model" of conversation, and argue as well that the Open Future perspective is neutral on the doctrine of modal realism.

Robert Kane (University of Texas at Austin): The Intelligibility Question For Free Will—Agency, Choice And Branching Time

In their important work, *Facing the Future* (Oxford 2001), Nuel Belnap and his collaborators, Michael Perloff and Ming Xu, say the following (p. 204): "We agree with Kane [1996] that … the question whether a kind of freedom that requires indeterminism can be made intelligible deserves … our most serious attention, and indeed we intend that this book contribute to what Kane calls 'the intelligibility question.' " I believe their book does contribute significantly to what I have called "the Intelligibility Question" for free will (which as I understand it is the question of how one might make intelligible a free will requiring indeterminism without reducing such a free will to either mere chance or to mystery and how one might reconcile such a free will with a modern scientific understanding of the cosmos and human beings). The theory of agency and choice in branching time that Belnap has pioneered and which is developed in detail in *Facing the Future* is just what is needed in my view as a logical foundation for an intelligible account of a free will requiring indeterminism, which is usually called libertarian free will. In the first two sections of this article, I explain why I think this to be the case. But the logical framework which Belnap et al. provide, though it is necessary for an intelligible account of an indeterminist or libertarian free will, is nonetheless not sufficient for such an account. In the remaining sections of the article (3–5), I then discuss what further conditions may be needed to fully address "the Intelligibility Question"for free will and I show how I have attempted to meet these further conditions in my own theory of free will, developed over the past four decades.

Peter Øhrstrøm (Aalborg University): What William of Ockham and Luis de Molina would have said to Nuel Belnap—A Discussion of some Arguments Against "The Thin Red Line"

According to A.N. Prior the use of temporal logic makes it possible to obtain a clear understanding of the consequences of accepting the ideas of indeterminism and free choice. Nuel Belnap is one of the most important writers who have contributed to the further exploration of these tense-logical ideas as seen in the tradition after Prior.

In some of his early papers Prior suggested the idea of the true future. Obviously, this idea corresponds to an important notion defended by classical writers such as William of Ockham and Luis de Molina.

Belnap and others have considered this traditional idea introducing the term, "the thin red line" (TRL), arguing that this idea is rather problematic. In this paper I argue that it is possible to respond to the challenges from Belnap and others in a reasonable manner. It is demonstrated that it is in fact possible to establish a consistent TRL theory. In fact, it turns out that there several such theories which may all be said to support the classical idea of a true future defended by Ockham and Molina.

Tomasz Placek (Jagiellonian University, Kraków): Branching for general relativists

The paper develops a theory of branching spatiotemporal histories that accommodates indeterminism and the insights of general relativity. A model of this theory can be viewed as a collection of overlapping histories, where histories are defined

as maximal consistent subsets of the model's base set. Subsequently, generalized (non-Hausdorff) manifolds are constructed on the theory's models, and the manifold topology is introduced. The set of histories in a model turns out to be identical with the set of maximal subsets of the model's base set with respect to being Hausdorff and downward closed (in the manifold topology). Further postulates ensure that the topology is connected, locally Euclidean, and satisfies the countable sub-cover condition.

Marek Sergot (Imperial College): Some examples formulated in a 'seeing to it that' logic—Illustrations, observations, problems

The paper presents a series of small examples and discusses how they might be formulated in a 'seeing to it that' logic. The aim is to identify some of the strengths and weaknesses of this approach to the treatment of action. The examples have a very simple temporal structure. An element of indeterminism is introduced by uncertainty in the environment and by the actions of other agents. The formalism chosen combines a logic of agency with a transition-based account of action: the semantical framework is a labelled transition system extended with a component that picks out the contribution of a particular agent in a given transition. Although this is not a species of the *stit* logics associated with Nuel Belnap and colleagues, it does have many features in common. Most of the points that arise apply equally to *stit* logics. They are, in summary: whether explicit names for actions can be avoided, the need for weaker forms of responsibility or 'bringing it about' than are captured by *stit* and similar logics, some common patterns in which one agent's actions constrain or determine the actions of another, and some comments on the effects that level of detail, or 'granularity', of a representation can have on the properties we wish to examine.

Niko Strobach (Westfälische Wilhelms-Universität Münster): In Retrospect. Can BST models be reinterpreted for what decisions, speciation events and ontogeny might have in common?

This paper addresses two interrelated topics: (1) a formal theory of biological ancestry (FTA); (2) ontological retrospect. The point of departure is a reinterpretation of Nuel Belnaps work on branching spacetime (BST) in terms of biological ancestry. Thus, Belnaps prior choice principle reappears as a principle of the genealogical unity of all life. While the modal dimension of BST gets lost under reinterpretation, a modal dimension is added again in the course of defining an indeterministic FTA where possible worlds are alternatives in terms of offspring. Indeterministic FTA allows to model important aspects of ontological retrospect. Not only is ontological retrospect a plausible account for the perspectival character of Thomason-style supervaluations, but it is shown to be a pervasive ontological feature of a world in development, since it is relevant for cases as diverse as speciation, the individual ontogeny of organisms and decisions of agents. One consequence of an indeterministic FTA which includes the idea of retrospect is that, contrary to what Kripke famously claims, species membership is not always an essential feature, but may depend on the way

the world develops. The paper is followed by a postscript by Martin Pleitz and Niko Strobach which provides a version of indeterministic FTA that is technically even closer to Belnap's BST than the one in this paper and which allows for a discussion of further philosophical details.

Martin Pleitz and Niko Strobach (Westfälische Wilhelms-Universität Münster): A Theory of Possible Ancestry in the Style of Nuel Belnap's Branching Space-Time

We present a general theory of possible ancestry that is a case of modal ersatzism because we do not take possibilities in terms of offspring as given, but construct them from objects of another kind. Our construction resembles Nuel Belnap's theory of branching space-time insofar we also carve all possibilities from a single pre-existing structure. According to the basic theory of possible ancestry, there is a discrete partially ordered set called a structure of possibilia, any subset of which is called admissible iff it is downward closed under the ordering relation. A structure of possibilia is meant to model possible living beings standing in the relation of possible ancestry, and the admissible sets are meant to model possible scenarios. Thus the Kripkean intuition of the necessity of (ancestral) origin is incorporated at the very core of our theory. In order to obtain a more general formulation of our theory which allows numerous specifications that might be useful in concrete biological modeling, we single out two places in our framework where further requirements can be implemented: Global requirements will put further constraints on the ordering relation; local requirements will put further constraints on admissibility. To make our theory applicable in an indeterminist world, we use admissible sets to construct the (possible) moments and (possible) histories of a branching time structure. We then show how the problem of ontological competition can be solved by adding an incompatibility partition to a structure of possibilia, and conclude with some remarks about how this addition might provide a clue for developing a variant of the theory of branching space-time that can account for the trousers worlds of general relativity.

Johan van Benthem and Eric Pacuit (University of Amsterdam and University of Maryland at College Park): Connecting Logics of Choice and Change

This paper is an attempt at clarifying the current scene of sometimes competing action logics, looking for compatbilities and convergences. Current paradigms for deliberate action fall into two broad families: dynamic logics of events, and STIT logics of achieving specified effects. We compare the two frameworks, and show how they can be related technically by embedding basic STIT into a modal logic of matrix games. Amongst various things, this analysis shows how the attractive principle of independence of agents' actions in STIT might actually be a source of high complexity in the total action logic. Our main point, however, is the compatibility of dynamic logics with explicit events and STIT logics based on a notion that we call 'control'—and we present a new system of dynamic-epistemic logic with control that has both. Finally, we discuss how dynamic logic and STIT face similar issues when including further crucial aspects of agency such as knowledge, preference, strategic behavior, and explicit acts of choice and deliberation.

Daniel Vanderveken (University of Quebec at Trois-Riviéres): Intentionality and minimal rationality in the logic of action

Philosophers have overall studied intentional actions that agents attempt to perform in the world. However the pioneers of the logic of action, Belnap and Perloff, and their followers have tended to neglect the intentionality proper to human action. My primary goal is to formulate here a more general logic of action where intentional actions are primary as in contemporary philosophy of mind. In my view, any action that an agent performs involuntarily could in principle be intentional. Moreover any involuntary action of an agent is an effect of intentional actions of that agent. However, not all unintended effects of intentional actions are the contents of unintentional actions, but only those that are historically contingent and that the agent could have attempted to perform. So many events which happen to us in our life are not really actions. My logic of action contains a theory of attempt, success and action generation. Human agents are or at least feel free to act. Moreover their actions are not determined. As Belnap pointed out, we need branching time and historic modalities in the logic of action in order to account for indeterminism and the freedom of action.

Propositions with the same truth conditions are identified in standard logic. However they are not the contents of the same attitudes of human agents. I will exploit the resources of a non classical predicative propositional logic which analyzes adequately the contents of attitudes. In order to explicate the nature of intentional actions one must deal with the beliefs, desires and intentions of agents. According to the current logical analysis of propositional attitudes based on Hintikka's epistemic logic, human agents are either perfectly rational or completely irrational. I will criticize Hintikka's approach and present a general logic of all cognitive and volitive propositional attitudes that accounts for the imperfect but minimal rationality of human agents. I will consider subjective as well as objective possibilities and explicate formally possession and satisfaction conditions of propositional attitudes. Contrary to Belnap, I will take into account the intentionality of human agents and explicate success as well as satisfaction conditions of attempts and the various forms of action generation. This paper is a contribution to the logic of practical reason. I will formulate at the end many fundamental laws of rationality in thought and action.

Ming Xu (Wuhan University): Group strategies and independence

We expand Belnap's general theory of strategies for individual agents to a theory of strategies for multiple agents and groups of agents, and propose a way of applying strategies to deal with future outcomes at the border of a strategy field. Based on this theory, we provide a preliminary analysis on distinguishability and independence, as a preparation for a general notion of dominance in the decision-theoretical approach to deontic logic.

Appendix B: On the History of *stit* and Branching Space-Times

Interview with Nuel Belnap, conducted at his home in Pittsburgh, March 15, 2013. Interviewer: Thomas Müller.

TM: Let's talk about the origins of *stit*. Jan Broersen, one of our authors, mentioned that you had told him about the history one evening over dinner, when you were in Utrecht a couple of years ago. You developed some of that in seminars, in the 1980s?

NB: It started with a seminar I taught on Charles Hamblin's book, *Imperatives*, as far as I recall. Maybe two seminars, maybe just the one. I certainly worked out a good bit about *stit* for the seminar, writing out a few pages each week.

TM: Hamblin's book came out in 1987, with your preface, so this must have been the mid-1980s. The first *stit* paper came out in 1988, so that would fit temporally. Rich Thomason, whose work on branching histories theories for indeterminism forms part of the formal background for *stit*, was your colleague at the University of Pittsburgh until 1999, when he moved to the University of Michigan. You have often remarked that you were amazed by how long this theory was lying dormant, with the initial paper from 1970 and the *Handbook of philosophical logic* chapter published in 1984—there was virtually nothing happening in between. Thomason has some remarks on the deontic aspects of his approach.

NB: He did work out some deontic ideas, yes.

TM: For *stit* you were mainly working with Mickey Perloff, right? And then some graduate students were attracted as the project was building up momentum—for example, Jeff Horty, Mitch Green, and Ming Xu. What I find interesting is the interaction between the two projects, clarifying the foundations of indeterminism through the application of indeterministic models in the logic of agency, and building up the logic of agency against the background of branching histories models for indeterminism. Your book, *Facing the future*, exemplifies this nicely.

NB: The book must be right.—Mickey took part in the Hamblin seminar; we worked together for many years afterwards.

TM: The branching times framework—I assume you knew about that from much earlier? When Alan Anderson was at Manchester to work with Prior in the mid-60s, he would have brought back some ideas about that?

NB: Yes, I think so. Prior visited Alan in 1965 or so, he came to a dinner party at his house. That's when he had decided not to come to the U.S. any more, because of the Vietnam war.

TM: So the branching time framework was basically sitting there to be used, and you made the connection, not working on issues in branching time, but when thinking about how to model the content of an imperative?

NB: Hamblin's book is on imperatives, yes. There's a mini-history of approaches to the modal logic of agency early in the book, *Facing the future*.

TM: When you started working on *stit*, was that working out the theory of a single agent first, with other agents entering the theory only later?

NB: No, the multi-agent case was in there from the beginning. The other agents didn't do anything, to begin with.

TM: There is the "independence of agents" axiom in multi-agent *stit*: "Something happens"; no matter what one agent chooses at a moment, all other possible choices of the other agents must be compatible with that. That was the nucleus of the project of branching space-times, I think Paul Bartha told me about that at one point?

NB: I do remember that I had the main ideas of branching space-time in the late 1980s, and I was shopping them around. Every visitor to the department got an hour of that. That was before the paper was published in 1992.

TM: Chris Hitchcock told me that he was there "when it happened".

NB: That was a small seminar, I think Chris and Philip Kremer were the only students in the class.—I don't have any records on what and who I was teaching. I had seven four-drawer file cabinets at the department, and when I retired a few years ago I just asked the secretary, Connie, to get rid of them.

TM: How did the main ideas come about?

NB: I learned about directed sets from Dana Scott. Not when he was at Carnegie Mellon University in Pittsburgh, but long before then. We overlapped at Oxford in 1970. Directed sets is really what made branching space-times go, it's the basis for the definition of a history. That idea had been with me for many years.

TM: This is a recurring theme in our discussions: It takes time. Ideas can take 20 years, and then they reappear, or become salient all of a sudden.

NB: They cook a long time.

TM: For me it's now 15 years since I first read the paper on branching space-times—and there's still a lot for me to discover, like the one footnote on topology that has driven a small industry over the last couple of years.

NB: I was just rereading it earlier the day, in order to see whether I could find the right platform for the method of extension and intension that we are working on now. I didn't get anywhere, though.

TM: It's good that you made that postprint, ten years after the first publication. That shows some progress.

NB: In the branching space-times paper in the beginning I had a substantial section on agency, which I was persuaded to dissever.

TM: There is a gap of more than ten years between the 1992 publication of the BST paper and your published work on agency in BST, starting around 2005.

NB: The connection was there from the start.

TM: Thanks, Nuel.

References

Belnap, N. 1992. Branching space-time. *Synthese* 92(3): 385–434 (see also the postprint 2003, available on philsci-archive, http://philsci-archive.pitt.edu/1003/).

Belnap, N. 1999. Concrete transitions. In *Actions, Norms, Values: Discussions with Georg Henrik von Wright*, ed. G. Meggle, 227–236. Berlin: de Gruyter.

Belnap, N. 2002a. Double time references: Speech-act reports as modalities in an indeterminist setting. In *Advances in modal logic*, vol. 3, eds. F. Wolter, H. Wansing, M. de Rijke and M. Zakharyaschev, 37–58. Singapore: World Scientific.

Belnap, N. 2002b. EPR-like "funny business" in the theory of branching space-times. In *Non-locality and modality*, ed. T. Placek and J. Butterfield, 293–315. Dordrecht: Kluwer.

Belnap, N. 2003. No-common-cause EPR-like funny business in branching space-times. *Philosophical Studies* 114: 199–221.

Belnap, N. 2005a. Agents and agency in branching space-times. In *Logic, thought and action*, ed. D. Vanderveken, 291–313. Berlin: Springer.

Belnap, N. 2005b. A theory of causation: Causae causantes (originating causes) as inus conditions in branching space-times. *British Journal for the Philosophy of Science* 56: 221–253.

Belnap, N. 2006. Bressan's type-theoretical combination of quantification and modality. In *Modality Matters. Twenty-five Essays in Honor of Krister Segerberg*, eds. H. Lagerlund, S. Lindström, and R. Sliwinski, 31–53. Department of Philosophy, Uppsala University, Uppsala.

Belnap, N. 2011. Prolegomenon to norms in branching space-times. *Journal of Applied Logic* 9(2): 83–94.

Belnap, N. 2012. Newtonian determinism to branching space-times indeterminism in two moves. *Synthese* 188(1): 5–21.

Belnap, N. and M. Green. 1994. Indeterminism and the thin red line. *Philosophical Perspectives* 8:365–388 (ed. J. Tomberlin).

Belnap, N., and T. Müller. 2013a. CIFOL: Case-intensional first order logic (I). Toward a logic of sorts. *Journal of Philosophical Logic*. doi:10.1007/s10992-012-9267-x.

Belnap, N. and T. Müller. 2013b. BH-CIFOL: Case-intensional first order logic. (II) Branching histories. *Journal of Philosophical Logic*. doi:10.1007/s10992-013-9292-4.

Belnap, N., and M. Perloff. 1988. Seeing to it that: A canonical form for agentives. *Theoria* 54:167–190. Corrected version in Kyburg Jr., H.E., R.P. Loui and G.N. Carlson, eds., 1990. *Knowledge representation and defeasible reasoning, Studies in cognitive systems*, vol. 5, 167–190. Dordrecht: Kluwer.

Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future. Agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Bressan, A. 1972. *A general interpreted modal calculus*. New Haven: Yale University Press. (Foreword by Nuel D. Belnap Jr.)

Carnap, R. 1947. *Meaning and necessity: A study in semantics and modal logic*. Chicago: University of Chicago Press.

Chellas, B.F. 1969. *The logical form of imperatives*. Stanford: Perry Lane Press.

Earman, J. 2008. Pruning some branches from branching space-times. In *The Ontology of Spacetime II*, ed. D. Dieks, 187–206. Amsterdam: Elsevier.

Einstein, A., B. Podolsky, and N. Rosen. 1935. Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 47(10): 777–780.

Fine, K. 2005. *Modality and Tense*. Oxford: Oxford University Press.

Hofer-Szabó, G., M. Rédei, and L.E. Szabó. 2013. *The Principle of the Common Cause*. Cambridge: Cambridge University Press.

Horty, J.F. 2001. *Agency and Deontic Logic*. Oxford: Oxford University Press.

Horty, J.F., and N. Belnap. 1995. The deliberative stit: A study of action, omission, ability and obligation. *Journal of Philosophical Logic* 24: 583–644.

Kamp, H. 1971. Formal properties of 'now'. *Theoria (Lund)* 37: 227–273.

Kaplan, D. 1989. Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals; and afterthoughts. In *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein, 481–563; 565–614. Oxford: Oxford University Press.
Kracht, M., and O. Kutz. 2007. Logically possible worlds and counterpart semantics for modal logic. In *Philosophy of logic, Handbook of the philosophy of science*, ed. D. Jacquette, 943–995. Amsterdam: Elsevier.
Lewis, D.K. 1986. *On the plurality of worlds*. Oxford: Blackwell.
MacFarlane, J. 2003. Future contingents and relative truth. *The Philosophical Quarterly* 53(212): 321–336.
MacFarlane, J. 2014. *Assessment sensitivity: Relative truth and its applications*. Oxford: Oxford University Press.
Mackie, J.L. 1980. *The cement of the universe. A study of causation*. Oxford: Oxford University Press.
Malpass, A., and J. Wawer. 2012. A future for the thin red line. *Synthese* 188(1): 117–142.
McCall, S. 1990. Choice trees. In *Truth or consequences. Essays in honor of Nuel Belnap*. J. Dunn and A. Gupta, eds., 231–244. Dordrecht: Kluwer.
McCall, S. 1994. *A model of the universe*. Oxford: Oxford University Press.
Müller, T. 2002. Branching space-time, modal logic and the counterfactual conditional. In *Non-locality and Modality*, ed. T. Placek and J. Butterfield, 273–291. Dordrecht: Kluwer.
Müller, T. 2005. Probability theory and causation: a branching space-times analysis. *British Journal for the Philosophy of Science* 56(3): 487–520.
Müller, T. 2007. Branch dependence in the "consistent histories" approach to quantum mechanics. *Foundations of Physics* 37(2): 253–276.
Müller, T. 2010. Towards a theory of limited indeterminism in branching space-times. *Journal of Philosophical Logic* 39: 395–423.
Müller, T. 2011a. Branching space-times, general relativity, the Hausdorff property, and modal consistency. Technical report, Theoretical Philosophy Unit, Utrecht University. http://philsci-archive.pitt.edu/8577/.
Müller, T. 2011b. Probabilities in branching structures. In *Explanation, prediction and confirmation. The Philosophy of Science in a European Perspective*, Vol. 2, eds. D. Dieks, W.J. Gonzalez, S. Hartmann, T. Uebel and M. Weber, 109–121. Dordrecht: Springer.
Müller, T. 2013a. Alternatives to histories? Employing a local notion of modal consistency in branching theories. *Erkenntnis*. doi:10.1007/s10670-013-9453-4.
Müller, T. 2013b. A generalized manifold topology for branching space-times. *Philosophy of science*, forthcoming; preprint URL = http://www.jstor.org/stable/10.1086/673895.
Müller, T., N. Belnap, and K. Kishida. 2008. Funny business in branching space-times: infinite modal correlations. *Synthese* 164: 141–159.
Penrose, R. 1979. Singularities and time-asymmetry. In *General relativity: an Einstein centenary survey*, ed. S.W. Hawking and W. Israel, 581–638. Cambridge: Cambridge University Press.
Placek, T. (2010). Bell-type correlations in branching space-times. In *The Analytic Way. Proceedings of the 6th European Congres of Analytic Philosophy*, eds. T. Czarneck, K. Kijania-Placek, O. Poller and J. Woleński, 105–144. London: College Publications.
Placek, T., and N. Belnap. 2012. Indeterminism is a modal notion: branching spacetimes and Earman's pruning. *Synthese* 187(2): 441–469.
Placek, T., and L. Wroński. 2009. On infinite EPR-like correlations. *Synthese* 167(1): 1–32.
Ploug, T. and P. Øhrstrøm. 2012. Branching time, indeterminism and tense logic. Unveiling the Prior-Kripke letters. *Synthese* 188(3): 367–379.
Pörn, I. 1977. *Action theory and social science: Some formal models*. Dordrecht: D. Reidel.
Prior, A.N. 1957. *Time and modality*. Oxford: Oxford University Press.
Prior, A.N. 1967. *Past, present and future*. Oxford: Oxford University Press.
Quine, W. 1980. *From a logical point of view. Nine logico-philosophical essays*. Cambridge: Harvard University Press. (2nd, revised edition.)

Rumberg, A., and T. Müller. 2013. Transitions towards a new account of future contingents. (Submitted.)
Steward, H. 2012. *A Metaphysics for Freedom*. Oxford: Oxford University Press.
Strobach, N. 2007. *Alternativen in der Raumzeit: Eine Studie zur philosophischen Anwendung multimodaler Aussagenlogiken*. Berlin: Logos.
Szabó, L., and N. Belnap. 1996. Branching space-time analysis of the GHZ theorem. *Foundations of Physics* 26(8): 989–1002.
Thomason, R.H. 1970. Indeterminist time and truth-value gaps. *Theoria* 36: 264–281.
Thomason, R.H. 1984. Combinations of tense and modality. In *Handbook of philosophical logic, vol. II: extensions of classical logic*, volume 165 of *Synthese Library, Studies in Epistemology*, eds. D. Gabbay and G. Guenthner, 135–165. Dordrecht: D. Reidel Publishing Company.
Thomson, J.J. 1990. *The realm of rights*. Cambridge: Harvard University Press.
Van Inwagen, P. 1983. *An essay on free will*. Oxford: Oxford University Press.
von Kutschera, F. 1986. Bewirken. *Erkenntnis* 24(3): 253–281.
von Wright, G.H. 1963. *Norm and action. A logical inquiry*. London: Routledge.
Weiner, M., and N. Belnap. 2006. How causal probabilities might fit into our objectively indeterministic world. *Synthese* 149: 1–36.
Wroński, L., and T. Placek. 2009. On Minkowskian branching structures. *Studies in History and Philosophy of Modern Physics* 40: 251–258.

Decisions in Branching Time

Paul Bartha

Abstract This chapter extends the deontic logic of Horty (*Agency and deontic logic*, 2001) in the direction of decision theory. Horty's deontic operator, the dominance ought, incorporates many concepts central to decision theory: acts, causal independence, utilities and dominance reasoning. The decision theory associated with dominance reasoning, however, is relatively weak. This chapter suggests that deontic logic can usefully be viewed as *proto-decision theory*: it provides clear foundations and a logical framework for developing norms of decision of varying strength. Within Horty's framework, deontic operators stronger than the dominance ought are defined for decisions under ignorance, decisions under risk, and two-person zero-sum games.

1 Introduction: Decision Theory and Deontic Logic

Consider the following two decision problems.

Example 1 (*Gambler*): An agent, α, is offered a gamble. If she accepts, she pays $5. A coin is then tossed: on *Heads* she wins $10; on *Tails* she wins nothing. If she declines the gamble, she simply keeps her $5.

Example 2 (*Matching Pennies*): Two agents, α and β, simultaneously choose whether to display a penny *Heads up* or *Tails up*. If the displayed sides of the two pennies match, then α wins $1 from β. If the two sides do not match, then β wins $1 from α.

Assuming that these agents value money positively and that there are no relevant external considerations, what should α do in these two scenarios? To find answers, we

P. Bartha (✉)
Department of Philosophy, University of British Columbia, 1866 East Mall, E370,
Vancouver, BC V6T 1Z1, Canada
e-mail: paul.bartha@ubc.ca

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_2,

might look to two distinct normative frameworks: decision theory[1] and deontic logic. Decision theory, despite its many paradoxes and controversies, provides our most successful formal account of rational choice. Deontic logic, with its own paradoxes and controversies, offers an alternative way to think about what α ought to do.

What is the relationship between decision theory and deontic logic? We might think of them as directed towards answering different questions. Decision theory rests on sharp assumptions about preferences and belief, but also upon less clear assumptions about causation, choice and counterfactuals. Deontic logic has no place for probabilities or probabilistic reasoning and does not pretend to offer a comprehensive theory of rational choice. Yet both theories can be applied to examples like *Gambler* and *Matching Pennies*. This suggests that they might, in some way, be rivals.

There is a third possibility. Rather than see them as unrelated or as rivals, we might regard deontic logic as a kind of *proto-decision theory*. Conceived in this way, deontic logic would play three roles. First, it would provide a rigorous logical framework for decision theory, a framework that clarifies foundational assumptions about causation, choice, counterfactuals and other relevant concepts. Second, stronger and weaker systems of deontic logic would be definable in this common framework. Third and finally, these systems of deontic logic would also be rudimentary decision theories: they would provide norms for choices by agents that are compatible with basic principles of decision theory.

When deontic logic is viewed in this way, the approach developed by Horty (2001) is exemplary. Horty's deontic logic (in contrast to many earlier approaches) is *prescriptive*: it is about choices by agents. It proposes semantics for what a particular agent *ought to do* at a particular moment in time. Horty's framework is built on top of Belnap's modal logic of agency, *stit* theory, a clear and rigorous logical and metaphysical account of agents making choices in indeterministic branching time.[2] *Stit* theory already takes us part way to causal decision theory because it incorporates causal notions: agents, branching time and a formulation of causal independence. Horty takes us further by incorporating utilities and dominance reasoning into his account, although he does not introduce probabilistic concepts. Still, his deontic logic does provide a weak decision theory: under reasonable assumptions, the set of obligations for an agent on the Horty semantics is a subset of the set of obligations that the agent has according to any sound principle of decision theory.

The main thesis of this chapter is that Horty's approach can be fruitfully enriched, first by a slight generalization of his account of causal independence and second by adding a 'thin' layer of probabilistic concepts.[3] The result is a framework in which deontic logic is even better suited to serve as *proto-decision theory* by playing the

[1] By 'decision theory' I mean to include the theory of decisions under ignorance, decisions under risk and normative game theory.

[2] This account is developed by Belnap and others in a series of articles, many of which are reprinted in (Belnap et al. 2001).

[3] In a similar spirit, Kooi and Tamminga (2008) show how Horty's framework can be supplemented to engage with game theory (though without introducing probabilistic ideas).

three roles mentioned above. There may be good reasons not to introduce full-blooded probability into the *stit* universe.[4] But it is plausible to introduce probabilities for mixed strategies by agents, and more generally for chance mechanisms (such as coin tosses and dice rolls). Horty's deontic logic can then be expanded fruitfully towards different branches of decision theory.

The chapter proceeds as follows. Section 2 reviews the basic ideas of *stit* (seeing-to-it-that). Section 3 proposes a generalization of Horty's account of causal independence. Horty's *Dominance Ought* is reviewed in Sect. 4, along with a slight modification corresponding to the generalization of Sect. 3. The remaining sections explore expansions of Horty's deontic framework to decisions under ignorance (Sect. 5), decisions under risk (Sect. 6) and elementary game theory (Sect. 7). While the focus of the chapter is on endowing deontic logic with resources from decision theory, I conclude with a brief discussion of return benefits for decision theory.

2 Seeing to it That (*stit*)

In a series of articles, Belnap and others have proposed semantics for the modal construction, "α sees to it that A," or [α *stit*: *A*] for short. To keep things brief, I pass over the philosophical motivation and simply review concepts that are crucial for this chapter. The best single source of information on *stit* is the volume of articles (Belnap et al. 2001).

2.1 Semantics for cstit with One Agent

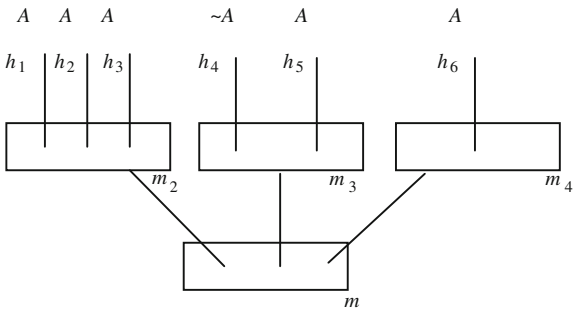
There are three accounts of [α *stit*: *A*] on offer: the Belnap "achievement stit" (*astit*), the Horty/von Kutschera "deliberative stit" (*dstit*) and the "Chellas stit" (*cstit*). Since the latter is employed by Horty in his deontic logic, this section presents, in cursory form, only the semantics for the Chellas stit, following notation that borrows from both (Belnap et al. 2001) and (Horty 2001).[5] The fundamental idea of [α *cstit*: *A*] is that *A* is guaranteed by a present choice of agent α ([α *dstit*: *A*] is more complex because it requires this same *positive condition*, together with the *negative condition* that *A* is not 'settled-true' in the sense of (3) below).

The framework begins with an indeterministic branching time structure $< Tree, < >$, where *Tree* is a non-empty set of *moments*, *m*, and $<$ is a tree-like partial ordering of those moments. A *history*, *h*, in *Tree* is a maximal chain of moments, i.e., a complete temporal evolution of the world. If *m* is a moment, write

[4] Belnap et al. (2001) consider and reject the idea that "α sees to it that *A*" should be modelled as "α's choice guarantees a high probability for *A*"; Broersen develops this very notion in (2011) and elsewhere. In this chapter, I am concerned not with a probabilistic version of *stit* but with the importance of probabilities for norms of choice.

[5] Both *dstit* and *cstit* have been useful in deontic logic (Belnap et al. 2001). On *dstit* versus *cstit* as an analysis of seeing-to-it-that, see (Chellas 1992) and (Horty 2001).

Fig. 1 Histories



$H_m = \{h/m \in h\}$ for the set of all histories containing (passing through) m. The situation is illustrated in Fig. 1, where the upward direction represents later moments.

In the picture, $H_m = \{h_1, \ldots, h_6\}$, while $H_{m2} = \{h_1, h_2, h_3\}$. The histories h_1 and h_2 are *undivided* at m because they share a later moment (m_2) in common; the histories h_1 and h_4 are *divided* at m.

Sentences are constructed from propositional variables $A, B, \ldots$ using the following operators:

(i) Truth-functional operators: $\sim$, $\vee$ (with abbreviations $\wedge$, $\supset$, $\equiv$)
(ii) Necessity operators: *Universally*:, *Settled*:
(iii) Tense operators: *Will*: and *Was*:
(vi) Agentive operator: [α *cstit*: __] where α denotes an individual agent

The truth of sentences is evaluated relative to a moment-history pair *m/h*, where m is a moment belonging to history h. A *model* M pairs the tree structure with an interpretation that maps each propositional variable A to a set of *m/h* pairs where A is true:

(1) $M, m/h \models A$ iff A is true at m/h.

In Fig. 1, for example, $M, m_2/h_1 \models A$ while $M, m_3/h_4 \not\models A$. The clauses for the truth-functional operators are standard. For the modal operators, *Universally*: represents truth throughout *Tree*, while *Settled*: represents truth throughout a moment. The relevant clauses are as follows:

(2) $M, m/h \models Universally$: A iff $M, m'/h' \models A$ for all m'/h' in $Tree$
(3) $M, m/h \models Settled$: A iff $M, m/h' \models A$ for all h' with $m \in h'$

For the tense operators, we have

(4) $M, m/h \models Will$: A iff $M, m'/h \models A$ for some m' in h with $m < m'$
(5) $M, m/h \models Was$: A iff $M, m'/h \models A$ for some $m' < m$.

In Fig. 1, *Will*: A is true at m/h_1 but false at m/h_4. *Settled*: A is true at m_2/h_1 but false at m_3/h_5.

Finally, we come to [α *cstit*: *A*]. This requires enriching the branching time framework of $< Tree, <\ >$ with a nonempty set *AGENT* of agents (denoted α, β, and so forth) and a function *Choice* that represents choices by agents. The most important idea is that of a *choice set* for α at moment *m*, which is a partition of the histories passing through *m* into *choice cells* (or simply *choices*) for α. α's power of choice consists in "constraining the course of events to lie within some definite subset of the possible histories still available". (Belnap et al. 2001, 33). That is, choice is identified with the selection of one cluster of histories. The agent picks the cluster, but cannot select a unique history within the choice cell. All of this is formalized in the following definitions.

(6) *stit frames*.

A *stit frame* $< Tree, <, AGENT, Choice >$ is a structure with *Tree* and $<$ as above, *AGENT* a nonempty set of agents, and *Choice* a function mapping agent α and moment *m* into a partition of H_m characterized as follows:

- $Choice^m_\alpha$ is a partition of H_m into mutually exclusive and exhaustive sets.
- Each member of $Choice^m_\alpha$ is called a *choice cell* (or *choice*) for α at *m*
- *h* and *h*′ are *choice-equivalent for* α *at m* (written $h' \equiv^\alpha_m h$) if they belong to the same choice cell for α at *m* (no choice that α can make at *m* tells them apart).

Choice is subject to the following condition (and one further condition, *Weak Independence of Agents*, to be described shortly):

(7) *No Choice between Undivided Histories*.

If *h* and *h*′ are undivided at *m*, then *h* and *h*′ must belong to the same choice cell for α at *m*.

A choice by some agent is the most obvious means by which histories are divided. Histories also divide as the result of chance processes in *Nature*. A coin toss serves as a paradigm example.

With this apparatus on board, we can define what it is for [α *cstit*: *A*] to hold at *m/h*:

(8) $M, m/h \models [\alpha\, cstit$: A] iff *A* is true at (m, h') for all h' with $h' \equiv^\alpha_m h$. (By choosing the cell containing h, α guarantees that *A* is true as *A* holds on all histories consistent with α's choice.)

Figure 2 provides the basic picture. In the picture, *m* is a moment in *Tree* that has been blown up to reveal the choice structure. The three boxes represent choice cells for α at *m*; each history in H_m belongs to exactly one box. The truth-value of *A* is shown for each moment–history pair. Here, [α *cstit*: *A*] holds just at m/h_1 and m/h_2 (Note that [α *cstit*: ∼*A*] is false throughout *m*; there is no *law of excluded middle* for seeing-to-it-that).

Fig. 2 [α *cstit*: *A*]

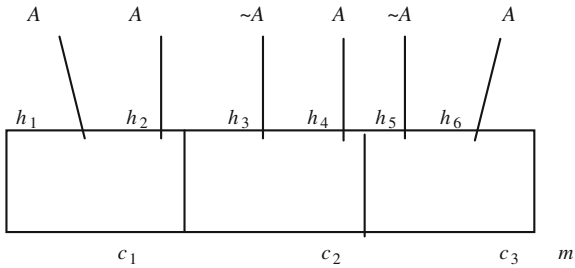
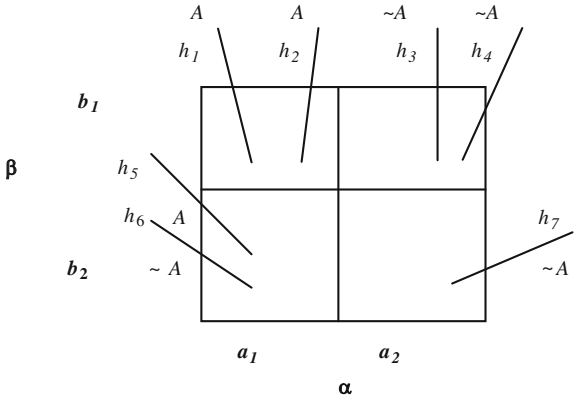


Fig. 3 Multiple agents



2.2 Multiple Agents: Independence and Joint Agency

The concept of *cstit* generalizes to groups of agents. Most of the ideas can be made clear by considering just two agents, α and β, making simultaneous choices. The simplest way to represent this is once again with a blown-up picture of the moment, this time two-dimensional, with the choices of α represented on the horizontal axis and the choices of β on the vertical axis, as in Fig. 3.

Here, α and β face non-trivial choices, formally specified by choice sets (partitions) $Choice^m_\alpha$ and $Choice^m_\beta$. In the picture, the possible choices are $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ for α, and $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ for β; thus, the choice *sets* are $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ and $\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$.

The key assumption made by Belnap and Perloff is that every combination of choices by α and β is possible:

(9) [*Weak*] *Independence of Agents*.

> For each moment and for each way of selecting one choice for every agent (in the set *AGENT*) from among that agent's set of possible choices at that moment, the intersection of all the choices selected must contain at least one history.

To formalize this condition, we follow Horty in defining a *selection function s at moment m* to be a mapping from *AGENT* into H_m that selects one action for each

agent: $s(\alpha) \in Choice^m_\alpha$ for each α. Let $Select_m$ be the set of all such functions. We re-state (9) as follows:

For each moment m and each selection function s in $Select_m$,

$$\bigcap_{\alpha \in AGENT} s(\alpha) \neq \phi.$$

Belnap comments (Belnap et al. 2001, p. 218) that while *Independence* is a "fierce" constraint (implying, for example, that no two agents can have the same possible choices at the same moment), it is also fairly weak ("banal" is Belnap's term): it would be strange indeed if without causal priority, one agent's choices could constrain what the other agent may choose.

The other important idea is *joint agency*. Again, the basic idea can be explained with just two agents. Let $\Gamma = \{\alpha, \beta\}$, where α and β are distinct agents. We want to define truth conditions for $[\Gamma\ cstit{:}\ A]$. The concept is illustrated by referring once again to Fig. 3. With the assignments given by our model M, neither α nor β can see to it that A on any history in m. However, M, $m/h_1 \models [\Gamma\ stit{:}\ A]$ because A holds at every history h' that belongs to the choice cell containing h_1 that is determined *jointly* by α and β. The formal definition is the same as (8) except that the condition invokes equivalence within the choice cell determined jointly by the agents in Γ.

3 Causal Independence

Both causal decision theory and Horty's deontic logic depend upon the concept of causal independence. Before examining how causal independence can be characterized in the *stit* framework, it is helpful to review its role in causal decision theory. I focus on the formulation due to Skyrms (1980).[6]

Skyrms's formulation requires the identification of a set of independent causal factors that provide the background for an agent's choices. Each independent causal factor is represented as a random variable X_i. For simplicity's sake, suppose that the set of possible outcomes $O_1, \ldots, O_n$ of interest to the agent is finite, that the set $X_1, \ldots, X_N$ of independent factors relevant to these possible outcomes is also finite, and that each variable X_i can take on finitely many values. Then the set $\boldsymbol{S}$ consisting of all possible combinations of assignments to these variables is also finite. This set constitutes a partition of the set of possible worlds into *causal background contexts* or *states* $S_1, \ldots, S_M$: each state S_i is obtained by specifying one possible value for each of $X_1, \ldots, X_N$. Suppose that the agent has a finite set $\{K_1, \ldots, K_m\}$ of available alternative acts. The crucial idea is that the causal factors *in conjunction* with these alternative acts determine relevant conditional chances of the possible outcomes: the

[6] There are numerous formulations of causal decision theory, including (Gibbard and Harper 1978), (Skyrms 1980) and (Joyce 1999). (Skyrms 1980) is in some ways the simplest and most relevant to our present concerns.

conditional chances $P(O_k \,/\, K_i \wedge S_j)$ are *constant* within each $K_i \wedge S_j$. Finally, the agent assigns a utility $u(O_k \wedge K_i \wedge S_j)$ to each outcome-act-state combination.

For a standard example, let K_i be the selection of a ball from one of m urns, let $O_1, \ldots, O_n$ represent n different colours of ball that may be drawn, and let $S_1, \ldots, S_M$ stand for M possible initial arrangements of coloured balls in the m urns. Then the probability $P(O_k \,/\, K_i \wedge S_j)$ is the conditional chance of drawing a ball of colour O_k, given arrangement S_j and the act K_i of drawing from urn i. The utility $u(O_k \wedge K_i \wedge S_j)$ depends upon the desirability of each combination (perhaps the agent has placed a bet in advance).

To allow for cases in which the agent is uncertain about the background context, Skyrms introduces a subjective probability distribution $prob(S_j)$ over all of the states. In the urn example, this represents your initial credence about the likelihoods of the different possible arrangements. The expected utility of act K_i is then given by the formula

(10) *Expected utility*.

$$\mathscr{U}(K_i) = \Sigma_j prob(S_j) \Sigma_k P(O_k / K_i \wedge S_j) \cdot u(O_k \wedge K_i \wedge S_j).$$

The thesis of causal decision theory is that *a rational agent maximizes expected utility* as given by (10). The equation highlights the importance of independent causal factors in the theory; the outer summation is over all possible states.[7]

Horty's deontic logic has a similar, but much weaker, guiding principle. Without conditional chances or credences, his analysis (outlined in the next section) is based solely upon the concept of dominance. Yet dominance reasoning shares with causal decision theory the need for a set of independent causal factors and a corresponding set of causal background contexts. *Stit* frames help to make these things precise by providing a concrete interpretation of possible worlds and a plausible way to identify some of the independent causal factors. I first review Horty's account and then propose a slight generalization.

Horty begins with an informal explication of causal independence:

> ...the basic intuition … is that a proposition is supposed to be causally independent of the actions available to a particular agent whenever its truth or falsity is guaranteed by a source of causality other than the actions of that agent. (2001, 82)

Recall postulate (9), *Weak Independence of Agents*: if agents in a group make simultaneous choices, then the intersection of the relevant choice sets is non-empty. Horty strengthens this in two ways. First, he assumes that all choices at m by agents other than α are independent causal factors for α's choice at m. This strengthened assumption, stated in counterfactual terms, applies to all moment–history pairs m/h and all agents α.

---

[7] In this chapter, for the sake of simplicity, we ignore the element of subjective probability represented by *prob*. Skyrms (1994) provides a good discussion. In the present framework, subjective probability could usefully be introduced to represent the agent's uncertainty about location, i.e., about which m is the moment of decision.

(11) *Strong Independence* (*of Agents*).

> Let S represent the intersection of all actual choices (i.e., choice cells) of all agents other than α at *m/h*. If α were to make a different choice than the one made at *m/h*, the other agents would still (collectively) choose S.

Second, Horty adopts a provisional simplifying assumption that I shall refer to as *Causal Completeness of AGENT*.

(12) *Causal Completeness* (*of AGENT*).

> Choices by agents in $AGENT\backslash\{\alpha\}$ (i.e., agents other than α) are the *only* independent causal factors relevant to α's choice.

Taken together, the two assumptions imply that the independent causal factors for α's choice are precisely choices by other agents.

To illustrate these ideas, consider Fig. 3 again. Suppose that the picture represents choices by two agents, α and β, to cooperate in moving a heavy box at moment m. At m/h_1 and m/h_2, the box is successfully moved (represented by the proposition A). Here, α chooses $\boldsymbol{a}_1$ and β chooses $\boldsymbol{b}_1$. The choice by β is causally independent of the choice by α (by *Strong Independence*): if α were to choose $\boldsymbol{a}_2$ (don't cooperate) at m, then β would still choose $\boldsymbol{b}_1$. Further, this is the only relevant independent causal factor for α's choice (by *Causal Completeness*).

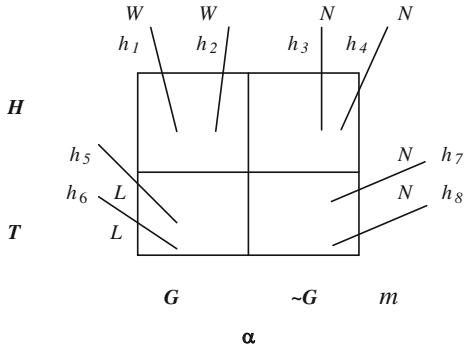
What defense can we give for assumptions (11) and (12)? For the first, the argument is that from *Weak Independence of Agents* and the simultaneity of choices by other agents, it is reasonable to infer *Strong Independence*. Simultaneous choices by agents must be causally independent of each other and of α's choices at m.[8] By contrast, *Causal Completeness* is offered merely as a useful "initial approximation" for Horty's deontic logic. Horty explicitly identifies two sources of independent causal influence that are not reflected in his account: "nonagentive sources" (*Nature*) and later choices by agents other than α (Horty 2001, 89–95). While I agree with Horty that fully to incorporate these influences into the analysis would be a "substantial research task", I believe that important special cases can be accommodated without great difficulty.

Consider first the case of *Nature*. Figure 4 illustrates a version of *Example 1* (*Gambler*), described at the start of this chapter.[9] At moment m, α has a choice of gambling (G) or not. The gamble costs $5. A fair coin toss is to be performed at m whether or not α gambles. If the coin comes up $Heads(H)$, α leaves with $10, but on *Tails* ($T$) she gets nothing. If α declines the gamble, she keeps her $5. The outcomes

[8] Talk of simultaneity suggests that there might be some gain in clarity by moving to a framework of *branching space-time* (Belnap 1992) instead of *branching time*. The added complexity of branching space-time is unnecessary, however, since for present purposes simultaneity is adequately characterized in a branching-time framework in terms of condition (7), *No Choice between Undivided Histories*. That is, it suffices that there exists a moment m such that for each agent, histories within the relevant choice cells are undivided at m while histories belonging to different choice cells are divided at m.

[9] The *Gambler* example is due to Horty (2001), who formulates and discusses a number of versions.

Fig. 4 Gambler (I)



are as shown, where *W* signifies a win, *L* a loss and *N* the *status quo* where α neither wins nor loses.

There are no other agents besides α in this picture.[10] Yet the coin toss has the characteristics of an independent causal factor. It occurs simultaneously with α's choice. It satisfies an analogue of *Weak Independence*: any choice by α is compatible with either result, *Heads* or *Tails*. Finally, it is reasonable to regard the coin toss as satisfying *Strong Independence*. Consider

(13) If α had gambled, he would have won.

We endorse the truth of (13) at m/h_3 and m/h_4, and its falsity at m/h_7 and m/h_8. To summarize these observations, we introduce a random variable *Toss* with values $\{Heads, Tails\}$. With respect to both *Weak Independence* and *Strong Independence*, *Toss* has characteristics analogous to those of an agent with choice set $\{Heads, Tails\}$.

The generalization proposed here is to extend Horty's account of causal independence to include not just agents but also *chance mechanisms operating simultaneously with* α's *choice*. By *chance mechanisms*, I mean well-understood processes such as those employed in games of chance: coin tosses, dice rolls, card drawings and so forth. These are singled out for two reasons. First, such processes have outcomes with well-defined and unproblematic probabilities.[11] Second, these processes are crucial in defining *mixed strategies*, which will be important later in this chapter. Each such mechanism can be modeled as a random variable X that may take different possible values $X = x_i$ at the moment m, i.e., at distinct moment history pairs m/h and m/h'. Let *VAR* be the set of independent variables representing chance processes.[12] We shall make assumptions about *VAR* that are entirely parallel to those for *AGENT*.

[10] It may be that agent β tosses the coin, but β does not choose the result *Heads* or *Tails*. So the clusters shown are not choice sets for β.

[11] Gillies (2000) argues that such processes have a distinguished role in accounts of objective chance. In particular, the problem of identifying an appropriate reference class is relatively insignificant.

[12] We could relativize to each moment, using VAR_m for the set of variables that represent chance processes operating at m. We avoid this relativization both because we shall only ever be concerned

First, each random variable X must satisfy an analogue of the *No Choice Between Undivided Histories* condition, representing the fact that the chance process operates at moment m (rather than at a later moment). A pair of definitions makes this clear.

(14) $Rng^m(X)$.

By analogy with $Choice^m_\alpha$, if X is a random variable, define $rng^m(X)$ as the partition of H_m corresponding to the possible values $X = x_i$ at m. (Two histories h_1 and h_2 belong to the same element of $rng^m(X)$ if for some x_i, both $M, m/h_1 \models X = x_i$ and $M, m/h_2 \models X = x_i$.) For simplicity, we shall assume that these values x_i are always real numbers.[13]

(15) *No Separation of Undivided Histories*.

Whenever h_1 and h_2 are undivided at m, X has the same value $X = x_i$ at both m/h_1 and m/h_2. That is, h_1 and h_2 must belong to the same element of $rng^m(X)$.

Condition (15) rules out random variables that partition H_m based on future processes.

Next, we need analogues for (9) *Weak Independence* and (10) *Strong Independence*. We want these analogues to apply to agents and chance processes taken together, which motivates the following definitions.

(16) *FACTOR*.

FACTOR is the union of *VAR* and the set of random variables representing choices by agents:

$$FACTOR = AGENT \cup VAR$$

(17) *Extended Selection Function*.

An *extended selection function* s at moment m is a mapping from FACTOR into H_m that selects a choice in $Choice^m_\alpha$ for each agent α in *AGENT*, and an element of $rng^m(X)$ for each variable X in *VAR*.

As before, we use the notation $Select_m$ for the set of all such functions.

It is sometimes convenient to regard the agents in *AGENT* as random variables, and to represent $Choice^m_\alpha$ as $rng^m(\alpha)$. This allows us to state a compact analogue of (9):

(Footnote 12 continued)
with a single moment and to maintain the analogy with *AGENT*, the set of agents fixed over all moments.

[13] For the purposes of this chapter and to maintain consistency with the definitions in Sect. 2, we assume that all statements $X = x_i$ can be represented in our language as propositional constants. My thanks to Thomas Müller for pointing this out.

(18) *Weak Independence of* FACTOR.

For each moment m and each extended selection function s in $Select_m$,

$$\bigcap_{X \in FACTOR} s(X) \neq \phi.$$

We also have an obvious formulation of *Strong Independence*:

(19) *Strong Independence of* FACTOR.

Let S represent the background state for $X = x_i$ at *m/h*, i.e.,

$$S = \bigcap_{Y \in FACTOR \setminus \{X\}} s(Y),$$

where $s(Y)$ is the element of $rng^m(Y)$ selected for Y at m/h. If a different value $X = x_j$ were selected at m, the other variable values and hence the background state S would remain the same.

[In particular, if any agent α were to make a different choice, all choices by other agents and all variable values would remain the same].[14]

On this account, causal independence extends to chance mechanisms that operate independently of each other and of agents, such as the coin toss in *Gambler* (I). We acknowledge this modification by extending our earlier definition of *stit frames*.

(20) *Extended stit frames*.

An *extended stit frame* is a structure $< Tree, <, FACTOR, Rng >$ that satisfies all earlier assumptions as well as (15), (18) and (19).

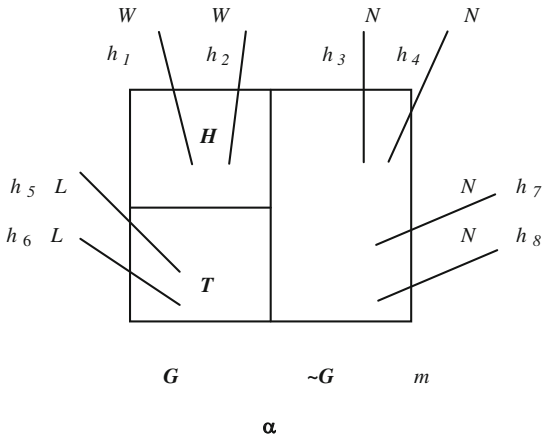
We are not quite done. A separate approach is needed to represent chance mechanisms *initiated* by agents. Consider a variation of *Gambler*: α tosses the coin if and only if she accepts the gamble; if she declines, there is no coin toss (Fig. 5).

In this case, it is inappropriate to model *Toss* as an independent causal factor. There is no independent partition of H_m; the toss does not happen if α declines to gamble. So both *Weak Independence* and *Strong Independence* fail.

The difficulty is that while *Gambler* (II) involves a well-understood chance mechanism that (to paraphrase Horty) represents a source of causality other than the actions of α, the mechanism does not operate independently of α and cannot be modelled as a random variable in *VAR*. An alternative approach, following Skyrms, is to represent

[14] Note that the plausibility of (19) depends upon the modest scope of the set *VAR*. The stated (though still undeniably vague) restriction is that the random variables in *VAR* are restricted to well-understood chance mechanisms, the sort of mechanisms that one could exploit in implementing a mixed strategy (see Sect. 7). In particular, I mean to exclude quantum phenomena.

Fig. 5 Gambler (II)



the operation of such mechanisms via conditional chances for *outcomes* within each causal background context. This approach will be developed below in Sect. 6.

4 Horty's Dominance Ought

Consider *Gambler* (I), as illustrated in Fig. 4. Substitute numerical values 10 in place of *W* (a winning gamble), 0 in place of *L* (a loss), and 5 in place of *N* (no gamble). These values represent the money that agent α possesses when the dust settles. We can also think of them as utilities that represent α's preferences.

Adding utilities to a *stit frame* gives us a *utilitarian stit frame*, defined by Horty as a structure of the form

$$< Tree, <, AGENT, Choice, Value >,$$

where $Tree$, $<$, $AGENT$ and *Choice* are as in Sect. 2, and *Value* is a function that assigns a real number $Value(h)$ to each history. A *utilitarian stit model* combines a utilitarian stit frame with an assignment of truth values to propositions. Figure 6 illustrates *Gambler* (I) in a utilitarian stit model.

Horty provides a semantics for statements of the form

$\odot[\alpha\ cstit\colon A]$ (α ought to see to it that A).[15]

The basic idea of his "dominance ought" is that α ought to see to it that *A* iff *A* is guaranteed by *each* optimal (non-dominated) choice. It takes care to make this

[15] Horty uses $\odot$ to distinguish his *dominance ought* from other obligation operators. I shall use $\preccurlyeq\cdot$ and $\prec\cdot$ to represent the corresponding dominance relations, described below. This helps to distinguish Horty's dominance ordering from variants to be introduced in later sections.

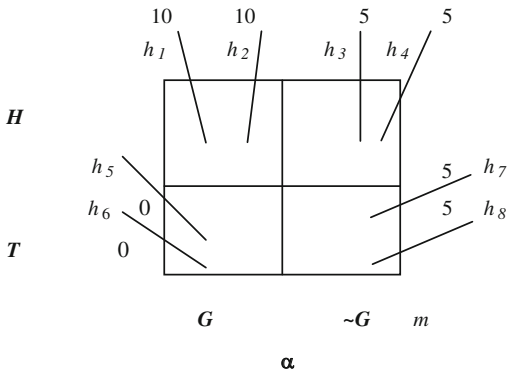


Fig. 6 Gambler (I) (utilitarian stit model)

precise and to handle cases where there are no optimal choices. The elements of Horty's account include *background states*, a *value ordering* on propositions at m (subsets of H_m), and the *dominance* relation between possible choices for an agent.

(21) *Dominance ordering on choices*.

- $State^{m}_{\alpha}$: the partition of histories through m into background causal contexts S for α's choice. For Horty, as we have seen, these background contexts are simply joint choices by all members of *AGENT* other than α.
- *Ordering relations* ($\leq$ and $<$) on propositions at moment m: If X and Y are two subsets of H_m, then (1) $X \leq Y$ if $Value(h) \leq Value(h')$ for each $h \in X$ and $h' \in$ Y, and (2) $X < Y$ if $X \leq Y$ and in addition, $Value(h) < Value(h')$ for some $h \in X$ and $h' \in Y$.
- *Dominance* relations ($\preccurlyeq\cdot$ and $\prec\cdot$) on $Choice^{m}_{\alpha}$: If K and K' are members of $Choice^{m}_{\alpha}$ (i.e., possible choices for α at m), then (1) $K \preccurlyeq\cdot K'$ (K' weakly dominates K) if $K \cap S \leq K' \cap S$ for each state S in $State^{m}_{\alpha}$, and (2) $K \prec\cdot K'$ (K' strongly dominates K) if $K \preccurlyeq\cdot K'$ and, in addition, $K \cap S < K' \cap S$ for some state S in $State^{m}_{\alpha}$.
- *Optimal acts*. If $K \in Choice^{m}_{\alpha}$ is a possible act for α at m, and there is no $K' \in Choice^{m}_{\alpha}$ such that $K \prec\cdot K'$, then K is an *optimal act* for α at m.

To illustrate these ideas, imagine that in Fig. 6, the result *Heads* or *Tails* is determined by another agent, β. In this case, $State^{m}_{\alpha}$ is {*Heads*, *Tails*}, and (G & *Tails*) $<$ $\sim G$ $<$ (G & *Heads*) on the propositional ordering. Neither [α *cstit*: G] nor [α *cstit*: $\sim G$] is a dominated act for α, so both are optimal.

In simple cases where α has only finitely many possible choices, Horty's account of obligation is as follows:

(22) *Horty Dominance Ought* (*Finite Choice case*).

$M, m/h \models \odot$[α *cstit*: A] iff $M, m/h' \models A$ for all h' belonging to any choice K that is optimal at m.

That is, α ought to see to it that *A* iff every optimal choice *guarantees A*. In the case of *Gambler*, the Horty account tells us that neither $\odot$[α *cstit*: *G*] nor $\odot$[α *cstit*: $\sim G$] is true. In the absence of probabilistic information, gambling and not gambling are both permitted.

There may be situations where α has infinitely many options, none of which is optimal. For instance, if α and β are playing the *greatest integer game*, where the person who names the largest integer wins, then there is no optimal choice. In such cases, there are still *dominated* choices and hence there are still obligations—for instance, the obligation to choose an integer greater than 1,000. To accommodate such cases, Horty provides a more general evaluation rule.

(23) *Horty Dominance Ought* (*general case*).

> $M, m/h \models \odot[\alpha\ cstit\colon A]$ iff for each choice $K \in Choice^{m}_{\alpha}$ that does not guarantee *A*, there is a choice $K' \in Choice^{m}_{\alpha}$ such that (1) $K \prec\cdot K'$ (K' strongly dominates *K*), (2) $M, m/h' \models A$ for all h' belonging to K', and (3) for every choice $K'' \in Choice^{m}_{\alpha}$ such that $K' \preccurlyeq\cdot K''$, $M, m/h'' \models A$ for all h'' belonging to K''.

The requirement for $\odot$[α *cstit*: *A*] is that any action *K* that does not guarantee *A* is dominated by an action K' that does guarantee *A* and is either optimal or dominated only by other actions that guarantee *A*.

Suppose we modify *Gambler* so that the values are as shown in Fig. 7. Once again, we imagine that the result *Heads* or *Tails* is determined by the choice of another agent, β, so that the background states for α's choice are {*Heads*, *Tails*}.

As before, gambling is not always better than not gambling: the value of h_3 and h_4 exceeds that of h_5 and h_6. But this time, the act of gambling dominates the act of not gambling, so that $\odot$[α *cstit*: *G*] is true at *m*.[16]

As a slight modification of Horty's account, let us bring in the generalization of causal independence introduced in the *extended stit frames* of Sect. 3. Suppose that, in Fig. 7, the result *Heads* or *Tails* is determined by a chance mechanism (a coin toss) rather than by another agent's choice. Since Horty's causal background contexts only take into account choices by other agents, his account gives us no partition between *Heads* and *Tails*, no dominance of [α *cstit*: *G*] over [α *cstit*: $\sim G$], and hence no obligation to gamble. The extension of *AGENT* to *FACTOR*, as explained in Sect. 3, remedies this problem by counting chance mechanisms such as coin tosses as independent causal factors on par with choices. If we allow this extension, then $State^{m}_{\alpha} = \{Heads, Tails\}$ in Fig. 7, which restores the dominance reasoning that leads to $\odot$[α *cstit*: *G*]. Henceforth, $\preccurlyeq\cdot$ and $\odot$ will be understood as incorporating this extended concept of factors and background states. Other than the change to $State^{m}_{\alpha}$, there is no formal modification required for definitions (21), (22) and (23).[17]

---

[16] Technically, true at all *m*, *h* pairs. Since the semantics guarantees that $\odot$[α *cstit*: *G*] is either settled true or settled false at a moment, however, we may speak of obligations as holding at a *moment*.

[17] We could similarly define *extended utilitarian stit frames* by substituting *FACTOR* for *AGENT* and *Rng* for *Choice*.

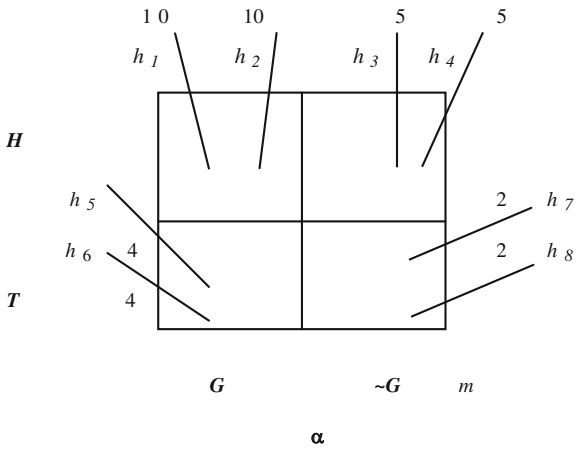


Fig. 7 Gambler III (utilitarian stit frame)

The distinctive feature of Horty's approach, in contrast to a great deal of earlier work in deontic logic, is that his semantics for obligation is based on an ordering on *choices*, rather than an ordering on *histories* or *worlds*. Horty's deontic logic gives us a weak decision theory, namely, the part of decision theory that corresponds to dominance reasoning. Let us say that an ordering $\preccurlyeq$ on choices K in $Choice^{\mathrm{m}}_{\alpha}$ is *admissible* if it extends Horty's dominance ordering: $K \preccurlyeq K'$ whenever $K \preccurlyeq\cdot\ K'$. Any decision principle based upon an admissible ordering on choices will preserve obligations that hold according to the *dominance ought*. But the reverse is not true: stronger decision principles justify assertions of obligation that fail on the Horty semantics. The remainder of this chapter shows how three of these stronger decision principles, and the corresponding notions of obligation, can be modeled by extensions within Horty's framework.

5 Decisions Under Ignorance: The Maximin Ought

In this section, I show how Horty's account might be extended to incorporate a principle that is sometimes used for making decisions under ignorance: *maximin*. The maximin rule tells the agent to compare *minimum* utilities possible for each available act, and to choose the act with the maximal minimum utility. The rationale behind this rule is *conservatism*: by following maximin, the agent guarantees the least bad outcome. In *Gambler* (I), for instance, the choice "Don't Gamble" guarantees a utility of 5, while "Gamble" allows possible outcomes with utilities of 0 and 10. Maximin thus prescribes the choice of not gambling. By contrast, as we have just seen, Horty's dominance ought prescribes nothing, since both gambling and not gambling are optimal choices.

There is an important ambiguity in the phrase "decisions under ignorance." Commonly, such decisions are characterized as those made "when it makes no sense to assign probabilities to the outcomes emanating from one or more of the acts" (Resnik 1987, 14). We can distinguish between cases of *total ignorance*, where the agent has no probabilistic information at all (not even knowledge of independence), and *ignorance of probabilities*, where the agent has no quantitative probabilistic information but does possess knowledge of causal independence. The latter case, in which the agent has exactly the same information as required for the dominance ought, is our focus. The objective is to provide a semantics for *maximin ought-to-do*, O_m, that strengthens Horty's dominance ought in the following sense:

(24) $\odot[\alpha\ cstit{:}\ A] \models O_m[\alpha\ cstit{:}\ A]$.

Both operators are formulated within utilitarian stit frameworks. The meaning of (24) is that for any utilitarian stit model M and for any m, h pair, if $M, m/h \models \odot[\alpha\ cstit{:}\ A]$ then $M,\ m/h \models O_m[\alpha\ cstit{:}\ A]$.

Unfortunately, our preliminary statement of the maximin rule is inconsistent with Horty's dominance ought. To see this, consider a kid-friendly version of *Gambler* that rewards a decision to gamble with \$10 if *Heads*, \$5 if *Tails*; a decision not to gamble yields \$5 regardless of outcome.[18] Gambling is plainly the dominant act. Yet the simple maximin rule regards gambling and not gambling as equally good because the worst outcome on either choice is \$5. This violation of dominance can be avoided by moving to a lexical version of maximin,[19] but an alternative approach will be offered below.

Another weakness of maximin as stated is its inability to handle a situation of infinite choices, such as the *greatest integer game* discussed in the preceding section. Even though no available act attains a maximal minimum value, it seems clear that maximin reasoning should license many of the same conclusions as dominance reasoning—for instance, that one ought to select an integer greater than 1,000.

We proceed in stages, starting with a new ordering on choices that *combines* maximin with the dominance ordering $\preccurlyeq\cdot$ defined in (21). The idea is to apply maximin only to pairwise comparisons where neither choice dominates the other.

(25) *Non-dominance*. If K and K' are members of $Choice^{\mathrm{m}}_{\alpha}$ (i.e., possible choices for α at m), write $K \not\lesssim K'$ if neither $K \preccurlyeq\cdot K'$ nor $K' \preccurlyeq\cdot K$.

(26) *Maximin ordering* ($\preccurlyeq_m$ and $\prec_m$) on $Choice^{\mathrm{m}}_{\alpha}$:

If K and K' are members of $Choice^{\mathrm{m}}_{\alpha}$ (i.e., possible choices for α at m), then

(1) $K \preccurlyeq_m K'$ if (i) $K \preccurlyeq\cdot K'$ or (ii) $K \not\lesssim K'$ and $inf\{Val(h)/h \in K\} \leq inf\{Val(h')/h' \in K'\}$; and

[18] Kid-friendly because the gambler never loses any money.

[19] See (Resnik 1987).

(2) $K \prec_m K'$ if (i) $K \prec \cdot K'$ or (ii) $K \nleqslant K'$ and $inf\{Val(h)/h \in K\} < inf\{Val(h')/h' \in K'\}$.[20]

(27) *Maximin ought*, O_m.

$M, m/h \models O_m[\alpha\ cstit{:}\ A]$ iff for each choice $K \in Choice^{\mathrm{m}}_{\alpha}$ that does not guarantee A, there is a choice $K' \in Choice^{\mathrm{m}}_{\alpha}$ such that (1) $K \prec_m K'$, (2) $M, m/h' \models A$ for all h' belonging to K', and (3) for every choice $K'' \in Choice^{\mathrm{m}}_{\alpha}$ such that $K' \preccurlyeq_m K''$, $M,\ m/h'' \models A$ for all h'' belonging to K''.

The relationship (24), that $\odot[\alpha\ cstit{:}\ A]$ entails $O_m[\alpha\ cstit{:}\ A]$, is clear because dominance is built into the definition of O_m. By way of example: in the kid-friendly version of *Gambler*, there is an obligation to gamble because gambling dominates not gambling. In the original version of *Gambler* (illustrated in Fig. 6) there is no dominant choice, but not gambling is superior to gambling on the maximin ordering; as a consequence, the agent has an obligation not to gamble ($O_m[\alpha\ cstit{:}\ \sim G]$). The same result holds when the result of *Heads* or *Tails* is achieved through placement of the coin by an independent agent. Finally, consider an infinite choice situation such as the *Greatest Integer Game*, where each possible choice of an integer is dominated by any choice of a larger integer. By (27), it is still true that one ought to choose an integer larger than 1,000.

The formulation of *maximin ought* in (27) has some advantages over the traditional formulation of *maximin* in decision theory. The first is its compatibility with dominance reasoning. The standard version of maximin, as noted earlier, does not always exclude dominated choices; the same problem applies to some forms of *lexical* maximin reasoning.[21] Other versions of lexical maximin, which respect dominance, are defined only for finite choice situations. By contrast, (27) is defined for arbitrary choice situations and is always compatible with dominance reasoning. A second advantage of the present formulation, indeed, is its ability to accommodate infinite choice situations, as noted in the preceding paragraph. In infinite choice situations where no individual choice is rational, we can still identify obligations. This highlights a general advantage of locating decision principles within deontic logic: whereas decision theory is focused specifically on rational *acts*, deontic logic provides truth conditions for all sentences of the form $O_m[\alpha\ cstit{:}\ A]$.

The point of this discussion is not to endorse the *maximin ought* over Horty's *dominance ought*. The weaknesses of maximin reasoning are well known.[22] There are two motives for developing O_m. The first is simply to flesh out the claim that the Horty semantics can be strengthened to yield a stronger decision theory. The second is that maximin reasoning plays an important role in game theory (Sect. 7).

[20] If S is a set of real numbers that is bounded below, then *inf* (S) refers to the *infimum* or greatest lower bound of S. Thus, I assume that the set of utility values within each K is bounded. The assumption of bounded utilities is standard in decision theory, to avoid problems such as the St. Petersburg paradox (see Resnik 1987, p. 107). Here, we require only the weaker assumption that utilities are bounded below within each possible choice.

[21] See (Resnik 1987).

[22] See (Resnik 1987) for discussion.

6 Decision Under Risk: Probabilistic Utilitarian stit Frames

This section extends Horty's account in a different direction by incorporating a simple type of probabilistic information – that which is related to chance mechanisms — into the semantics of obligation. This results in a strengthening of the dominance ought that is incompatible with maximin (just as expected utility reasoning is incompatible with maximin reasoning in decision theory). For simplicity, we ignore other agents; we have a single agent, α, making choices. We assume that α has finitely many choices and that there are only finitely many relevant independent causal factors, so both $Choice^{m}_{\alpha}$ and $State^{m}_{\alpha}$ are finite.

Let us begin with *Gambler* (I) as depicted in Figs. 4 (without utilities) and 6 (with utilities). Suppose that we have a coin toss with known probabilities 0.5 for *Heads* and *Tails*. In the theory of decision under risk, a straightforward application of expected utility reasoning yields a tie: gambling and not gambling have equal expected utility (EU = 5). A similar analysis yields a tie for *Gambler* (II) as depicted in Fig. 5, where the coin toss only occurs if α decides to gamble. But it is clear in these examples that slight changes to the utilities would tip the decision one way or the other. To extend Horty's account to such cases, we need to add some concepts to utilitarian *stit* frames. It suffices to add two additional concepts: *outcomes* and *conditional chances*.

We shall assume a finite set $O_1, \ldots, O_n$ of *outcomes* of interest. These are propositions at moment m (subsets of H_m) that constitute a partition of H_m and which, in conjunction with the background contexts and the agent's choices, influence the assignment of conditional chance and utility. In particular, they allow us to represent probabilistic information about chance processes initiated by agents; such processes cannot be treated as independent causal factors, as explained at the end of Sect. 3. In *Gambler*, the outcomes may be described as $\{Win, Lose, Neither\}$.

For the *conditional chance* function on H_m, the simplest approach is to take the underlying algebra[23] of subsets of H_m to consist of all finite unions of sets $K_i \wedge S_j \wedge O_k$, where $K_1, \ldots, K_m$ are available acts, $S_1, \ldots, S_M$ are the background contexts, and $O_1, \ldots, O_n$ are the outcomes. Probabilistic information about chance mechanisms is given by a conditional chance function P, assigning values $P(O_k/K_i \wedge S_j)$ that we take as primitive. P must satisfy the standard axioms of the probability calculus. Since the algebra is finite, P need only be finitely additive.[24]

The following two assumptions would allow easy extension of Horty's framework to handle probabilistic choices:

(a) *Uniform conditional chances.* $P(O_k/K_i \wedge S_j)$ is constant for each relevant outcome $O_1, \ldots, O_n$ within each $K_i \wedge S_j$.
(b) *Uniform utilities.* $Val(h) = Val(h')$ for all h, $h' \in K_i \wedge S_j \wedge O_k$, for all i, j, k.

[23] An algebra of subsets of X is a family $\mathscr{F}$ of subsets that includes X and the empty set, and is closed under finite unions, intersections and complementation.

[24] The function P may, but need not, assign unconditional chances to elements of the algebra, including α's own actions. Since P represents objective chance, difficulties alleged to exist for the assignment of subjective probabilities to one's current choices are not relevant; see (Levi 1997) and (Spohn 1977).

Given these assumptions, we can "import" decision theory into the *stit* framework. We can use expected utility maximization as the criterion for what agent α ought to do at *m*, in simple cases such as *Gambler* (I) and (II).

But these assumptions need not always hold. First, assumption (a) might fail. Some of the objective chances used in decision theory are not conditional chances associated with chance mechanisms. For example, they may be derive from observed frequencies. So there may be information about objective conditional chances that is not represented in the *stit* framework. Second, assumption (b) might fail. Within a single choice-state combination, we might find histories with different utilities based (for example) on future choices by agents or future events. In general: if it is impossible to find a set of outcomes satisfying (a) and (b), then it is impossible to apply straightforward expected utility maximization. To keep matters simple, however, I shall assume that condition (a) is satisfied but that (b) may fail.

This leads us to a definition of *probabilistic utilitarian stit frames*.

(28) A *probabilistic utilitarian stit frame* is a structure of the form.

$$< Tree, <, FACTOR, Rng, Value, Outcome, P >,$$

where $Tree$, $<$, *FACTOR*, *Rng* and *Value* are as in Sects. 2, 3, and 4, *Outcome* is a function that assigns to each moment *m* a partition $\{O_1, \ldots, O_n\}$ of H_m and $P(\cdot/\cdot)$ is a conditional probability function that assigns a value $P(O_k/K_i \wedge S_j)$ for each outcome O_k, choice K_i and state S_j.[25]

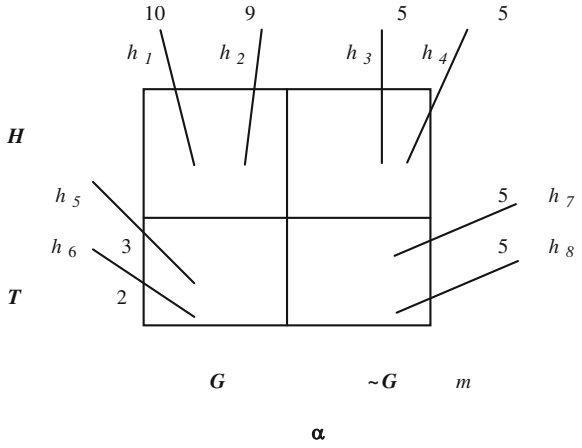
A *probabilistic utilitarian stit model* combines a probabilistic utilitarian stit frame with an assignment of truth values to propositions.

In order to formulate the concept of obligation in probabilistic utilitarian stit frames, consider the following case (Fig. 8). In this example, the coin toss is an independent factor and, as usual, *Heads* and *Tails* have fixed conditional chances of 0.5 regardless of whether α gambles. The outcomes are *Win*, *Lose* and *Neither*, but this time the utilities of *Win* and *Lose* are not fixed (i.e., assumption (b) fails). So there is no sharp value for the expected utility of gambling. Still, we can see that the expected utility of gambling is at least $(0.5)(9) + (0.5)(2) = 5.5$, which exceeds the expected utility of not gambling. Thus, we ought to gamble.

This motivates the following account of obligation, replacing dominance with *dominating expectation* in the Horty semantics. We continue to assume that both $Choice^{m}_{\alpha}$ and $State^{m}_{\alpha}$ are finite.

[25] One other assumption is necessary: the utilities given by *Value* (or *Val*) represent an interval scale (i.e., they are unique up to a positive linear transformation). This assumption guarantees that the expected utility ordering on choices, defined below, is invariant under allowable changes in representation of the utilities. Consider Fig. 8 below, which depicts utilities assigned by a particular function *Val*. The expected utility calculation given below, which shows that we ought to gamble, fails if the agent's utilities are equally well represented by a function Val' that assigns 7 to h_1, 6 to h_2, and keeps all other values the same as *Val*. Although *Val* and Val' agree on their ordinal ranking of histories, they are not related by a positive linear transformation. If $Val' = aVal + b$ for $a > 0$, however, then they induce the same ordering on choices.

Fig. 8 Gambler (IV)



(29) *Dominating expectation ordering* ($\preccurlyeq_d$ and $\prec_d$) on $Choice^{\mathrm{m}}_{\alpha}$.

If K and K' are members of $Choice^{\mathrm{m}}_{\alpha}$ (i.e., possible choices for α at m), then

(1) $K \preccurlyeq_d K'$ if for all choices h_{jk}, $h_{\mathrm{jk}}{}'$ of histories in $K \cap S_j \cap O_k$ and $K' \cap S_j \cap O_k$ respectively, $\Sigma_{\mathrm{j,k}} Val(h_{jk}) P(O_k/K \wedge S_j) \leq \Sigma_{\mathrm{j,k}} Val(h_{jk'}) P(O_k/K' \wedge S_j)$; and

(2) $K \prec_d K'$ if $K \preccurlyeq_d K'$ but not $K' \preccurlyeq_d K$.

This principle says that act K' is better than act K if the expectation of K' dominates the expectation of K.

Corresponding to this new ordering on choices, we have a new concept of obligation defined analogously to the earlier definitions (23) and (27). It states, roughly, that $O_d[\alpha\, cstit\colon A]$ if A is guaranteed by all choices whose expectation is not dominated.

(30) *Dominating expectation ought*, O_d.
$M, m/h \models O_d[\alpha\ cstit\colon A]$ iff for each choice $K \in Choice^{\mathrm{m}}_{\alpha}$ that does not guarantee A, there is a choice $K' \in Choice^{\mathrm{m}}_{\alpha}$ such that (1) $K \prec_d K'$, (2) $M, m/h' \models A$ for all h' belonging to K', and (3) for every choice $K'' \in Choice^{\mathrm{m}}_{\alpha}$ such that $K' \preccurlyeq_d K''$, $M,\ m/h'' \models A$ for all h'' belonging to K''.

It should be clear that O_d is a strengthening of Horty's $\odot$, since $K \preccurlyeq_d K'$ whenever $K \preccurlyeq\cdot K'$.[26] On the other hand, just as we would expect, O_d diverges from the *maximin ought* O_m. In the version of *Gambler* depicted by Fig. 8, we have $O_d[\alpha\, cstit\colon G]$ while $O_m[\alpha\, cstit\colon \sim G]$.

[26] Proof: if $K \preccurlyeq\cdot K'$, then for any state S_{j}, $Val(h) \leq Val(h')$ for each $h \in K \cap S_{\mathrm{j}}$ and $h' \in K' \cap S_j$. Then $\Sigma_{\mathrm{k}} Val(h_{jk}) P(O_k/K \wedge S_j) \leq \Sigma_{\mathrm{k}} Val(h_{jk'}) P(O_k/K' \wedge S_j)$ for all choices $h_{\mathrm{jk}} \in K \cap S_j \cap O_k$ and $h_{jk'} \in K' \cap S_j \cap O_k$, and the result follows.

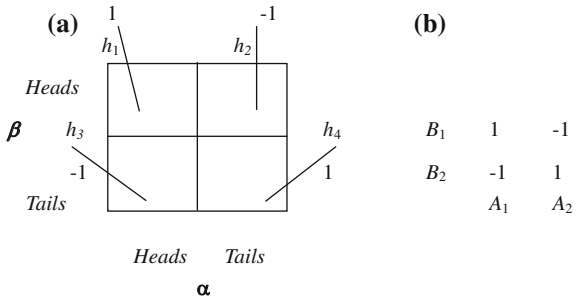


Fig. 9 Matching pennies **(a)** stit picture; **(b)** game theory

7 Game Theory and Mixed Strategies

Finally, I consider briefly how Horty's account might be extended to handle obligations in the setting of game theory. To do this in general would introduce many complications, including the need for separate *Value* functions to keep track of each agent's utilities.[27] My interest here lies mainly in showing how we might use the probabilistic utilitarian stit frames of the preceding section to make sense of *mixed strategies* in game theory. For this reason, I limit the discussion to *two-person zero-sum games*. The *Value* function represents α's utilities, while utilities for the other agent, β, are exactly the negative of α's utilities. We initially assume that finitely many choices—*pure strategies* in game theory—are available to both agents, and that there are no additional independent causal factors. Thus, the background contexts in $State^m_\alpha$ are simply β's possible choices.

By way of motivation, notice that a very simple game, *Matching Pennies* (introduced at the outset of this chapter), generates difficulties for Horty's account. In this game, both α and β simultaneously display a coin with one side up. If the two displayed sides match (both *Heads* or both *Tails*), then β pays \$1 to α; if the sides do not match, then α pays \$1 to β. The situation is illustrated in Fig. 9, with the *stit* and game-theoretic representations side by side.

In this situation, neither choice by α is optimal. Consequently, on Horty's account, we then have neither $\odot[\alpha\, cstit : Heads]$ nor $\odot[\alpha\, cstit: Tails]$. That is reasonable if the only choices available are the pure strategies [*Display*] *Heads* or [*Display*]*Tails*. But Horty's conclusion is not plausible if we allow mixed strategies of the form

Display Heads with probability *p* and *Display Tails* with probability 1 – *p*, abbreviated as

$$[p\, Heads, (1 - p)\, Tails]$$

or more simply as

[27] See Kooi and Tamminga (2008) for an account developed along these lines.

B_1	4	2
B_2	5	6
	A_1	A_2

Fig. 10 A zero-sum game

$$p\,Heads.$$

In game theory, this type of problem is solved by finding a *Nash equilibrium*: a pair of choices such that neither player can do better by unilaterally changing his or her choice. In *Matching Pennies*, there is a unique Nash equilibrium where both agents adopt the mixed strategy: 1/2 *Heads*. This is the unique rational choice on the assumption that each player has full knowledge of the game and adopts the best possible strategy. In the remainder of this section, we suggest one way in which Horty's account can be expanded to accommodate mixed strategies, and then propose a semantics that yields the obligation to adopt an equilibrium strategy.

But first let's consider a preliminary question. How well does Horty's account fare if we limit ourselves to two-person zero-sum games with only pure strategies? Consider the following game (Fig. 10), with utilities for α shown.

Here, α chooses between the left (A_1) and right (A_2) columns, while β chooses between the top (B_1) and bottom (B_2) rows. From α's point of view, neither choice is dominant: A_1 does better if β chooses B_1, while A_2 does better if β chooses B_2. So Horty's "dominance ought" yields no obligation: neither $\odot[\alpha\, cstit\colon A_1]$ nor $\odot[\alpha\, cstit\colon A_2]$ is true. However, both players will recognize that the top row is the dominant choice for β (whose utilities are the negative of those in Fig. 10). Given that β will choose B_1, α ought to choose A_1 (Indeed, A_1 and B_1 constitute the unique Nash equilibrium for this game). This example shows that even without mixed strategies, Horty's dominance ought is inadequate for game theory.

One promising possibility might be the *maximin ought* (O_m) of Sect. 5, which combines dominance with *maximin* reasoning. Applied to Fig. 10, O_m gives the correct result: A_1 guarantees the maximal minimum, so $O_m[\alpha\, cstit\colon A_1]$. Further encouragement comes from a standard result of game theory (Resnik 1987, p. 130):

> *Minimax equilibrium test*. In a two-person zero-sum game, a necessary and sufficient condition for a pair of (pure) strategies to be in (Nash) equilibrium is that the payoff determined by them equal the minimal value of its (column) and the maximal value of its (row).[28] The values for all such equilibrium pairs are the same.

It is easy to establish the following proposition:

[28] Most expositions of game theory represent the utilities of the *Row* player in zero-sum games. Following *stit* conventions, I represent instead the utilities of the column player. The statement of the *Minimax equilibrium test* in (Resnik 1987) thus reverses *row* and *column*.

B_1	2	1
B_2	0	5
	A_1	A_2

Fig. 11 Another zero-sum game

(31) *Proposition*: If there is a pure-strategy equilibrium pair in a two-person zero-sum game and if A is true at all non-dominated choices K for α that belong to such an equilibrium pair, then $O_m[\alpha\, cstit\colon A]$.

Proof:

For each such K, $K' \preccurlyeq_m K$ for all choices K'; thus, these K are *optimal* with respect to the ordering $\preccurlyeq_m$ and we have $O_m[\alpha\, cstit\colon A]$.

(31) shows that whenever there is a pure Nash equilibrium, the *maximin ought* correctly prescribes choices that are part of such an equilibrium.

Despite this success, *maximin ought* appears to overshoot the mark. It prescribes choices even when there is no pure equilibrium. Consider the following example (Fig. 11). Here there are no dominant choices for either player and no equilibrium. Nevertheless, *maximin ought* prescribes A_2 for α and B_1 for β, since these choices maximize minimal utility.

It might be interesting to consider whether there is any merit to these prescriptions. It might also be worth investigating whether there is a notion of obligation, intermediate in strength between $\odot$ and O_m, that corresponds precisely to acts that comprise a Nash equilibrium. I pass over such investigations for the following reason: once we allow *mixed strategies*, the problem of capturing Nash equilibria with a Horty-style account of obligation is solved through an interesting combination of *maximin* reasoning and the weak concept of expected utility introduced in Sect. 6.

The first task is to give an analysis of mixed strategies. In game theory, a mixed strategy is commonly characterized as the use of a chance mechanism to select a pure strategy, followed by acting on the selected strategy. The details may not matter much in game theory, but they matter a great deal in the *stit* framework. If the chance mechanism operates at a moment prior to the choice of the pure strategy, then the analysis of a mixed strategy will involve both the prior moment when the mechanism operates and alternative later moments at which the agent chooses a pure strategy. To make things worse, the *stit* picture for the later moment will be identical to the original 'pure strategy' picture. If we evaluate the obligation at that later moment, it is unclear how the earlier operation of a chance mechanism can make any difference.

Perhaps the simplest approach, and the one which will be adopted here, is to represent each available mixed strategy as a separate choice existing at the same moment as the pure strategies. It is the choice of a chance mechanism whose possible outcomes are identical in structure with the pure strategies. Strictly speaking, this

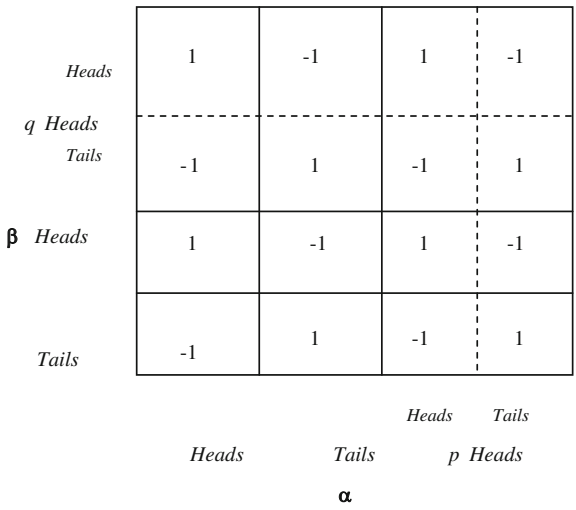


Fig. 12 Matching pennies with mixed strategies

analysis requires that we modify $Choice^{m}_{\alpha}$ by adding one additional choice for each available chance distribution over the (finitely many) pure strategies. In practice, it usually suffices to represent all of the pure strategies plus a single choice that stands for an arbitrary mixed strategy (incorporating probabilistic parameters) or, on occasion, for a particular mixed strategy. Figure 12 illustrates *Matching Pennies* with mixed strategies. Dotted lines are used to separate outcomes for the case of choices that involve chance mechanisms.

None of the concepts of obligation described above gives the correct result here, namely, the obligation to choose 1/2 *Heads*. According to the *dominance ought*, there are no obligations. The same is true for the *maximin ought*, since each choice has the same worst case. The *dominating expectation ought* is not even defined for settings involving multiple agents.

A helpful way to obtain the right choice ordering and the right concept of obligation is to exploit a well-known result from game theory (Resnik 1987, p. 136):

> *Maximin Theorem for two-person zero-sum games*.
>
> For every two-person zero-sum game there is at least one strategy (mixed or pure) for Row and at least one strategy for Col that form an equilibrium pair. If there is more than one such pair, their expected utilities are equal.

The expected utility for the equilibrium pair is referred to as the *security level* for both players because, by playing the equilibrium strategy, each player maximizes his or her minimum expected utility. The security level in *Matching Pennies* is 1/2, which can be guaranteed by playing 1/2 *Heads*. In contrast to our earlier discussion of zero-sum games with pure strategies, the inclusion of mixed strategies ensures the existence of an equilibrium.

The right ordering, then, is that one mixed strategy is preferable to another if its minimal expected utility exceeds that of the other. This ordering can be defined

within the setting of *probabilistic utilitarian stit frames*. Write P_α and P_β for α's and β's choice of mixed strategy. P_α and P_β are probability distributions over the choices available to α and β, respectively. That is, if $K_1, \ldots, K_m$ are the available pure strategies for α, i.e., the members of $Choice^{\mathrm{m}}_\alpha$, then $P_\alpha(K_i) = p_i$ with $\Sigma p_i = 1$; similarly, $P_\beta(B_j) = q_j$ for pure strategies $B_1, \ldots, B_n$. The choice of a pure strategy K_i or B_j is just the special case where $p_i = 1$ or $q_j = 1$. Let $^*Choice^{\mathrm{m}}_\alpha$ be the set of mixed strategies P_α based on the pure strategies in $Choice^{\mathrm{m}}_\alpha$.

(32) *Equilibrium ordering* ($\preccurlyeq_e$ and $\prec_e$) on $^*Choice^{\mathrm{m}}_\alpha$.

If P_α and P'_α are members of $^*Choice^{\mathrm{m}}_\alpha$ (i.e., mixed strategies for α at m), then

(1) $P_\alpha \preccurlyeq_e P'_\alpha$ if
$\inf\{\Sigma_{\mathrm{i,j}} Val(h_{ij}) P_\alpha(K_i) P_\beta(B_j) / h_{ij} \in K_i \cap B_j$ and P_β a mixed strategy for $\beta\} \leq \inf\{\Sigma_{\mathrm{i,j}} Val(h_{ij}) P'_\alpha(K_i) P_\beta(B_j) / h_{ij} \in K_i \cap B_j$ and P_β a mixed strategy for $\beta\}$;

and

(2) $P_\alpha \prec_e P'_\alpha$ if $P_\alpha \preccurlyeq_e P'_\alpha$ but not $P'_\alpha \preccurlyeq_e P_\alpha$.

The mixed strategy P'_α is better than P_α if it has greater minimal expected utility. The ordering $\preccurlyeq_e$ is admissible in the following special sense: if $P_\alpha \preccurlyeq\cdot P'_\alpha$ where both P_α and P'_α are *pure strategies*, then $P_\alpha \preccurlyeq_e P'_\alpha$.[29]

(33) *Equilibrium ought*, O_e.
$M, m/h \models O_e[\alpha\, cstit\colon A]$ iff for each $P_\alpha \in {}^*Choice^{\mathrm{m}}_\alpha$ that does not guarantee A, there is a $P'_\alpha \in {}^*Choice^{\mathrm{m}}_\alpha$ such that (1) $P_\alpha \prec_e P'_\alpha$, (2) $M, m/h' \models A$ for all h' belonging to P'_α, and (3) for every $P''_\alpha \in {}^*Choice^{\mathrm{m}}_\alpha$ such that $P'_\alpha \preccurlyeq_e P''_\alpha$, $M, m/h'' \models A$ for all h'' belonging to P''_α.

In two-person zero-sum games where an equilibrium exists, (33) states that $O_e[\alpha\, cstit\colon A]$ if A is guaranteed by all equilibrium mixed strategies. O_e is a strengthening of Horty's ⊙, and it gives the right answer in the case of *Matching Pennies*: $O_e[\alpha\, cstit\colon P_{1/2}]$.

Summarizing: mixed strategies can be defined in Horty's framework, and we can give an ordering on mixed strategies that yields the correct account of what agents ought to do in two-person zero-sum games. Extending these ideas to games involving more than two agents and to cooperative games may or may not be feasible.

8 Conclusion

Horty observes that his account of obligation "closes the gap" between deontic logic and act utilitarianism. That gap existed so long as deontic logic was viewed as an

[29] Proof: $P_\alpha(K_i) = 1$ for some i, and $P'_\alpha(K_{i'}) = 1$ for some i′. If $P_\alpha \preccurlyeq\cdot P'_\alpha$, then $Val(h_{\mathrm{ij}}) \leq Val(h_{\mathrm{i'j}})$ for any h_{ij} in $K_i \cap B_j$ and $h_{\mathrm{i'j}}$ in $K_{i'} \cap B_{\mathrm{j}}$. From this it follows that $P_\alpha \preccurlyeq_e P'_\alpha$.

account of classifying *states of affairs* as right or wrong, while utilitarianism was concerned with classifying *actions*. Horty's *dominance ought* clearly goes a long way towards closing another gap as well: the one between deontic logic and decision theory.

Because of the weakness of dominance reasoning, however, Horty's account seems of limited value as a theory of choice. This chapter suggests how, with modest extensions, Horty's framework can move beyond dominance into the three main branches of the theory of decision: decisions under ignorance, decisions under risk and game theory. This leads to a motivational question: what is the point of trying to bring deontic logic "up to speed" if we already have a successful decision theory? I close by suggesting two main ways in which deontic logic provides return benefits for decision theory.

The first, noted at the outset of this chapter, is by offering rigorous analysis of foundational notions: causation, choice, counterfactuals and background states. That such analyses matter should be clear to anyone who has followed the history of decision theory as formulated by Savage, modified by Jeffrey, and re-formulated by causal decision theorists. For example, we claimed here that the *states* of decision theory are causal background contexts and provided an analysis of causal independence and background contexts within *stit* models. By contrast, Joyce (1999, 61) writes that *states* include all "aspects of the world that lie outside the decision maker's control". He tells us that future choices and events, if relevant to our present decision problem, must be incorporated into the background states for that decision. Now it is harmless to incorporate future choices and events into the background states if they are causally independent of the agent's present choice, but not so harmless if their future occurrence is contingent upon present choices. *Stit* frames take care of this automatically: histories belonging to distinct states at m must be *divided* at m. This rules out treating future choices or processes as constituents of states at m. Future chance processes must be incorporated into decisions via conditional chances for outcomes (as described in Sect. 6). To handle sequential choices in the stit framework requires something like Horty's *strategic ought* (2001, Chap. 7), which takes us beyond the present discussion.

As a second benefit, deontic logic offers a model for thinking about problems where decision theory and game theory cannot offer clear, uncontroversial solutions. One source of such problems is *infinite decision theory*, comprising decision problems in which an agent has to deal with infinite utilities, an infinity of possible acts, or both.[30] Some of these problems are genuinely paradoxical and have no clear solution. In other cases, however, there are clear prescriptions, yet decision theory is silent because there is no optimal act. Because deontic logic is concerned with the truth of obligation sentences $O[\alpha\, cstit\colon A]$ even where A does not describe an act, it has the resources to offer advice in such cases.

One example of this kind, noted earlier, is the *greatest integer game*. Decision theory cannot recommend the choice of any particular integer, but our deontic logics tell us that $\odot[\alpha\, cstit\colon A_n]$ and $O_m[\alpha\, cstit\colon A_n]$ where A_n is the proposition "α

[30] See (Sorensen 1994) for examples.

chooses an integer larger than n". As a similar example, imagine that α is a perfectionist attempting to finish a journal article. Suppose that α represents his position to himself as an infinite choice situation where A_0 stands for not submitting the chapter at all, A_1 for submitting the current version as is, and A_2, A_3, ... for producing and submitting polished versions, each A_{n+1} slightly better than A_n. Suppose that all of the relevant utilities are bounded above by some fixed limit. In such a case, no act A_n is optimal. But our deontic logics still give us $\odot[\alpha\, cstit\colon \sim A_0]$ and $O_m[\alpha\, cstit\colon \sim A_0]$, representing the obligation to submit the chapter.

Decision theory need not always be concerned about the metaphysical details of choice, or the precise characterization of the acts and background states needed to specify a decision problem. But at the boundaries of decision theory, where those details matter, *stit*-based deontic logic, made possible thanks to Belnap's rigorous analysis of agency, provides a wonderful resource.

Acknowledgments I acknowledge support from the Australian National University, in the form of a Visiting Research Fellowship.

References

Belnap, N. 1992. Branching space-time. *Synthese* 92(3): 385–434.

Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future: agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Broersen, J. 2011. Modeling attempt and action failure in probabilistic stit logic. Proceedings of the twenty-second international joint conference on, artificial intelligence, 792–8.

Chellas, B. 1992. Time and modality in the logic of agency. *Studia Logica* 51: 485–517.

Gibbard, A., and W.L. Harper. 1978. Counterfactuals and two kinds of expected utility. In Foundations and applications of decision theory, of western ontario series in philosophy of science, eds. C.A. Hooker, J.J. leach and E.F. McClennen, vol. 13, 125–162. Dordrecht: D. Reidel Publishing.

Gillies, D. 2000. *Philosophical theories of probability*. New York: Routledge.

Joyce, J. 1999. *The foundations of causal decision theory*. Cambridge: Cambridge University Press.

Horty, J.F. 2001. *Agency and deontic logic*. Oxford: Oxford University Press.

Kooi, B., and A. Tamminga. 2008. Moral conflicts between groups of agents. *Journal of Philosophical Logic* 37: 1–21.

Levi, I. 1997. *The covenant of reason: rationality and the commitments of thought*. Cambridge: Cambridge University Press.

Resnik, M. 1987. *Choices*. Minneapolis: University of Minnesota Press.

Skyrms, B. 1980. *Causal necessity: a pragmatic investigation of the necessity of laws*. New Haven: Yale University Press.

Skyrms, B. 1994. Adams conditionals. In *Probability and conditionals: belief revision and rational decision*, ed. E. Eells, and B. Skyrms. Cambridge: Cambridge University Press.

Sorensen, R. 1994. Infinite decision theory. In *Gambling on God: essays on Pascal's wager*, ed. J. Jordan. Savage: Rowman and Littlefield.

Spohn, W. 1977. Where Luce and Krantz do really generalize Savage's decision model. *Erkenntnis* 11: 113–134.

Internalizing Case-Relative Truth in CIFOL+

Nuel Belnap

Abstract CIFOL is defined in Belnap and Müller 2013 (*J Phil Logic* 2013) as the first-order fragment of Aldo Bressan's higher-order modal typed calculus MC^{ν}. Bressan based his calculus on Carnap's "method of extension and intension": In CIFOL, truth is relative to "cases," where cases play the formal role of "worlds" (but with less pretension). CIFOL+ results by following Bressan in adding term-constants **t** for the true and **f** for the false, and a single predicate constant, P_0, which together with a couple of simple axioms enable the representation of "sentence Φ is true in case x" by means of a defined expression, $T(\Phi, x)$, where Φ is the sentence of CIFOL+ in question and where x ranges over a defined family of "elementary cases." (Whereas being a *case* is defined in the semantic metalanguage, *elementary cases* are squarely in the (first order) domain of CIFOL+.) A suitable suite of axioms guarantees that one can prove (in CIFOL+) that there is exactly one elementary case, x, such that x happens (i.e., such that $x = \mathbf{t}$), a fact that underlies the equivalence of $\Box(x = \mathbf{t} \rightarrow \Phi)$ and $\Diamond(x = \mathbf{t} \ \wedge \ \Phi)$. (Proofs are surprisingly intricate for first order modal logic.) One can then go on to show that $T(\Phi, x)$ is well-behaved in terms of its relation to the connectives of CIFOL+, a result required for ensuring that $T(\Phi, x)$ is properly read as "that Φ is true in elementary case x."

1 Introduction

Belnap and Müller 2013 (BM2013) defined and discussed the first-order fragment, CIFOL,[1] of Aldo Bressan's splendid but little-known higher-order modal typed calculus MC^{ν} (Bressan 1972).[2] Since higher-order type theories are intrinsically pow-

[1] "CIFOL" is an acronym for "case-intensional first order logic." Thanks to Thomas Müller for his help with this chapter.

[2] Bressan's logic is rooted in "the method of intension and extension" due to Carnap 1947.

N. Belnap (✉)
1001 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.
e-mail: belnap@pitt.edu

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_3,

erful, it is perhaps not surprising that MC^{ν} contains a truth concept for itself. CIFOL by design is "case-intensional," meaning that the basic truth concept is "true in a case," which plays the same role as "true according to a world" or the like plays in the most common quantified modal logics. One might not expect, however, that a certain modest *first-order* extension of CIFOL, which we will call CIFOL+, can contain a powerful kind of truth concept for itself. It is this surprising result, which is based on Bressan 1972, that we present here. (The proof is the most intricate of any of which we know of a theorem *in*—rather than a theorem *about*—quantified modal logic.) We say a little about CIFOL in this introductory section, but by and large we presuppose acquaintance with BM2013.

1.1 Grammar and Semantics

Grammatically, CIFOL is the first order quantified modal logic with identity detailed in BM2013. (CIFOL+ results from adding two more axioms to CIFOL, as we see in Sects. 1.4 and 2.1.) Its proof theory is largely—but not entirely—a simple combination of **S5** with first order predicate calculus with identity (with predication intensional, but with replacement of identicals restricted to extensional contexts), conservatively extended by both definite descriptions, $\imath x \Phi$, and lambda abstracts, $\lambda x \Phi$, governed by transparent principles.[3] (This chapter uses both the $\imath$ and the λ of CIFOL, but we also use λ metalinguistically.) It is the semantics of CIFOL that sets it apart. CIFOL is a "case-intensional" logic, meaning the following. There is a set of *cases*, Γ, which is formally like a set of "worlds" in the jargon of much contemporary modal logic. Following Carnap (1947), each expression, ξ, of each type, be it individual expression α (whether constant c, variable x, or complex, predicate constant P, operator constant f, or sentence Φ) has an *extension* in each case, $\gamma \in \Gamma$, written $ext_{\gamma}(\xi)$; in particular, truth of sentences is case-relative. Furthermore, each expression, ξ, has an *intension*, written $int(\xi)$, which is not something extra, but is explicable as the pattern of its extensions $ext_{\gamma}(\xi)$, as γ varies over Γ.

Fact 1 (Intension-extension connection) *Using lambda-abstraction, we may say, where γ ranges over the set of cases, Γ, that $int(\xi) = \lambda\gamma(ext_{\gamma}(\xi))$, and that $ext_{\gamma}(\xi) = (int(\xi))(\gamma)$.*

There is an "individual domain," D, which harbors all the possible extensions of singular terms, α, and there is a parallel "sentential domain," $\mathbf{2} = \{\mathbf{T}, \mathbf{F}\}$, containing the standard truth values to serve as the extensions of sentences. Where $X \mapsto Y$ is the set of functions from X into Y, an *intension* is always a function in $(\Gamma \mapsto Y)$, for appropriate Y. A CIFOL *interpretation*, $\mathcal{I}$, endows each atomic expression with an intension of the appropriate type.[4] Then recursive clauses come along to guarantee

[3] CIFOL includes the Barcan permutation of possibility and the existential quantifier.

[4] Significantly, CIFOL adopts Bressan's interpretation of predicate constants, which renders predication "intensional," in dramatic contrast to other quantified modal logics. Intensional predication

that singular terms α, sentences Φ, operators f, and predicates P each have the right type of extension and intension. Along with sentences and singular terms, we illustrate only one-place operators and predicates.

Fact 2 (Types of intensions and extensions)
$ext_\gamma(\alpha) \in D$.
$int(\alpha) \in (\Gamma \mapsto D)$.
$ext_\gamma(\Phi) \in \mathbf{2}$.
$int(\Phi) \in (\Gamma \mapsto \mathbf{2})$.
$ext_\gamma(P) \in ((\Gamma \mapsto D) \mapsto \mathbf{2})$.
$int(P) = \mathcal{I}(P) \in (\Gamma \mapsto ((\Gamma \mapsto D) \mapsto \mathbf{2}))$.
$ext_\gamma(f) \in ((\Gamma \mapsto D) \mapsto D)$.
$int(f) = \mathcal{I}(f) \in (\Gamma \mapsto ((\Gamma \mapsto D) \mapsto D))$.

Identity is an important special case: Unlike predication, its semantics is extensional, so that the truth value of $\alpha = \beta$ in γ depends only on the extensions in case γ of α and of β.

CIFOL invokes δ as an assignment of intensional values (that is, values in $\Gamma \mapsto D$) to the individual variables, so that free individual variables and individual constants have exactly the same semantics. Then BM2013 explains a CIFOL "model" as a triple $\mathcal{M} = \langle \Gamma, D, \mathcal{I} \rangle$.[5] There is a special constant, $*$, whose case-relative extensions in D are also called $*$. In Frege's way, $*$ helps process definite descriptions that do not satisfy the standard unique-existence clause. It is assumed that D contains something other than $*$. Among modal logics, the hallmark of CIFOL is that truth of sentences is defined relative to "cases," which are a common generalization of worlds, times, and so on. When $\mathcal{M}$ is understood, the fundamental semantic (metalinguistic) truth-locution has the form "Φ is true in case $\gamma \in \Gamma$," with Φ denoting a sentence, and corresponding to the more familiar phrase, "Φ is true in world w." We sometimes use $\gamma \models \Phi$ to say that Φ is true in case γ. For example, with reference to a certain model, $\mathcal{M}$, the semantic clause for necessity proclaims that $\gamma \models \Box\Phi$ just in case $\gamma' \models \Phi$ for all $\gamma' \in \Gamma$.

1.2 Finding "True in a Case" in CIFOL+

What we seek is a sentence of CIFOL itself which can reasonably be read in English as "that Φ is true in case x," with Φ taking the place of (rather than denoting) a sentence; CIFOL is not, however, adequate in this respect. We repair the inadequacy in four steps. (1) We enrich CIFOL with two new axioms, labeling the result "CIFOL+."

(Footnote 4 continued)
lies behind the unusual power of CIFOL. Montague (1973) does not feature intensional predication in the present sense; however, by a somewhat more complicated device, Montague attains the same end, rendering his system as powerful as Bressan's.

[5] We have suppressed the parameter δ, since it isn't needed in this chapter.

(2) We take Φ as *being* rather than *denoting* a CIFOL+ sentence, and we take x as an individual variable that one can interpret as ranging over an "internal" representation of the set, Γ, of all the cases in a certain model $\mathcal{M}$. (3) We formulate *within* CIFOL+ an "internal" concept of cases, which intuitively are in one-one correspondence with the set Γ of "external" cases. Theorem 1 testifies that we have succeeded in this endeavor. Finally, (4) we define *within* CIFOL+ a rich (but paradox-free) concept of truth via a locution having the force of "that Φ is true in case x."[6]

1.3 Paths not Taken

We pause to contrast our path with nearby paths that we do not take. We are after defining what Curry calls a "mixed nector," the English "that Φ is true in case x," to be written $T(\Phi, x)$, where the character Φ takes the place of a sentence and where x takes the place of an individual variable ranging over internal cases. The comparable Tarskian goal would be to define a locution having the form, "s is true in case x," written $T(s, x)$, where s denotes a sentence and T is a genuine predicate. (So s *denotes* a sentence, whereas Φ *is* a sentence.) Consider the non-case-relative truth predicate for a moment: If one wished to exhibit a Tarskian form with Φ, one would need to write $T(\text{`}\Phi\text{'})$ rather than $T(\Phi)$, so that the grammatical argument of T would be the name of a sentence rather than a sentence. If one wrote $T(\Phi)$, T would be a connective; and given a Tarski-like schema $T(\Phi) \leftrightarrow \Phi$, T would have to be the trivial identity connective. The contrast is that given case-relativity, $T(\Phi, x)$ is by no means trivial; we shall have to engage in honest toil to find a schema that will serve. We might think merely to add the "true in" form to CIFOL, but that would rightly be judged as theft.

1.4 Extending CIFOL

Suppose we are looking for a truth predicate for some language, L. Tarski explained that a language with a truth predicate for L must of necessity be stronger than L. What about a true-in schema for L? No such general result is available: It is obvious that there are languages with case-dependent semantics that can consistently contain their own true-in schema. But what about CIFOL in particular? It would seem—but I don't offer a proof—that CIFOL, as it stands, does not permit an appropriate definition of "true in" for CIFOL. What is much more important, however, is that the modest extension of CIFOL to CIFOL+ by an extra pair of first order principles does permit the definition of a "true in" schema.

[6] Note that we avoid paradox by introducing a truth concept without ascending to a metalanguage. Theorem 2, our final result, serves this purpose.

To begin, we add to CIFOL a pair of individual (not sentential) constants, **t**, **f**, that can be used to tag sentences true in a case with **t** and sentences false in a case with **f**.[7] We can ensure that **t** and **f** do their jobs properly by postulating that they are necessarily distinct, and that some intension is possibly (extensionally) equal to each:

Axiom 1 $(\mathbf{tf}) \vdash \Box(\boldsymbol{t} \neq \boldsymbol{f} \ \wedge \ \exists x[\Diamond(x = \boldsymbol{t}) \ \wedge \ \Diamond(x = \boldsymbol{f})])$

Evidently there must be at least two cases, a low-grade fact indeed. It is noteworthy that Axiom 1 is the only information that we have concerning **t** and **f**; we have no information about their extensions in any case, γ, except that their extensions are not the same. In particular, there is no assumption that **t** and **f** are "rigid designators."

1.5 Picturing Intensions

In picturing the intensions of **t** and **f**, however, it is helpful *to imagine* first, that in every model, **t** and **f** are not only individual constants, but are also members of the extensional domain, D, and second, that in every case, $\gamma \in \Gamma$, **t** has itself as its own extension, and likewise for **f**. So in our imagination, **t** and **f** each has a constant extension, namely, the symbol itself in every case.[8] As a further mental prop, we imagine that the set of cases, Γ, comes as a sequence, $\gamma_1, \gamma_2, \ldots, \gamma_i, \ldots$, so that

$$int(\mathbf{t}) = \mathbf{tt}\ldots, \mathbf{t}, \ldots$$

and

$$int(\mathbf{f}) = \mathbf{ff}\ldots, \mathbf{f}, \ldots \ .$$

Of course, the intent of the pictures is not to limit the cardinality or structure of Γ in any way. We may then picture the intension of a sentence, Φ, as a sequence of occurrences of **t** and **f**, with the former marking those cases in which $\gamma \models \Phi$, and the latter marking the cases in which $\gamma \not\models \Phi$. Suppose, for example, that Φ is true in the odd cases and not true in the even cases. Recalling that in his semantics, Carnap defined "the range of Φ" as the set of cases[9] in which Φ is true, the intension

$$int(x) = \mathbf{tftftf}, \ldots$$

could be a first-order (intensional) representation of the range of some Φ.

---

[7] In fact Bressan presses into service the arithmetic constants 0 and 1, which are borrowed from an appropriately higher type at which they can be proved necessarily distinct. In order to remain first order, we will postulate rather than prove.

[8] Each of **t** and **f** is thereby imagined as ferociously autonymic.

[9] Not, of course, Carnap's word.

2 Theory of Internal Ranges

In order to find an internal representation of "that Φ is true in case γ," it is necessary to find an internal representation of γ. A natural thing to try is to represent γ by an intension, that is, by a function in $\Gamma \mapsto D$, that picks out γ uniquely. Standard set-theory tells us that a function $f \in (\Gamma \mapsto D)$, such that $f(\gamma) = ext_{\gamma}(\mathbf{t})$, while $f(\gamma') = ext_{\gamma'}(\mathbf{f})$ for every $\gamma' \in \Gamma$ other than γ, would adequately fill that bill. We just need a way of saying this in CIFOL. How can we describe x such that $int(x) \in \Gamma \mapsto D$ in such a way that $int(x)$ picks out a unique case γ? First, we want $int(x)$ to be pictured as having in each case $\gamma \in \Gamma$ either $\mathbf{t}$ or $\mathbf{f}$.[10] Let us describe such an x as a "range," since it does the work of a Carnapian range. In the language of CIFOL, we can carry "in each case" by the necessity modality, and "in some case" by the possibility modality. We make a definition of "proper range" that includes the requirement that the picture of x contains at least one $\mathbf{t}$ and at least one $\mathbf{f}$:

Definition 1 (Proper range, *PR*)

$$\forall x[PR(x) \leftrightarrow_{df} (\Box(x = \boldsymbol{t} \ \vee \ x = \boldsymbol{f}) \ \wedge \ \Diamond(x = \boldsymbol{t}) \ \wedge \ \Diamond(x = \boldsymbol{f}))].$$

Of course "=" in Definition 1 is extensional identity, telling us only that in each case either $\gamma \models x = \mathbf{t}$ or $\gamma \models x = \mathbf{f}$, but that is quite enough for our purpose.

Second, we want a way of saying that there is one and only one case γ such that $\gamma \models x = \mathbf{t}$. (This cannot be said by directly using the "exactly one x" quantifier; it is cases that need counting, not intensions and not extensions.) "In at least one case" is easy: $\Diamond(x = \mathbf{t})$, and we have already included it as part of the definition of *PR*. To say "there is at most one case" is a bit more work. Begin by finding a way to say that one range, x, is a "subrange" of another range, y, meaning that in every case in which $x = \mathbf{t}$ is true, so also is $y = \mathbf{t}$; but not necessarily conversely (the picture is easily imagined).

Definition 2 (Subrange, *SubR*)

$$\forall x \forall y[SubR(x, y) \leftrightarrow_{df} (PR(x) \ \wedge \ PR(y) \ \wedge \ \Box(x = \boldsymbol{t} \rightarrow y = \boldsymbol{t}))].$$

Lastly, define an "elementary range" as a minimal subrange (that is, a proper range without a proper subrange).

Definition 3 (Elementary range, *ElR*)

$$\forall x[ElR(x) \leftrightarrow_{df} (PR(x) \ \wedge \ \forall y[SubR(y, x) \rightarrow \Box(x = y)])].$$

In a picture, it has to be that

[10] That is the picture. Literally, we are saying that in each case, either $ext_{\gamma}(x) = ext_{\gamma}(\mathbf{t})$ or $ext_{\gamma}(x) = ext_{\gamma}(\mathbf{f})$. Or, equivalently, either $\gamma \models x = \mathbf{t}$ or $\gamma \models x = \mathbf{f}$.

Table 1 *PR* (proper range)

γ_1	γ_2	γ_3
ttf	**ttf**	**ttf**
tft	**tft**	**tft**
tff	**tff**	**tff**
fft	**fft**	**fft**
ftf	**ftf**	**ftf**
ftt	**ftt**	**ftt**

Table 2 *ElR* (Elementary range)

γ_1	γ_2	γ_3
tff	**tff**	**tff**
ftf	**ftf**	**ftf**
fft	**fft**	**fft**

$$int(x) = \mathbf{ff}\ldots\mathbf{fftff}\ldots. \tag{1}$$

That is, the picture of Eq. (1) shows exactly one **t** among all the **f**s. Please sit still for an interpretive hint: No matter what happens in an elementary case, exactly one case happens. In case-intensional logic, we want to say that a certain case happens. Let x be an (intensional) individual variable. Then for x to represent that a particular case, γ, happens, x must code an elementary range, and it must be that $ext_\gamma(x) = ext_\gamma(\mathbf{t})$. The trick, such as it is, comes to "identifying" cases.

The following can be verified by eye simply by noting that each atomic part of each of the three definiens occurs within the scope of a modal connective.

Fact 3 (Status of *PR*, *SubR*, and *ElR*) *PR, SubR, and ElR are modally constant; that is, their extensions do not vary with the case.*

Intuitively, in each model, the elementary ranges are in one-one correspondence with the cases. Given a case, γ, let its mate be the unique intension, x, such that $ext_\gamma(x) = ext_\gamma(\mathbf{t})$, and for $\gamma' \neq \gamma$, $ext_{\gamma'}(x) = ext_{\gamma'}(\mathbf{f})$; and going the other way, given x with an intension such that $ElR(x)$, let its mate be the one and only case γ such that $ext_\gamma(x) = ext_\gamma(\mathbf{t})$. So we should be able to say what we want to say about cases in CIFOL+, that is, indirectly, by instead speaking of elementary ranges. Let us use pictures of intensions to give an (abstract) example. Let $\Gamma = \{\gamma_1, \gamma_2, \gamma_3\}$. The following three tables exhibit the possible extensions (a) of a proper range, *PR*, (b) of an elementary range, *ElR*, and (c) of the property, $\lambda x(ElR(x) \wedge x = \mathbf{t})$, of being an elementary range that happens.

In Table 1, you can see that *PR* is modally constant, and in each case contains all **t**-**f** sequences with at least one **t** and at least one **f**.

Table 2 shows that *ElR* is modally constant, and given $ElR(x)$, that the extension of x in each γ_i contains exactly one **t**.

Table 3 $\lambda x(ElR(x) \wedge x = \mathbf{t})$ (Elementary range that happens)

γ_1	γ_2	γ_3
tff	**ftf**	**fft**

Table 3 on p. 64 is the most interesting, just because it is *not* modally constant. In case γ_i, it contains some or all elementary ranges—that is, *the* elementary range—with **t** in the column for γ_i, and **f** in each other column.

2.1 CIFOL+ and Elementary Ranges

There is still work to do in order to have a viable theory of elementary ranges. It is a trivial truth in our semantic metalanguage that for each $\gamma \in \Gamma$, there is an individual intension (in $\Gamma \mapsto D$) that is an elementary range whose extension in γ is identical to the extension of **t** in γ. This is what Table 3 portrays in miniature. The problem is to try to bring this down to the language of CIFOL+ itself, rather than leaving it to pictures, or descriptions in the semantic metalanguage. Already a solution is almost possible: A CIFOL + Ax. 1 sentence that seems to have the right form is the following:

$$\Box\exists x[ElR(x) \ \wedge \ x = \mathbf{t}]. \tag{2}$$

Equation 2 may be read informally as saying that necessarily there is an elementary range (a case) that happens, relying on a convention that a case that happens is marked with **t**. So CIFOL + Ax. 1 has the expressive power to say what needs to be said, a fact that might seem to solve the problem. However, the question is whether we can *prove* Eq. 2 in CIFOL + Ax. 1 itself; the answer is "not quite." Bressan shows, however, that it can be proven in MC^{ν} with the help of a certain modest second-order axiom (his Axiom 12.19):

Bressan axiom 12.19.

$$\vdash \Box\exists P\forall x[(Px \leftrightarrow \Phi) \ \wedge \ (\Box Px \leftrightarrow \Diamond Px)].$$

According to this axiom, for each case, γ, for each sentence, Φ, presumably with x free, there is a property, P, that applies to any x if and only if x satisfies the condition Φ in case γ, and furthermore is modally constant. So for each case, γ, the range of P is the same in every case, γ', and is precisely the range of Φ with respect to x in the particular case, γ. In a picture, P picks up the range of Φ with respect to x in case γ, and duplicates that range in every case. The first conjunct of 12.19 is standard in classical second order logic; it is the addition of the second conjunct that is distinctive. It arises neither out of second order quantification theory nor out of **S5** considerations.

We cannot merely add second-order 12.19 to CIFOL + Ax. 1, which is intended to be first order. For the purpose of proving that some elementary range happens,

it suffices to define CIFOL+ by adding to CIFOL + Ax. 1 a single further first-order axiom involving a single "reserved predicate constant," P_0, corresponding to instantiating Φ in 12.19 with $(x = \mathbf{t} \ \wedge \ \Box(x = \mathbf{t} \ \vee \ x = \mathbf{f}))$:

Axiom 2 (12.19 instance)

$$\forall x[(P_0 x \leftrightarrow (x = \boldsymbol{t} \ \wedge \ \Box(x = \boldsymbol{t} \ \vee \ x = \boldsymbol{f}))) \ \wedge \ (\Diamond P_0 x \leftrightarrow \Box P_0 x)].$$

This P_0 consequence of Bressan's axiom 12.19 is first order, and we count it as the second and last axiom of CIFOL+:

$$\text{CIFOL+} = \text{CIFOL + Ax. 1,2.}$$

Observe that Axiom 2 does not begin with $\Box$. As a first-order work-around of the second-order axiom 12.19, we are to think of it as only contingent, true in some particular case—a second-class axiom, if you like. In proofs, we mark only "first-class" formulas with the customary "$\vdash$".

It is helpful to keep in mind that Axiom 2 can be seen as coming by second-order existential instantiation of P in a demodalized version of 12.19 by P_0, all in the interest of keeping to the first order. We give bite to this mental picture of the axiom by imposing two requirements on CIFOL+: Axiom 2 must be the only postulate that mentions this predicate constant; and P_0 cannot occur in the last line of any complete proof in CIFOL+. (The requirements are the same as imposed on the result of "existential instantiation" in Belnap 2009.) To repeat, it is understood that although Axiom 2 may be used in a proof in CIFOL+, the last line must not contain P_0—that is, P_0 must be discharged. This requirement (and the requirement that no other axiom may contain an occurrence of P_0) distinguishes the *logic* CIFOL+ from a CIFOL+ *theory* with Axiom 2 as a non-logical axiom. The payoff is that we will now be able to prove Eq. 2. Indeed, we can prove not only the existence claim, but also existence and (strict) uniqueness.

Theorem 1 (Unique existence of an elementary range that happens)

$$\vdash \Box\exists x[ElR(x) \ \wedge \ x = \boldsymbol{t} \ \wedge \ \forall y[(ElR(y) \ \wedge \ y = \boldsymbol{t}) \rightarrow \Box(y = x)]].$$

By Definition 4 coming up, this may be written as

$$\vdash \Box\exists_1^{\Box} x[ElR(x) \ \wedge \ x = \mathbf{t}].$$

It is essential to observe that Theorem 1 contains no occurrence of P_0. That means that we can count Theorem 1, once we prove it, as a theorem of logic, rather than merely as a consequence of Axiom 1 and the contingent Axiom 2.

3 Proving Theorem 1

The proof of the theorem will invoke three CIFOL+ abbreviative definitions. We use the standard notation for syntactic replacement; that is, $[y/x]\Phi$ stands for the expression obtained by replacing all free occurrences of x by y. We will also employ the CIFOL definition of definite descriptions: The extension of the term $\imath x\Phi$ in a case γ is $d \in D$ iff d is the extensionally unique witness fulfilling Φ in case γ, and $*$ otherwise.

Definition 4 *Unique existence, strict unique existence, extensionality*

$$\exists_1 x\Phi \leftrightarrow_{df} \exists x[\Phi \ \wedge \ \forall y[[y/x]\Phi \rightarrow (y = x)]].$$

$$\exists_1^{\Box} x\Phi \leftrightarrow_{df} \exists x[\Phi \ \wedge \ \forall y[[y/x]\Phi \rightarrow \Box(y = x)]].$$

$$(extnl\ x)\Phi \leftrightarrow_{df} \forall y[(\Phi \ \wedge \ x = y) \rightarrow [y/x]\Phi]$$

These three definitions provide the respective CIFOL+ renderings of "there is a unique x such that Φ," "there is a strictly unique x such that Φ," and "Φ is extensional with respect to x."

Fact 4 *If Φ is extensional with respect to x (i.e., if $(extnl\ x)\Phi$, so that Φ supports replacement of identicals), then $\vdash \exists_1 x\Phi \rightarrow \Phi(\imath x\Phi)$.*

The proof of the Fact is straightforward.

We now turn to the proof in CIFOL+ of Theorem 1 stating that a strictly unique elementary case happens; the proof, which has five parts, occupies the rest of Sect. 3. We note that lines of the proof that contain any of P_0, Θ, or θ, the latter two of which are defined in terms of P_0, do not have the status of theorems of CIFOL+. Neither Axiom 2 nor any such line begins with the sign of necessity.[11] We emphasize the distinction when we mark proper theorems with the customary turnstile, $\vdash$. The last line of the proof is 4.12, which can be seen by inspection to contain no notation that relies on P_0 for its definition. For convenience, we break up the proof of Theorem 1 into five parts. Parts Ia and Ib prove that the individual constant θ denotes a proper range, and is (extensionally) equal to **t**—which says that θ happens. The conclusion of this part, since it contains θ, depends on Axiom 2 as a hypothesis. Only at the end of Part IV can we discharge Axiom 2 as a hypothesis. The annotation "C.P." signifies "Conditional proof," and "$\forall$" advertises a quantifier rule. Watch for the role of θ.

[11] So what is their status? Lines that contain any of P_0, Θ, or θ are certainly not offered as logical truths. Metaphorically each serves as a Wittgensteinian crutch that is to be thrown away. More literally, they may be taken to be proved under the hypothesis Axiom 2. The constant θ plays an especially critical role.

Part Ia

1a.1	$\forall x[(P_0x \leftrightarrow (x = \mathbf{t} \ \wedge \ \Box(x = \mathbf{t} \ \vee \ x = \mathbf{f}))) \wedge$ $(\Diamond P_0x \leftrightarrow \Box P_0x)]$	Ax. 2
1a.2	$\Theta \leftrightarrow_{df} ((\exists x[P_0x \ \wedge \ x = \mathbf{f}] \ \wedge \ y = \mathbf{f}) \vee$ $(\neg\exists x[P_0x \ \wedge \ x = \mathbf{f}] \ \wedge \ y = \mathbf{t}))$	Def.
1a.3	$\Box\exists_1 y\Theta \ \wedge \ \Box(extnl \ y)\Theta$	1a.2, **S5**
1a.4	$\theta =_{df} \imath y\Theta$	Def.
1a.5	$\Box[(\exists x[P_0x \ \wedge \ x = \mathbf{f}] \ \wedge \ \theta = \mathbf{f}) \vee$ $(\neg\exists x[P_0x \ \wedge \ x = \mathbf{f}] \ \wedge \ \theta = \mathbf{t})]$	1a.2, 1a.3, 1a.4, Fact 4
1a.6	$\forall x[P_0x \rightarrow x = \mathbf{t}]$	1a.1
1a.7	$\neg\exists x[P_0x \ \wedge \ x = \mathbf{f}]$	Ax. 1, 1a.6, Reductio
1a.8	$\theta = \mathbf{t}$	1a.5, 1a.7
1a.9	$\Diamond(\theta = \mathbf{t})$	1a.8, **S5**
1a.10	$\Box(\theta = \mathbf{f} \ \vee \ \theta = \mathbf{t})$	1a.5
1a.11	$\Diamond(x_0 = \mathbf{t}) \ \wedge \ \Diamond(x_0 = \mathbf{f})$	Ax. 1; choose x_0
1a.12	$\Psi_1 \leftrightarrow_{df} [(x_0 = \mathbf{t} \ \wedge \ x = \mathbf{t}) \vee$ $(\neg(x_0 = \mathbf{t}) \ \wedge \ x = \mathbf{f})]$	Def.
1a.13	$\Psi_2 \leftrightarrow_{df} [(\neg(x_0 = \mathbf{t}) \ \wedge \ x = \mathbf{t}) \vee$ $(x_0 = \mathbf{t} \ \wedge \ x = \mathbf{f})]$	Def.
1a.14	$\psi_n =_{df} \imath x\Psi_n \ [n = 1, 2]$	Def.
1a.15	$\Box(\exists_1 x\Psi_n \ \wedge \ (extnl \ x)\Psi_n) \ [n = 1, 2]$	Ax. 1, 1a.12, 1a.13
1a.16	$\Box[\psi_n/x]\Psi_n \ [n = 1, 2]$	Fact 4, 1a.14, 1a.15

Part Ib

1b.1	$\Box(x_0 = \mathbf{t} \rightarrow (\psi_1 = \mathbf{t} \ \wedge \ \psi_2 = \mathbf{f}))$	1a.16
1b.2	$\Box(\neg(x_0 = \mathbf{t}) \rightarrow (\psi_1 = \mathbf{f} \ \wedge \ \psi_2 = \mathbf{t}))$	1a.16
1b.3	$\Box(x_0 = \mathbf{t} \ \vee \ \neg x_0 = \mathbf{t})$	Excl. mid.
1b.4	$\Box((\psi_1 = \mathbf{t} \ \wedge \ \psi_2 = \mathbf{f}) \ \vee \ (\psi_1 = \mathbf{f} \ \wedge \ \psi_2 = \mathbf{t}))$	1b.1–1b.3
1b.5	$\Box(\psi_1 = \mathbf{t} \ \vee \ \psi_1 = \mathbf{f}) \ \wedge \ \Box(\psi_2 = \mathbf{t} \ \vee \ \psi_2 = \mathbf{f})$	1b.4
1b.6	$\Box(\psi_1 = \mathbf{t} \ \vee \ \psi_2 = \mathbf{t})$	1b.4
1b.7	$(\psi_1 = \mathbf{t} \ \wedge \ \Box(\psi_1 = \mathbf{t} \ \vee \ \psi_1 = \mathbf{f})) \vee$ $(\psi_2 = \mathbf{t} \ \wedge \ \Box(\psi_2 = \mathbf{t} \ \vee \ \psi_2 = \mathbf{f}))$	1b.5, 1b.6
1b.8	$P_0(\psi_n) \leftrightarrow (\psi_n = \mathbf{t} \ \wedge \ \Box(\psi_n = \mathbf{t} \ \vee \ \psi_n = \mathbf{f}))$	$[n = 1, 2]$; 1a.1
1b.9	$P_0(\psi_1) \ \vee \ P_0(\psi_2)$	1b.7, 1b.8
1b.10	$\Box P_0(\psi_1) \ \vee \ \Box P_0(\psi_2)$	1b.9, 1a.1, *MConst*
1b.11	$\Diamond(\psi_n = \mathbf{f}) \ [n = 1, 2]$	1a.11, 1b.1, 1b.2, **S5**
1b.12	$\Diamond(P_0(\psi_1) \ \wedge \ \psi_1 = \mathbf{f}) \ \vee \ \Diamond(P_0(\psi_2) \ \wedge \ \psi_2 = \mathbf{f})$	1b.10, 1b.11, **S5**
1b.13	$\Diamond\exists x[P_0x \ \wedge \ x = \mathbf{f}]$	1b.12
1b.14	$\Diamond(\theta = \mathbf{f})$	1b.1, 1a.5, **S5**
1b.15	$PR(\theta)$	1a.9, 1a.10, 1b.14 Def. *PR*
1b.16	$\theta = \mathbf{t} \ \wedge \ PR(\theta)$	1a.8, 1b.15

Note that y is free in line 1a.2, and that x is free in lines 1a.12, 1a.13. "*MConst*" in line 1b.10 refers to the part of Axiom 2 saying that P_0 is modally constant.

So θ is a proper range that is extensionally equal to **t**; but that conclusion of Part I is insufficiently strong. What is wanted is that θ is not just a proper range, but an elementary range, which is proved in Part III. Part II is chiefly "housekeeping" on the way to facts 2.12 and 2.13, required for Part III. Throughout Part II, Φ is any sentence, and ρ_Φ may be thought of as the internal representation of the range of Φ.

Part II

2.1	$\rho_\Phi =_{df} \imath x[(\Phi \ \wedge \ x = \mathbf{t}) \vee (\neg\Phi \ \wedge \ x = \mathbf{f})]$	Def., Φ any sent.
2.2	$\Box(\Phi \rightarrow (\rho_\Phi = \mathbf{t}))$	2.1
2.3	$\Box(\neg\Phi \rightarrow (\rho_\Phi = \mathbf{f}))$	2.1
2.4	$\Box(\rho_\Phi = \mathbf{t} \ \vee \ \rho_\Phi = \mathbf{f})$	2.1, 2.2, 2.3
2.5	$\Phi \rightarrow P_0(\rho_\Phi)$	1a.1, 2.1, 2.4
2.6	$\Phi \rightarrow \Box P_0(\rho_\Phi)$	2.5, 1a.1, *MConst*
2.7	$\Phi \rightarrow \Box(\neg\Phi \rightarrow P_0(\rho_\Phi))$	2.6, **S5**
2.8	$\Phi \rightarrow \Box(\neg\Phi \rightarrow (\rho_\Phi = \mathbf{f}))$	2.3
2.9	$\Phi \rightarrow \Box(\neg\Phi \rightarrow (P_0(\rho_\Phi) \ \wedge \ (\rho_\Phi = \mathbf{f})))$	2.7, 2.8
2.10	$\Phi \rightarrow \Box(\neg\Phi \rightarrow \exists x[P_0(x) \ \wedge \ (x = \mathbf{f})])$	2.9, ρ_Φ/x
2.11	$\Phi \rightarrow \Box(\neg\Phi \rightarrow (\theta = \mathbf{f}))$	2.10, 1a.5
2.12	$\forall x[x = \mathbf{t} \rightarrow \Box(x = \mathbf{f} \rightarrow \theta = \mathbf{f})]$	Ax. 1, 2.11 $(x = \mathbf{t})/\Phi$
2.13	$\forall x[x = \mathbf{f} \rightarrow \Box(x = \mathbf{t} \rightarrow \theta = \mathbf{f})]$	Ax. 1, 2.11 $(x = \mathbf{f})/\Phi$

Part III is chiefly a "conditional proof," from 3.1 to 3.12, the purpose of which is to serve as a premiss for the use of the rule of conditional proof in justifying 3.13. Part III ends with establishing that θ is an elementary range equal to **t**, thus providing material for the existence portion of the desired conclusion, Theorem 1, at line 4.12.

Part III

3.1	$\underline{SubR}(x_0, \theta)$	Hypothesis, choose x_0
3.2	$x_0, \theta \in PR$	Def. *SubR*
3.3	$\Box(x_0 = \mathbf{f} \ \vee \ \theta = \mathbf{t})$	Def. *SubR*
3.4	$\Diamond(x_0 = \mathbf{t})$	3.2
3.5	$x_0 = \mathbf{f} \rightarrow \Diamond(x_0 = \mathbf{t} \ \wedge \ \theta = \mathbf{f})$	2.13, 3.4, **S5**
3.6	$x_0 = \mathbf{f} \rightarrow \neg\Box(x_0 = \mathbf{f} \ \vee \ \theta = \mathbf{t})$	3.2, 3.5, **S5**
3.7	$\neg(x_0 = \mathbf{f})$	3.2, 3.3, 3.6
3.8	$x_0 = \mathbf{t}$	3.2, 3.7
3.9	$\Box(x_0 = \mathbf{f} \rightarrow \theta = \mathbf{f})$	2.12, 3.8
3.10	$\Box(x_0 = \mathbf{t} \ \vee \ x_0 = \mathbf{f})$	3.2
3.11	$\Box(\theta = \mathbf{t} \ \vee \ \theta = \mathbf{f})$	3.2
3.12	$\Box(x_0 = \theta)$	3.10, 3.11, 3.9, 3.3
3.13	$\forall x[SubR(x, \theta) \rightarrow \Box(x = \theta)]$	3.1–12, C.P., $\forall$
3.14	$\theta = \mathbf{t} \ \wedge \ ElR(\theta)$	1b.16, 3.13, Def. *ElR*

Part III might be taken to establish θ, introduced at line 1a.4, as a kind of logical constant; but is it unique? That is the job of the next and last part of the proof: Part IV establishes that θ is not only an elementary range extensionally equal to **t**, but is strictly (or intensionally) unique in that respect. Then existential generalization yields Theorem 1 at line 4.10, which, aside from the abbreviated and necessitated version at line 4.12, finishes our work: We will have established that elementary ranges, *ElR*, are suitable surrogates for "cases."

Part IV

4.1	$\underline{y_0 = \mathbf{t} \ \wedge \ ElR(y_0)}$	Hypothesis, choose y_0
4.2	$\Box(y_0 = \mathbf{f} \rightarrow \theta = \mathbf{f})$	4.1, 2.12
4.3	$\Box(\neg(y_0 = \mathbf{f}) \ \vee \ \theta = \mathbf{f})$	4.2
4.4	$\Box(y_0 = \mathbf{f} \ \vee \ y_0 = \mathbf{t})$	4.1, Def. *ElR*
4.5	$\Box(y_0 = \mathbf{t} \ \vee \ \theta = \mathbf{f})$	4.3, 4.4
4.6	$SubR(\theta, y_0)$	3.14, 4.1, 4.5, Def. *SubR*
4.7	$\Box(y_0 = \theta)$	4.6, 4.1, Def. *ElR*
4.8	$\forall y[(y = \mathbf{t} \ \wedge \ ElR(y)) \rightarrow \Box(y = \theta)]$	4.1–7, C.P., $\forall$
4.9	$\theta = \mathbf{t} \ \wedge \ ElR(\theta) \ \wedge$ $\forall y[(y = \mathbf{t} \ \wedge \ ElR(y)) \rightarrow \Box(y = \theta)]$	3.14, 4.8
4.10	$\vdash \exists x[(x = \mathbf{t} \ \wedge \ ElR(x)) \ \wedge$ $\forall y[(y = \mathbf{t} \ \wedge \ ElR(y)) \rightarrow \Box(y = x)]]$	4.9, θ/x
4.11	$\vdash \exists_1^{\Box} x[x = \mathbf{t} \ \wedge \ ElR(x)]$	4.10, Def. $\exists_1^{\Box}$
4.12	$\vdash \Box\exists_1^{\Box} x[x = \mathbf{t} \ \wedge \ ElR(x)]$	4.11, **S5**; = Thm. 1 ■

Observe in particular that 4.10–4.12 contain no occurrences of P_0 nor of any notation defined in terms of P_0; accordingly, we are entitled to count 4.12 = Theorem 1 in particular as a theorem of logic.

4 The concept of case-relative truth

Theorem 1 told us that necessarily there is an intensionally unique elementary range—our internal surrogate for "case"—that happens. This takes us part way to finding the idea of "true in a case" inside of CIFOL+. Let us step back and think for a minute about "true in a case." That phrase is special, partly because there is very little intuitive support for the phrase "true in a world," even though one can be led by well-worked-out formal machinery to prize the idea. A substitute phrase, "truth according to a world" seems marginally more idiomatic, although hardly an expression of conversational English. In contrast, "true in a case" is somewhat more idiomatic. I don't mean the truth of sentences, which isn't idiomatic at all, but rather certain informal "true that" expressions relativized to cases:

- It's true that we shall get wet in case it rains, but otherwise not.
- I shall be sad in case you find fault with my examples.

- There are two cases in which it would be true that Mary will bake pies, but there is no case in which it would be true that I do the baking.

If we idealize by ruling out subcases, such a concept of truth can find a (formal) home in CIFOL+. The idea is to carry the true-in locution with a mixed nector (Curry), with one input place for a sentence and another for a term, and with the output being a sentence. So the locution that we are after in CIFOL+ will have the form

$$T(\Phi, \alpha)$$

with Φ a CIFOL+ sentence and α a CIFOL+ term that we can take as standing for a case. The CIFOL+ sentence, $T(\Phi, \alpha)$, is intended to be read in English with the pattern "that Φ is true in α," presupposing that α is an elementary range (no proper subranges). $T(\Phi, \alpha)$ must be "found" (i.e., defined) in CIFOL+, not merely added.

We begin by noting that in any CIFOL+ model $\mathcal{M} = \langle \Gamma, D, \mathcal{I} \rangle$, for each case, γ, we can find the set of intensions representing elementary ranges that satisfy $ElR(x)$ in γ. We can gain some purchase on this in the semantic metalanguage of CIFOL+ by way of "$\gamma \models (ElR(x) \wedge x = \mathbf{t})$," which can be read as saying that x stands for *the* elementary case that holds in case γ. Simple logic tells us that if there is exactly one European king alive today, then to say that *some* European king alive today is bald is to say the same thing as saying that *every* European king alive today is bald. Just so, since intensionally speaking, in each model and in each case γ, there is exactly *one* elementary range that happens, we should expect that if $ElR(x)$, then $\Diamond(x = \mathbf{t} \wedge \Phi)$ (that Φ is true in *some* elementary range (= case) that happens) and $\Box(x = \mathbf{t} \rightarrow \Phi)$ (that Φ is true in *every* elementary range that happens) equally express that Φ is true in *the* elementary range that happens in γ.[12] Furthermore, Lemma 1 below suggests that provided $ElR(x)$, each of $\Diamond(x = \mathbf{t} \wedge \Phi)$ and $\Box(x = \mathbf{t} \rightarrow \Phi)$ is a suitable candidate for the CIFOL+ representation of "Φ is true in case x." We choose the former, and then from time to time mention the availability of the latter.

Definition 5 *(Φ* **is true in case** *x)*

$$\forall x[ElR(x) \rightarrow (T(\Phi, x) \leftrightarrow_{df} \Diamond(x = \mathbf{t} \wedge \Phi))].$$

Definition 5 is intended as a conditional definition, the thought being that one can apply the equivalence of $T(\Phi, x)$ and $\Diamond(x = \mathbf{t} \wedge \Phi)$ only when x is an elementary range.

We need to verify the equivalence between $\Diamond(x = \mathbf{t} \wedge \Phi)$ and $\Box(x = \mathbf{t} \rightarrow \Phi)$, under the hypothesis that $ElR(x)$, for that equivalence is required for showing that the appropriate clauses for $T(\Phi, x)$ hold as they should. That is, we prove Lemma 1 as an essential step in verifying that "that Φ is true in case x" can be found in CIFOL+.

[12] Throughout we assume that the variable, x, that we are taking to range over *ElR* does not occur free in Φ.

Lemma 1 (Fundamental equivalence)

$$\vdash \forall x[\mathit{ElR}(x) \rightarrow (\Diamond(x = \boldsymbol{t} \ \wedge \ \Phi) \leftrightarrow \Box(x = \boldsymbol{t} \rightarrow \Phi))].$$

PROOF.

5.1	$(\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t}) \rightarrow \Box(x_0 = \theta)$	Pt. IV, 4.8, choose x_0
5.2	$(\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t}) \rightarrow \Box(\neg\Phi \rightarrow x_0 = \theta)$	5.1, **S5**
5.3	$\Phi \rightarrow \Box(\neg\Phi \rightarrow (\theta = \mathbf{f}))$	Pt. II, 2.11
5.4	$\vdash (\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t} \ \wedge \ \Phi) \rightarrow \Box(\neg\Phi \rightarrow x_0 = \mathbf{f})$	5.2, 5.3, **S5**
5.5	$\vdash (\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t} \ \wedge \ \Phi) \rightarrow \Box(\neg(x_0 = \mathbf{f}) \rightarrow \Phi)$	5.4, **S5**
5.6	$\vdash \Box(x_0 = \mathbf{t} \rightarrow \neg(x_0 = \mathbf{f}))$	Ax. 1, **S5**
5.7	$\vdash (\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t} \ \wedge \ \Phi) \rightarrow \Box(x_0 = \mathbf{t} \rightarrow \Phi)$	5.5, 5.6, **S5**
5.8	$\vdash \Box((\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t} \ \wedge \ \Phi) \rightarrow \Box(x_0 = \mathbf{t} \rightarrow \Phi))$	5.7, **S5**
5.9	$\vdash \Diamond(\mathit{ElR}(x_0) \ \wedge \ x_0 = \mathbf{t} \ \wedge \ \Phi) \rightarrow \Box(x_0 = \mathbf{t} \rightarrow \Phi)$	5.8, **S5**
5.10	$\vdash (\mathit{ElR}(x_0) \ \wedge \ \Diamond(x_0 = \mathbf{t} \ \wedge \ \Phi)) \rightarrow \Box(x_0 = \mathbf{t} \rightarrow \Phi)$	5.9, **S5**, Fact 3
5.11	$\vdash \mathit{ElR}(x_0) \rightarrow (\Diamond(x_0 = \mathbf{t} \ \wedge \ \Phi) \rightarrow \Box(x_0 = \mathbf{t} \rightarrow \Phi))$	5.10
5.12	$\vdash \mathit{ElR}(x_0) \rightarrow \Diamond(x_0 = \mathbf{t})$	Def. *ElR*
5.13	$\vdash (\Diamond(x_0 = \mathbf{t}) \rightarrow (\Box(x_0 = \mathbf{t} \rightarrow \Phi) \rightarrow \Diamond(x_0 = \mathbf{t} \ \wedge \ \Phi)))$	**S5**
5.14	$\vdash (\mathit{ElR}(x_0) \rightarrow (\Box(x_0 = \mathbf{t} \rightarrow \Phi) \rightarrow \Diamond(x_0 = \mathbf{t} \ \wedge \ \Phi)))$	5.12, 5.13
5.15	$\vdash (\mathit{ElR}(x_0) \rightarrow (\Diamond(x_0 = \mathbf{t} \ \wedge \ \Phi) \leftrightarrow \Box(x_0 = \mathbf{t} \rightarrow \Phi)))$	5.11, 5.14
5.16	$\vdash \forall x[\mathit{ElR}(x) \rightarrow (\Diamond(x = \mathbf{t} \ \wedge \ \Phi) \leftrightarrow \Box(x = \mathbf{t} \rightarrow \Phi))]$	5.15, $\forall$ ■

Lemma 1, which grounds our indifference between the two ways to express the "that Φ is true in case x" locution in CIFOL+—subject, of course to the assumption that $\mathit{ElR}(x)$—encourages us to verify that the "true in" locution in CIFOL+ behaves properly with respect to the connectives of CIFOL+. That is, we show that the various metalinguistic semantic clauses can be approximated within CIFOL+ itself using the predicate *ElR* and the mixed nector $T(\Phi, x)$.[13] The result, Theorem 2, is evidence for the conceptual coherence of CIFOL+ and its semantics. This is a striking result given that neither truth nor satisfaction (that is, "true on an assignment to the variables") is available within CIFOL+, on pain of contradiction.[14]

Theorem 2 (Internal semantic clauses) *Each of the following is provable in CIFOL+. We assume that x has no free occurrence in* Φ, Φ_1, *or* Φ_2. *All clauses are subject to the condition,* $\mathit{ElR}(x)$, *stating that x is an elementary range— which is the internal representation of "x is a case" (or* $x \in \Gamma$*). Here we are using "*Φ*" to take the place of CIFOL+ sentences in schemata, so that the instances of e.g.* $(\neg)$ *are*

[13] In the CIFOL+ locution, "$T(\Phi, x)$," the role of x is that of a proper CIFOL+ variable. In contrast, the expression Φ serves as a schematic variable only.

[14] Not too much, however, should be made of a comparison between CIFOL+'s "true in" and the classical truth predicate. The grammars differ in important ways. For example, as we observed in Sect. 1.3, were we to try to make sentences Φ instead of expressions denoting sentences the complements of "truth" without paying attention to case-relativity, in this way turning a truth predicate into a truth connective, the theory of truth would be trivial.

particular CIFOL+ sentences. (In contrast, in giving the metalinguistic semantics of CIFOL, as in Belnap and Müller 2013, "Φ" denotes rather than takes the place of a CIFOL+ sentence.)

$(\wedge)\ \vdash \forall x[ElR(x) \to (T((\Phi_1 \ \wedge \ \Phi_2), x) \leftrightarrow (T(\Phi_1, x) \ \wedge \ T(\Phi_2, x)))]$
$(\vee)\ \vdash \forall x[ElR(x) \to (T((\Phi_1 \ \vee \ \Phi_2), x) \leftrightarrow (T(\Phi_1, x) \ \vee \ T(\Phi_2, x)))]$
$(\neg)\ \vdash \forall x[ElR(x) \to (T(\neg\Phi, x) \leftrightarrow \neg T(\Phi, x))]$
$(\forall y)\ \vdash \forall x[ElR(x) \to (T(\forall y\Phi, x) \leftrightarrow \forall y T(\Phi, x))]$
$(\exists y)\ \vdash \forall x[ElR(x) \to (T(\exists y\Phi, x) \leftrightarrow \exists y T(\Phi, x))]$
$(\Box)\ \vdash \forall x[ElR(x) \to (T(\Box\Phi, x) \leftrightarrow \forall z[ElR(z) \to T(\Phi, z)])]$
$(\Diamond)\ \vdash \forall x[ElR(x) \to (T(\Diamond\Phi, x) \leftrightarrow \exists z[ElR(z) \ \wedge \ T(\Phi, z)])]$

These "conditioned equivalences" are structurally like conditional definitions: *ElR* appears as a presupposition of $T(\Phi, x)$ rather than as an implicate, just as it would if we were to take the list as (say) giving the clauses of an inductive definition.

PROOF. In each of the following, we assume that x does not occur in Φ, and as definiens of $T(\Phi, x)$ we use whichever of $\Box(x = \mathbf{t} \to \Phi)$ or $\Diamond(x = \mathbf{t} \ \wedge \ \Phi)$ is more convenient, relying on Lemma 1 as the warrant for our indifference.

$(\wedge)$ It suffices that in quantified **S5**, $\vdash \forall x[ElR(x) \to (\Box(x = \mathbf{t} \to (\Phi_1 \ \wedge \ \Phi_2)) \leftrightarrow (\Box(x = \mathbf{t} \to \Phi_1) \ \wedge \ \Box(x = \mathbf{t} \to \Phi_2)))]$.

$(\vee)$ It suffices that in quantified **S5**, $\vdash \forall x[ElR(x) \to (\Diamond(x = \mathbf{t} \ \wedge \ (\Phi_1 \ \vee \ \Phi_2)) \leftrightarrow (\Diamond(x = \mathbf{t} \ \wedge \ \Phi_1) \ \vee \ \Diamond(x = \mathbf{t} \ \wedge \ \Phi_2)))]$.

$(\neg)$ By the fundamental equivalence of Lemma 1,
$\vdash \forall x[ElR(x) \to (\Diamond(x = \mathbf{t} \ \wedge \ \neg\Phi) \leftrightarrow \Box(x = \mathbf{t} \to \neg\Phi))]$.
By **S5**, $\Box(x = \mathbf{t} \to \neg\Phi)$ is interchangeable with $\neg\Diamond(x = \mathbf{t} \ \wedge \ \Phi)$.
So using Def. 5 twice, we have the required conditional equivalence:

$$\vdash \forall x[ElR(x) \to (T(\neg\Phi, x) \leftrightarrow \neg(T(\Phi, x)))]$$

$(\forall y)$ It suffices that in quantified **S5**,
$\vdash \forall x[ElR(x) \to (\Box(x = \mathbf{t} \to \forall y\Phi) \leftrightarrow \forall y[\Box(x = \mathbf{t} \to \Phi)])]$.

$(\exists y)$ It suffices that in quantified **S5**,
$\vdash \forall x[ElR(x) \to (\Diamond(x = \mathbf{t} \ \wedge \ \exists y\Phi) \leftrightarrow \exists y[\Diamond(x = \mathbf{t} \ \wedge \ \Phi)])]$.

$(\Box)$ It suffices that in quantified **S5**,
$\vdash \forall x[ElR(x) \to \Diamond(x = \mathbf{t} \ \wedge \ \Box\Phi) \leftrightarrow \forall z[ElR(z) \to \Diamond(z = \mathbf{t} \ \wedge \ \Phi)]]$.
We supply detail, first proving 6.19 as a lemma.

6.1	$\underline{\Diamond}\neg\Phi$	Hypothesis
6.2	$\Box\exists x[ElR(x) \wedge x = \mathbf{t}]$	Thm. 1, **S5**
6.3	$\Diamond(\exists x[ElR(x) \wedge x = \mathbf{t} \wedge \neg\Phi])$	6.1–2, **S5**
6.4	$\exists x[\Diamond(ElR(x) \wedge x = \mathbf{t} \wedge \neg\Phi)]$	6.3, Barcan
6.5	$\Diamond(ElR(x_0) \wedge x_0 = \mathbf{t} \wedge \neg\Phi)$	6.4, choose x_0
6.6	$\Diamond(ElR(x_0) \wedge x_0 = \mathbf{t} \wedge \neg\Phi) \rightarrow$ $\Box(x_0 = \mathbf{t} \rightarrow \neg\Phi)$	5.9, $[\neg\Phi/\Phi]$
6.7	$\Box(x_0 = \mathbf{t} \rightarrow \neg\Phi)$	6.5, 6.6
6.8	$\neg\Diamond(x_0 = \mathbf{t} \wedge \Phi)$	6.7, **S5**
6.9	$ElR(x_0)$	6.5, Fact 3
6.10	$\exists x[ElR(x) \wedge \neg\Diamond(x = \mathbf{t} \wedge \Phi)]$	6.8, 6.9
6.11	$\vdash \Diamond\neg\Phi \rightarrow \exists x[ElR(x) \wedge \neg\Diamond(x = \mathbf{t} \wedge \Phi)]$	6.1–10, C.P.
6.12	$\vdash \forall x[ElR(x) \rightarrow \Diamond(x = \mathbf{t} \wedge \Phi)] \rightarrow \Box\Phi$	6.11, **S5**
6.13	$\underline{\Box}\Phi$	Hypothesis
6.14	$\quad \underline{ElR}(x_0)$	Hyp., choose x_0
6.15	$\quad \Diamond(x_0 = \mathbf{t})$	6.14, Def. *ElR*
6.16	$\quad \Diamond(x_0 = \mathbf{t} \wedge \Phi)$	6.13, 6.15, **S5**
6.17	$\forall x[ElR(x) \rightarrow \Diamond(x = \mathbf{t} \wedge \Phi)]$	6.14–16, C.P., $\forall$
6.18	$\vdash \Box\Phi \rightarrow \forall x[ElR(x) \rightarrow \Diamond(x = \mathbf{t} \wedge \Phi)]$	6.13–17, C.P.
6.19	$\vdash \Box\Phi \leftrightarrow \forall x[ElR(x) \rightarrow \Diamond(x = \mathbf{t} \wedge \Phi)]$	6.12, 6.18

Now we are ready to prove what is needed for the "truth-condition" clause ($\Box$) as line 7.13.

7.1	$\underline{ElR}(x_0)$	hyp., choose x_0
7.2	$\quad \underline{\Diamond}(x_0 = \mathbf{t} \wedge \Box\Phi)$	hyp.
7.3	$\quad\quad \underline{ElR}(z_0)$	hyp., choose z_0
7.4	$\quad\quad \Box\Phi$	7.2, **S5**
7.5	$\quad\quad \forall z[ElR(z) \rightarrow \Diamond(z = \mathbf{t} \wedge \Phi)]$	7.4, 6.19
7.6	$\quad\quad \Diamond(z_0 = \mathbf{t} \wedge \Phi)$	7.3, 7.5
7.7	$\quad \forall z[ElR(z) \rightarrow \Diamond(z = \mathbf{t} \wedge \Phi)]$	7.3–7.6, $\forall$
7.8	$\quad \underline{\forall} z[ElR(z) \rightarrow \Diamond(z = \mathbf{t} \wedge \Phi)]$	hyp.
7.9	$\quad \Box\Phi$	7.8, 6.19
7.10	$\quad \Diamond(x_0 = \mathbf{t})$	7.1, Def. *ElR*
7.11	$\quad \Diamond(x_0 = \mathbf{t} \wedge \Box\Phi)$	7.9, 7.10, **S5**
7.12	$\quad \Diamond(x_0 = \mathbf{t} \wedge \Box\Phi) \leftrightarrow$ $\forall z[ElR(z) \rightarrow \Diamond(z = \mathbf{t} \wedge \Phi)]$	7.2–.7, 7.8–.11, $\leftrightarrow$
7.13	$\forall x[ElR(x) \rightarrow (\Diamond(x = \mathbf{t} \wedge \Box\Phi) \leftrightarrow$ $\forall z[ElR(z) \rightarrow \Diamond(z = \mathbf{t} \wedge \Phi)]]$	7.1–7.12, $\forall$

Lastly, ($\Diamond$). It suffices to dualize the argument for $\Box$, which concludes the proof of Theorem 2, and in this way speaks to a fit between CIFOL+ and its internal theory of "true in a case." ■

5 Summary

To put these results in a suitable context, we repeat that our aim has been to find a way of properly formalizing "that Φ is true in case γ" in the first-order extension, CIFOL+, of CIFOL, which is itself a first-order distillation from Bressan's higher-order modal calculus, MC^{ν}. This aim is intermediate between the futile aim of formalizing naive disquotation in CIFOL+, "'Φ' is true iff Φ," and the too-easy aim of formalizing the tautological "that Φ is true iff Φ."

We began by giving an extremely brief account of the chief features of CIFOL, described in Belnap and Müller 2013, and we indicated the wisdom of extending CIFOL to CIFOL+ by including a first-order trace of a certain second-order principle. We introduced a way of understanding a pair of CIFOL+ singular terms, **t** and **f**, as playing the role of internal truth values. Then we defined the CIFOL+ predicate, *ElR* (elementary range), as a suitable surrogate for "case," and $(ElR(x) \wedge x = \mathbf{t})$ as a surrogate for "x is a case that happens," giving a detailed proof that these CIFOL+ concepts are provably adequate representations of their respective intuitive ideas. Finally, we showed that $\Diamond(x = \mathbf{t} \wedge \Phi)$ adequately represents in CIFOL+ itself "that Φ is true in case x," where Φ is a CIFOL+ sentence and x denotes a CIFOL+ surrogate case, thus successfully threading our way between the impossible task of formalizing "'Φ' is true" and the trivial task of formalizing "that Φ is true."

References

Belnap, N. 2009. Notes on the art of logic (Unpublished). http://www.pitt.edu/~belnap.

Belnap, N., and T. Müller. 2013. Case-intensional first order logic. (I) Toward a theory of sorts. *Journal of Philosophical Logic*. doi:10.1007/s10992-012-9267-x.

Bressan, A. 1972. *A general interpreted modal calculus*. New Haven, CT: Yale University Press.

Carnap, R. 1947. *Meaning and necessity: a study in semantics and modal logic* (Enlarged edition 1956). Chicago, IL: University of Chicago Press.

Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to natural language: Proceedings of the 1970 Stanford workshop on grammar and semantics*, eds. Hintikka, J., Moravcsik, J., Suppes, P., 221–242. D. Reidel, Dordrecht (reprinted as Chap. 8 of Montague 1974).

Montague, R. 1974. *Formal philosophy: Selected papers of Richard Montague*, ed. Thomason, R.H., Yale University Press, New Haven.

A *stit* Logic Analysis of Morally Lucky and Legally Lucky Action Outcomes

Jan Broersen

Abstract Moral luck is the phenomenon that agents are not always held accountable for performance of a choice that under normal circumstances is likely to result in a state that is considered bad, but where due to some unexpected interaction the bad outcome does not obtain. We can also speak of 'moral misfortune' in the mirror situation where an agent chooses the good thing but the outcome is bad. This paper studies formalizations of moral and legal luck (and moral and legal misfortune). The three ingredients essential to modelling luck of these two different kinds are (1) indeterminacy of action effects, (2) determination on the part of the acting agent, (3) the possibility of evaluation of acts and/or their outcomes relative to a normative moral or legal code. The first, indeterminacy of action, is modelled by extending *stit* logic by allowing choices to have a probabilistic effect. The second, deliberateness of action, is modelled by (a) endowing *stit* operators with the possibility to specify a lower bound on the change of success, and (b) by introducing the notion of attempt as a maximisation of the probability of success. The third, evaluation relative to a moral or legal code, is modelled using Anderson's reduction of normative truth to logical truth. The conclusion will be that the problems embodied by the phenomenon of moral luck may be introduced by confusing it with legal luck. Formalizations of both forms are given.

1 Introduction

Agents may be morally lucky (Williams 1982) in several ways. Nagel (1979) explains how an agent can be morally lucky due to circumstance: if circumstances would have been different, for instance, if an agent would have had the opportunity to steal

J. Broersen (✉)
Intelligent Systems Group, Department of Information and Computing Sciences,
Faculty of science, Universiteit Utrecht, P.O. Box 80089, 3508 TB UTRECHT, The Netherlands
e-mail: broerser@cs.uu.nl

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_4,

somebodies money without anybody noticing, then the agent would have done so and would have gone morally wrong. Another category is moral luck due to character, such as when an agent would have committed an immoral act of retribution in case it would be less forgiving than it actually is. But these classes of moral luck are not as interesting, in my opinion, as the class related to non-determinate outcomes of actions.[1] That is, cases where an agent is morally lucky because his agency interferes with nature or the agency of other agents in such a way that its immoral beheavior does not lead to an outcome that is considered morally bad.

If moral luck has to be taken serious as a principle of ethical reasoning, that is, if we really agree that, at least to a certain extent, lucky outcomes secure the acting agent from blame, then moral luck presents us with a problem. This problem is that moral luck conflicts with another principle of ethical reasoning, namely, that agents are not morally responsible for actions or outcomes that are not under their control. A case where this conflict between moral principles comes to the foreground, is the following variant on the well-known drinking and driving examples.

Example 1.1 *In* scenario 1 *a man drinks 10 beers. He drives home, knows he is taking a risk, and drives very carefully. After 1 km, he is held by the police. The alcohol percentage in his blood is measured to be 0.15 %. He gets fined 300 Euros.* Scenario 2 *is only slightly different. A man drinks 10 beers. He drives home, knows he is taking a risk, and drives very carefully. However, now after 1 km, he fatally hits somebody crossing the road. After the accident the alcohol percentage in his blood is measured to be 0.15 %. He gets convicted for involuntary manslaughter.*

If the moral responsibility in both scenarios of Example 1.1 is different, than this justifies the difference in legal treatment of the two scenarios; the view would be that driving drunk and killing somebody is morally despicable and is better than driving drunk without hitting anybody. However, the problem is that it is not clear that indeed moral responsibility is different in the two scenarios. The example is designed to make it clear that in scenario 1 and 2 both agents took the same *risk*; until 1 km of driving, the period in which the agents exercised their agency, the two scenarios are entirely identical. And, that the scenarios are indeed identical is to a certain extent also proven to a judge having to decide on the verdicts in both scenarios, since there is objective evidence that in both cases the percentage of alcohol was 0.15 %. So if the agency involved in both scenarios is identical, and if moral responsibility is proportional to the agency or control exercised over a situation, the conclusion should be that the moral responsibility in both scenarios is the same. The conflict is then between two principles: if we commit ourselves to the standpoint that there is a difference in moral responsibility for the two scenarios, than this difference must be linked to the different outcomes of the scenarios. So, in that case we commit ourselves to the principle of (the possibility of) moral luck. The second, conflicting principle is that agents cannot be morally responsible for what is not under their

[1] I do appreciate however Nagel's argument that if we try to deal with the problem of moral luck relative to action effects by simply denying it, the only thing we do is to push the problem to deeper levels of consideration like those concerning character or intention.

control, which implies that, given their set-up, there should not be a difference in assignment of moral responsibility between the two scenarios.

In this article, in Sects. 4, 5 and 6, I will defend the standpoint that moral luck with respect to action outcomes, as described above, does not exist. I will then explain the difference in legal treatment of the two scenarios as resulting from differences between the legal and the moral evaluation of actions: in a moral evaluation they are identical, in a legal evaluation they are not. After all, legal responsibility is not the same as moral responsibility. A father can be legally responsible for wrongdoings of his children without being morally responsible for these same wrongdoings. Or a company can be legally responsible for polluting the environment without being morally responsible (because it is not clear at all if moral responsibility can be lifted to groups or organisations of agents). The problem with this 'way out' is that it raises the question of how much moral and legal responsibility can diverge. It is clear that legal responsibility always reflects at least some underlying level of moral responsibility. First of all, our laws cannot be too different from what people belief to be morally justified, otherwise our social choice mechanisms will correct them. Second, to a certain extent the father is morally responsible, since in a remote way, being responsible for their upbringing, his children's acts derive from his own acts. And the company is maybe also morally responsible in some derived sense since somehow the company's acts are acts of the agents working for the company. These observations ask for clarification, the kind of clarification that in my opinion can best be provided by formalisation.

Before explicating the goal of this paper further, I want to take away a possible misconception. The described problem with moral luck is mostly seen as conflicting with Kantian ethics. Kantian ethics links morality to the action and not to its outcome. So, if in practice we assign morality to outcomes instead of actions, as we seem to do in many cases, we are operating in conflict with Kantian ethics. It might seem that we have to conclude that our practice of assigning morality to outcomes is an argument in favour of consequentialist ethics. However, that would be too hasty a conclusion. Moral luck is just as much a problem for consequentialist ethics as it is for Kantian ethics. Consequentialist ethics takes the position that badness of acts *derives* from the badness of their outcomes. But this dependency on the status of outcomes does not transfer to cases where outcomes are uncertain, that is, to cases where moral luck can be involved. For instance, a consequentialist can argue that his killing is justified or even the good thing to do if it safes more lives than it costs. Now also for the ethical reasoning in judging this consequentialist killer moral luck poses a problem, since if it turns out that this agent's killing does *not* safe the lives of several others, the moral luck (in this case moral misfortune) phenomenon will raise its head and the killer will have a hard time justifying his killing. That is, the killing agent is likely to be judged for the unlucky outcome even though before the killing he was correct to expect that his act would safe more lives than it would cost and was therefore, according to his consequentialist ethical reasoning, the good thing to do.

I strongly belief that we can get a much clearer picture of the situation by giving fomalizations of agency, failure, attempt, negligence and (normative) luck. So this

is what I aim to do in this paper. In order to formalize luck in a normative context, we should resolve three core issues. We need:

1. a way to represent indeterminacy of action
2. a way to express the determination of an agent (risk avoidance, risk taking, negligent acting or refraining, attempt, etc.)
3. a way to represent the moral or legal value of an act relative to some (implicit) normative code.

The first problem we solve by resorting to probabilistic *stit* (Broersen 2011c). In this form of *stit* theory, effects of choices are no longer guaranteed but are obtained with a certain probability. The second problem we solve by endowing the probabilistic *stit* operators with lower bounds on the chance of success and by defining attempt as a maximization of the chance of success (Broersen 2011a). Finally, the third problem is tackled by introducing violation constants in this context (Anderson 1958; Broersen 2011b). If a violation occurs, an agent does not behave according to some moral or legal code that, in this paper, is not made explicit. We can see the paper then as a combination of the ideas put forward in Broersen (2011a, b, c). We will present essential parts of the material from those papers here and then discuss the application to the formalization of moral and legal luck.

2 Modeling Indeterminacy of Action

In the following two sub-sections we first introduce the base stit logic and then extend this logic to allow for non-determinate effects.

2.1 Determinate Action: XSTIT^p

In this section we define the base logic, which is a variant of the logic XSTIT that we call XSTIT^p (Broersen 2013). The difference with XSTIT is embodied by an axiom schema concerning modality-free propositions p, which explains the name. The semantics uses h-relative effectivity functions, which specialize the notion of effectivity function from Coalition Logic (Pauly 2002) by defining choices relative to histories.

Definition 2.1 *Given a countable set of propositions P and $p \in P$, and given a finite set Ags of agent names, and $ag \in Ags$, the formal language $\mathcal{L}_{XSTIT^p}$ is:*

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [ag\ \mathtt{xstit}]\varphi \mid X\varphi$$

Besides the usual propositional connectives, the syntax of XSTIT^p comprises three modal operators. The operator $\Box\varphi$ expresses 'historical necessity', and plays

the same role as the well-known path quantifiers in logics such as CTL and CTL* (Emerson 1990). Another way of talking about this operator is to say that it expresses that φ is 'settled'. We abbreviate $\neg\Box\neg\varphi$ by $\Diamond\varphi$. The operator $[ag\ \texttt{xstit}]\varphi$ stands for 'agent ag sees to it that φ in the next state'. We abbreviate $\neg[ag\ \texttt{xstit}]\neg\varphi$ by $\langle ag\ \texttt{xstit}\rangle\varphi$. The third modality is the next operator $X\varphi$. It has a standard interpretation as the transition to a next state.

Definition 2.2 *An XSTIT^p-frame is a tuple $\langle S, H, E\rangle$ such that*[2]*:*

1. *S is a non-empty set of static states. Elements of S are denoted s, s', etc.*
2. *H is a non-empty set of 'backwards bundled' histories. A history $h \in H$ is a sequence $\ldots s, s', s'' \ldots$ of mutually different elements from S. To denote that s' succeeds s on h we use a successor function succ and write $s' = succ(s, h)$. The following constraint on the set H ensures that if different histories share a state, they are bundled together in the past direction:*
   a. *if $s = succ(s', h)$ and $s = succ(s'', h')$ then $s' = s''$*
3. *$E : S \times H \times Ags \mapsto 2^S$ is an h-effectivity function yielding for an agent ag the set of next static states allowed by the choice exercised by the agent relative to a history. We have the following constraints on h-effectivity functions:*
   a. *if $s \notin h$ then $E(s, h, ag) = \emptyset$*
   b. *if $s' \in E(s, h, ag)$ then $\exists h' : s' = succ(s, h')$*
   c. *if $s' = succ(s, h')$ and $s' \in h$ then $s' \in E(s, h, ag)$*
   d. *$E(s, h, ag_1) \cap E(s, h', ag_2) \neq \emptyset$ for $ag_1 \neq ag_2$.*

In Definition 2.2 above, we refer to the states s as 'static states'. This is to distinguish them from 'dynamic states', which are combinations $\langle s, h\rangle$ of static states and histories. Dynamic states function as the elementary units of evaluation of the logic. This means that the basic notion of 'truth' in the semantics of this logic is about dynamic conditions concerning choices. This distinguishes *stit* from logics like Dynamic Logic and Coalition Logic whose central notion of truth concerns static conditions holding for static states.

The name 'h-effectivity functions' for the functions defined in item **3** above is short for 'h-relative effectivity functions'. This name is inspired by similar terminology in Coalition Logic whose semantics is in terms of 'effectivity functions'. Condition **3.a** above states that h-effectivity is empty for history-state combinations that do not form a dynamic state. Condition **3.b** ensures that next state effectivity as seen from a current state s does not contain states s' that are not reachable from the current state through some history. Condition **3.c** expresses the well-known *stit* condition of 'no choice between undivided histories'. Condition **3.d** above states that simultaneous choices of different agents never have an empty intersection. This represents a constraint on the independence of choices by different agents.

[2] In the meta-language we use the same symbols both as constant names and as variable names, and we assume universal quantification of unbound meta-variables.

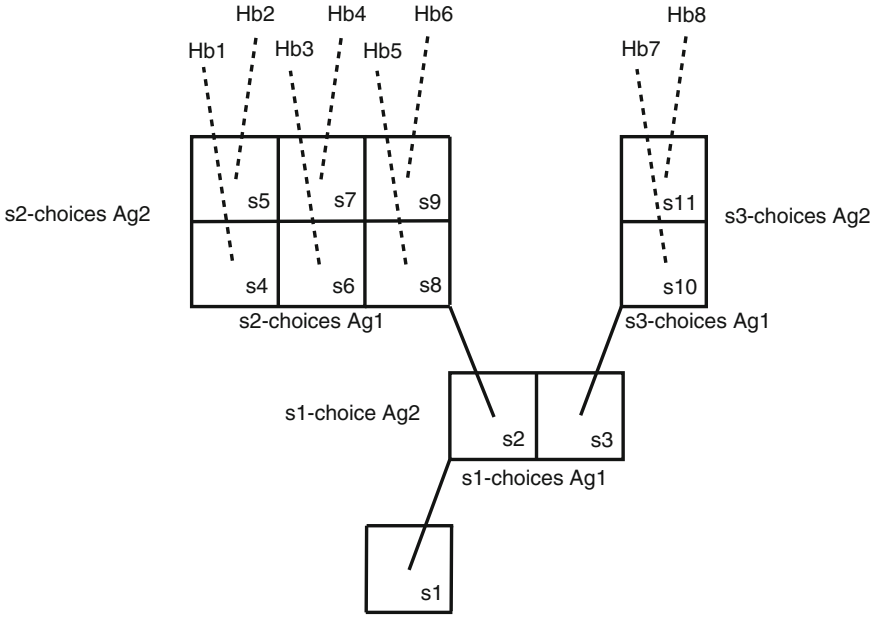


Fig. 1 Visualization of a partial two agent XSTIT^p frame

Figure 1 visualizes a frame of the type defined by Definition 2.2. The columns in the games forms linked to each state are the choices of agent ag_1 and the rows are the choices of agent ag_2. Independence of choices is reflected by the fact that the game forms contain no 'holes' in them. Choice taking in this 'bundled' semantics is thought of as the separation of two bundles of histories: one bundle ensured by the choice exercised and one bundle excluded by that choice. The pictures of the frames suggest more constraints than are actually specified by Definition 2.2. For instance, the technical definition of the frames does not exclude that the choices of an agent ag are mutually disjoint. However, since they result in much tidier pictures, in the visualizations of the frames we assume such conditions.

We now define models by adding a valuation of propositional atoms to the frames of Definition 2.2. We impose that all dynamic states relative to a static state evaluate atomic propositions to the same value. This reflects the intuition that atoms, and modality-free formulas in general do not represent dynamic information. Their truth value should thus not depend on a history but only on the static state. This choice does however make the situation non-standard. It is a constraint on the models, and not on the frames.

Definition 2.3 *A frame $\mathcal{F} = \langle S, H, E\rangle$ is extended to a model $\mathcal{M} = \langle S, H, E, V\rangle$ by adding a valuation V of atomic propositions:*

- *V is a valuation function $V : P \longrightarrow 2^S$ assigning to each atomic proposition the set of static states relative to which they are true.*

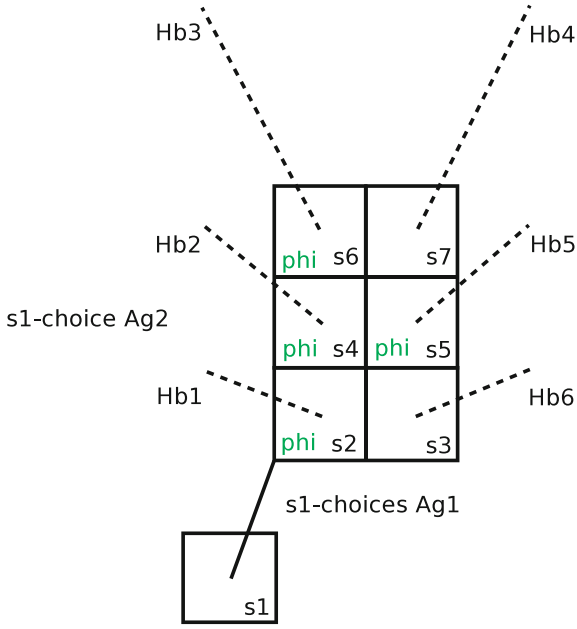


Fig. 2 Visualization of a partial two agent XSTIT^p model

We evaluate truth with respect to dynamic states built from a dimension of histories and a dimension of static states.

Definition 2.4 *Relative to a model* $\mathcal{M} = \langle S, H, E, V\rangle$*, truth* $\langle s, h\rangle \models \varphi$ *of a formula* φ *in a dynamic state* $\langle s, h\rangle$*, with* $s \in h$*, is defined as:*

$$
\begin{array}{ll}
\langle s,h\rangle \models p & \Leftrightarrow s \in V(p) \\
\langle s,h\rangle \models \neg\varphi & \Leftrightarrow \textit{not } \langle s,h\rangle \models \varphi \\
\langle s,h\rangle \models \varphi \wedge \psi & \Leftrightarrow \langle s,h\rangle \models \varphi \textit{ and } \langle s,h\rangle \models \psi \\
\langle s,h\rangle \models \Box\varphi & \Leftrightarrow \forall h' : \textit{if } s \in h' \textit{ then } \langle s,h'\rangle \models \varphi \\
\langle s,h\rangle \models X\varphi & \Leftrightarrow \textit{if } s' = succ(s,h) \textit{ then } \langle s',h\rangle \models \varphi \\
\langle s,h\rangle \models [ag\ \texttt{xstit}]\varphi & \Leftrightarrow \forall s', h' : \textit{if } s' \in E(s,h,ag) \textit{ and} \\
 & \quad s' \in h' \textit{ then } \langle s',h'\rangle \models \varphi
\end{array}
$$

Satisfiability, validity on a frame and general validity are defined as usual.

Note that the historical necessity operator quantifies over one dimension, and the next operator over the other. The *stit* modality combines both dimensions. Now we proceed with the axiomatization of the base logic.

Figure 2 gives an example model that we can use to discuss the evaluation of formulas. Relative to static state s_1 and the history h_5 that is part of the bundle of histories Hb_5 we do not have that the choice by agent ag_1 ensures that φ holds, since the other agent has two choices (the bottom one and the top one) for which φ will not be true. So in this model we have that $\langle s_1, h_5\rangle \not\models [ag_1\ \texttt{xstit}]\varphi$. However, relative

to, for instance, a history in the bundle Hb_1, the agent ag_1 does guarantee that φ obtains as the result of the choice it excerts independent of what agent ag_2 choses simultaneously: for all three choices of the other agent φ is the result. So we have that $\langle s_1, h_1\rangle \models [ag_1\ \texttt{xstit}]\varphi$.

Definition 2.5 *The following axiom schemas, in combination with a standard axiomatization for propositional logic, and the standard rules (like necessitation) for the normal modal operators, define a Hilbert system for* ***XSTIT****^{p}:*

(p)	$p \rightarrow \Box p$ *for p modality free*
	S5 for $\Box$
(D)	$\neg[ag\ \texttt{xstit}]\bot$
(Lin)	$\neg X\neg\varphi \leftrightarrow X\varphi$
(Sett)	$\Box X\varphi \rightarrow [ag\ \texttt{xstit}]\varphi$
(XSett)	$[ag\ \texttt{xstit}]\varphi \rightarrow X\Box\varphi$
(Agg)	$[ag\ \texttt{xstit}]\varphi \wedge [ag\ \texttt{xstit}]\psi \rightarrow [ag\ \texttt{xstit}](\varphi \wedge \psi)$
(Mon)	$[ag\ \texttt{xstit}](\varphi \wedge \psi) \rightarrow [ag\ \texttt{xstit}]\varphi$
(Dep)	$\Diamond[ag_1\ \texttt{xstit}]\varphi \wedge \ldots \wedge \Diamond[ag_n\ \texttt{xstit}]\psi \rightarrow$
	$\Diamond([ag_1\ \texttt{xstit}]\varphi \wedge \ldots \wedge [ag_n\ \texttt{xstit}]\psi)$
	for $Ags = \{ag_1, \ldots, ag_n\}$

Theorem 2.1 (Broersen 2011b) *The Hilbert system of Definition 2.5 is complete with respect to the semantics of Definition 2.4.*

2.2 Action with Non-Determinate Effect: XSTIT.Prob

The *stit* logic of the previous section was based on the idea of acting as ensuring a certain condition. In the present section we put forward a theory that relaxes this assumption. Now actions are no longer necessarily successful. We are going to assume we measure success of action against an agent's beliefs about the outcome of its choice. So, the perspective is an internal, subjective one, and the criterion of success is formed by an agent's beliefs about its action. To represent these beliefs we choose here to use probabilities. In particular, we will represent beliefs about simultaneous choices of other agents in a system as subjective probabilities. Several choices have to be made. We will assume that an agent can never be mistaken about its own choice, but that it can be mistaken about choices of others. The actual action performed results from a simultaneous choice of all agents in the system. Then, if an agent can be mistaken about the choices of other agents (including possibly an agent with special properties called 'nature'), the action can be unsuccessful.

We introduce operators $[ag\ \texttt{xstit}^{\geq c}]\varphi$ with the intended meaning that agent ag exercises a choice for which it believes to have a chance of at least c of bringing about φ. We assume that numbers c are between 1 and 0 and that the set of possible c's is at least countable (that is, a subset of $\mathbb{Q}$). Roughly, the semantics for this new operator

is as follows. We start with the multi-agent *stit*-setting of the previous section. Now to the semantic structures we add functions such that in the little game-forms, as visualized by Fig. 1, for each choice of an agent ag we have available the subjective probabilities applying to the simultaneous choices of other agents in the system. For an agent ag the sum of the probabilities over the choices of each particular other agent in the system adds up to one. So, the probabilities represent agent ag's beliefs concerning what choices are exercised simultaneously by other agents. We use this subjective probability function to define for each choice the change of success to obtain a condition φ: relative to the choice we add up the probabilities for each of the choices of all other agents in the system leading to a situation obeying φ.

For the definition of the probabilistic frames, we first define an augmentation function returning the choices a group of agent has in a given state.

Definition 2.6 *The range function* $Range : S \times Ags \mapsto 2^{2^S \setminus \emptyset} \setminus \emptyset$ *yielding for a state s and an agent ag, the choices this agent has in s is defined as:*
$Range(s, ag) = \{Ch \mid \exists h : s \in h \text{ and } Ch = E(s, h, ag)\}$

A range function is similar to what in Coalition Logic (Pauly 2002) is called an 'effectivity function'. Now we are ready to define the probabilistic *stit* frames.

Definition 2.7 *A probabilistic* XSTIT^p*-frame is a tuple* $\langle S, H, E, B\rangle$ *such that:*

1. $\langle S, H, E\rangle$ *is an* XSTIT^p*-frame*
2. $B : S \times Ags \times Ags \times 2^S \mapsto [0, 1]$ *is a subjective probability function such that* $B(s, ag_1, ag_2, Ch)$ *expresses agent 1's belief that in static state s agent 2 performs a choice resulting in one of the static states in Ch. We apply the following constraints.*
   a. $B(s, ag, ag', Ch) = 0$ *if* $ag \neq ag'$ *and* $Ch \notin Range(s, ag')$
   b. $B(s, ag, ag', Ch) > 0$ *if* $ag \neq ag'$ *and* $Ch \in Range(s, ag')$
   c. $\sum_{Ch \in Range(s, ag')} B(s, ag, ag', Ch) = 1$ *if* $ag \neq ag'$
   d. $B(s, ag, ag, Ch) = 1.$

The conditions in Definition 2.7 are variations on the Kolmogorov axioms for probability. Condition **2.a** says that agents only assign non-zero subjective probabilities to choices other agents objectively have. Condition **2.b** says these probabilities are strictly larger than zero. Condition **2.c** says that the sum of the subjective probabilities over the possible choices of other agents add up to 1. Condition **2.d** says that agents always know what choice they exercise themselves. We may call this property the 'free will' property. In an *objective* view on the choices of an agent, the probabilities for the choices in the agent's repertoire have to add up to 1. From such an objective view point for each of the possible choices we get a chance somewhere between 0 and 1, and the standard Kolmogorov conditions apply. But from the perspective of the agent itself, that is, from the subjective viewpoint taken in this logic, the standard conditions do not apply. Conditional on what choice an agent actually takes, we can say that subjectively, the agent is 100 % sure about what choice it is

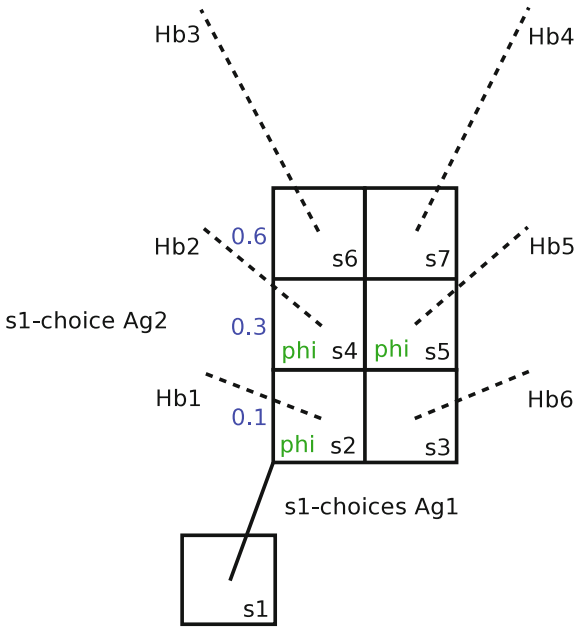


Fig. 3 Visualization of a partial two agent probabilistic XSTIT^p model

taking. And at the same time it has the free will to take any of the choices open to him/her. No agent would regard its own choice taking as a matter of chance: it has the free will to choose and if it chooses something it is 100 % sure about the fact that it is taking that choice. That is what free will demands. So from a subjective viewpoint and taking into account the free will of the agents, we have reason to violate the third Kolmogorov axiom and allow for the possibility that an agent's subjective probabilities concerning its own possibilities for taking choices add up to infinity (if the number of choices is infinite).[3]

Figure 3 extends the earlier example model with subjective probabilities for agent ag_1's belief concerning the choice agent ag_2 exercises simultaneously. We see that agent ag_1 beliefs the the chance that agent ag_2 chooses the top row is 0.6, that the chance for the middle row is 0.3 and the chance for the bottom row is 0.1. It is easy to check that this model satisfies all the conditions discussed above.

In the sequel we will need an augmentation function yielding for an agent and an arbitrary next static state the chance an agent ascribes to the occurrence of this state (given its belief, i.e., subjective probabilities about simultaneous choice taking of other agents). For this, we first need the following proposition. To guarantee that the proposition is true, we need the extra condition that choices do not overlap, which we can safely add to the semantics.

[3] We can also give brief explanations of the determinist and compatibilist positions in this context: a determinist would argue that objectively, the chance for one of the choices is one while for the other choices, they are 0. A compatibilist would then argue that this is compatible with the agent assigning itself probability 1 to any of the choices: it cannot know that its choice is actually determined.

Proposition 2.2 *For any pair of static states s and s' for which there is an h such that $s' = succ(s, h)$ there is a unique 'choice profile' determining for each agent ag in the system a unique choice $Ch = E(s, h, ag)$ relative to s and s'.*

Now we can define the subjective probabilities agents assign to possible system outcomes. Because of the idea of independence of choices, we can multiply the chances for the choices of the individual agents relative to the system outcome (the resulting static state). Note that this gives a new and extra dimension to the notion of (in)dependence that is not available in standard *stit* theories.[4]

Definition 2.8 *$BX : S \times Ags \times S \mapsto [0, 1]$ is a subjective probability function concerning possible next static states, defined by*

$$BX(s, ag, s') = \prod_{ag' \in Ags} B(s, ag, ag', E(s, h, ag')) \text{ with } s' = succ(s, h) \text{ for some } h$$

Note that $BX(s, ag, s')$ expresses agent ag's belief in state s that its choice ends up in s' modulo the assumption that ag actually chooses such as to make s' a possible outcome; if ag chooses such that s' is excluded by its choice, the chance for s' is of course 0.

Now before we can define the notion of 'seeing to it under a minimum bound on the probability of success' formally as a truth condition on the frames of Definition 2.7 we need to do more preparations. First we assume that the intersection of the h-effectivity functions of all agents together yields a unique static state. We can safely assume this, because this condition is not modally expressible. This justifies Definition 2.9 below, that establishes a function characterizing the static states next of a given state that satisfy a formula φ relative to the current choice of an agent.

Definition 2.9 *The 'possible next static φ-states' function $PosX : S \times H \times Ags \times \mathcal{L} \mapsto 2^S$ which for a state s, a history h, an agent ag and a formula φ gives the possible next static states obeying φ given the agent's current choice determined by h, is defined by: $PosX(s, h, ag, \varphi) = \{s' \mid s' \in E(s, h, ag)$ and $\langle s', h' \rangle \models \varphi$ for all h' with $s' \in h'\}$.*

Now we can formulate the central 'chance of success' (CoS) function that will be used in the truth condition for the new operator. The chance of success relative to a formula φ is the sum of the chances the agent assigns to possible next static states validating φ.

Definition 2.10 *The chance of success function $CoS : S \times H \times Ags \times \mathcal{L} \mapsto [0, 1]$ which for a state s and a history h an agent ag and a formula φ gives the chance the agent's choice relative to h is an action resulting in φ is defined by: $CoS(s, h, ag, \varphi) = 0$ if $PosX(s, h, ag, \varphi) = \emptyset$ or else $CoS(s, h, ag, \varphi) = \sum_{s' \in PosX(s,h,ag,\varphi)} BX(s, ag, s')$.*

[4] I believe however, that there is a glitch in the terminology surrounding the phenomena of dependence in *stit* theory. I now prefer to talk about "independence of choices" and belief this corresponds to "dependence of agency".

Extending the probabilistic frames of Definition 2.7 to models in the usual way, the truth condition of the new operator is defined as follows.

Definition 2.11 *Relative to a model* $\mathcal{M} = \langle S, H, E, B, V \rangle$*, truth* $\langle s, h \rangle \models [ag\ \mathtt{xstit}^{\geq c}]\varphi$ *of a formula* $[ag\ \mathtt{xstit}^{\geq c}]\varphi$ *in a dynamic state* $\langle s, h \rangle$*, with* $s \in h$*, is defined as:*

$$\langle s, h \rangle \models [ag\ \mathtt{xstit}^{\geq c}]\varphi \Leftrightarrow CoS(s, h, ag, \varphi) \geq c$$

Using the example model of Fig. 3 we can now discuss truth evaluations on probabilistic *stit* models. As we saw earlier, relative to static state s_1 and the history h_5 that is part of the bundle of histories Hb_5 we do not have that the choice by agent ag_1 ensures that φ holds, since the other agent has two choices (the bottom one and the top one) for which φ will not be true. So in this model we have that $\langle s_1, h_5 \rangle \not\models [ag_1\ \mathtt{xstit}^{\geq 1}]\varphi$. But we do have that $\langle s_1, h_5 \rangle \models [ag_1\ \mathtt{xstit}^{\geq 0.3}]\varphi$ since ag_1 believes that with a chance of 0.3 agent ag_2 exercises the choice of the middle row. But, relative to histories in, for instance the bundle $Hb1$, agent ag_1 has better chances to see to it that φ will be true. In particular we have that $\langle s_1, h_1 \rangle \models [ag_1\ \mathtt{xstit}^{\geq 0.4}]\varphi$, because it can add up the chances of the bottom two rows. Note that this is also true relative to the histories in bundle Hb_3 for which the result is $\neg\varphi$. Here we have a situation where the agent saw to it that φ with a chance of success of at least 0.4, but failed. Also note that situations like these show that it is consistent in the logic to have, for instance, that $[ag_1\ \mathtt{xstit}^{\geq c}]\varphi \wedge [ag_2\ \mathtt{xstit}^{\geq c}]\neg\varphi$, that is, if c is not 1.

The probabilistic *stit* operator we gave in Definition 2.11 faithfully generalizes the *stit* operator of our base XSTIT^p system: the objective *stit* operator $[ag\ \mathtt{xstit}]\varphi$ discussed in Sect. 2.1 comes out as the probabilistic *stit* operator assigning a probability 1 to establishing the effect φ. This is very natural. Where in the standard *stit* setting we can talk about 'ensuring' a condition, in the probabilistic setting we can only talk about establishing an effect with a certain lower bound on the probability of succeeding.

We now give a Hiblert system for the probabilistic stit logic. The system is parametric is probabilistic variables c and k. This means that the system encodes infinitely many axioms, since there can be infinitely many values for c and k. To obtain a standard Hilbert system we can pose a prior limit to the possible values of probabilities.

Definition 2.12 *Relative to the semantics following from Definitions 2.4 and 2.11 we define the following Hilbert system. We assume all the standard derivation rules for the normal modalities X and* $\Box$*. Furthermore, we assume the standard derivation rules for the weak modality* $[ag\ \mathtt{xstit}^{\geq c}]\varphi$*, like closure under logical equivalence.*

(p) $p \rightarrow \Box p$ *for* p *modality free*
S5 for $\Box$
(D) $\neg[ag\ \texttt{xstit}^{\geq c}]\bot$ *for* $c > 0$
(Triv) $[ag\ \texttt{xstit}^{\geq 0}]\varphi$
(Lin) $\neg X \neg\varphi \leftrightarrow X\varphi$
(Sett) $\Box X\varphi \rightarrow [ag\ \texttt{xstit}^{\geq c}]\varphi$
(XSett) $[ag\ \texttt{xstit}^{\geq 1}]\varphi \rightarrow X\Box\varphi$
(Min) $[ag\ \texttt{xstit}^{\geq c}]\varphi \rightarrow [ag\ \texttt{xstit}^{\geq k}]\varphi$ *for* $c \geq k$
(Add) $[ag\ \texttt{xstit}^{\geq c}]\varphi \wedge [ag\ \texttt{xstit}^{\geq k}]\psi \rightarrow [ag\ \texttt{xstit}^{\geq c+k-1}](\varphi \wedge \psi)$ *for* $c + k > 1$
(Mon) $[ag\ \texttt{xstit}^{\geq c}](\varphi \wedge \psi) \rightarrow [ag\ \texttt{xstit}^{\geq c}]\varphi$
(Dep) $\Diamond[ag_1\ \texttt{xstit}^{\geq c}]\varphi \wedge \ldots \wedge \Diamond[ag_n\ \texttt{xstit}^{\geq k}]\psi \rightarrow$
$\Diamond([ag_1\ \texttt{xstit}^{\geq c}]\varphi \wedge \ldots \wedge [ag_n\ \texttt{xstit}^{\geq k}]\psi)$ *for* $Ags = \{ag_1, \ldots, ag_n\}$

Proposition 2.3 (Broersen 2011c) *The Hilbert system is sound relative to the semantics.*

Proposition 2.4 (Broersen 2011c) *The Hilbert system reduces to the complete Hilbert system for xstit after substitution of 1 for the parameter c.*

Note that all axioms for xstit have a natural generalization in the above Hilbert system. The most interesting one is agglomeration that generalizes from the standard normal modal logic axiom (Agg) to the set of weak modal scheme's (Add).

3 Modeling the Determination in Action

The second ingredient of moral luck is the determination of an agent. In particular, moral luck can be described as a moral judgement on the difference between the determination of an agent and the indeterminacy of the result of his action. We define two ways in which to represent the determination of an agent's action. In the first sub-section we argue that the lower bounds given in the previous section already express constraints on the determination on the part of the agent. In the second sub-section we discuss the definition of attempt.

3.1 Risk in Action

Operators of the form $[ag\ \texttt{xstit}^{\geq k}]\varphi$ to a limited extent already express determination on the part of an agent ag. They express that agent ag currently performs a choice where it estimates that k is a lower bound to the probability that φ will obtain. Clearly, this does *not* model determination of the agent in the stronger sense of *intentional* action. But, one can say that the formula gives an 'outer constraint' on the determination of the agent from which some information about the agent's

intentions can be abduced. For instance, if the formula $[ag\ \texttt{xstit}^{\geq 0.8}]p$ is true (the agent chooses an action where it beliefs the chance to achieve p is at least 0.8), while at the same time the formula $\Diamond[ag\ \texttt{xstit}^{\geq 0.1}]p$ is true (the agent *could* have chosen an action where it beliefs the chance to achieve p can be much lower) then we might explain this by abducing that the agent prefers the higher chance to see to it that p. Of course, from this information we cannot deduce that it is the agent's aim to do p; that would be jumping to conclusions. But what this small example shows is that the formulas of the logic of the previous section can already be used to specify constraints on the agent's determination. In the next section we will take this one step further by modeling the notion of 'attempt'.

3.2 Attempt

We see an attempt for φ as exercising a choice that is maximal in the sense that an agent assigns the highest chance of achieving φ to it (Broersen 2011a). So we aim to model attempt as a comparative notion. This means, that in our formal definition for the attempt operator $[ag\ \texttt{xatt}]\varphi$ that we introduce here, we drop the absolute probabilities. Let us first go back briefly to Fig. 3 to explain the intended semantics of attempt. We have that for agent 1, the right choice is not an attempt for φ, since the left choice has a higher probability (0.4 vs. 0.3) of obtaining φ. So we have that $\langle s_1, h_5\rangle \models X\varphi \wedge \neg[Ag_1\ \texttt{xatt}]\varphi$ and $\langle s_1, h_2\rangle \models [Ag_1\ \texttt{xatt}]\varphi$. We can also see in the picture that an attempt is not necessarily successful: $\langle s_1, h_3\rangle \models X\neg\varphi \wedge [Ag_1\ \texttt{xatt}]\varphi$.

We now give the formal definition. The truth condition for the new operator $[ag\ \texttt{xatt}]\varphi$ is as follows.

Definition 3.1 *Relative to a model* $\mathcal{M} = \langle S, H, E, B, \pi\rangle$*, truth* $\langle s, h\rangle \models [ag\ \texttt{xatt}]\varphi$ *of a formula* $[ag\ \texttt{xatt}]\varphi$ *in a dynamic state* $\langle s, h\rangle$*, with* $s \in h$*, is defined as:*

$$\begin{array}{l}\langle s, h\rangle \models [ag\ \texttt{xatt}]\varphi \Leftrightarrow \\ \forall h' : \textit{if } s \in h' \textit{ then } CoS(s, h', ag, \varphi) \leq CoS(s, h, ag, \varphi) \\ \textit{and} \\ \exists h'' : s \in h'' \textit{ and } CoS(s, h'', ag, \varphi) < CoS(s, h, ag, \varphi)\end{array}$$

This truth condition explicitly defines the comparison of the current choice with other choices possible in that situation. In particular, if and only if the chance of obtaining φ for the current choice is higher than for the other choices possible in the given situation, the current choice is an attempt for φ. The 'side condition' says that there actually must be a choice alternative with a strictly lower chance of success.

Proposition 3.1 (Broersen 2011a) *Each instance of any of the following formula schemas is valid in the logic determined by the semantics of Definition 3.1.*

(Cons) $\neg[ag\ \texttt{xatt}]\bot$
(D) $[ag\ \texttt{xatt}]\neg\varphi \rightarrow \neg[ag\ \texttt{xatt}]\varphi$
(Dep-Att) $\Diamond[\{ag1\}\ \texttt{xatt}]\varphi \wedge \Diamond[\{ag2\}\ \texttt{xatt}]\psi \rightarrow \Diamond([\{ag1\}\ \texttt{xatt}]\varphi \wedge [\{ag2\}\ \texttt{xatt}]\psi)$
(Sure-Att) $[ag\ \texttt{xstit}]\varphi \wedge \Diamond\neg[ag\ \texttt{xstit}]\varphi \rightarrow [ag\ \texttt{xatt}]\varphi$

The D-axiom says that the same choice cannot be at the same time an attempt for φ and $\neg\varphi$. This is due to the presence of the 'side condition' in Definition 3.1. The side condition says that a choice can only be an attempt if there is at least one alternative choice with a strictly lower chance of success. Now we see immediately why the D-axiom holds: this can never be the case for complementary effects, since these have also complementary probabilities. In *stit* theory, side conditions are used to define 'deliberative' versions of *stit* operators (Horty and Belnap 1995). And indeed the same intuition is at work here: a choice can only be an attempt if it is 'deliberate'.

The (Indep-Att) schema says that attempts of different agents are independent. Attempts are independent, because maximizing choice probabilities from the perspective of one agent is independent from maximizing choice probabilities from the perspective of some other agent.

Finally, the (Sure-Att) schema reveals the relation between the *stit* operator of our base language and the attempt operator. We saw that we can associate the operator $[ag\ \texttt{xstit}]\varphi$ with a probabilistic *stit* operator with a chance of success of 1. Now, if such a choice qualifies as an attempt, it can only be that there is an alternative to the choice with a probability strictly lower than 1 (due to the side condition in Definition 3.1). In the base language we can express this as the side condition $\Diamond\neg[ag\ \texttt{xstit}]\varphi$ saying that φ is not ensured by ag's choice. This results in the property (Sure-Att) that says that if ag ensures φ with a chance of success of 1, and if ag could also have refrained (i.e., ag took a chance higher than 0 for $\neg\varphi$), then ag *attempts* φ. This again reveals the relation between the notion of attempt and the notion of 'deliberate choice' from the *stit* literature (Horty and Belnap 1995).

4 Moral Obligations, Prohibitions and Luck

The third ingredient of a formalization of both moral luck and legal luck is the normative aspect. In this section we will formalise obligation and prohibition in the moral sense. In the next section we do the same for obligation and prohibition in the legal sense. Adapting the approach put forward in Broersen (2011b) to the case of probabilistic action, we will use Anderson's reduction of normative truth to logical truth (Anderson 1958) to express either the legal or the moral evaluation of the result of an action against ethical or legal normative codes. We do not explicitly represent these normative codes. However, for future research it will be interesting to investigate how moral luck might depend on the specific moral (ethical) normative code used to evaluate actions. The reduction enables us to express normative assertions

about the good or bad determination in an agent's action. We give four definitions. The first is about being morally forbidden to take a risk.

Definition 4.1 *A (moral) prohibition for agent ag to perform a choice by which ag believes to take a risk of at least k to obtain φ, denoted $Forb_{Mor}[ag\ \mathtt{xstit}^{\geq k}]\varphi$, is defined by:*

$$Forb_{Mor}[ag\ \mathtt{xstit}^{\geq k}]\varphi =_{def} \Box([ag\ \mathtt{xstit}^{\geq k}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}]Viol)$$

The definition makes a link between action results in two different realms. The condition φ is an action effect in the physical realm that is subject to the moral prohibition $Forb[ag\ \mathtt{xstit}^{\geq k}]\varphi$. The condition $Viol$ can be thought of as an action effect in *social* reality (Searle 1995). The definition defines prohibition by relating the effects in both realities. In the examples below we will discuss why this formalises *moral* prohibition rather then *legal* prohibition. First we define other deontic modalities. The pattern of Definition 4.1 is repeated in the definitions below. The first is about being morally obliged to preserve a given lower bound on the chance on success. The second and third definition are about being morally forbidden and being morally obliged to attempt.

Definition 4.2 *A (moral) obligation for agent ag to perform a choice by witch ag believes to have a chance of at least k to obtain φ, denoted $Obl_{Mor}[ag\ \mathtt{xstit}^{\geq k}]\varphi$, is defined by:*

$$Obl_{Mor}[ag\ \mathtt{xstit}^{\geq k}]\varphi =_{def} \Box(\neg[ag\ \mathtt{xstit}^{\geq k}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}]Viol)$$

Definition 4.3 *A (moral) prohibition for agent ag to attempt φ, denoted $Forb_{Mor}$ $[ag\ \mathtt{xatt}]\varphi$, is defined by:*

$$Forb_{Mor}[ag\ \mathtt{xatt}]\varphi =_{def} \Box([ag\ \mathtt{xatt}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}]Viol)$$

Definition 4.4 *A (moral) obligation for agent ag to attempt φ, denoted Obl_{Mor} $[ag\ \mathtt{xatt}]\varphi$, is defined by:*

$$Obl_{Mor}[ag\ \mathtt{xatt}]\varphi =_{def} \Box(\neg[ag\ \mathtt{xatt}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}]Viol)$$

We can also make meaningful variants of the definitions where conditions $\neg[ag\ \mathtt{xatt}]\varphi$ are replaced by $[ag\ \mathtt{xatt}]\neg\varphi$. The notions resulting are weaker from a normative perspective, since for these variants the agent is only in violation if it explicitly sees to the bad thing happening. For a non-probabilistic setting, nuances like these are explained in Broersen (2011b).

Note that obligations and prohibitions are moment determinate. This means that their truth value is the same for any history through a specific static state. This is due to the presence of the historical necessity operator '$\Box$' as the first operator in all the

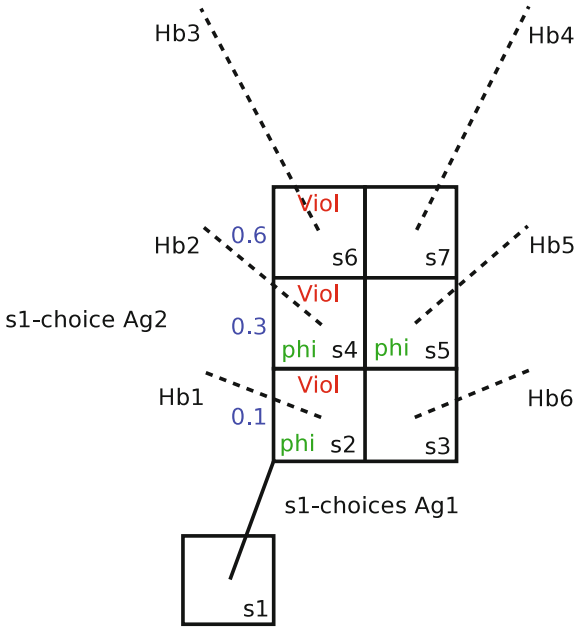


Fig. 4 In state s_6 the agent is lucky that $\neg\varphi$

definitions. So it is assumed here that what an agent is obliged or is forbidden does not depend on what it is doing or on what others are doing.

Before discussing the moral character of the prohibitions and obligations defined, let us look at some example formulas. We discuss these normative formulas by interpreting them in the model of Fig. 4, which adds violation constants to the earlier discussed model of Fig. 3. Our first example formula, which is satisfied in dynamic state $\langle s_1, h_3\rangle$ in Fig. 4 is $Forb[Ag_1\ \texttt{xstit}^{\geq 0.4}]\varphi \wedge [Ag_1\ \texttt{xstit}^{\geq 0.4}]\varphi \wedge X\neg\varphi$. The formula expresses that agent Ag_1 is forbidden to choose in such a way that it believes to have a risk of at least 0.4 to obtain φ, while at the same time agent Ag_1 is actually doing such an action, but, where it is lucky since what is actually happening is that $\neg\varphi$ is obtained (state s_6 in the model). The second example formula is very closely related. It is also satisfied in dynamic state $\langle s_1, h_3\rangle$ of Fig. 4. The formula is $Forb[Ag_1\ \texttt{xatt}]\varphi \wedge [Ag_1\ \texttt{xatt}]\varphi \wedge X\neg\varphi$ and it expresses that agent Ag_1 is forbidden to attempt φ, while at the same time that is actually what it is doing, but, where it is lucky since what is actually happening is that $\neg\varphi$ is obtained (again, state s_6 in the model).

So, for both example formulas state s_6 is a state of luck. The condition φ the agent according to the normative part of the formula is supposed to avoid, is indeed not true in it, but that is not due to the determination of the agent, but due to the lucky coincidence that the other agent took the top row as its choice. The agent is lucky in both example formula situations; the first formula specifies how the agent deliberately took a significant risk for φ and the second formula specifies how the agent even explicitly attempted φ. So definitely, the state s_6 is a state of luck. But, it

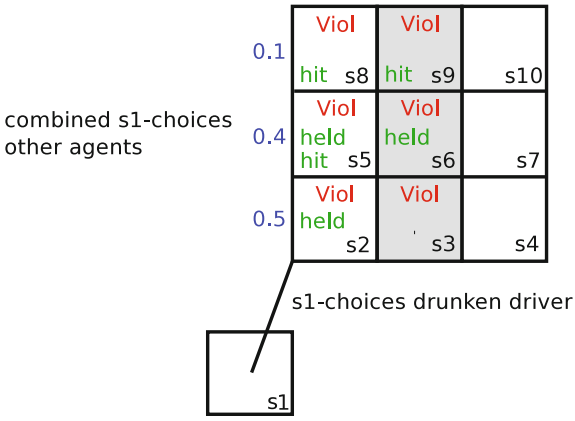


Fig. 5 Moral violations by the drunken driver

is not of the moral kind. Even thought it is justified to say that the agent is lucky, there is still a violation. The violation is due to the fact that the agent exercised the wrong *choice*; one that went against its moral obligations or prohibitions. So, although the outcome is not bad, there is still a violation, representing that the agent was morally wrong. But if that is the interpretation, then in this setting moral luck does not exist. This observation connects to one made by Bernard Williams himself. Williams said (Williams 1993) that when he coined the term 'moral luck' he thought it would be an oxymoron. This formalisation represents that initial opinion; agents are *not* morally lucky if the outcome by coincidence is according to the moral obligations and prohibitions, because in that case there is still a *moral violation*.

We are now in the position to come back to the drunken driver example. We discuss the two scenario's of the example by relating them to the model of Fig. 5. First there are several things to explain about this model's representation of the scenario. In the model, the middle 'grey' column choice is the choice taken by the driver; the choice to drive after drinking. The short hand *hit* stands for the driver's car hitting a person. The short hand *held* stands for the driver being held by the police. Now the first important thing to be explained about the modelling of the scenarios by the pictured model concerns the chosen granularity of the actions. In the model, the exact granularity of the choices is left unspecified. Of course, the choice of the driver to take the risk and drive is not what game theorists would call a 'one shot' action; it typically involves several choices in a row. The driving itself, that at least takes place for 1 km in the scenarios, is an action/process with duration that stretches out from the initial moment of choosing to the moment of hitting/being held by the police. It can be that during this period the agent reconsidered but saw no possibility to undo its choice, or it can be that the agent reinstated its decision, being reinforced by the idea that so far everything went smoothly. But all this is abstracted away from. Game theorists would say that the 'extensive' game form is normalised to a 'normal' game form. Actually, the possibility to look at scenarios more abstractly by normalising the situation is the big advantage of *stit* formalisms. If we would

have to model the same scenario in a dynamic logic (Harel et al. 2000) or situation calculus (McCarthy 1963) formalism, or if we would develop a Davidsonian event-based theory (Davidson 1980), we would have to commit to some bottom level of action/event description. *Stit* theory, on the other hand, allows us to take *any* level of granularity in the description of action. And here we assume exactly the right level of description for the problem at hand.

A second modelling choice to explain are the choices of the other agents in the scenario. There are at least two other agents: a police man and the person risking being hit. In the picture we combine their choices. Actually, we will not be clear about what the exact combined choices of these two other agents are. We see that the model assigns three such choices to this sub-group (the three rows), without making explicit what combined choices they represent. In particular, we do not assign probabilities to the choices of the drunken driver as seen from the perspective of the other two agents. Without these probabilities, we cannot say much about the character of the three row choices. But that is not a problem. The example is about the choices of the drunken driver, and the choices for the drunken driver are clear. There is the right-most choice, which is the choice of avoiding being held by the police and avoiding hitting any person (maybe taking a taxi). There is the middle choice of taking the risk to be held by the police (subjective chance of 0.4) and hitting a person (subjective chance of 0.1). The reason the agent takes this risky choice is that he hopes to end up in state s_3, that is, in the state where he reaches home without any problems. The left-most choice is one with an optimal chance of either being held by the police or hitting a person, or both. This choice is there in order to make clear that the middle choice is not one that optimises the chances of either hitting a person or being held by the police, that is, the agent does not attempt to hit a person, and does not attempt to be held by the police; it is just being negligent. A more precise model of the situation, with all relevant choices of all three involved agents explicitly represented would be much more extensive. Here we only represent the choices of the drunken driver to discuss the phenomenon of moral luck.

The question raised by the possibility of moral luck is whether or not we should judge occurrence of the situation s_6 and occurrence of the situation s_9 differently. In practice we seem to do that, as the drunken driver example aims to portray. If it is s_6 that comes out of the agent's choice, it is fined 300 Euro, and if it is s_9 that comes out, it is convicted for manslaughter, even though both are the result of exercising the middle 'grey' choice in the figure, which means that the agency involved in both possibilities is exactly the same. And of course, as the model shows, the agent can be even more lucky and end up in state s_3. This possibility is the reason that the agent exercises this choice in the first place. Then, in state s_6, the agent is lucky and not lucky at the same time. It is lucky, because it could have ended up in state s_9 which would be worse. But it is also unlucky, because it could have ended up in state s_3, which would have been much better. But again we can ask the question: what kind of luck are we considering here? If it would be genuine moral luck, and if violations are moral violations, then the modelling is problematic, since our definitions demand that all three states s_3, s_6 and s_9 are violation states. The solution I want to suggest is that there can be a justification different from a moral justification for the fact that

there is a difference in the legal treatment of the different outcomes. The justification can be found by making clear the purpose of our legal systems. Before I give the argument, we will look at the formalisation of legal prohibition and obligation.

5 Legal Obligations, Prohibitions and Luck

The drunken driver example is built around the phenomenon that the legal evaluation of the different outcome states is likely to be different. Of course, the exact differences depend on the legal normative code relative to which we evaluate the outcomes. But, this will not be of our concern. Instead our concern will be the formalisation of legal obligation and prohibition. For the formal definition of the legal versions of these deontic operators we will have to consider different violation conditions. Figure 6 gives a possible model for such legal violation conditions in the drunken driver example. We see two differences with the model of Fig. 5: first, violations for the decision to drink and drive only occur if indeed the agent is either held by the police or hits a person, and second, the violations for being held and hitting a person are different, which reflects that legal systems evaluate such outcomes differently. We can now adapt Definition 4.1 and the other definitions for moral obligations and prohibitions to the legal context. The formal definitions reflect the dependency on outcomes by introducing an extra condition on effects in the normative realm (which is part of social reality), namely that the effect indeed must have occurred. This would bring us to the following characterisation of the legal prohibition to take a risk.

$$Forb_{Leg}[ag\ \mathtt{xstit}^{\geq k}]\varphi = \Box([ag\ \mathtt{xstit}^{\geq k}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}](\varphi \rightarrow Viol))$$

However, we can use the logic to simplify this characterisation. Since the operator $[ag\ \mathtt{xstit}^{\geq 1}]\varphi$ is normal (which is not the case for other values of k), it obeys the K axiom. Furthermore, we have that $[ag\ \mathtt{xstit}^{\geq 1}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq k}]\varphi$. Using these properties we can show that the above definition, for any specific instantiation of the propositional meta-variable φ, is equivalent to the characterisation as in the following definition.

Definition 5.1 *The legal prohibition to exercise a choice for which there is a subjective risk of at least k that it has φ as an outcome, denoted $Forb_{Leg}[ag\ \mathtt{xstit}^{\geq k}]\varphi$, is defined by:*

$$Forb_{Leg}[ag\ \mathtt{xstit}^{\geq k}]\varphi =_{def} \Box([ag\ \mathtt{xstit}^{\geq 1}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}]Viol)$$

Here we see how in a legal definition of prohibition that is strictly based on outcomes, the subjective element, in the context of this prohibition represented by the number k, is eliminated from the definition; the only things that count are if φ

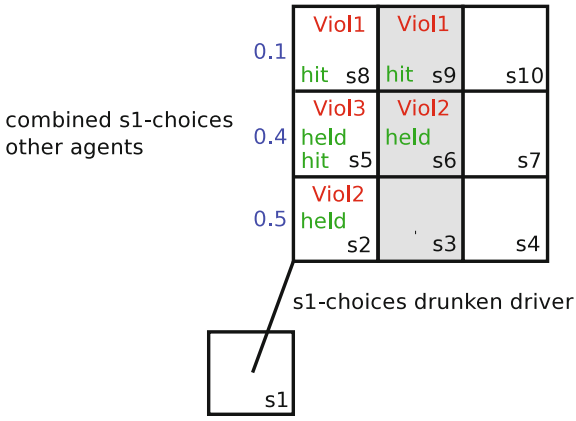


Fig. 6 Possible legal violations by the drunken driver

indeed occurs and if it occurs due to the involvement of agent ag. Using a similar line of reasoning we come to the following characterisation of legal obligation.

Definition 5.2 *The legal obligation to exercise a choice for which there is a subjective chance of at least k that it has φ as an outcome, denoted $Obl_{Leg}[ag\ \mathtt{xstit}^{\geq k}]\varphi$, is defined by:*

$$Obl_{Leg}[ag\ \mathtt{xstit}^{\geq k}]\varphi =_{def} \Box(\neg[ag\ \mathtt{xstit}^{\geq 1}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}]Viol)$$

For coming to definitions of 'the legal prohibition to attempt' and 'the legal obligation to attempt', we cannot assume a property like $[ag\ \mathtt{xstit}^{\geq 1}]\varphi \rightarrow [ag\ \mathtt{xatt}]\varphi$. This property does not hold, because an attempt cannot be an attempt if there is no alternative (see Proposition 3.1). This means that we cannot perform the same elimination as for the 'risk' versions of the operators, as given above. We come to the following definitions.

Definition 5.3 *The legal prohibition to attempt, denoted $Forb_{Leg}[ag\ \mathtt{xatt}]\varphi$, and the legal obligation to attempt, denoted $Obl_{Leg}[ag\ \mathtt{xatt}]\varphi$, are defined by:*

$$Forb_{Leg}[ag\ \mathtt{xatt}]\varphi =_{def} \Box([ag\ \mathtt{xatt}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}](\varphi \rightarrow Viol))$$

$$Obl_{Leg}[ag\ \mathtt{xatt}]\varphi =_{def} \Box(\neg[ag\ \mathtt{xatt}]\varphi \rightarrow [ag\ \mathtt{xstit}^{\geq 1}](\neg\varphi \rightarrow Viol))$$

The definitions do not reflect that violations for different outcomes are different; they only reflect that a violation depends on the occurrence of a bad outcome as such. But different violations for different bad outcomes are easily added to the picture. We can work with separate violation constants for violations of separate prohibitions and obligations.

6 Discussion

The two formalizations I have given, the one of Sect. 4 and the one of Sect. 5 represent two extremes. In the formalisation of moral deontic operators in Sect. 4 violations are the result of the moral evaluation of an agent's subjective choices. If an agent attempts something wrong, or is negligent by taking a risk, it is in violation, independent of the outcome. In the formalisation of Sect. 5 we have the other extreme: the choices themselves are not evaluated, but only their outcomes. And this reflects legal practice; legal systems cannot inspect the subjective considerations accompanying choices of agents and have to rely on outcomes. But, of course, this does not yet explain why in our example hitting a person and being held by the police are evaluated differently by the legal system while there is objective evidence (0.15 % alcohol) that the subjective risk-taking behind both scenarios is the same. There are several possible explanations for this phenomenon. The first is that legal evidence of similarity in risk taking is still 'only' evidence. It does not *proof* with 100 % certainty that the risk-taking in both situations was the same. And this means that there will always be some influence from the actual outcome on the evaluation of the level of risk-taking; we can gather as much evidence about the similarity in risk-taking as we can find, still there will always be the suspicion that in the situation where the outcome was worse, the risk-taking was higher. A second explanation for the phenomenon is that legal systems are not so much directed at the regulation of individual behaviour but at the regulation of societies of agents. Legal systems are not always fair towards individuals; they sacrifice fairness towards the individual to the general benefits of regulation for the society as a whole. It is generally felt that it gives the wrong signal to other possible offenders to let somebody who drank ten beers and fatally hits a person get away with a 300 Euro fine. And it does not make a difference if it is true that if instead this agent would have been held by the police, that same 300 Euro would have been the fine for drinking and driving. Furthermore, if specialists like police investigators, lawyers and judges will have difficulty assessing the similarity in risk taking for the two scenarios, then certainly the general public will. The society as a whole will simply demand higher penalties for outcomes that are less lucky, so that is what the laws of our legal systems reflect. Indeed, this argument ultimately boils down to the observation that our legal systems cannot avoid a certain level of scapegoat justice.

Given the observed differences between the moral and legal evaluation of actions, there is an obvious explanation for the problem introduced by the phenomenon of moral luck: our views about the moral assessment of actions are influenced and obscured by our legal views on the matter. So, if that is true, then moral luck does indeed not exist, and the luck involved in the normative assessment of outcomes is always of the legal kind. The confusions surrounding the concept of moral luck are then due to the influence of our legal views on our views on morality.

7 Conclusion

In this paper we considered a formal approach to the understanding of the problem of moral luck. We had to take three steps: (1) we had to account for indeterminacy of action effects, (2) we had to account for the determination in agentive choice, and (3) we had to define the normative evaluation of action. Luck was described as a difference between an agent's determination (i.e., aspect 2) and the outcome of his action (i.e., aspect 1) in the light of a normative assessment of the situation (i.e., aspect 3). The first result of these efforts is a logic framework where we can reason with moral prohibitions and obligations and legal prohibitions and obligations in the context of risk taking actions. The second result is an explanation for the phenomenon of (the belief in) moral luck. Arguably we could also have found this explanation without the formalisation of the notions involved. But I think the formalisation has helped to arrive at the explanation sooner, and helps to argue for its plausibility in a more convincing way.

Several things are left to investigate. A first thing to do would be to find out what the logics are of the defined operators. But more interesting would be to relate the theory put forward here to legal theories about risk taking. For instance, an agent can go legally wrong if it takes risks a 'normal' person would not take. Here we see that a third form of probability is involved; the probability associated with 'normal' risk taking behaviour. We can then say that three forms of probability are relevant for the theory: objective probabilities, that is, likelihood information about what is objectively going on; subjective probabilities, that is, information in probabilistic form about the risks an acting agent believes to be taking; and 'normal' probabilities, that is, probabilistic information about the risk a normal or average person would be taking (this is best thought of as the probabilistic version of common belief). In particular the relation between the latter two forms is essential for the legal assessment of actions and their outcomes. We will have to leave this interesting subject to future research.

References

Anderson, A.R. 1958. A reduction of deontic logic to alethic modal logic. *Mind* 67:100–103.

Broersen, J.M. 2011a. Modeling attempt and action failure in probabilistic stit logic. In *Proceedings of twenty-second international joint conference on artificial intelligence (IJCAI 2011)*, ed. Toby Walsh, 792–797.

Broersen, J.M. 2011b. Deontic epistemic stit logic distinguishing modes of mens rea. *Journal of Applied Logic* 9(2):127–152.

Broersen, J.M. 2011c. Probabilistic stit logic. In *Proceedings 11th european conference on symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU 2011),* vol. 6717 of *Lecture notes in artificial intelligence*, ed. W. Liu, 521–531. Springer.

Broersen, J.M. 2013. Probabilistic stit logic and its decomposition. *International Journal of Approximate Reasoning* 54:467–477. Elsevier.

Davidson, D. 1980. *Essays on actions and events*. Oxford: Clarendon Press.

Emerson, E.A. 1990. Temporal and modal logic. In *Handbook of theoretical computer science, volume B: Formal models and semantics*, ed. J. van Leeuwen, Chap. 14, 996–1072. Amsterdam: Elsevier Science.

Harel, D., D. Kozen, and Tiuryn, J. 2000. *Dynamic logic*. Cambridge: The MIT Press.

Horty, J.F., and N.D. Belnap. 1995. The deliberative stit: A study of action, omission, and obligation. *Journal of Philosophical Logic* 24(6):583–644.

McCarthy, John. 1963. Situations, actions, and causal laws. *Technical report*. Stanford: Stanford University.

Pauly, Marc. 2002. A modal logic for coalitional power in games. *Journal of Logic and Computation* 12(1):149–166.

Searle, John. 1995. *The construction of social reality*. New York: The Free Press.

Thomas, Nagel. 1979. Moral luck. In *Mortal questions*, 24–38. New York: Cambridge University Press.

Williams, Bernard. 1982. Moral luck. In *Moral luck*, 20–39. Cambridge: Cambridge University Press.

Williams, Bernard. 1993. Postscript. In *Moral luck*, ed. D. Statman. Albany: State University of New York Press.

Worlds Enough, and Time: Musings on Foundations

Mark A. Brown

Abstract Belnap's work on stit theory employs an Ockhamist theory of branching time, in which the fundamental possibilia within models are commonly taken to be moments of time, connected into a tree-like branching structure. In the semantics for alethic modal logic, necessity is characterized by quantification over relevant possible worlds within a model, yet Belnap refers to an entire model of branching time as our world, seemingly leaving no room for non-trivial quantification over worlds within a single model. This chapter explores the question how the notion of possible worlds should be understood in relation to an Ockhamist framework, in order to be able to combine an account of alethic modalities with an account of branching time and stit theory. The advantages and drawbacks of several alternative approaches are examined.

1 Core Features of Ockhamist Branching Time

Systems of logic based on Ockhamist models of branching time[1] offer rich opportunities for the representation of concepts that are not as readily or as sensitively representable in systems based on possible worlds. There are a number of such systems based on models of branching time, suitable for interpreting a variety of formal languages, and models for different systems will differ in various ways, having different components, different constraints on their structure, and different associated satisfaction conditions.

[1] I distinguish these from models of branching spacetime, which have a more complex structure. Some of the remarks made here about models of branching time will have analogous application to models of branching spacetime, however.

M. A. Brown (✉)
Department of Philosophy, Syracuse University, 541 Hall of Languages,
Syracuse, NY 13244, USA
e-mail: mabrown@syr.edu

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_5,

Some broad features of the semantics, however, are held in common. A branching time model $\mathcal{M}$ for a language $\mathcal{L}$ will include among its components some non-empty set **M** of moments, a binary ordering relation $<$ among the moments, and a valuation $\mathcal{V}$, assigning an extension to each atomic formula of the language and, more generally, an extension to each non-logical constant of the language, at each point of valuation in the model.[2] Some structural constraints are also held in common: in particular, $<$ is constrained to be a strict partial ordering with no backwards branching:

(1) $\neg(m < m)$ $(m \in \mathbf{M})$

(**irreflexivity**)

(2) $m_1 < m_2 \,\&\, m_2 < m_3 \Rightarrow m_1 < m_3$ $(m_1, m_2, m_3 \in \mathbf{M})$

(**transitivity**)

(3) $m_1 < m_3 \,\&\, m_2 < m_3 \Rightarrow m_1 < m_2 \vee m_1 = m_2 \vee m_2 < m_1$ $(m_1, m_2, m_3 \in \mathbf{M})$

(**no backwards branching**)

The partial ordering induces a tree structure and within any tree a maximal linearly ordered subset of moments is called a *history*. We shall let **H** be the set of all histories in the model. For the logic of action, other components may be added, most commonly a non-empty set **A** of *agents* and an associated *choice function* **C**, assigning to each agent a, at each moment m, a partition $\mathbf{C}_a^m$ of the set $\mathbf{H}_m$ of all histories passing through m. In such cases, two further constraints are imposed:

(4) $m < m' \in h_1 \cap h_2 \Rightarrow c_1 = c_2$ $(c_1, c_2 \in \mathbf{C}_a^m; h_1 \in c_1; h_2 \in c_2)$

(**no choice between undivided histories**)

(5) $a \neq b \Rightarrow c_1 \cap c_2 \neq \varnothing$ $(a, b \in \mathbf{A}; c_1 \in \mathbf{C}_a^m; c_2 \in \mathbf{C}_b^m)$

(**independence of agents**)

Pairs m/h in which moment m falls within history h are called *points of evaluation*.[3] The semantics for branching time systems is normally Ockhamist: formulas are evaluated at *points of evaluation* rather than simply at moments, with the result that a future-tense statement, in particular, may be true at a given moment relative to one of its histories into the future, but false at the same moment relative to another. As I stand here at a given moment deliberating whether to visit my mother, along some histories my mother will soon be happy because I visited her, while along others she will soon be disappointed because I didn't. There is, at this moment, no simple fact of the matter about whether she will soon be made happy or soon be disappointed.

[2] We shall soon examine the question just what these points of valuation should be.

[3] We distinguish these, at least temporarily, from points of *valuation*.

Some propositions will, however, be *settled* at a given point of evaluation: *settled true* if true at the given moment along *each* of the histories through that moment; *settled false* if false along each.

The Ockhamist satisfaction conditions for the most basic operators, then, are these:

(**SC**H)	$m/h, \mathcal{M} \models \mathsf{H}p$	iff $(\forall m^* < m)[m^*/h, \mathcal{M} \models p]$;	(∀ past)
(**SC**P)	$m/h, \mathcal{M} \models \mathsf{P}p$	iff $(\exists m^* < m)[m^*/h, \mathcal{M} \models p]$;	(∃ past)
(**SC**G)	$m/h, \mathcal{M} \models \mathsf{G}p$	iff $(\forall m^*: m < m^* \in h)[m^*/h, \mathcal{M} \models p]$;	(∀ future)
(**SC**F)	$m/h, \mathcal{M} \models \mathsf{F}p$	iff $(\exists m^*: m < m^* \in h)[m^*/h, \mathcal{M} \models p]$;	(∃ future)
(**SC**Sett)	$m/h, \mathcal{M} \models \mathsf{Sett}\, p$	iff $(\forall h^*: m \in h^*)[m/h^*, \mathcal{M} \models p]$.	(settled)

In our discussion in this chapter, we will assume that our models do have, or have the functional equivalent of, a non-empty set of moments, an ordering relation obeying constraints 1–3, and a valuation, that they have Ockhamist satisfaction conditions, and that if they also have agents and a choice function (or some functional equivalent), these obey constraints 4 and 5 above. We observe, moreover, that branching time models seem to reflect arrays of possibilities of various sorts, and consequently it is not evident that any additions to the models would be required—or even that any would be appropriate—in order to support alethic modalities.

These rather minimal assumptions leave open a number of options. There are four notably unspecified details of interest to us as we use such systems as the foundation of a logic of action and as we compare such systems with ones based on possible worlds:

(i) does it make sense to suppose distinct moments can be simultaneous?
(ii) are the points of valuation moments or are they instead moment/history pairs?
(iii) in such models, what should count as a possibility—i.e. as a possible world?
(iv) do we require that the set of moments in a model be pastwards connected?

2 Newton Versus Einstein

In common discourse, we expect to be able to say where we would be *at this time* if we had taken a different road. Einstein says that strictly speaking, we can't: that there is no such thing as absolute simultaneity. Our common parlance is based on intuitions which are more Newtonian than relativistic, and the Newtonian view includes a linear conception of time, rather than a branching one. One may wonder, however, whether the non-relativistic aspect of Newtonian physics might be separable from its assumption of temporal linearity. Is it reasonable to entertain an account of branching time which makes room for simultaneity between moments on distinct branches, and thus in this respect is Newtonian rather than relativistic?

Some systems of branching time logic include an equivalence relation **I** among moments. In such models the equivalence classes under **I** are called *instants*, and

appropriate constraints are imposed to ensure that each history intersects each instant at exactly one moment, and that the ordering of moments on any history induces a linear ordering among instants that is independent of the choice of history. Such systems are undoubtedly internally consistent. The question, however, is whether they are conceptually coherent and, further, whether they are of use in a relativistic world despite their non-relativistic character.

To answer this, we may begin by noting that Newtonian mechanics remains of great utility even in a relativistic world. At normal velocities, over normal distances, for everyday purposes, relativistic effects are for the most part negligible. Using the more complex apparatus of relativity theory to solve normal engineering problems would be needlessly complex and inexcusably inefficient. Similarly, we may hold, a logic of branching time with instants may for most ordinary purposes be only negligibly inaccurate compared with a relativistic logic of spacetime, and might be rewardingly more efficient to use.

There remains, however, some doubt about whether the addition of instants to a theory of branching time actually produces benefits in proportion to the additional complexity it introduces. Thus far, instants have found application chiefly in the characterization of one operator in the logic of action, the achievement *stit*, or *astit*, operator.[4] This is used to express the claim that an agent has by her action achieved the state expressed by a certain sentence *s*, and explains this as the claim that she had, at some prior moment, a choice which, exercised as it was, assured that whatever else intervened, *s* would be true at this instant, but exercised differently might under some conditions have left *s* false at this very instant. While the notion of an instant plays an essential role in this definition, it is also the weak point in the concept, from the point of view of applicability, since we seldom know, or care, what would have been true *at this very time* (even in the relativistically innocent sense we might call *nominal time*: *when the clock there reads the same as the clock here does now*) along other histories. Rather, we are likely to be concerned about what would have been true *at approximately this time*, or even in some cases just *eventually*.

When asked at 3 o'clock who shut the door, I claim that I did, some five minutes ago. In doing so, I need not claim that, given what I did, and no matter what happened between 2:55 and now, the door would have been shut when the clock struck 3. I need only maintain that no matter what happened between the time when the clock read 2:55 and the time the clock read 3:00 the door would have been shut *for a while* and that, as it happens, it has remained shut until now. There was perhaps some risk that someone would open it again in the interval, but in fact nobody did. This claim makes no use of the notion of the same time in other circumstances and thus doesn't require instants. Accordingly, it appears that instants are of little practical value as a component of a system of logic of branching time.

This suggests the following strategy: to omit instants, leaving a system which, though not fully relativistic, is at least not in conflict with relativity, and in those infrequent cases in which it seems desirable to compare moments chronologically

[4] Here, '*stit*' is an acronym for 'sees to it that', used in naming any of a variety of operators in logics of action based on branching time.

across histories, to do so only by comparing the states of clocks and calendars, or other relevant indicia, at those moments.

It suggests, as well, that another temporal operator, often used in the literature on linear time, might also be kept available in the discussion of branching time, to express the notion *for an interval (however brief) of time*. The satisfaction condition could be given as

(**SCT**) $m/h, \mathcal{M} \models \mathsf{T}p$ iff $(\exists m'' > m)(\forall m'\colon m < m' < m'')[m'/h, \mathcal{M} \models p]$ (for a time)

3 The Enigmatic Present

In the indeterminist philosophy which the logic of branching time is intended to capture, we consider the past settled, but the future unsettled: there is only one path back, but there are many forward. The present, however, is somewhat enigmatically situated on the cusp between these two: is it settled, like the past, or unsettled, like the future? (Often "unsettling", to be sure, but *unsettled*?).

The technical issue associated with this query is this: should the valuation $\mathcal{V}$ in models for branching time assign values to atomic sentences at *moments* or should it instead assign such values at *moment/history pairs*? We might call these the **static** and **dynamic** views of moments, respectively.

When we consider moments to be analogs of possible worlds, each associated with a maximal consistent set of basic facts and their consequences, we expect the atomic formulas to be true or false at a moment, even though more complex formulas involving tense operators must be evaluated at moment/history pairs. This is the view that a moment is a state through which some histories pass, and that the atomic formulas of the language are purely stative and should therefore be determinately true or false at any given state. On such a static view, the valuation should assign a truth value to each atomic formula at each moment, rather than at each point of evaluation. This has the odd result that some formulas get truth values at moments, but most only get truth values at moment/history pairs. We then have an odd contrast between points of *valuation* (moments) and points of *evaluation* (moment/history pairs), and a correspondingly odd contrast in treatment between atomic and non-atomic sentences.

If, on the other hand, we hold that formulas—*all* formulas—have truth values only with respect to a point of evaluation—a moment/history pair—then in the absence of further constraints the valuation is free to assign different values to the same atomic formula at the same moment, but along different histories. Then, for example, "the cat is alive" can be true at m/h_1 and yet "the cat is dead" be true at m/h_2. If the cat in question is Schroedinger's, suffering from that malaise known as quantum uncertainty, this might seem a desirable feature of the system.

Similarly, "I choose c_1" and "I choose c_2" could be true at the same moment—the moment of choice—but along distinct histories. On this view, then, atomic formulas do not (or perhaps do not all) express states. They are (some of them, at least)

more dynamic than static. On this dynamic view, the present is then in some degree unsettled, just as the future is, because the present sometimes prefigures aspects of the future.

(We might entertain the thought that there really is no present—only the past and the future and the *distinction* between the two. In that case, however, the very existence of the present tense, ubiquitous though it is in language, would have to be considered metaphysically misleading. Moreover, and more important for our limited topic here, the notion of moments of time would lose its place in our theory, with intervals assuming the dominant role.)

By way of contrast we see that on the static view an *atomic* formula expressing the claim "I choose c_1" really has no place in the language. It cannot be true at the very moment at which there is a real choice between c_1 and c_2, because then at any later point of evaluation m/h in a history within choice c_2 it would be true both that I had and that I had not chosen c_1. To avoid such a contradiction we would have to hold that there are no moments at which I choose—only moments at which I have chosen.

Earlier, I said that we consider the past settled and the future unsettled. But strictly speaking, if the future is unsettled the past is as well, in the sense that not all assertions about the past will be settled true. For example if a future tense formula $\mathsf{F}p$ is not settled (is neither settled true nor settled false) at a point of evaluation m/h, then the formula $\mathsf{PF}p$ is not settled either. If $\mathsf{F}p$ is satisfied at m/h but not at m/h', for example, then $\mathsf{PF}p$ will likewise be satisfied at m/h but not at m/h'. So what, then, is settled about the past? The original thought was that the facts of the *pure* past (uncluttered by embedded references to the future) are settled, because there is no pastwards branching to provide alternate pasts. If individual moments may be unsettled, even with respect to sentences including no overt reference to the future, then this thought might seem to fall apart. Actually, however, it is not quite as bad as that.

Suppose that, at a moment m, an atomic sentence s is true along some histories through m, but false along others, and consider how things will look from the perspective of a later moment m'. Various histories run through m'. In order to be able to say that the past is settled, we need to be assured that at m' the truth value of a past-tense sentence such as $\mathsf{P}s$ will be the same no matter which history through m' we are considering. So let's consider two such histories, h_1 and h_2. Since m' is later than m, both these histories run through m as well. If s is true at m/h_1 but false at m/h_2, then $\mathsf{P}s$ will be true at m'/h_1 but might be false at m'/h_2. To avoid this risk, and save the settledness of the past, all that would be required would be to add a constraint reminiscent of the principle *no choice between undivided histories*, namely:

(4*) $m < m' \in h_1 \cap h_2 \Rightarrow \mathcal{V}(s, m/h_1) = \mathcal{V}(s, m/h_2)$ (for s atomic)

(**no momentary differentiation between undivided histories**)

This would still permit us to have $\mathcal{V}(s, m/h_1) \neq \mathcal{V}(s, m/h_3)$ for any history h_3 that doesn't run through m', and thus would be compatible with holding that the present can be unsettled.

This begins to feel like a very artificial and unmotivated position, however. Perhaps the uneasiness it provokes can be seen this way: If it is possible to have $\mathcal{V}(s, m/h_1) \neq \mathcal{V}(s, m/h_3)$, then it is no longer clear what it means to say that the two histories h_1 and h_3 are merged at that moment m or, to put it the other way around, what it means to say that m is really the same moment in both histories. The contingent truths at m/h_1 might be entirely distinct from those at m/h_3. The very idea of branching time then seems to fall apart, leaving us merely with a set of potentially unrelated histories. In short, to preserve the view of time as branching, we need to preserve the picture of moments as states, like still frames in a film, described by atomic formulas which give the basic facts, and with the dynamics of the model arising only from the connection of moments into histories, just as the dynamics of the movie arise from the sequence of the static pictures in its individual frames.

On the whole, then, it seems preferable to give up present tense atomic formulas of the special form *a chooses c* rather than to give up our general picture of what binds histories together at a moment, and accordingly it seems to be the norm in the literature that valuations are assumed to assign truth values at a moment rather than at a moment/history pair.[5]

If, then, we are not to abandon altogether the thought that sentences such as "John chooses the left fork in the road" can be expressed in our formal language, we must recognize them as complex in some way—not simple, and perhaps not truly present-tense. It might seem that this would call for introducing a new operator of some sort, but in fact the well-known deliberative *stit*, or *dstit*, operator will serve perfectly well here: John deliberatively sees to it that he takes the left fork, i.e. he makes a choice which assures that he takes the left fork, while there is another choice available which would not assure this outcome. Along some histories through the present moment, this will be true, while along others it will be false, just as we might expect.

From here on, we will assume that the valuation $\mathcal{V}$ assigns values at moments, rather than at points of evaluation. Schroedinger's cat will have to fend for itself.

4 What is a World?

As we compare systems based on possible worlds with ones based on an Ockhamist logic of branching time, we sense that the systems based on branching time are more fine-grained and hence potentially more sensitive. But this implies that in branching time systems we should be able to do, or mimic, any of the things one could do with possible worlds. In particular, we ought to be able to introduce the alethic modalities: possibility and necessity. In possible worlds models, the possibilities are represented by the possible worlds. The question then arises: What should be considered to

[5] We could, of course, declare that we assign values to formulas only at moment/history pairs, but add the constraint that in the case of atomic formulas, the valuation must be the same at pairs that share the same moment. This would simply be a way of accepting the static account, while offering the surface appearance of uniformity of treatment between atomic and non-atomic formulas.

represent possibilities within branching time models? Pondering this, we discover we have an embarrassment of riches.

We notice first that the moments, by virtue of their being the units of construction within branching time models, play a role analogous to that of possible worlds. Moreover if, as we've suggested should be the case, the valuation $\mathcal{V}$ assigns values to atomic formulas at moments rather than at moment/history pairs, then each moment is associated with a maximal consistent set of present-tense formulas: the atomic formulas specified by $\mathcal{V}$, together with their logical consequences. This bears some analogy with the way in which possible worlds are associated with maximal consistent sets of non-modal formulas in a typical possible worlds model for, say, **S4**. But the analogy is limited. Whereas in a model for **S4** there is also a maximal consistent set of formulas, *including* modal formulas, associated with each world, there will be no such set associated with a moment, since in a branching time model future tense formulas, in particular, will get their truth values only at moment/history pairs, not at moments.

Since the history, not just the moment, is crucial, we might consider points of evaluation—moment/history pairs—as candidates for the role of possibilities. Each of these will be associated with its own maximal consistent set of formulas which, collectively, will fully describe a possible historical situation—past, present and (within its history) future. This matches up with the fact that in possible worlds models the points of evaluation are just the individual worlds. So far, so good, but there is an oddity here: each moment/history pair within a given history will essentially represent the same possibility as each of the others, though from the perspective of a different moment in the history. So there is really only an indexical difference between points of evaluation within the same history, a difference concerning which moment within the history is thought of as the present, while the sequence of events will be the same.

This observation suggests that we should consider whole histories as the fundamental possibilities within branching time models, each such possibility specifying one way the complete current state of the world could be at each moment throughout its history. To be sure, there will be no single maximal consistent set of formulas associated with a history, since the values of present tense formulas, for example, will change from moment to moment within a single history. But we could accept this as just another indexical phenomenon, with the values of sentences involving the term 'now' to be resolved by reference to an index, just as sentences involving the term 'me' must be.[6] Given that we are permitting tensed language, we would face the same problem in possible worlds theory as well. Evaluation of indexical sentences will require that we supplement the specification of the world with a value for each of the needed indices, including a temporal index.

We now begin to have a basis for distinguishing what we might call internal and external possibilities. Within a given history perhaps the light switch is sometimes on, sometimes off. If so then that history includes the internal possibility of the light

[6] It won't be just present-tense sentences that will need a temporal index for their evaluation, of course: *all* tensed statements will be indexical relative to a history.

switch's being on, as well as the internal possibility of its being off. This contrasts with the external possibilities such as that there *at some point* be (as in the given history) or (as in others) *never* be a light switch in the room at all. A contrast between internal and external possibilities, if it withstands further scrutiny, provides a basis for two layers of alethic modality, requiring two distinct sets of alethic operators. That is a prospect worth investigating to see whether it can be put to good use in some way. That investigation would take us beyond the scope of this essay, however.

Going further, we might consider whole trees to be correlates of possible worlds. In doing so, we would be acknowledging that some possible worlds are, by virtue of their branching, laden with potential for choice, and full of *internal* possibilities. Now we would need to treat both moment and history as indices to be specified in addition to the specification of the tree, in order to induce a value for most sentences. Indeed we might find ourselves with three layers of alethic modality based on three different world-like units of construction: moment/history pairs, histories, and trees, respectively.

We'll reflect further on this in the coming sections.

5 Chronological Unity and Belnap's World(s)

There remains the question whether our models should require that time be pastwards connected. The constraint in question would be this (with $\leq$ defined in terms of $<$ in the obvious way):

(6) $(\exists m_0 \in \mathbf{M})[m_0 \leq m_1 \,\&\, m_0 \leq m_2]$ $(m_1, m_2 \in \mathbf{M})$

(**pastwards connection**)

Constraints 1–3 ensure that the moments in a model are organized into trees. Adding constraint 6 would ensure that there is only one such tree per model.

We focus first on models which are pastwards connected. One interesting observation is that if, in such models, we take the histories as correlates of possible worlds, we find that to determine whether two individuals appearing in different histories within a single model are or are not identical, we need only trace them back in time to see whether they have a common origin at some moment included in both the histories. So provided we are able to trace identity back through time, we get a natural solution to what would have been the problem of trans-world identity but which is now recast as the non-problem of trans-history identity.

But from another point of view, because the various histories in a branching time model are all connected it is reasonable to consider, as Belnap does, that the entirety of the structure in one pastwards-connected branching time model represents "our world". Such a model depicts a world rich with *internal* possibilities, past and present, and rich with alternative histories, each of them a possible history of the actual world, rather than an actual[7] history of a different possible world.

[7] Actual according to an indexical understanding of that notion, that is.

From that point of view, distinct branching time models represent genuinely different ways a world might be, each with its own branching structure, its own histories, its own internal possibilities. These pastwards-connected branching time models—*Belnapian worlds*, as we may call them[8]—will be in some respects much like the possibilities represented by the Kripkean worlds of normal systems of modal logic. They will differ from such worlds in at least two respects, however. First, of course, each Belnapian world has a rich internal structure which Kripkean worlds lack. But also, these Belnapian worlds are isolated from one another in separate models, whereas Kripkean worlds coexist within the same model. Accordingly, we might recognize the possibilities internal to a Belnapian world as *real* possibilities (relative to that world) and recognize the possibilities represented by the availability of other models as merely *nominal* possibilities—other ways the language might have been given application. We will examine this thought again in Sect. 7.

At this level, the problem of trans-world identity might be thought to surface again, but our having noted that there is no special problem of trans-history identity within a branching time structure can be the occasion for reconsidering the role different worlds play. If different Belnapian worlds only appear in different models, i.e. if we impose on models the requirement of historical connection, then perhaps we can profit from some reflection on the general nature of formal models, examining the comparative role of different models, and therefore of different Belnapian worlds.

So at the risk of being pedantic, let us review the basic features of models.

6 The General Character of Models

First of all, we note that a model for a logical system is always a model *with respect to a language*, and that this language is held fixed for the whole class of models for a given system.

Second, we note that each model will have two distinguishable components—a structure and an interpretive scheme. The structure is our stand-in for the world; the interpretive scheme links the language to the structure and in doing so represents the way the language is understood to connect with the world. The structure in turn has two sub-components: an ontology, which varies from model to model, and a set of structural constraints which remain constant across models. The interpretive scheme also has two sub-components: a valuation, which varies from model to model, and a set of satisfaction conditions which is invariant across models.

The ontology gives the array of types of entities that are taken to be explanatorily fundamental, for the level and style of explanation undertaken by the system. All other types of entities acknowledged by the system are constructed from, or

[8] *We* may, but Belnap may not approve. There is, after all, only one world, *our* world, and the real possibilities are all included within it. However the alternative worlds we consider here are logically possible—logically consistent—, and even metaphysically possible: consistent with the metaphysical commitments built into our definition of models and our satisfaction conditions.

possibly in some cases supervenient on, this ontological base. Typically, for Belnapian worlds the ontology may include such kinds of items as moments, agents, acts, and perhaps instants. Fundamental relations and functions built into models—the relation of temporal precedence among moments or a choice function for agents, for example—might be construed either as part of the ontology or as contributing to the structural constraints, but generally can be and are treated as part of the ontology. *There is* a relation $<$ between moments; that sounds like ontology. The relation is transitive; that's structural.

It is important to note that in the general specification of the class of models for a system the ontology is not normally constrained to include any specific entities, i.e. does not include any distinguished objects. The definition of models may specify that there are to be a number of agents in each, but it will not normally identify any of them; it will not normally specify that I am to be among them, for example, nor that any other specific entity is. On the other hand, it is equally important to note that any given model in the accepted class of models will have a specific ontology, filled with specific entities. Model 1 may have a set $\mathbf{A}_1$ of agents, for example, and model 2 a set $\mathbf{A}_2$. But nothing is said, normally, about whether sets $\mathbf{A}_1$ and $\mathbf{A}_2$ share any elements. Similarly with moments and other types of entities. In general, then, two models of the class may share some items in their ontology or they may not. Normally there are no constraints on models that will rule out either of these alternatives. This is one aspect[9] of what we might call our *metaphysical diffidence*: we restrain ourselves from pretending to present all the details of the constitution of reality.

The structural constraints represent those fundamental assumptions about the nature of the universe that go beyond answering the question what kinds of things there are, to answer questions about how the entities of the universe are organized in relation to one another. In Belnapian worlds these constraints include such constraints as that time does not branch pastwards and, even more fundamentally, that the relation of temporal precedence is irreflexive and transitive. I tend to think of these constraints as metaphysical commitments, though the distinction between metaphysics and physics might blur here. Taken together, however, the class of models satisfying the structural constraints reflects the metaphysical foundations taken as underpinning the language.

The valuation assigns to each non-logical atomic component of the language an extension of an appropriate type in, or constructible from, the ontology. Finally, the satisfaction conditions exploit this assignment to make it possible to calculate truth values for sentences of the language at each point of evaluation, and in doing so constrain the meanings of the logical constants of the language.

Given this general understanding of the nature and role of models, a little reflection indicates that different models represent different logical possibilities, each internally consistent and each consistent with the metaphysical foundations reflected in the ontology and structure that defines the class of models. The necessity which can be defined via such models will be of just one sort: *logical* necessity *relative to the*

[9] The most fundamental reflection of our metaphysical diffidence is the fact that we entertain a whole class of models, and do not designate one of these as the *real* model.

constraints. The necessary truths are the logically necessary consequences of the acceptance of the satisfaction conditions and the acceptance or the imposition of those constraints.

7 Comparing Belnapian Worlds

When we compare two Belnapian worlds, they will be alike in the broad outline of their ontology and their metaphysics. But although they will agree on what *kinds* of entities are available for discussion, they will commonly differ concerning which *particular* entities of a given kind are involved. We may require that each model have a non-empty set of agents, for example, but we don't presume to specify *what* set, and as a reflection of this metaphysical diffidence different models needn't have the *same* set. On the other hand, they needn't have distinct sets, either.

Because each model in a given system will be a model for the same language, the models will agree about what names are available to be given to items in the ontology, but will typically not agree about which objects bear which names. So if 'Belnap' appears in the lexicon of the language, and is constrained by its lexical category to name an agent, it may name our favorite logician in one model closely corresponding to the actual world, but in another model might name some seventeenth-century nun who developed, say, an irrelevance logic. Names that are assigned distinct denotations in one model might be two names for a single entity in another. So although we can trace names from world to world, we cannot use those names to trace entities outside the boundaries of a single world. We can say that the name 'Belnap' is used differently in different worlds, but we cannot on that basis say that Belnap himself—*our* Belnap—even occurs in those worlds, much less that he has different properties. On the other hand, Belnap himself *will* occur in some of those worlds, though it is anybody's guess what name(s) the language assigns him there.

One consequence of all this is that a form of the problem of trans-world identity *does* arise across the class of Belnapian worlds, even though there is little problem of trans-history identity within a given Belnapian world. Another consequence seems to be that it makes little sense to even contemplate judging specific causal connections using alternative Belnapian worlds. In both these cases, however, the problem may be less compelling than it seems at first. Let us look more closely.

It is true that the agent called Belnap in one Belnapian world may bear no relation to the agent so called in another, and that therefore if in one Belnapian world the sentence 'Belnap is a seventeenth century nun' is true, that does not by itself establish the possibility that Belnap could have been a seventeenth century nun. But if we contemplate the full array of all Belnapian worlds, we will find many in which Belnap himself does appear, and among those many in which he is accorded the name 'Belnap'. Moreover, among these, there is (barring special restrictions on the class of models) at least one identical in all internal details to the one in which 'Belnap is a seventeenth century nun' is true, except that the object there named 'Belnap' really is our *real* Belnap, not merely some name-sake, and so is the basis

for saying, within *any* Belnapian world, that it is possible that Belnap could have been a seventeenth century nun. We may not be able to trace identity from world to world, but the richness of the array of Belnapian worlds renders that limitation of no practical consequence for the evaluation of sentences about possibility and necessity. The fact that we cannot be sure that *this* Belnapian world has the Belnap we are talking about doesn't matter: there will be another Belnapian world descriptively just like it that does, and that uses the same name for him. *It's* possibilities are therefore *his* possibilities. Similar remarks will apply to tracing causal connections: connections that can be traced by name will also be traceable through cases (and there will be such cases) in which the name is applied to the same object.

But in what sense of 'possibility' can possibility be attributed here? Normally, in Kripkean systems, we count only worlds within the same model as providing a basis for claims of possibility. We don't acknowledge worlds from other models as relevant. At best, they provide for the *logical* possibility of the truth of certain claims, an acknowledgement that the language *could* have been used that way without contradiction, and without violating the metaphysical assumptions which underly it, whether it is in fact applied that way or not. So to the extent that we see Belnapian worlds as isolated from one another in separate models, we make them irrelevant to one another for alethic purposes, determining real possibility, though they remain relevant for purposes of determining logical possibility, validity and logical truth.

Suppose, now, that we nonetheless contemplate setting Belnapian worlds to some of the tasks normally assigned to Kripkean worlds, such as making it possible to introduce alethic possibilitation and necessitation operators into the language. In Kripkean systems, the ontology includes possible worlds, and various such worlds will be gathered together into a single simple Kripke model for, say, the system **S5**. So we typically have many worlds in a Kripkean model for **S5**, and understand a sentence p to be necessarily true at world w in the model iff p itself is true at each world in the model. Here, incidentally, our metaphysical diffidence manifests itself again in the fact that we do not presume to specify how many, nor which, possible worlds there are, and we permit (nay, *require*) the array of different models to offer a full array of different answers to such questions.

Introducing explicit possibilitation and necessitation operators $\Diamond$ and $\Box$ into the language for **S5** enables us to create formulas $\Diamond p$ and $\Box p$ to express the claims that the claim expressed by p is possibly true or necessarily true, respectively. Now if we try to put Belnapian worlds into an otherwise Kripkean model where there would previously have been Kripkean worlds, we get a sort of *supermodel* whose ontology includes Belnapian worlds which are themselves Ockhamist models, each with their own internal ontology. Can we make any sense at all of this? Let's think it through.

First, we will continue to have a single common language for each of the Belnapian worlds in the supermodel, and indeed across the various supermodels appropriate to our system. The language will, presumably, be the language we would have used for the Belnapian worlds themselves, but augmented by the new alethic operators. Since at the moment we're thinking only of what we might call a Belnapian **S5**,

no additional structural constraints seem needed.[10] However, when we turn to the valuation and the satisfaction conditions, things begin to look a bit more complex. The first question we face is: what are we to take as the points of evaluation in a supermodel? The analogy with Kripkean models for **S5** would make each Belnapian world a point of evaluation, while the analogy with Belnapian models would make moment/history pairs within Belnapian worlds the points of evaluation.

What at first blush seems a reasonable hybrid, namely to accept moment/history/world triples *m*/*h*/*w* as points of evaluation might turn out not to be a hybrid at all. This depends on how we view the sets of moments in distinct Belnapian worlds. If we assume (or require) that no moment occurs in more than one world then a given moment/history pair can only occur in a single Belnapian world. Then we are effectively back to just moment/history pairs, since settling on values for *m* and *h* will force a value for *w*; then the satisfaction conditions for $\Box p$ at a given point of evaluation in the supermodel will naturally be simply that *p* be true at *each* point of evaluation in the supermodel. No obvious problem here. As in **S5**, it will be automatic that whenever a formula *p* is valid, the corresponding formula $\Box p$ will also be valid. However if *p* is not valid, the formula $\Box p$ can be true in one supermodel while remaining untrue in another—*really* necessary, as far as that world is concerned, but not *logically* necessary.

That's if we assume that distinct worlds have disjoint sets of moments. Our metaphysical diffidence would suggest, however, that we might wish to restrain ourselves from such an assumption. Indeed, despite our misgivings about identifying times across separated histories within a Belnapian world, it doesn't initially seem absurd or unnatural to speak about the same time in different worlds. For linear time, this could be made to work out very simply, with each world using the same moments as every other, and with the same ordering in each world. However with branching time, much of the point would be lost if we supposed the trees in different worlds were all isomorphic to one another. And if they are not isomorphic to one another, then it is hard to see how the very same moment that occurs in one could occur in another in any meaningful way, i.e. in any way that made use of that identity. *A fortiori*, the same is true for moment/history pairs. Accordingly, let us overcome this bit of our metaphysical diffidence and assume that in supermodels distinct Belnapian worlds have disjoint sets of moments, and thus that a given point of evaluation in the supermodel will occur in exactly one world.[11]

When we look at the possibility of using Belnapian worlds in a supermodel to support other normal alethic modal operators, as in system **K**, or **S4**, or **S4.3**, for example, the direct analogy calls for us to complicate the supermodels with a relevance relation between Belnapian worlds, and (for any except the weakest normal system, **K**) add constraints on this relation. It would also, I would submit, be important to provide an interpretation of the relevance relation, and to justify any constraints on that

[10] We'll consider normal systems other than **S5** in a moment.

[11] This would not rule out the possibility that moments from different worlds, though distinct, might be comparable with respect to the truth values of certain canonical forms of chronological sentences about, say, clocks and calendars.

relation in terms of this interpretation. Unless we were engaging in a purely technical investigation, we should have some story to tell about what makes one Belnapian world relevant to another—some account of what one world would have to be like in order to be relevant to another.[12]

However, given that (as we are now assuming) each point of evaluation will occur in only one world, we could consider a relevance relation directly between points of evaluation. In the closest analogy to standard models, instead of having world w relevant to world w', we could have all w's points of evaluation relevant to each of w''s. Then $\Box p$ would be true at a given point of evaluation m/h iff p is true at each point m'/h' relevant to m/h. Once we contemplate such point-to-point relevance, however, we should at least consider the possibility that relevance could be more selectively defined, so that perhaps only some of w's points of evaluation would be relevant to ones in w', and perhaps only to selected points of evaluation in w'. This would call for rethinking the relevance relation, to provide an interpretation which could reasonably be understood to be so selective. Depending on what that interpretation might be, we might also contemplate the possibility that the relevance relation could hold between selected pairs of points of evaluation within the same Belnapian world. In principle, these relaxations of the relevance relation open up a whole new dimension of potential sensitivity for systems based on such supermodels—a dimension surely worthy of at least preliminary technical exploration. We shall not explore it further here, however.

Looking in a different direction: instead of seeking to pursue the analogy with standard models, we could consider pursuing an analogy with neighborhood models based on possible worlds. In classical models the relevance relation relates a world to relevant neighborhoods, i.e. sets, of worlds. One common rationale for doing so is to take advantage of the fact that in possible worlds semantics, any given proposition will naturally be associated with a uniquely determined set of worlds: the worlds at which the proposition is true. The neighborhood is then used to represent the comprehensive proposition which captures all that is true throughout the neighborhood but which is false at all other worlds. Other interpretations similarly associate sets of worlds with events, or with actions. In each such case, worlds are gathered into neighborhoods in their capacity as points of evaluation, and so the apt analog for our supermodels would be neighborhoods made up of moment/history pairs, rather than of worlds. The default view would be that the neighborhoods could, and typically would, include points of evaluation from different worlds: the proposition *that p*, for example, would be represented in the supermodel by the set of all points m/h at which p was true.[13]

---

[12] One classic illustration is the specification, in standard deontic logic, that world w' is to be considered *deontically* relevant to world w iff w' is *normatively ideal* by the ethical standards in force at w, i.e. iff those standards are all actually met at w'.

[13] Assuming that propositions transcend the given language, so that there can be true propositions not expressible in the language, the neighborhood consisting of all points at which p is true is likely to represent a stronger proposition than simply the proposition *that p*. Nonetheless, this neighborhood is the best we can do by way of representing the proposition *that p* short of having a completely expressive language. Even with an expressively complete language as our syntax, neighborhoods

Another way in which neighborhoods have been put to use is in Lewis's logic of counterfactuals. There, given any world w, the worlds of a model are gathered into concentric neighborhoods relevant to w, with the interpretation that worlds in a neighborhood are more similar to w than are worlds outside it. A counterfactual conditional $p \mathrel{\Box\!\!\rightarrow} q$ is then vacuously true at w if there is no world in any neighborhood relevant to w at which p is true; otherwise $p \mathrel{\Box\!\!\rightarrow} q$ is true at w iff there is a neighborhood of w throughout which $p \rightarrow q$ is true and within which there are worlds at which p is true.

In a system using supermodels, there will be a basis for more than one sort of counterfactual conditionals, falling into two broad categories which we might call *external* and *internal* counterfactuals. External counterfactuals will, like Lewis's, involve comparing one Belnapian world with others with respect to some measure of similarity, and will be saddled with all the problems of explaining that notion of similarity. Internal counterfactuals will instead compare one history or one point of evaluation with others within a single Belnapian world.

For the internal counterfactuals it will be possible to draw on a natural sense of similarity, based on the distance back in time one must go to find a moment common to the two histories. Histories which have split off from one another only recently will in one natural sense be more similar than ones which diverged at some still earlier moment. It seems plausible to suppose that many counterfactual conditionals that occur in ordinary reasoning are best considered as internal rather than external counterfactuals, and therefore are correspondingly more intelligible than they might otherwise appear to be.

One kind of internal counterfactual can be based on a particular agent's choices. For a sentence like

If John had gone left at the fork in the road, he would have come to Paris

we could expect this to be true at m/h iff at some earlier point m_0/h one of John's choices included histories in each of which he takes the left fork, and at all points m'/h with $m_0 \leq m' < m$ at which John's choices include choosing the left fork, then in each of the histories in which he takes that choice he subsequently comes to Paris.

Other counterfactuals tacitly involve agents' choices, but not for a specific agent, and so do not involve specific choices. So to evaluate a sentence like

If Cheney had not been Vice President, the U.S. wouldn't have invaded Iraq

we could look for earlier branch points at which there are histories in which Cheney does not become Vice President, and verify that the U.S. does not subsequently invade Iraq in any of the most recently diverging such histories.

(Footnote 13 continued)
would represent propositions only up to logical equivalence. These subtleties, though important, do not normally deter us from considering neighborhood semantics useful, however.

There remains the possibility that for some examples, as with external counterfactuals, the relevant degree of similarity between histories within a tree might take some other form than simple distance in time back to their nearest common moment. One notorious example which appears to be of this type arises if I accidentally leave my coat behind in the cloakroom at the close of a conference session, and return the next day to find it still there.[14] Knowing that there were dubious characters in the neighborhood when I left, and that many individuals had access to that cloakroom during my absence, I would reject as false the sentence:

If my coat had been stolen, it would have been the most recent person to visit the cloakroom who would have stolen it.

No doubt the first really dubious character who came by after I left was likely to take my coat, pre-empting any opportunity that the most recent nefarious visitor might have had.

For such an example, it is difficult to say what the relevant sense of similarity between histories might be, but it seems clear nonetheless that no comparison with histories from other worlds is particularly apt.

For external counterfactuals, we have a choice: we could base our account on a similarity relation between worlds or on a similarity relation between points of evaluation. For a sentence such as

If the match were struck, it would light

it might be most appropriate to compare moment/history pairs (without regard to what world they were in) and focus on the ones whose factual conditions were most similar to those at the point of evaluation. On the other hand, for Lewis's example

If kangaroos didn't have tails, they would fall over backwards

it might be best to refer to similar worlds, particularly since there are probably no suitably similar moment/history pairs in our world at which kangaroos lack tails, and it would take a very different world to include such situations in a coherent way.

The moral of all this rumination about supermodels is that they open up considerable new prospects for exploration and exploitation. We now begin to glimpse the plausibility of supposing, for example, that different kinds of counterfactuals call for different accounts, and we see that supermodels might provide an environment friendly to such fine-grained discriminations. Moreover, it is not unreasonable to suppose something similar might be the case with accounts of causation, particularly if we suppose that counterfactuals play a major role in, or are in some other way closely connected with, an account of causation.

[14] Ironically, this actually happened to me at the conference at which I first heard mention of this type of example.

8 Belnapian Multi-Worlds

So far, we have been considering the uses to which Belnapian worlds might be put. Belnapian worlds involve the constraint of pastwards connection. Now, however, let us consider the consequences of relinquishing that requirement of pastwards connection. With this constraint gone, we get models within which there may be multiple trees of moments, unconnected to one another. Because each independent tree within such a model will be in many respects very like a Belnapian world, let us call such models *Belnapian multi-worlds*.[15]

A Belnapian multi-world will not be just like an arbitrary set of Belnapian worlds, because the multi-world will have a single specification of its entities, and a single valuation assigning names to those entities, rather than having separate specifications of these for each tree. Thus a Belnapian multi-world is like a *coordinated* set of Belnapian worlds—worlds coordinated with respect to their ontology and their assignment of names—, but we must still wonder, as in standard models of alethic logic, whether it makes sense to suppose the same entity can occur in more than one world, i.e. in more than one tree.

There is no need to *assume* that no moment occurs in distinct worlds within a multi-world, because the constraints on the ordering relation $<$ will assure us of this. This is another respect in which a Belnapian multi-world differs from a supermodel or a mere set of Belnapian worlds, since although it seemed overwhelmingly appropriate to assume that a given moment could not occur in distinct Belnapian worlds within a supermodel, we did have to treat this as an assumption.

If the choice function works, as usual, to assign a set of choices to each agent in the model at each moment in the model, it would appear at first that each agent is presumptively active in each of the trees—each of the worlds—in a given multi-world. But further reflection suggests that this need not be so, if (as seems unavoidable, if we are to be realistic) the "choice" given an agent at certain moments is the "Hobson's choice" consisting of just one alternative: the set of all histories through that moment. Surely this is the sort of "choice" the agent has at moments when unconscious, for example. Moreover, we might want to assign this trivial choice to agents at all moments before their birth and all moments after their death. Doing so would provide a convenient way of representing the finitude of the lives of agents while allowing a sense of their existence, and therefore their availability for reference, outside their lifespan. In 'Socrates was Greek', the name 'Socrates' will continue to have a referent even after that philosopher is no longer alive.

If the choice function can assign the trivial choice at some moments, then there is nothing to prevent its assigning trivial choices to a given agent at *every* moment throughout a given tree, thus effectively excluding that agent from participation in that world. As a result it is possible for worlds within a multi-world to have effectively disjoint sets of agents.[16] On the other hand, of course, it is possible for such worlds

[15] Belnapian worlds will, of course, be special cases of Belnapian multi-worlds.

[16] Indeed, this has nothing special to do with Belnapian multi-worlds. In any Belnapian world, unless we introduce a constraint not normally imposed, it is possible for a given agent to be present in

to share an active agent, and if they do, the agent will bear the same name(s) in each world within the model.

A system based on Belnapian multi-worlds would seem to provide a natural basis for an alethic necessitation operator: $\Box p$ will be true at m/h in a multi-world iff p is true at each point of evaluation in the model. If no further complications are added, this will automatically be an **S5** sense of necessity. Logical truths will, of course, be necessary truths, on this reading, but as usual the converse will not hold in general: $\Box p$ may be true at each point of evaluation in one multi-world model, but fail throughout another.

It might appear that there should also be room for what we might call situationally necessary truths—claims p which are necessarily true at some points of evaluation in a multi-world, but not at others. Indeed, we *will* have something a little like this: true claims about the past will be settled true, but might not be true at points of evaluation on other histories; and some claims about the future will be settled true at some sufficiently late points along a given history, but might not be settled true at earlier ones. But of course such cases are handled by the *settled true* operator Sett, and need not be considered cases of true necessity: $\Box p$ will express a stronger claim than Sett p.

If a suitable rationale can be found for doing so, it would be technically possible to add a relevance relation between points of evaluation, with suitable constraints on this relation, so as to have the necessitation operator be an **S4** operator, or some other normal necessitation operator.

Since there are typically multiple worlds in a multi-world model, we again have room for various types of counterfactual operators, both internal and external. The internal operators would be just like the ones in supermodels or in single Belnapian worlds. The external counterfactual operators available within a multi-model would depend on a relation of similarity among the worlds contained within that multi-world. We must not forget that there will be more than one model—more than one Belnapian multi-world. We could contemplate assembling super-multi-worlds within each of which we gather a set of multi-worlds, but the motivation for contemplating super-multi-worlds seems weak: we already have, as in standard possible worlds models, room for more than one world per model, and so don't need to gather multi-worlds into super-multi-worlds to get worlds collected together.

9 The Making of an Agent

Typical propositional systems of the logic of action, constructed along Belnapian lines, focus on agency and the truth conditions for agentive sentences, rather than

(Footnote 16 continued)
the ontology, but not active at any moment in any history, and thus to be for all practical purposes non-existent. It's a bit difficult to see how to interpret such a situation. This perhaps argues for the introduction of such a constraint, particularly for the usual systems whose models are single Belnapian worlds. It is also possible that an agent should be active only in some histories and not in others, e.g. that there might be some histories in which the agent was, and others in which she was not, born.

on the agents themselves. A non-empty set of agents is postulated, and the choice function indicates what choices each will have at each juncture in branching time. But little else is normally said about the nature of the agents and about what distinguishes one agent from another other than brute non-identity.

Belnap has remarked[17] that in a branching spacetime system, it is possible to associate each agent with a unique set of point events, the set of those at which the agent is present. This is made possible by the fact that distinct agents cannot occupy the same place at the same time. Unfortunately, if we look merely at branching time, with no basis for discussion of spatial dimensions, no such simple account of the identity of agents is possible.

However, the usual constraints imposed on the choice function, including in particular the constraint *independence of agents*, do make it possible for us to look at agents in new ways. In particular, we can give some formal substance to, and gain some new insight into, the view that an agent is the sum of the choices she makes and thus that an agent is a work in progress, existentially shaping her character and her very identity through her choices.

In a Belnapian world, the constraint *independence of agents* assures that, strictly speaking, no two agents will be presented with the same choices at a given moment in time. Of course at the restaurant both may be choosing between having the scallops and having the mussels, but that is only to say that the choices facing one may be descriptively like those facing the other. For one thing, even if they both choose the scallops, for example, they will not get the same scallops. But more significantly, agent *a* is choosing what *a* will order (and, presumably, eat), not what agent *b* will order. So even confining attention to a single moment, agents are normally[18] differentiated from one another by the choices they face.

If we shift attention to the choices agents *make*, not just the choices they *face*, this becomes even clearer: even at a single moment, provided only that the agents are active, they are differentiated by their activity, which is to say: by the choices they make.

Widening our perspective to scan a history within which a moment falls, we find the agent making a succession of choices which cumulatively help define that very history, setting it apart, choice by choice, from others that had been available. The totality of those choices will be absolutely unique to a given agent, no matter whether we focus on the menu of choices the agent faces or on the choices the agent selects from that menu. Seeing this accumulation of choices along the history as uniquely associated with one agent, we can begin to consider it as *constituting* that agent, in which case as we survey the history we now get a strong sense of what it might mean to say that the agent is creating herself by her choices. In a different history, pursued by making different choices, she would have become a different person.

---

[17] At the ΔEON'10 Conference, Fiesole, Italy, 2010.

[18] There is one kind of exception to this generalization: at a given moment two agents might be given exactly the same choice by being given no choice at all. If both are asleep, for example, the choice function will presumably assign the whole of $\mathbf{H}_m$ as the one "Hobson's choice" available to each at moment *m*.

Each agent will correspond to a unique set of partitions of histories at choice points, and along any history, there will be a particular set of choices the agent makes and without which the agent's life would not have followed that history. So from the point of view of that history, it will seem that the agent's choices will have accumulated to make the agent the individual she has become. From this point of view, a simple version of existentialism seems vindicated.

But it is more complex than that, because the choices of others also influence the history one takes, and therefore the subsequent choices with which one is faced. In the end, then, the individual one becomes is profoundly influenced by the choices of other agents, as well as by her own, and this is where the simple existentialist picture fails.

And there is another layer of complexity added when we look beyond a single history. The totality of the agent's choices throughout the Belnapian world is also uniquely associated with that agent, and reflective not merely of what the agent does become (along this history or that) but also of the agent's *potential*, which we may consider is equally essential to their identity. At this scale we begin to see the agent as a unique collection of possibilities for action.

When we survey this larger picture, balancing the agent's tree of possibilities against the developments of those possibilities along a given history, we find a new perspective on the old nature/nurture debate: we need not choose between nature and nurture as the sources of one's character, and we must indeed add a third factor—will. Nature has its role in providing our potential as seen in the tree-wide totality of the choices available to us; will has its role through our choices; nurture has its role in the choices of others which influence and limit our choices. Together they work, along any given history, to create a uniquely matured version of the agent, different from what they could have become had they not had that potential, different from what they would have become had they made different choices, and different from what they would have become had others chosen differently, as well.

Another aspect of the focus on agency rather than on agents, as we have begun to see, is that there is nothing in the models for such systems that rules out the possibility that a given agent has been making choices throughout time, and will continue to make choices throughout time along each history. This is, of course out of keeping with the fact that the logic is intended to reflect the situation of mortal agents and agents who are not perpetually active.

If we undertake the task of constructing a quantified logic of action we can expect to be able to remedy this situation, perhaps by introducing a distinguished predicate *is alive* into the language and making appropriate provisions for its interpretation in our models. The many quandaries associated with quantified alethic modal logic tend to make us shy away from a quantified logic of action (which could hardly present fewer such challenges) at least for the moment. But let us contemplate a slightly enhanced propositional logic of action and use this to get at least a preliminary look at some of the challenges involved in taking account of the finitude of agents.

Suppose that in addition to the other more or less standard components of models for our logic of action we include a function **Q** (for *quick*, in the sense of *alive*)

from moments to subsets of the set of agents, with the intended interpretation that for each moment m, $\mathbf{Q}(m)$ will be the set of agents alive at m. To make this work in the intended way, we would need to add some constraints. One constraint would express the *continuity of life* for agents:

(7) if $m_1 < m_2 < m_3$ then $\mathbf{Q}(m_1) \cap \mathbf{Q}(m_3) \subseteq \mathbf{Q}(m_2)$

(**continuity of life**)

Another might express the principle that *dead agents don't choose*:

(8) if $\alpha \notin \mathbf{Q}(m)$ then $\mathbf{H}_m \in \mathbf{C}_a^m$

(**dead agents don't choose**)

This would assure that only a live agent ever has more than one choice, which is to say a dead (or an unborn) agent has no active influence on the course of affairs.

Note that nothing here requires that live agents have non-trivial choices at any given moment. An agent may very well be asleep, or simply inactive, at a given moment in her life.

We might also contemplate some further constraints:

(9) if $\alpha \in \mathbf{Q}(m_1) \cap \mathbf{Q}(m_2)$ then $(\exists m_0\colon m_0 \leq m_1 \& m_0 \leq m_2)[\alpha \in \mathbf{Q}(m_0)]$

(**uniqueness of origins**)

This would assure that the same agent isn't born independently in two different histories.

(10) if $\alpha \in \mathbf{Q}(m)$ then $(\exists m_1, m_2\colon m_1 \leq m \leq m_2)[\alpha \notin \mathbf{Q}(m_1) \,\&\, \alpha \notin \mathbf{Q}(m_2)]$

(**mortality**)

This assures that no agent has been alive from all time, and that none lives forever.

Such refinements of our models would probably not have profound effects on the core logic of action, i.e. on the class of valid formulas involving only the action operator, and from that limited point of view they are probably not very important. Their value lies chiefly in their ability to reflect a little more fully the underlying picture which brings us to the logic of branching time in the first place, and in opening the prospect of additional operators which would enrich our formal language in interesting ways.

Perhaps the most conspicuous such possibilities arise when a logic of action based on branching time is pressed into service in a system of logic of ability or a system of deontic logic. In a logic of ability it will certainly commonly—but perhaps not always and in every sense of ability—be a condition of ability that one be alive. To be able to drive a car requires that one be alive. To be able to evoke happy memories perhaps does not.

In deontic logic, it will be important to note that, for example, death frees one from many obligations. If I am dead, I can have no obligation to visit my father on his next birthday. Similarly if he is dead. Developing deontic logic to the point that it is able to deal sensitively with questions about abortion, homicide, accidental death, etc., could all be expected to be aided by this sort of enrichment of our models.

10 Conclusion

The basic framework of the logic of branching time, and of the logic of action based on branching time offers rich opportunities for refinement and elaboration in a variety of dimensions, and in ways that deserve exploration. There is prospect for new insights into alethic modal logic, the logic of counterfactuals, and deontic logic, to cite only a few areas. Such opportunities deserve at least preliminary exploration, but such investigations will have to be reserved to another time.

The works provided in the bibliography contain material which forms the background for the discussions undertaken here.

Bibliography

Belnap, Nuel. 1991. Backwards and forwards in the modal logic of agency. *Philosophy and Phenomenological Research* 51: 777–807.

Belnap, Nuel. 1991. Before refraining: Concepts for agency. *Erkenntnis* 34: 137–169.

Belnap, Nuel, and Michael Perloff. 1988. Seeing to it that: A canonical form for agentives. *Theoria* 54: 175–199.

Belnap, Nuel, Michael Perloff and Ming Xu. 2001. *Facing the Future*. Oxford: Oxford University Press.

Chellas, Brian F. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Horty, John F. 2001. *Agency and deontic logic*. Oxford: Oxford University Press.

Horty, John F., and Nuel Belnap. 1995. The deliberative stit: A study in action, omission, ability and obligation. *Journal of Philosophical Logic* 24: 583–644.

Lewis, David K. 1973. *Counterfactuals*. Cambridge: Harvard University Press.

Thomason, Richmond. 1981. Deontic logic as founded on tense logic. In *New studies in deontic logic: Norms, actions, and the foundations of ethics*, ed. Hilpinen, Risto, 165–176. Dordrecht: D. Reidel Publishing Company.

Xu, Ming. 1997. Causation in branching time (I): Transitions, events, and causes. *Synthese* 112: 137–192.

Open Futures in the Foundations of Propositional Logic

James W. Garson

Even in the analysis of Greek and Latin (where the 'future' like the 'present' and the 'past' is realized inflexionally), there is some reason to describe the 'future tense' as partly modal.
John Lyons (1968, p. 306)

Abstract This chapter weaves together two themes in the work of Nuel Belnap. The earlier theme was to propose conditions (such as conservativity and uniqueness) under which logical rules determine the meanings of the connectives they regulate. The later theme was the employment of semantics for the open future in the foundations of logics of agency. This chapter shows that on the reasonable criterion for fixing meaning of a connective by its rule governed deductive behavior, the natural deduction rules for classical propositional logic do not fix the interpretation embodied in the standard truth tables, but instead express an open future semantics related to Kripke's possible worlds semantics for intuitionistic logic, called natural semantics. The basis for this connection has already been published, but this chapter reports new results on disjunction, and explores the relationships between natural semantics and supervaluations. A possible complaint against natural semantics is that its models may disobey the requirement that there be no branching in the past. It is shown, however, that the condition may be met by using a plausible reindividuation of temporal moments. The chapter also explains how natural semantics may be used to locate what is wrong with fatalistic arguments that purport to close the door on a open future. The upshot is that the open future is not just essential to our idea of agency, it is already built right into the foundations of classical logic.

J. W. Garson (✉)
Philosophy Department, University of Houston, Houston, TX 77204-3004, USA
e-mail: garson@Central.UH.EDU

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_6,

1 Introduction

This chapter weaves together two themes in the work of Nuel Belnap. The earlier theme, launched in "Tonk, Plonk, and Plink" (1962), was to propose conditions under which logical rules determine the meanings of the connectives they regulate. The later theme was the employment of semantics for the open future in the foundations of logics of agency. The first theme leads to the second in the following way. Consider the natural deduction rules for standard propositional logic, which by the lights of "Tonk, Plonk and Plink" successfully define meanings for the connectives &, →, ∼, and ∨. What are the meanings so assigned? Is there a way to give a semantical characterization of what the demand that a connective obey its rules says about the connective's meaning? It will be shown here that on at least one reasonable criterion for fixing meaning of a connective by its rule governed deductive behavior, the natural deduction rules for (classical) propositional logic do not fix the interpretation embodied in the standard truth tables, but instead express an open future semantics related to Kripke's S4 possible worlds semantics for intuitionistic logic. Part of the basis for this connection has already been worked out (Garson 1990, 2001). This chapter reports new results on disjunction, and explores the relationships between this open future semantics and supervaluations. The upshot is that the semantics actually expressed by the standard natural deduction rules for propositional logic is a semantics for an open future. Since *that* semantics is the one the rules of propositional logic actually fix, it is reasonable to think that that is the interpretation of the connectives that we have (secretly?) employed all along. It is not just that the conception of an open future is built into our idea of agency, it is already found in the foundations of classical logic. It will be no surprise then, when this interpretation shows itself to be useful for locating what is wrong with fatalistic arguments that attempt to close the door on a open future.

2 What Rules Express

The idea of a sentence (or group of sentences) expressing a condition on a model should be familiar. For example, the sentence $\exists x \exists y \sim x = y$ expresses that the domain of a model contains at least two objects, for $\exists x \exists y \sim x = y$ is true on a model exactly when its domain meets that condition. So in general, sentence A expresses property P of models iff A is true on a model exactly when that model has property P.

How should this idea be generalized to the case of what is expressed by a rule of logic? The generalization we are seeking involves two dimensions. The first has to do with how we conceive of logic rules. It is natural to think of a traditional logical rule as a function taking one or more sentence (forms) into a new sentence (form). However, it will be important for this chapter to accommodate natural deduction rules, for they have expressive powers that traditional rules lack. Natural deduction (ND) systems allow the introduction of ancillary hypotheses and subproofs. In that case, a rule amounts to a function that takes an *argument* or set of arguments into a new

argument. For example, the rule of Conditional Proof takes the argument H, A / C (which asserts that C follows from the ancillary hypothesis A along with other hypotheses H) to the new argument H / A→C (which asserts that the conditional A→C follows from hypotheses H).

The second aspect of the generalization concerns how to define what we mean by a model of a rule. A model of a *sentence* is one where the sentence is true. However, we have decided that a rule is a function that takes an argument or arguments to a new argument. So what does it mean to say that a *rule* holds in a model? One answer is to say that a model *satisfies* an argument iff whenever the model makes its premises true it makes the conclusion true. Then a model of a rule would be any model that preserves satisfaction of its arguments. However, this idea is not sufficiently general. There are rules (such as Necessitation in modal logic and Universal Generalization in predicate logic) that do not preserve satisfaction. Therefore, preservation of *validity* rather than preservation of truth should be used to define what a rule expresses. The upshot will be that the definition of what a rule expresses has it that a ND rule expresses a property P iff P holds exactly when the rule preserves validity. The following series of definitions implements this basic idea.

Let an *argument* H / C be composed of a (possibly empty) set of wffs H (called the hypotheses), and a single wff C called the conclusion. Let a *valuation* be *any* function from the set of wffs of propositional logic (PL) to the set {t, f} of truth-values such that it assigns f to at least one wff. (The last requirement ensures that valuations be minimally consistent.) Valuation v satisfies a set H of wffs H (written v(H) = t) iff v(B) = t for each member B of H. Valuation v satisfies *an argument* H ⊢ C iff whenever v(H) = t, v(C) = t. Let a *model* V be any set of valuations. (The valuations in V play the role of possible worlds in models for modal logic.) An argument H ⊢ C is *V-valid* iff it is satisfied by every member of V. A set V is *a model of a ND rule* iff whenever its inputs are V-valid then so is its output. An ND-rule R *expresses* property P iff V is a model of R exactly when property P holds of V.

3 What Intuitionistic Logic Expresses

With the notion of what a rule expresses in hand, we can explore the conditions under which a given collection of rules defines the meaning of the connectives they govern. When a system of rules S expresses a property ||S|| that qualifies as truth conditions for its connectives, it is reasonable to claim that ||S|| gives the meanings defined by those rules. When this occurs, I call ||S|| the natural semantics for S. Garson (2001) reports results on the natural semantics expressed by ND rules for intuitionistic logic, which will be briefly reviewed here. The concern in this chapter will be to extend these results to standard propositional logic with special emphasis on the interpretation of disjunction. It will then be possible to reflect on the relationships between this semantics and models for an open future.

We will assume that the ND systems discussed in this chapter obey the following structural rules. When argument H / C is provable for a given system S we write:

'$H \vdash_S C$', but we suppress the subscript S when it is clear what S is from the context. Therefore the symbol '/' is in the object language and '$\vdash$' in the metalanguage.

The Structural Rules for Natural Deduction

(Hypothesis) $H \vdash C$, provided C is in H.

(Reiteration)
$$\frac{H \vdash C}{H, A \vdash C}$$

(Restricted Cut)
$$\frac{H \vdash A \quad H, A \vdash C}{H \vdash C}$$

(Permutation and Contraction come for free, since H is taken to be set.) Natural deduction rules for a system PL for propositional logic follow.

Natural Deduction Rules for PL

S &: (& In)
$$\frac{H \vdash A \quad H \vdash B}{H \vdash A\&B}$$
(& Out)
$$\frac{H \vdash A\&B}{H \vdash A} \qquad \frac{H \vdash A\&B}{H \vdash B}$$

S →: (→ In)
$$\frac{H, A \vdash B}{H \vdash A \rightarrow B}$$
(→ Out)
$$\frac{H \vdash A \quad H \vdash A \rightarrow B}{H \vdash B}$$

S∼ : (∼ In)
$$\frac{H, A \vdash B \quad H, A \vdash \sim B}{H \vdash \sim A}$$
(∼ Out)
$$\frac{H, \sim A \vdash B \quad H, \sim A \vdash \sim B}{H \vdash A}$$

Sv: (v In)
$$\frac{H \vdash A}{H \vdash A \vee B} \qquad \frac{H \vdash B}{H \vdash A \vee B}$$
(v Out)
$$\frac{H \vdash A \vee B \quad H, A \vdash C \quad H, B \vdash C}{H \vdash C}$$

For the purposes of this chapter, it is best to begin with an intuitionistic logic I¬ that lacks disjunction. So let the system I¬ be identical to PL save that the connective v, and the rules for v are missing, and (∼Out) is replaced with the following rule (EFQ) for ex falso quodlibet:

(EFQ)
$$\frac{H \vdash B \quad H \vdash \sim B}{H \vdash A}$$

Results of Garson (1990 and 2001) are sufficient to show that I¬ expresses the following truth conditions for the connectives &, →, and ∼. (For a more unified treatment, see Sect. 6.4 of Garson (2013). Here the truth conditions are expressed as properties of a model V, the metavariables 'v' and 'v′' are understood to range over V, and the relation ≤ is defined by (Def ≤).

(Def ≤)	$v \leq v'$ iff for each propositional variable p, if $v(p) = t$ then $v'(p) = t$.
||& ||	$v(A\&B) = t$ iff $v(A) = t$ and $v(B) = t$.
||→||	$v(A \rightarrow B) = t$ iff for all $v' \in V$, if $v \leq v'$ then $v'(A) = f$ or $v'(B) = t$.
||∼||	$v(\sim A) = t$ iff for all $v' \in V$, if $v \leq v'$ then $v'(A) = f$.

Note that the interpretation for & induced by I¬ is the standard one, while in the case of ||→||, and ||∼|| we have truth conditions reminiscent of Kripke's S4 semantics for intuitionistic logic. In that semantics, a model <W, ⊆, **a**> is a triple, where W is a non-empty set (of possible worlds), **a** is an assignment function taking each world **w** in W and propositional variable p, into a truth value $\mathbf{a_w}(p)$, and ⊆ is a transitive and reflexive relation over W, such that for **w** and **w′** in W, if $\mathbf{w} \subseteq \mathbf{w}'$, then if $\mathbf{a_w}(p) = t$ then $\mathbf{a_{w'}}(p) = t$, for each propositional variable p. In this semantics, the relation ⊆ is understood to represent the historical process of the addition of new mathematical results by the community of mathematicians.

Let ||I¬|| be the semantics expressed by I¬, that is, the conjunction of ||&||, ||→||, and ||∼||. It is a straightforward matter to show Garson (2013) that any model V that obeys ||I¬|| is isomorphic to a corresponding Kripke model <**W**, ⊆, **a**> where W is simply V, ⊆ is the relation ≤ defined by (Def ≤), and $\mathbf{a_v}(p) = t$ iff $v(p) = t$. Therefore, the condition expressed by I¬ is that the connectives &, →, and ∼ obey their corresponding truth behavior in Kripke semantics for intuitionistic logic. An important lemma in the proof of this result follows for v and v′ in V.

Persistence Lemma.
$v \leq v'$ iff for a every wff A, if $v(A) = t$ then $v'(A) = t$.

This means that the relation ≤ holds for v and v′ when the set of sentences true in v is a subset of those true in v′. So the possible worlds (or valuations) can be understood as representing states of mathematical knowledge expressed as consistent sets of (atomic and complex) sentences, and $v \leq v'$ holds when v′ represents a (possible) extension of mathematical knowledge from what is reflected in v.

The fact that I¬ expresses the above truth conditions is a powerful result, for we know that *any* model for the rules of I¬ will have to give the connectives these interpretations. The rules of I¬ *exactly* determine the S4 reading for the connectives. It is a simple matter to exploit this finding to obtain completeness results for I¬ with respect to ||I¬||. In fact, there is a general result that whenever a system S expresses a natural semantics ||S||, then S must be complete with respect to ||S||. (See Chapter 12 of Garson (2013).)

4 Open Future Semantics

It is a fundamental presupposition of a theory of action that there are some things that lie within, and some that lie outside, our control. The events of the past are settled, and nothing we can do will change them. Therefore, the sentences that report those events are not the "targets" of agency. If sentence A reports an event in the past, then the claim that person p brings it about (now) that A is automatically false. There are also sentences reporting future events that defy agency as well, for example, tautologies and contradictions. However, within the class of future contingent sentences A, there are at least some where we have control over the events they describe, so that it would be true to say that person p brings it about that A. There is a strong intuition that when I act to bring about A, whether A is true or not is up to me. Therefore both A and ∼A must have been possible before I acted. So, the future offers me a collection of possibilities, which we represent in a tree, with my choices at each of the branch points. An essential intuition related to this vision is that some future contingent sentences are not yet settled, though they may be at a future time. Therefore, if I act in a way that settles A, then neither A nor ∼A could have been settled before I act.

Some philosophers claim that the asymmetry between past and future reflected in our ideas about action is actually bogus, and there is only one temporal stream, where both past and future are fixed. Others insist that our freedom to choose is an illusion. But even if one had good reasons for accepting such a views, it wouldn't change the fact that a natural model for the way we actually *do* understand agency treats future possibilities as a forward facing tree with our choices at the branch points. Whatever the fate of the concept of agency, its pervasive use motivates the development of a semantics that does justice to its basic intuitions.

The intuitionistic semantics ||I¬|| built into propositional logic has a natural application to this project. Consider a language whose atomic sentences report dated events. That will mean that atomic sentences are temporally closed in the sense that their truth-values are insensitive to their time of evaluation. A possible world (or valuation) v assigns t to those sentences that report events that are so far settled; so when A reports a past event, or something in the future that is inescapable (for example, something tautologous), v ought to assign it the value t. The relation $\leq$ then keeps track of the way in which valuations are extended as the passage of time settles more and more sentences. So $v \leq v'$, indicates that v' is a possible extension of the sentences that are settled in v, that is, v' is one of the ways that choices might be eventually be settled given the choices available in v.

To capture these ideas formally, some definitions are in order. We will say that A is *settled true* at valuation v (written: $v(A) = T$) iff the value of A is t in every extension of v (including v itself). Similarly, A is *settled false* at v (written: $v(A) = F$) iff the value of A is f (untrue) in every extension of v. If A is settled true or settled false at v then we say A is *settled,* and if A is not settled we call it *unsettled* (written $v(A) = U$). So, we have the following official definitions:

(DefT) $v(A) = T$ iff for all $v' \in V$, if $v \leq v'$, then $v'(A) = t$.
(DefF) $v(A) = F$ iff for all $v' \in V$, if $v \leq v'$, then $v'(A) = f$.
(DefU) $v(A) = U$ iff neither $v(A) = T$ nor $v(A) = F$.

Some useful facts about these definitions are worth noting.

(Fact 1) $v(A) = T$ iff $v(A) = t$.

The proof of (Fact 1) is by the Persistence Lemma and the reflexivity of $\leq$. It says that being true (t) and settled true (T) are extensionally equivalent.

(Fact 2) $v(\sim A) = t$ iff $v(A) = F$.

(Fact 2) follows from the (DefF), and the truth condition $||\sim||$. It allows a quick calibration of the difference between classical negation and intuitionistic negation. For classical negation $\sim$A's being true entails that A has the value f, while in the intuitionistic semantics, the truth of $\sim$A entails the stronger claim that A is settled false (F).

(Fact 3) $v(A) = U$ iff $v(A) = f$ and $v(\sim A) = f$.

The proof of (Fact 3) follows from the definition of U, (Fact 1), and $||\sim||$. The idea is that an unsettled sentence is simply one for which neither it nor its negation is true. In light of (Fact 1), this makes sense, for when A or $\sim$A are true they are settled true, and by (Fact 2) when $\sim$A is settled true, A must be settled false. Therefore for A to be unsettled, neither A nor $\sim$A can be true. The upshot is that the intuitionistic semantics has the resources to capture the idea that some sentences are unsettled and so count as possible "targets" for agency.

(Fact 4) $v(A) \neq F$ iff $v(A) = T$ if A is settled at v.

(Fact 4) follows because settled true, settled false, and unsettled are exhaustive categories. Therefore, v(A) is not F if and only if v(A) is U or T, or to put it another way, $v(A) = T$ if A is settled at v. So when $v(A) \neq F$ we might say that A is *quasi-true* at v, and write $v(A) = qT$ to indicate that A is settled true at v, if settled there at all. Then Fact (4) entails (Fact 5).

(Fact 5) $v(A) \neq F$ iff $v(A) = qT$.

5 What Propositional Logic Expresses

We have shown that the framework for an open future semantics is expressed by intuitionistic logic I¬. However, this chapter is concerns standard propositional logic. So let us explore what is expressed by a system of classical propositional logic PL- (without disjunction). Since PL- may be obtained by adding (DN) the law of double negation to I¬, the question of what PL- expresses amounts to finding what (DN) expresses.

$$\text{(DN)}\ \frac{H \vdash \sim\sim A}{H \vdash A}$$

It turns out that the corresponding condition expressed by (DN) is $||\sim\sim||$. (See Humberstone 1981 p. 318, who calls a related condition Refinability.)

$||\sim\sim||$ If $v(A) = f$ then for some v', $v \leq v'$ and $v'(A) = F$.

This condition can be read off from the validity of the classical argument $\sim\sim A \vdash A$ using $||\sim||$. It amounts to saying that whenever a sentence is false at a valuation, there is always some extension of that valuation where it is settled false. Garson (1990, p. 163) shows that the system PL- expresses exactly the semantics $||PL-||$, which is the conjunction of $||I\neg||$ with $||\sim\sim||$.

Since we are thinking of $||PL-||$ as an open future semantics, it is worth looking at what $||\sim\sim||$ says more carefully. Consider the sentence $A\rightarrow A$, where A reports some future event over which someone has control. So, for example, A might report a sea battle at a date in the future. Presumably A is unsettled, and so at the present situation v, $v(A) = v(\sim A) = f$. Though one can control A, it does not seem correct to assert that one has control over $A\rightarrow A$. The reason is that $A\rightarrow A$ is inevitable, that is, it will turn out true no matter how A gets settled, and so my actions have nothing to do with settling it. Therefore, although A is unsettled, we want $A\rightarrow A$ to be settled true, since it is inevitable. This is exactly what $||\sim\sim||$ entails, for it amounts to the claim that all inevitable sentences are true, and hence settled true.

To demonstrate that, we will need an official definition of inevitability—the notion that a sentence is true at every point in the future where it is settled. Here it helps to deploy the concept of quasi-truth, for the inevitable sentences are simply those that are quasi-true at every possibility for the future. Humberstone (2011, p. 896) calls this weak inevitability.

(INV) A is inevitable at v iff for all $v' \in V$, if $v \leq v'$, then $v'(A) = qT$.

It is now easy to prove that $||\sim\sim||$ is equivalent to $||IT||$, the claim that all inevitable sentences are settled true.

||IT|| If A is inevitable at v, then v(A) = T.

Theorem: V obeys ||∼∼|| iff V obeys ||IT||.

Proof. The contrapositive of || ∼∼ || amounts to ||C∼∼||, and when the definition of inevitability is unpacked in ||IT||, we have ||IT′||.

||C ∼∼ || If for all v′ ∈ V, if v ≤ v′ then v′(A) ≠ F, then v(A) = t.
||IT′|| If for all v′ ∈ V, if v ≤ v′, then v(A) = qT, then v(A) = T.

These are equivalent in light of (Fact 5) and (Fact 1).

The upshot of this theorem is that the semantic contribution of the law of double negation to the intuitionistic semantics amounts to exactly the requirement that inevitable sentences are settled true. This fits nicely with our intuitions about when agency is possible for situations described by sentences about the future.

6 What Natural Deduction Rules for Disjunction Express

It is natural to express our choices using disjunction. In this chapter, a discussion of disjunction has been postponed because of difficulties that arise for it in intuitionist logic. Taken alone, the system S∨ consisting of (∨ In) and (∨ Out) expresses a condition ||S∨|| on models that does not appear to provide properly recursive and non-circular truth conditions for the connective ∨. Although some possible solutions for the problem are suggested (Garson 2001, p. 126–127 and Garson 2013, Chapter 7), these are not fully satisfactory, for they require relaxing the standards for when a set of rules expresses connective meaning, or the addition of additional *ad hoc* semantical structure. One symptom of the problem is that the Persistence Lemma no longer holds when ||S∨|| is added to ||I¬||.

A main result of this chapter is to show how these problems are resolved in standard propositional logic, where the classical condition ||∼∼|| is expressed. When ||∼∼|| holds, the system S∨ of disjunction rules expresses the following relatively straightforward truth condition, which we call *the quasi-truth interpretation* for disjunction.

||q∨||v(A∨B) = t iff for all v′ ∈ V, if v ≤ v′ then v′(A) = qT or v′(B) = qT.

So the truth condition for ∨ expressed by a classical logic states that A∨B is true when one of its disjuncts is quasi-true in every possible future. Though developed independently, this treatment of disjunction has already been deployed by

Humberstone (1981) for a possibilities logic where situations are treated as sets of worlds, or time intervals. It is a simple matter to show that the open futures semantics given here is isomorphic to the propositional fragment of Humberstone's semantics.

The quasi-truth interpretation ||qv|| for disjunction is more than a random artifact of a search for what propositional logic expresses. It is well suited for matching intuitions about sentences of the form of Excluded Middle such as A∨∼A, when A reads: 'there is a sea battle tomorrow at t′ and 't' refers to a time in the future. The reason A∨∼A is settled true, and so not a target for agency, is that in every possible future, either A is true if settled or ∼A is true if settled. That follows directly from the fact that being settled true, being settled false and being unsettled are exhaustive categories. So the truth condition ||qv|| explains nicely how it can be that A∨∼A is settled true at a time when both of its disjuncts is unsettled. Therefore ||qv|| both accepts Excluded Middle and leaves room for unsettled sentences. To put it another way, the semantics obeys the dictum: "no choice before its time" (Belnap 2005, Sect. 3.1), since disjuncts of A∨∼A may remain unsettled. However, *disjunctions* may be settled well before the time their disjuncts are settled.

It may appear to the reader that we could simplify ||qv|| by saying that v(A∨B) = t iff either A or B is inevitable at v. However, ||qv|| does not say that the truth of A∨B entails that one of its disjuncts is inevitable. (Pay attention to the relative scopes of 'or' and of the universal quantifier on the right hand side of ||qv||.) Were that to be true, ||qv|| would collapse to the classical truth condition, since the inevitability of a disjunct is equivalent to its being t, by ||IT||. It is crucial to the very nature of ||qv|| that the condition expressed by Sv not be classical, for were that to be true, the acceptance of A∨ ∼A, would entail that either v(A) = T or v(A) = F, leaving no room for unsettled sentences. This in turn would convert the truth conditions for each of the connectives into its classical counterpart.

We are ready to report on the main result. Let the language of PL include the connectives &, →, ∼, and ∨, and let PL be PL- plus Sv, the ND rules for disjunction. Let ||PL|| be the semantics for the language of PL that results from adding ||qv|| to ||PL−||. Then PL expresses ||PL||, and so ||PL|| qualifies as a natural semantics for PL.

Theorem 1. PL expresses ||PL||.

The proof of this theorem is found in Appendix A. This result immediately entails the completeness of PL for ||PL||. (See Garson (1990), p. 159.)

7 No Past Branching

The accessibility relation in an open future semantics is ordinarily taken to be reflexive, transitive and antisymmetric.

(Antisymmetric) For all v, v′ in V, if v ≤ v′ and v′ ≤ v, then v = v′.

The relation $\leq$ of ||PL|| obeys these three properties. However, it is also presumed that the set of open possibilities has the structure of a forward facing tree, with branching towards the future, but none in the past. Belnap, Perloff and Xu (2001, p. 185ff) argue that no branching in the past is essential to our concept of agency. So if ||PL|| were to count as a full-blooded open futures semantics, we would expect it to satisfy the following condition, for all v, v′ and u in V.

(No Past Branching) If $v \leq u$ and $v' \leq u$ then $v \leq v'$ or $v' \leq v$.

However there are models V that obey ||PL|| where (No Past Branching) fails. Nothing said so far rules out the possibility that two valuations v and v′ might extend to the same valuation u even though the two are not comparable, that is neither $v \leq v'$ nor $v' \leq v$. So one might object that ||PL|| does not really qualify as a semantics of the open future, since it does not treat the past properly. However, the problem can be repaired by constructing a finer individuation of the set of possibilities. Instead of taking the "moments" in our model to be valuations, think of them instead as pairs <c, v> where c is a *past* for v, that is, a connected set of valuations u that are earlier than v in the ordering $\leq$. Given any set of valuations V obeying ||PL||, it is possible to construct a *past model* P = <W, $\subseteq$, u> for V by letting the members of W be pairs <c, v> where c is a past for v in V, rather than the valuations themselves. (We could also require a past c to be a past *history* for v, where c must be a *maximal* connected set, but all that does is to complicate the result given below.) This idea matches the intuition that were there to be two moments where all the same sentences were true but with different pasts, we would count them non-identical. By defining the relation $\subseteq$ and the assignment function u for P in the appropriate way, it will be possible to show that a past model for V has a relation $\subseteq$ that obeys (No Past Branching), and P preserves the truth-values for valuations in V, in a sense to be made clear below. Therefore, a set of valuations V has the resources to set up a truth preserving structure that qualifies as a full-fledged semantics for an open future.

Here are the relevant definitions, where it is presumed that $\leq$ is defined by ($\leq$) above.

(Connected) Relation $\leq$ is *connected* for set s iff for every v and $v' \in s$, $v \leq v'$ or $v' \leq v$.
(Chain) A *chain* c (for V) is a subset of V such that $\leq$ is connected for c.
(Past for v) c is a *past* for v iff c is a chain for V, $v \in c$ and for every $u \in c$, $u \leq v$.

(Past Model for V) The *past model* P = <W, $\subseteq$, u> for V is defined as follows:

$$W = \{<c, v> : c \text{ is a past for } v \text{ and } v \in V\}.$$

To save eyestrain, we abbreviate pairs '<c, v>' to 'cv'.
The relation $\subseteq$ is defined for cv and $c'v' \in W$, as follows.

($\subseteq$) $cv \subseteq c'v'$ iff $v \leq v'$ and $c = \{u : u \in c' \text{ and } u \leq v\}$

So $cv \subseteq c'v'$ holds when $v < v'$ and c and c' agree on the past up to v.
The assignment function u is defined for $cv \in W$, so that

$$u(cv, p) = v(p), \text{ for propositional variables p.}$$

The function u is extended to the complex sentences by the following analogs of truth conditions in ||PL||, for arbitrary w in W.

||u&|| $u(w, A\&B) = t$ iff $u(w, A) = t$ and $u(w, B) = t$.
||u→|| $u(w, A \rightarrow B) = t$ iff for all $w' \in W$, if $w \subseteq w'$, then $u(w', A) = f$ or $u(w', B) = t$
||u ∼ || $u(w, \sim A) = t$ iff for all $w' \in W$, if $w \subseteq w'$, then $u(w', \sim A) = f$.
||uq∨|| $u(w, A \vee B) = t$ iff for all $w' \in W$, if $w \subseteq w'$, then for some $w'' \in W$, $w' \subseteq w''$ and either $u(w'', A) = t$ or $u(w'', B) = t$.

Now that the past model for V is defined, it is possible to show that Reflexivity, Transitivity, Antisymmetry and (No Past Branching) all hold in this model. So, in that sense, V generates a full-fledged open future semantics. We can also show that the past model for V is truth preserving in the sense that $u(cv, A) = v(A)$ for all wffs A and any past c for v. The intuition behind this result is that the truth conditions "face the future" and so are insensitive to adjustments to past structure created by past models.

Past Model Theorem. Let V be any set of valuations that obeys ||PL||. Then the past model $P = <W, \subseteq, u>$ for V is such that $u(cv, A) = v(A)$ for all wffs A, and any past c for v, and the frame $<W, \subseteq>$ is reflexive, transitive, antisymmetric, and obeys (No Past Branching).

The proof of this theorem appears in Appendix B. The ability of V to generate past models is important because it shows that V has the resources for defining a frame $<W, \subseteq>$ with the right structure for an open future. Furthermore, when any set of sentences H is satisfied by V, we know that it is also satisfied in the past model for V. As a result, any argument H / C is V-valid for all V obeying ||PL|| iff it is valid for all past models for V.

8 Open Future Semantics and Supervaluations

The reader may complain that open futures semantics for PL is nothing new. The existence of non-classical interpretations for classical propositional logic has been well-known since the invention of supervaluation semantics (van Fraassen 1969). Supervaluations may be used to show how a sentence of the form $A \vee \sim A$ can be validated in three-valued scheme that allows the values of A and ∼A to be unsettled. So, supervaluations can already serve the role of providing for a logic of an open future.

Although it is granted that ||PL|| and supervaluations have some strong points of similarity, there are crucial points of difference, and these argue for the superiority

of the open futures approach embodied in ||PL||. To make the issues clear, a brief account of supervaluation semantics is in order,

The fundamental idea behind supervaluations is to allow some sentences to remain undetermined, but only provided that would be compatible with truth-values fixed by classical truth tables. When the atomic constituents of a sentence A are not defined, the value of A is t if all ways of filling in the missing values using classical truth tables would assign A t, and the value of A is f if every way of filling the missing values yields f. Otherwise A is left undefined.

Let us present the idea more formally following (McCawley 1993, 334ff.). Let H be any consistent set of sentences. Let H $\models$c A mean that every classical valuation that satisfies H (assigns t to every member of H) also satisfies A (assigns t to A). Then *the supervaluation s_H induced by set H* is the assignment of truth-values T, F, and U (neither or undefined) such that (SVT), (SVF) and (SVU). hold.

(SVT) $s_H(A) = T$ if H $\models$c A.
(SVF) $s_H(A) = F$ if H $\models$c $\sim$A.
(SVU) $s_H(A) = U$ if neither $s_H(A) = T$ nor $s_H(A) = F$.

The relation $\models$s of supervaluation validity is now defined as follows. H $\models$s C holds iff every supervaluation induced by a consistent set of sentences that satisfies H also satisfies C. A well-known result concerning supervaluations is that the notion of validity defined by the class of supervaluations is equivalent to classical validity, and so PL is sound and complete for supervaluation semantics.

Similarities between supervaluations and ||PL|| are obvious. Think of an inducing set H as defining a corresponding valuation v_H such that $v_H(A) = t$ exactly when H $\models$c A. Now note the parallels between the conditions (SVT), (SVF), (SVU) and their counterparts (DefT), (DefF), (DefU) in ||PL||, which we write in equivalent forms with the help of (Fact 1) and (Fact 2) to emphasize the correspondence.

(DefT) $v(A) = T$ iff $v(A) = t$. (Fact 1)
(DefF) $v(A) = F$ iff $v(\sim A) = t$. (Fact 2)
(DefU) $v(A) = U$ iff neither $v(A) = T$ nor $v(A) = F$.

This idea provides the basis for a result showing a 3-valued preserving isomorphism between the set of all supervaluations and the canonical model V_{PL} of PL, which is defined as the set of all valuations v which are closed under deduction in PL, that is, such that whenever $v(H) = t$ and H $\vdash$ C, $v(C) = t$. (See (Garson 2013, Section 9.2) for details.) This means we can translate from talk of valuations in V_{PL} to talk of supervaluations at will.

A related point of similarity has to do with the partial truth tables for the connectives in the two schemes. A partial truth table records the 3-valued output (T, F, or U) for a connective as a function of its 3-vauled inputs in a 3 by 3 matrix. The tables for the binary connectives are not entirely functional (hence the term 'partial'), since it is only possible to fix 8 of the 9 values uniquely, leaving one cell where

two values are possible. Garson (2013, Section 9.3) shows that the partial tables for supervaluations and those for sets of valuations that satisfy ||PL|| are identical.

Despite these similarities, there are fundamental and crucial differences between ||PL|| and supervaluation semantics. Not only is supervaluation semantics not a legitimate interpretation for PL, it fails to define any meanings for the connectives at all.

One point should be clear at the outset. ||PL|| provides an alternative semantics for PL by providing intensional truth conditions for the connectives with the help of a structure $<V, \leq>$ that can be read as defining a temporal/modal order. Supervaluations simply lack this structure, so they do not qualify as semantics for an open future. Furthermore, it is far from clear that supervaluation semantics offers any particular account of connective truth conditions. Granted, a statement of connective truth conditions is implicit in the consequence relation $\models_c$ where classical conditions are chosen. However, it would not change the outcome in any way were we to define $\models_c$ using ||PL|| or even proof theoretically, so that $H \models_c C$ iff $H \vdash_{PL} C$. All that matters for success of the supervaluation tactic is that the relation $\models_c$ pick out the valid arguments of propositional logic, and this can be done with an alternative semantics or even using syntactic means. Therefore, supervaluation semantics radically underdetermines the meaning of the connectives, if it gives them any meanings at all.

A second major point of difference is that supervaluations do not preserve the validity of the PL rules. Supervaluation semantics is not sound for PL, so it can hardly count as a way of interpreting its rules. Early on (van Fraassen (1969, p. 81) noted that the following classical rule is unsound for some classes of supervaluations that are subsets of SV.

$$\frac{A \vdash B}{\sim B \vdash \sim A}$$

This failure is pervasive. All classical ND rules that discharge hypotheses fail as well: for example ($\rightarrow$ In), ($\sim$ In), ($\sim$ Out) and ($\vee$ Out). Williamson (1984, p. 120) takes this to be a profound betrayal of our ordinary deductive practices, and argues that therefore supervaluations are not up to the task of providing a coherent account of vagueness. Analog complaints against treating supervaluation semantics as an account of the open future seem equally compelling.

The upshot of this is that the pathological behavior of supervaluations is massive. While supervaluation semantics accepts as valid the valid *arguments* of PL, that is, the arguments PL *asserts*, it does not respect the deductive behavior of $\rightarrow$, $\sim$, and $\vee$ as embodied in their natural deduction rules. So, it disagrees fundamentally with the *use* to which the connectives are put.

This underscores an important moral. A theory that attempts to define connective meaning by which arguments are *accepted*, faces problems of underdetermination. As Garson (2013) shows, traditional systems built from axioms and rules defined over sentences faces massive underdetermination results. They simply cannot define any coherent meanings for the connectives. Supervaluations fail to give meanings to the

connectives for a similar reason: they simply fail to do justice to the uses to which the connectives are put in the process of reasoning from one argument form to another. On the other hand, a theory that takes seriously the deductive roles connectives play, by exploring constraints that arise from assuming that the rules preserve validity, may fix a unique interpretation of the connectives, as does ||PL||. For those who adopt the natural deduction rules of PL to guide their reasoning, ||PL|| tells us what the connectives mean. It should come as no surprise that ||PL|| is useful, since it is the interpretation most of us have been employing all along whether we know it or not.

9 Defeating Fatalism

The reader may have serious worries about ||PL||. (Fact 1) entails that truth and settled truth are the same thing.

(Fact 1) If $v(A) = t$ then $v(A) = T$.

Furthermore, since PL is classical, the Law of Excluded Middle is a theorem. The concern is that these two features do not leave room for unsettled values in the semantics. Arguments related to this concern have surfaced at many points in the literature. Two notable examples are Taylor (1962) famous argument for fatalism, and Williamson's purported demonstration that supervaluation semantics has no room for unsettled values (1994, p. 300).

Here a basic argument form concerning ||PL|| will be examined with an eye to uncovering the flaw in its reasoning. Once the main idea is in place, the same solution may be applied wherever arguments of this kind arise. Here is the basic argument form:

Ur Argument for Fatalism

A or not-A.	Excluded Middle
If A, then it is settled that A.	(Fact 1)
If not-A, then it is settled that not-A.	(Fact 1)
Therefore , either it is settled that A or settled that not-A.	

The argument has the form of (∨ Out), so it is classically valid. The premises appear indisputable, since adopting classical logic gives us Excluded Middle and (Fact 1) was proven for ||PL||. It appears to follow that there is no room in ||PL|| for any unsettled sentences, for when it is settled that not-A, that is, $v(\sim A) = T$, we have $v(A) = F$, so that the conclusion of the argument asserts that A must be settled true or settled false, hence settled.

The problem with this reasoning is that it does not take proper care in distinguishing the object language from the metalanguage. Therefore, the English renderings

of the premises of the Ur Argument are ambiguous. Let us attempt to rewrite the argument more accurately using the notation: 'v(A) =' in which (Fact 1) is actually written. Here we assume v is an arbitrary member of V.

v Argument for Fatalism

$v(A \vee \sim A) = t$.	Excluded Middle is V-valid
If $v(A) = t$, then $v(A) = T$.	(Fact 1)
If $v(\sim A) = t$, then $v(\sim A) = T$.	(Fact 1)
Therefore $v(A) = T$ or $v(\sim A) = T$.	

It should clear right away that this argument is invalid. The problem is that we need:

$$(\text{or} \sim)\ v(A) = t \text{ or } v(\sim A) = t.$$

rather than what we see in the first premise: $v(A \vee \sim A) = t$ in order for it to have the form of ($\vee$ Out) in the metalanguage. So let us replace the first premise with (or $\sim$).

Or $\sim$ Argument for Fatalism

$v(A) = t$ or $v(\sim A) = t$.	(or $\sim$)
If $v(A) = t$, then $v(A) = T$.	(Fact 1)
If $v(\sim A) = t$, then $v(\sim A) = T$.	(Fact 1)
Therefore $v(A) = T$ or $v(\sim A) = T$.	

This will not help matters, since here is no reason to accept (or $\sim$). As (or $\sim$) does not have the form of Excluded Middle, there is no classical argument in its favor. Furthermore, it begs the question, because (or $\sim$) just amounts to the claim that there are no unsettled values. Even worse, (or $\sim$) is demonstrably false. We can find models V that obey ||PL|| where there are unsettled values. For example, consider the set V* of valuations v that respect deductive closure in PL, that is, if $v(H) = t$ and $H \vdash_{PL} C$ then $v(C) = t$. It is near trivial to prove that V* is a model of PL, and since ||PL|| is expressed by PL, V* must obey PL. However, there are many members of V*, notably the valuation $v\vdash$ that assigns t to all and only theorems of PL which allow $v\vdash (p) = v\vdash (\sim p) = f$.

Perhaps a second variation of this argument form might work by changing the first premise to a claim with the form of Excluded Middle that is true of ||PL||, where the disjunction and negation are expressed in the metalanguage, and the third premise is modified to guarantee that the argument has the form of ($\vee$Out):

Or Not Argument for Fatalism

v(A) = t or not v(A) = t.	Metalanguage Excluded Middle
If v(A) = t, then v(A) = T.	(Fact 1)
If not v(A) = t, then v(∼A) = T.	?????
Therefore v(A) = T or v(∼A) = T.	

However, the third premise is no longer supported by (Fact 1), and it is demonstrable that this claim is false for some valuations in models that obey ||PL||. If not v(A) = t, then v(A) = f. But this, as we have just argued, is compatible with v(∼A) = f, thus undermining v(∼A) = T. (See Brown ann Garson (in preparation) for the deployment of this tactic to show that ||PL|| can overcome problems Williamson lodges against supervaluations.)

The upshot of this is that (Fact 1), acceptance of Excluded Middle, and the existence of unsettled sentences are demonstrably compatible with each other. In fact ||PL||, the very semantics that tells us what is expressed by classical rules, shows how this is possible. The secret is that the quasi-truth interpretation of disjunction makes room for accepting A∨∼A when the value of A is unsettled.

This realization has direct applications to a variety of arguments that purport to show that there cannot be an open future. Take a simplification of Taylor's famous argument (Taylor 1962, p. 129 ff.) for fatalism. Here Q abbreviates "A naval battle will occur", and O abbreviates "I issue the order for the battle", and it is presumed that O is necessary and sufficient for Q.

Q is true or not-Q is true.
If Q, then O is out of my control.
If not-Q, then not-O is out of my control.
Therefore, O is out of my control or not-O is out of my control.

Given the strategy of ||PL|| semantics, it may appear that the argument has a valid form, and that all premises must be accepted. ||PL|| would apparently support the second premise, because if Q is true, Q is settled true, and whatever is settled true entails the settled truth of any sentence (such as O) necessary for Q. Therefore O is settled and therefore not the subject of my control despite its being in the future. Similar reasoning can be given to support the third premise. It appears ||PL|| yields fatalist conclusions.

However, it is easy to see what has gone wrong when care is taken to present the argument with sufficient notational detail. If we take its form to be the analog of the v Fatalist argument, we have the following, which has a true first premise and an invalid form:

v Argument for Fatalism

v(Q∨∼Q) = t.	Excluded Middle
If v(Q) = t, then v(O) = T.	
If v(∼Q) = t, then v(∼O) = T.	
Therefore v(O) = T or v(∼O) = T.	

Modifying the first premise yields a valid form:

Or Argument for Fatalism

v(Q) = t or v(∼Q) = t.	?????
If v(Q) = t, then v(O) = T.	
If v(∼Q) = t, then v(∼O) = T.	
Therefore v(O) = T or v(∼O) = T.	

However, the first premise no longer has the form of Excluded Middle, and in fact begs the question by claiming that Q is determined, something that can be refuted in ||PL||.

Suppose we attempt to fix this by expressing the negation in the object language and modifying the third premise to maintain the form of (v Out).

Or Not Argument for Fatalism

v(Q) = t or not v(Q) = t.	Metalanguage Excluded Middle
If v(Q) = t, then v(O) = T.	
If not v(Q) = t, then v(∼O) = T.	?????
Therefore v(O) = T or v(∼O) = T.	

Now the third premise is the problem, for it is demonstrably false.
When v(Q) is not t, it is f. Since Q is necessary and sufficient for O, O is also f, and its being f is compatible with O's being unsettled, and hence a target for agency.

The conclusion to be drawn is that because classical logic essentially takes on an open futures interpretation, it automatically has the resources to undermine arguments for fatalism, and this despite its acceptance of Excluded Middle and the seemingly fatalist proposal that truth amounts to settled truth.

Appendix A

Here we provide a proof of Theorem 1.

Theorem 1. PL expresses ||PL||.

The first task is to verify that the Persistence Lemma holds for ||PL||.

$\leq$ Lemma. If V obeys ||PL||, then $v \leq v'$ iff for a every wff A,
if v(A) = t then v'(A) = T.

Proof. Let us define: $v \leq_A v'$ to mean that for every wff A, if $v(A) = t$ then $v'(A) = t$. Let V be any set of valuations such that ||PL|| holds. It will be sufficient to show that $v \leq v'$ iff $v \leq_A v'$. The proof from right to left is trivial. Now assume $v \leq v'$, and show that for any wff A, if $v(A) = t$ then $v'(A) = t$ by mathematical induction on the length of A. For the base case we must show that when A is a propositional variable p, if $v(A) = t$ then $v'(A) = t$. This is guaranteed by the definition of $\leq$. For the inductive case, assume the inductive hypothesis for wffs B and C, and show that it holds when A has one of the forms B&C, B→C, ∼B, and B∨C as follows.

A has the form B&C. Assume $v(B\&C) = t$, from ||&|| it follows that $v(B) = t$ and $v(C) = t$. By the inductive hypothesis, $v'(B) = t$ and $v'(C) = t$, and so $v'(B\&C) = t$ by ||&||.

A has the form B→C. Assume $v(B \rightarrow C) = t$, and establish $v'(B \rightarrow C) = t$ using $||\rightarrow||$. Assume that v'' is an arbitrary member of V such that $v' \leq v''$ and show that $v''(B) = f$ or $v''(C) = t$ as follows. From $v \leq v'$, $v' \leq v''$ and the transitivity of $\leq$, it follows that $v \leq v''$. Given $v(B \rightarrow C) = t$, it follows by $|| \rightarrow ||$ that either $v''(B) = f$ or $v''(C) = t$ as desired.

A has the form ∼A. Proof similar to the preceding case.

A has the form B∨C. Assume $v(B \vee C) = t$, and establish $v'(B \vee C) = t$ as follows. By ||q∨||, it will be sufficient to show that for all v'' if $v' \leq v''$, then $v''(B) = qT$ or $v''(C) = qT$. So assume that $v' \leq v''$ for any valuation v''. Then by transitivity of $\leq$, $v \leq v''$, and by ||q∨|| and $v(B \vee C) = t$, it follows that $v''(B) = qT$ or $v''(C) = qT$ as desired.

Now for the main theorem.

Theorem 1. PL expresses ||PL||.

Proof. To show PL expresses ||PL||, it must be shown that V obeys ||PL|| iff the rules PL are V-valid. For the proof of this from left to right, assume V obeys ||PL|| and show that the rules preserve V-validity as follows. The demonstration for rules other than those for disjunction is found in (Garson 1990, Theorems 1-3 pp. 21 ff.). What remains to show is that (∨ In) and (∨ Out) preserve V-validity.

(∨In). Assume $H \models_V A$ and show that $H \models_V A \vee B$ by assuming that v is any member of V such that $v(H) = t$ and proving that $v(A \vee B) = t$ as follows. In light of ||q∨||, $v(A \vee B) = t$ will follow if we demonstrate that whenever $v \leq v'$, $v'(A) = qT$ or $v'(B) = qT$. So let v' be any member of V such that $v \leq v'$. From $H \models_V A$ and $v(H) = t$, it follows that $v(A) = t$. Hence by $v \leq v'$ and the $\leq$ Lemma, $v'(A) = t$. From this it follows immediately by (Fact 1) that $v'(A) = T$, and hence $v'(A) = qT$. The proof that if $H \models_V B$ then $H \models_V A \vee B$ is similar.

(∨Out). Assume (1) $H \models_V A \vee B$ (2) H, A $\models_V$ C and (3) H, B $\models_V$ C, and show that $H \models_V C$, by assuming the opposite and deriving a contradiction. From $H \not\models_V C$ it follows that for some v in V, $v(H) = t$ and $v(C) = f$. Given (1) $H \models_V A \vee B$, it follows that $v(A \vee B) = t$. By $|| \sim\sim ||$ and $v(C) = f$, it follows that for some v' such that $v \leq v'$, $v'(C) = F$. By ||q∨||, $v(A \vee B) = t$, and $v \leq v'$, it follows that $v'(A) = qT$ or $v'(B) = qT$. Suppose it is $v'(A)$ that is qT. By (Fact 5), $v'(A) \neq F$,

and so by (DefF), there must be some $v'' \in V$ such that $v' \leq v''$ and $v''(A) = t$. By $v(H) = t$, the transitivity of $\leq$, and the $\leq$ Lemma, we have that $v''(H) = t$. From $v''(A) = t$ and (2) H, A $\models_V$ C, we have $v''(C) = t$. But $v' \leq v''$ and $v'(C) = F$ entails that $v''(C) = f$ a contradiction. Similarly a contradiction follows from (3) H, B $\models_V$ C assuming $v''(B) = t$. Either way, we have the desired contradiction.

The next stage of the proof is to show that when the rules of PL preserve V-validity, V obeys ||PL||. The proof that V obeys the truth conditions of ||PL−|| is given in Garson (1990 pp. 121ff.), and the demonstration for ||∼∼|| is given in Garson (2001, Lemma 4.4 p. 164), so all that remains is to show that when both (∨ In) and (∨ Out) preserve V-validity, V obeys ||q∨|| as follows.

Proof of ||q∨|| *from left to right*. Assume that $v(A \vee B) = t$, and $v \leq v'$ for an arbitrary member v' of V, and show that $v'(A) = qT$ or $v'(B) = qT$ as follows. Since v' is a valuation, there is a sentence C such that $v'(C) = f$. Let $H_{v'}$ be the set of sentences assigned t by v'. It follows that $H_{v'} \models_V A \vee B$ by the following reasoning. Let u be any member of V such that $u(H_{v'}) = t$ and show $u(A \vee B) = t$ as follows. Since $u(H_{v'}) = t$, it follows that $v' \leq u$. But we had that $v(A \vee B) = t$, $v \leq v'$, and $v' \leq u$, so $u(A \vee B) = t$ by the $\leq$ Lemma and the transitivity of $\leq$. This establishes $H_{v'} \models_V A \vee B$; but we also have $v'(H_{v'}) = t$ and $v'(C) = f$, so $H_{v'} \not\models_v C$. It follows from the fact that (∨ Out) preserves V-validity that either $H_{v'}, A \not\models_v C$ or $H_{v'}, B \not\models_v C$. In the first case, there must be a valuation v'' such that $v''(H_v, A) = t$ and $v''(C) = f$. Because $v''(H_{v'}) = t$, $v' \leq v''$. Since $v''(A) = t$, and $v' \leq v''$, $v'(A) \neq F$, hence by (Fact 5), $v'(A) = qT$. Therefore $v'(A) = qT$ or $v'(B) = qT$ as desired. The second case, where $H_{v'}, B \not\models_v C$, is similar.

Proof of ||q∨|| *from right to left.* Assume that for all $v' \in V$, if $v \leq v'$, then $v'(A) = qT$ or $v'(B) = qT$, and show that $v(A \vee B) = t$ as follows. By the contrapositive ||C ∼∼|| of ||∼∼||, it will be sufficient for proving $v(A \vee B) = t$ to show that for any $v' \in V$ if $v \leq v'$, then $v'(A \vee B) \neq F$.

$$||C \sim\sim|| \text{ If for all } v' \in V, \text{ if } v \leq v' \text{ then } v'(A) \neq F, \text{ then } v(A) = t.$$

So let v' be any member of V such that $v \leq v'$ and show $v'(A \vee B) \neq F$ as follows. By our initial assumption, we have $v'(A) = qT$ or $v'(B) = qT$. Suppose that $v'(A) = qT$. Then by (Fact 5), $v'(A) \neq F$ and for some v'', $v' \leq v''$ and $v''(A) = t$. Since the V-validity of the rules of PL is preserved, all provable arguments of PL are V-valid including A ⊢ A∨B and B ⊢ A∨B. Therefore for any valuation $v \in V$, if either $v(A) = t$ or $v(B) = t$, $v(A \vee B) = t$. By $v''(A) = t$, it follows that $v''(A \vee B) = t$ and so $v'(A \vee B) \neq F$ as desired. When $v'(B) = qT$, the reasoning is similar.

This completes the proof of the theorem.

Appendix B

Here we prove the Past Model Theorem.

Past Model Theorem. Let V be any set of valuations that obeys ||PL||. Then the past model P = <W, ⊆, u > for V is such that u(cv, A) = v(A) for all wffs A, and any past c for v, and the frame <W, ⊆> is reflexive, transitive, antisymmetric and obeys (No Past Branching).

We begin with a few lemmas.

Lemma 1. If <W, ⊆, u> is a past model for V, then <W, ⊆> is reflexive, transitive, antisymmetric and obeys (No Past Branching).

Proof. It is easy to verify that <W, ⊆> is reflexive, transitive and antisymmetric. To show that it obeys (No Past Branching), let cv, c′v′, and c″v″ be any members of W, such that c′v′ ⊆ cv and c″v″ ⊆ cv and demonstrate that c′v′ ⊆ c″v″ or c″v″ ⊆ c′v′ as follows. We have from the definition of ⊆ that v′ ≤ v, v″ ≤ v, c′ = {u : u ∈ c and u ≤ v′} and c″ = {u : u ∈ c and u ≤ v′}. When c is a past for v, it follows that v ∈ c. Therefore, v′ ∈ c′ and v″ ∈ c″. It follows from c′ = {u : u ∈ c and u ≤ v′} and c″ = {u : u ∈ c and u ≤ v″} that v′ ∈ c and v″ ∈ c. Since c is connected, it follows that v′ ≤ v″ or v″ ≤ v′. In the first case, it is possible to show that c′ = {u : u ∈ c″ and u ≤ v′} from which it follows immediately that c′v′ ⊆ c″v″. To show that c′ = {u : u ∈ c″ and u ≤ v′} simply show the following.

$$u \in c' \text{ iff } u \in c'' \ \& \ u \leq v'$$

The proof of this from right to left follows from c′ = {u : u ∈ c and u ≤ v′} and c″ = {u : u ∈ c and u ≤ v″}. For the other direction, use the same two facts, and v′ ≤ v″. In case v″ ≤ v′, it follows that c″v″ ⊆ c′v′ by similar reasoning.

Lemma 2. If v ≤ v′ and cv ∈ W, then for some c′v′ ∈ W cv ⊆ c′v′.

Proof. Suppose v ≤ v′ and cv ∈ W. Then c is a past for v, hence v ∈ c, c is connected and for every u ∈ c, u ≤ v. Let c′ = c″ ∪ {v′}. Then c′ is a past for v′, because v′ ∈ c′ and for every u ∈ c′, u ≤ v′, and c′ is connected. The reason that c′ is connected is that cv ∈ W entails c is connected. The only additional member of c′ beyond the members of c is v′. But u ≤ v′ for all u ∈ c′. Therefore adding v′ to the connected set c results in a new connected set c′. Set c is clearly {u: u ∈ c′ and u ≤ v}, so by the definition of ⊆, cv ⊆ c′v′, and c′v′ is the desired member of W such that cv ⊆ c′v′.

Now we are ready to prove the Past Model Theorem.

Proof of the Past Model Theorem. To show that the frame <W, ⊆> is reflexive, transitive, antisymmetric and obeys (No Past Branching), simply appeal to Lemma 1. The proof that u(cv, A) = v(A) for all wffs A, and every past c for v is by structural induction on A. The base case and the case for & are straightforward.

In the case of negations ∼B show u(cv, ∼B) = v(∼B) by showing that the right hand side of ||∼|| and the right hand side of ||u ∼ || are equivalent given the

hypothesis of the induction: u(cv, B) = v(B), for any member cv of W. So we must show ||∼||r iff ||u∼||r.

$$||\sim||r \text{ For all } v' \in V, \text{ if } v \leq v' \text{then } v'(B) = f.$$

$$||u\sim||r \text{ For all } w' \in W, \text{ if } cv \subseteq w' \text{ then } u(w', B) = f.$$

For the proof from ||∼||r to ||u ∼ ||r, assume cv ⊆ w′ for any w′ ∈ W, and prove u(w′, B) = f as follows. Since w′ ∈ W, w′ = c′v′ for some v′ ∈ V. Since cv ⊆ c′v′, v ≤ v′. Hence v′(B) = f by ||∼||r. By the hypothesis of the induction, we have u(c′v′, B) = f as desired. For the other direction, assume v ≤ v′ and prove v′(B) = f as follows. We know cv ∈ W and v ≤ v′, so by Lemma 2, we have cv ⊆ c′v′ for some member c′v′ of W. From ||u∼||r, it follows that u(c′v′, B) = f. The hypothesis of the induction yields v′(B) = f as desired.

The case for → is similar.

The case for disjunctions B∨C will follow from showing that the following two conditions are equivalent, given the hypothesis of the induction.

||q∨||r For all v′ ∈ V, if v ≤ v′, then for some v″ ∈ V, v′ ≤ v″
and either v″(B) = t or v″(C) = t.
||uq∨||r For all w′ ∈ W, if cv ⊆ w′, then for some w″ ∈ W, w′ ⊆ w″ and either
u(w″, B) = t or u(w″, C) = t.

For the proof from ||q∨||r to ||uq∨||r, assume cv ⊆ w′ for any w′ ∈ W, and show that for some w″ such that w′ ⊆ w″, either u(w″, B) = t or u(w″, C) = t as follows. By the definition of W, w′ = c′v′ for some v′ ∈ V, and by cv ⊆ c′v′, we obtain v ≤ v′. From ||q∨||r, it follows that for some v″ ∈ V, v′ ≤ v″ and v″(B) = t or v″(C) = t. By Lemma 2, there is a member c″v″ of W such that c′v′ ⊆ c″v″. By the hypothesis of the induction u(c″v″, B) = t or u(c″v″, C) = t. So c″v″ is the desired w″ ∈ W such that w′ ⊆ w″ and either u(w″, B) = t or u(w″, C) = t.

For the proof from ||uq∨||r to ||∨||r, assume v ≤ v′, and find a v″ in V such that v′ ≤ v″ and either v″(B) = t or v″(C) = t as follows. Since cv ∈ W, it follows from v ≤ v′ by Lemma 2 that for some c′v′ in W, cv ⊆ c′v′. By ||uq∨||r, there is a member w″ of W such that c′v′ ⊆ w″ and either u(w″, B) = t or u(w″, C) = t. Since w″ must be c″v″ for some v″ ∈ V, we have by the hypothesis of the induction that v″(B) = t or v″(C) = t. We have c′v′ ⊆ c″v″, so v′ ≤ v″, hence v″ is the desired valuation such that v′ ≤ v″ and v″(B) = t or v″(C) = t.

This completes the proof of the theorem.

References

Belnap, N. 1962. Tonk, plonk, and plink. *Analysis* 22: 130–134.
Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future*. New York: Oxford University Press.

Belnap, N. 2005. Branching histories approach to indeterminism and free will. In Truth and probability, essays in honour of Hugues Leblanc, ed. B. Brown, and F. Lepage, 197–211. London: College Publishing.

Brown J.D.K., and J. Garson. The natural semantics of vagueness (in preparation).

Garson, J. 1990. Categorical semantics. In *Truth or consequences*, ed. M.J. Dunn, and A. Gupta, 155–175. Dordrecht: Kluwer.

Garson, J. 2001. Natural semantics. *Theoria* 67: 114–139.

Garson, J. 2013. *What logics mean: from proof theory to model-theoretic semantics*. Cambridge: Cambridge University Press.

Humberstone, L. 1981. From worlds to possibilities. *Journal of Philosophical Logic* 10: 313–339.

Humberstone, L. 2011. *The connectives*. Cambridge: MIT Press.

Lyons, J. 1968. *Introduction to theoretical linguistics*. London: Cambridge University Press.

McCawley, J. 1993. *Everything that linguists have always wanted to know about logic (but were ashamed to ask)*. Chicago: University of Chicago Press.

Taylor, R. 1962. Fatalism. *The Philosophical Review* 71: 56–66, also in Thinking about logic, ed. S. Cahn, R. Talise, and S. Aikin. Westview Press, Boulder, Colorado (Page numbers cited here are to the latter volume.).

Williamson, T. 1984. *Vagueness*. New York: Routledge.

van Fraassen, B. 1969. Presuppositions, supervaluations and free logic. In *The logical way of doing things*, ed. K. Lambert, 67–91. New Haven: Yale University Press.

Chapter 7
On Saying What Will Be

Mitchell Green

Abstract In the face of ontic (as opposed to epistemic) openness of the future, must there be exactly one continuation of the present that is what *will* happen? This essay argues that an affirmative answer, known as the doctrine of the Thin Red Line, is likely coherent but ontologically profligate in contrast to an Open Future doctrine that does not privilege any one future over others that are ontologically possible. In support of this claim I show how thought and talk about "the future" can be made intelligible from an Open Future perspective. In so doing I elaborate on the relation of speech act theory and the "scorekeeping model" of conversation, and argue as well that the Open Future perspective is neutral on the doctrine of modal realism.

1 Branching Time and Ontic Frugality

Our best current theory of the physical world implies that certain events occur in an irreducibly indeterministic way. For instance, if a radioactive atom decays, then its doing so is not the result of a prior sufficient physical condition. Instead, its decay is an irreducibly probabilistic process about which the most that can be said is that the atom's decay was something very likely to occur within a certain interval of time. At no time, however, was its decay physically determined to occur. So too, on certain views about freedom of will, in some cases agents act or choose freely, and according

I am grateful to participants at the, "What is Really Possible?" Workshop, University of Utrecht, June, 2012, for their comments on an earlier draft of this chapter. Research for this chapter was supported in part by Grant #0925975 from the National Science Foundation. Any opinions expressed herein are solely those of the author and do not necessarily reflect those of the National Science Foundation..

M. Green (✉)
Department of Philosophy, University of Connecticut, 101 Manchester Hall, 344 Mansfield Road, Storrs, CT 06269-1054, USA
e-mail: mitchell.green@uconn.edu

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_7,

to such views, this means that their free action or choice is not an event that had any prior, physical, sufficient condition.

On our best theory of the physical world, then, and on some views about free action or choice, there are points in time at which the future is ontologically open. This ontological openness is logically independent of epistemic openness. In many situations the future is epistemically open but ontologically closed. If the toss of a coin is a deterministic process, then how the coin will land may be epistemically open from the point of view of the person tossing it: she does not know whether it will come up heads or tails. By contrast, how the coin will land is not ontologically open. Things become more complicated when we ask whether the future can ever be epistemically closed but ontologically open. The atom's decay is ontologically open: nothing in the current physics of the situation determines whether it will decay within the next hour. Might its decay nevertheless be epistemically closed? Might, that is, there still, at least in principle, be an omniscient being who knows what the future holds even when the future is not physically determined by the present? How we settle this question in turn depends on how we settle another. When the future is ontologically open, will it always nevertheless be the case that one of the potentially many possible continuations of affairs from the present is *the one that is going to happen*?

The view that in a situation of ontic openness, exactly one of the many possible continuations of affairs from the present is the one that is going to happen, has come to be known as the doctrine of the Thin Red Line (TRL). This usage originates with Belnap and Green (1994), who delineated the above characterization, offered an alternative view of the ontic status of the future in the face of ontic openness, and argued that the doctrine of the TRL is of dubious coherence. Belnap et al. (2001) develop these lines of thought in greater detail. However, in the two ensuing decades, innovative research on the semantics of tense and related topics has made it plausible that the TRL is, contrary to Belnap and Green, technically tractable (Øhrstrøm 2009; Malpass and Wawer 2012). A workable formal semantics for tensed statements, including those about "the future" can be developed that achieves various benchmarks for logical adequacy.

These developments do not, however, immediately settle all questions we may have about the adequacy of the TRL. For the TRL involves positing one—of many possible futures that are physically possible continuations from an earlier moment—as distinguished from those others as being what is going to happen. A more parsimonious view would treat all those physically possible continuations as on a par. Assuming that the former, TRL approach can be spelled out in a logically coherent way, we may still ask whether parsimony counsels against it. That will be our strategy below. Our question will not be whether the TRL is coherent, but whether it is justified by the ontological, semantic or other pertinent facts. More precisely, we will ask whether we can eschew the TRL doctrine while doing all we need to do in making sense of our talk and thought about time, the future, and the openness thereof. My argument will be that positing a TRL is coherent but unnecessary in making sense of talk and thought about the future, and that therefore parsimony enjoins us to eschew it.

2 Some Concepts from Speech Act Theory

Before proceeding it will be helpful to have on hand some concepts from the theory of speech acts.

2.1 Speech Acts Versus Acts of Speech

An act of speech is simply an act of uttering a word, phrase or sentence. One performs acts of speech while testing a microphone or rehearsing lines for a play. By contrast 'speech act' is a quasi-technical term referring to any act that can be performed by saying that one is doing so (Green 2013a). Promising, asserting, commanding and excommunicating are all speech acts; insulting, convincing, and winning are not. One can perform an act of speech without performing a speech act. One can also perform a speech act without performing an act of speech: imagine a society in which a marriage vow is taken by virtue of one person silently walking in three circles around another. Similarly, among Japanese gangsters known as Yakuza, cutting off a finger in front of a superior is a way of apologizing for an infraction. A sufficiently stoic gangster can issue an apology in this manner without making a sound. Also, speech acts can be performed by saying that one is doing so, but need not be. One can assert that the window is open by saying, "I assert that the window is open." But one also can simply say, 'The window is open," and if one does so with the appropriate intentions and in the right context, one has still made an assertion.[1]

Another feature distinguishing speech acts from acts of speech is that the former may be retracted but the latter may not be. I can take back an assertion, threat, promise, or conjecture, but I cannot take back an act of speech (Green 2013b). Of course, I cannot on Wednesday change the fact that on Tuesday I made a certain claim, promise, or threat. However, on Wednesday I can retract Tuesday's claim with the result that I am no longer at risk of having been shown wrong, and no longer obliged to answer such challenges as, "How do you know?" This pattern recurs with other speech acts such as compliments, threats, warnings, questions, and objections. By contrast, with speech acts whose original felicitousness required uptake on the part of an addressee, subsequent retraction mandates that addressee's cooperation. I cannot retract a bet with the house without the house's cooperation, and I cannot take back a promise to Mary without her releasing me from the obligation that the promise incurred.

[1] Failure to keep in view a distinction between speech acts and acts of speech can lead to mischief. For instance, R. Langton begins her 'Speech acts and unspeakable acts' (1993) as follows, "Pornography is speech. So the courts declared in judging it protected by the First Amendment. Pornography is a kind of act. So Catharine MacKinnon declared in arguing for laws against it. Put these together and we have: pornography is a kind of speech act." Although Langton's conclusion may be correct, the reasoning she uses to arrive at it is fallacious: the most that her premises establish is that pornography is an act of speech.

2.2 Saying Versus Asserting

In light of the speech act/act of speech distinction, it should also be plausible that one can say that P (for some indicative sentence P) without asserting P. One's saying that P may not even be a speech act, as in the microphone case above. Or one might put forth P as a conjecture, guess, or supposition for the sake of argument instead of asserting P. All this would be too banal to merit mention were it not for the fact that some prominent authors have used these terms in idiosyncratic ways. For instance, Grice uses 'say' in such a way that one who says that P must also speaker mean that P. (This is why he treats ironical utterances ("Nice job!" said to a server who drops a bowl of calamari on my lap) as cases of *making as if to say*, rather than as cases of saying; otherwise Grice would fall in line with more common usage according to which the speaker *said*, "Nice job!" but *meant* something else (Grice 1989)).

2.3 Two Levels of Determination

An indicative sentence may, relative to a context of utterance, express a proposition, which in turn may be asserted. The first (syntactic) level underdetermines the second (semantic) level, which in turn underdetermines the third (pragmatic) level. How does the syntax of a sentence underdetermine its semantics? The sentence might be either lexically or structurally ambiguous. Even if a precise syntactic characterization resolves structural ambiguities (such as those found in 'Every boy loves a girl.') it will not disambiguate all ambiguous words. Furthermore, even an unambiguous sentence can fail to express a proposition in the absence of a context of utterance. 'I am hungry,' is not ambiguous, but only expresses a proposition in a context of utterance containing a speaker. (For other context-sensitive terms such as 'here', 'now', 'recently', and 'you', the context must also supply a location, a time, a past and an addressee, respectively.) Suppose then that ambiguity has been banished from our sentence and that a context of utterance has been supplied. We now have the sentence expressing a semantic content, but it will still not be determined whether it is being used in a speech act, assertoric or otherwise. For that, we would need to determine that the speaker is intending to commit herself in a certain way.

2.4 Assertion Proper and the Assertive Family

Assertion is only one of many speech acts aimed at conveying information. Keeping these other types of act in view will help us bring out assertion's distinctive features. To help ensure clarity I will distinguish between the *assertive family* and

Table 1 Speech acts, what they express, and in what light they show it

Speech act	Expresses	As
Assertion that p	Belief that p	Justified appropriate for knowledge
Conjecture that p	Belief that p	Justified
Educated guess that p	Acceptance or belief that p	Justified
Guess that p	Acceptance of p	n/a
Presumption of p	Acceptance of p	Justified for current conversational purposes
Supposition of p (for argument)	Acceptance of p	Aimed at the production of justification for some related content r

assertion proper. The *assertive family* is that class of actions in which a speaker undertakes a certain commitment to the truth of a proposition. Examples are conjectures, assertions, presuppositions, presumptions and guesses. The type of commitment in question is known as word-to-world *direction of fit*. Members of the assertive family have word-to-world direction of fit, but we still do well to distinguish some of its members, such as conjectures, from *assertion proper*. We may begin to do so by noting that only assertion proper is expressive of belief. Were assertion not expressive of belief, it would not be absurd to assert, 'P but I don't believe it.' By contrast, it is not absurd to say, 'P but I don't believe it' when P is put forth as a guess, conjecture, or presupposition. These other members of the assertive family are thus not expressive of belief, although they may express other psychological states.

What is more, one who makes an assertion is open to the challenge, "How do you know?", whereas this would not be an appropriate challenge to one who issues the same content with the force of a conjecture or a guess. Instead, an appropriate challenge to a conjecture would be to ask whether the speaker has any reason at all for her conjecture; another would be strong grounds for believing the conjecture to be incorrect. By contract, one can appropriately guess without having any reason for the guess at all. (To the challenge that there must be something that made the speaker guess one thing rather than something else, we may reply: such a cause need not be a reason.) Here again we see grounds for distinguishing assertion proper from the assertive family.

A more general pattern emerges upon inspection of Table 1:

The second and third columns describe what felicitous speech acts express, and in what way they express it. While all six speech acts considered here involve commitment to a propositional content, only two require belief for their sincerity condition. Guesses, presumptions, and suppositions require only acceptance for their sincerity condition *sensu* Stalnaker (1984); educated guesses can go either way.

3 Assertion and Scorekeeping

A good case can be made for the claim that an assertion is, at least in part, a proffered contribution to conversational common ground. Suppose we have a set S of interlocutors. Then S will have a common ground, CG_S, which will be a (possibly empty) set of propositions that all members of S take to be true, and such that it is common knowledge that all members of S take them to be true. When a proposition $P \in CG_S$, speakers can felicitously presuppose P in their speech acts. For instance, if P is the proposition that Susan owns a zebra, then Frederick's utterance of *Susan is late for work today because her zebra is ill*, will be felicitous. If P is not in CG_S, then at best, Frederick's utterance will update common ground only if his interlocutors accommodate him by adding P to CG_S. Similarly, if $P \in CG_S$, then members of S can presuppose P in deliberating on courses of action.

Being a proffered contribution to conversational common ground is not, however, a sufficient condition for a speech act's being an assertion. Other members of the assertive family meet this condition without being assertions: for instance, an educated guess is also characteristically a proffered contribution to conversational common ground, but is not to be confused with assertion proper. So too for conjectures and perhaps even suppositions for the sake of argument. In order to distinguish assertion proper from other members of the assertive family, we need to note that assertion has normative properties that other members of its family do not share. One making an assertion puts forth what she does as justified above a certain level. By contrast, one making an educated guess, or for that matter a sheer guess, puts that same content forth with a lower expectation of justification.

Assertions, conjectures, suggestions, guesses, presumptions and the like are cousins sharing the property of commitment to a propositional content. They also share the property of being used, characteristically, to contribute to conversational common ground. Yet these speech acts differ from one another in the norms by which they are governed, and thereby in the nature of the commitment they generate for those who produce them. An assertion (proper) puts forth a proposition as something for which the speaker has a high level of justification; by contrast, a guess might put forth a proposition as true but need not present it as having any justification at all. (Educated guesses, by contrast, seem to be closer to conjectures, which require a higher level of justification than do guesses, but not as high as do assertions.) Correspondingly, a speaker incurs a distinctive vulnerability for each such speech act–including a liability to a loss of credibility and, in some cases, a mandate to defend what she has said if appropriately challenged.[2]

The development of common ground is typically only a means to other conversational ends. Many interlocutors work toward the development of common ground on their way to such larger aims as answering a question or forming a plan. In for

[2] This observation prompts a comparison between some speech acts and the phenomenon of handicaps as discussed in the evolutionary biology of communication. Green (2009) develops this analogy.

instance an *inquiry* a group of speakers undertake to answer a question to which none of them has, or takes herself to have, an answer. Characteristically, such an inquiry is embarked upon as follows. One interlocutor may raise a question, and others may respond by accepting it as a worthwhile issue for investigation. (This is often marked by such replies as "Good question," or more informally, "No idea; let's figure it out.") Once that has been done, the conversation now has a question Q *on the table*, and by definition has become an inquiry. Inquiries have distinctive norms. Participants in inquiries are to make assertions that are complete or, barring that, partial answers to the question on the table, and so long as no participant in the conversation demurs from those answers the interlocutors will make progress on their question. The level of informativeness required of inquirers flows from the content of the question on the table together with what progress has been made on that question thus far. If an inquiry has question Q on the table and thus far by offering and accepting assertions interlocutors have ruled out all but a few answers to Q, then all that remains is to determine which of the remaining answers is correct. Each interlocutor is to make assertions that will with the greatest efficiency, and in conjunction with the contents of common ground, rule out all but one of the answers that remain. Once that is done the question on the table will have been settled and this particular conversational task attained.

For those conversation that are also inquiries, then, a "scorekeeping" approach mandates keeping track not only of common ground, but of how its development moves interlocutors toward answering a question that is on the table.

4 Future-Directed Speech Acts

Assertions are not the only type of speech act that can raise questions about the reasonableness of talk of the future. We also conjecture, guess, suppose, and comport with other members of the assertive family while speaking of the future. As with assertion, so too with, say, conjectures: one might conjecture that the world's oceans will rise by an inch by the end of the decade. Here, too, we want to be able to say that such a conjecture may well be justified even if we are aware that the future is sufficiently open to leave alive the ontic possibility that things will not go this way.

Sometimes assertions in the face of an ontically open future are reasonable. An example is a case in which there is a genuine but small chance of something occurring, such as a series of fifty consecutive heads on a fair but ultimately indeterministic coin. Perhaps I can assert reasonably that the coin will not come up heads on fifty consecutive tosses. On the other hand, imagine we are faced with what we know to be a fair coin, and consider the prospect of its being flipped. Here it is hard to see how it would be reasonable to assert that the coin will come up heads. It might be slightly more reasonable to *conjecture* that it will. By contrast one can easily see how it would be reasonable to *guess* that it will come up heads.

It is sometimes reasonable to make assertions about what we know to be an open aspect of the future. Such reasonableness can be accounted for by the fact that these

assertions are well supported by currently available evidence. On the other hand, it can be reasonable to perform other acts within the assertive family about aspects of the future that are as likely to occur as not. Guesses about the tossing of a fair coin are a case in point. However, the reasonableness of such acts, be they assertions, guesses, conjectures, or suppositions for the sake of argument, does not have to be accounted for by appeal to a future that is privileged over all the others that are objectively possible. Instead, we may make sense of their reasonableness by adverting to the fact that it can be useful to commit oneself in such a way that what one says will turn out to be right or wrong depending on how things eventuate.

Why would it be useful to so commit oneself? There are at least two reasons. First of all, in so committing myself, I might enable us to answer a question that is on the table, and on that basis help us make a decision as to what to do. My prediction of tomorrow's rain will, if accepted, help us to decide what to wear out of doors. On questions of less practical significance, I might still wish to commit myself for the purpose of burnishing my reputation in the event that what I say is borne out. Upon vindication I might proclaim, "See, I was right!" With enough such successes I might establish myself as an authority on a subject and reap the privileges that such a status affords. Predicting is in this respect like investing.

5 The Assertion Problem

In 'Indeterminism and the Thin Red Line,' Belnap and I described what we termed the Assertion Problem as an issue that needs to be faced by anyone who theorizes about thought and action directed toward an open future. The problem was as follows. It would seem that one can make assertions about what one knows to be an open future, and in particular about aspects of that open future that are not yet settled by what has transpired thus far. One can assert that the coin will land heads, knowing full well that, as we may now suppose, the coin-tossing process is a fundamentally indeterministic one. But in the absence of a TRL, it is difficult to see how we can provide truth values to such assertoric contents as 'The coin will land heads' when one history branching out of the moment of utterance contains a moment on which the coin lands head, and another history branching out of the moment of utterance contains a moment on which the coin lands tails. Were we to posit a TRL, then we could think of it as privileging one of these histories as the one that will happen, and thereby give us a state of affairs that settles the truth of the future-directed assertion. Barring that, it is not clear how we might characterize the context of utterance, or the circumstances of evaluation in such a way as to tell us whether the assertion is true. The context of utterance might provide values for indexical expressions, but it is less clear how the context of utterance selects one history from among all those that might be how things go.

The intuitive datum seems, then, to be twofold: one can (i) reasonably, and, (ii) felicitously, make an assertion about an aspect of the future that is ontically open and thus not settled by what has gone thus far.

The TRL approach has no difficulty making sense of this twofold datum. How can one do so when one abjures the TRL? In hopes of answering this question, let's proceed more cautiously with our characterization of the data that need to be accounted for:

1. Rational speakers make predictions about the future, and often with an awareness that the future is ontically open.
2. Some of these predictions have the force of assertions, others the force of conjectures, while yet others have the force of other members of the assertive family.
3. Some of the aforementioned acts would seem reasonable, as for instance when one guesses that the coin will come up heads.
4. Some future-directed speech acts end up, in the fullness of time, being vindicated or impugned as the case may be.

In ITRL we considered an explanation of the above datum that supposes that all speech acts need for their evaluation to be relativized to a history. This history will then give a truth value for the content of the assertion, and thereby can help explain why such an assertion can be reasonable.

We also gave an account of future-directed speech acts in terms of liability to credit or blame. According to this view, an assertion that the coin will land heads is vindicated on those histories in which the coin lands heads; impugned on all other histories. This perspective is elegantly explained and motivated in Perloff and Belnap (2012).

Some authors have expressed dissatisfaction with the above "pragmatic" construal of assertion as a response to the assertion problem. Their core intuition seems to be as follows. Since an assertion on Tuesday about a future event on Wednesday can be fully formed—intelligible, felicitous, etc.—then the propositional content of that assertion must be "fully formed" as well, and thus that content must have a truth value at the time at which the assertion is made.

Thus baldly stated, the above reasoning rests on a fallacy of division that is easy to discern. However, while most authors will likely avoid such a fallacy, the conclusion of this reasoning seems to be seductive. For instance, Malpass and Wawer write,

> To us, this move to pragmatics seems to be no help. We are concerned with the way that truth-values are given to predictions of future contingents in Priorian-Ockhamism. The basic problem is that utterances occupy single moments but many histories. Since we have to have both to ascribe a truth-value to a prediction (according to Priorian-Ockhamism), there are many non-trivial ways in which we can evaluate a given prediction. It can be true and false, at the same time, that there will be a sea battle tomorrow. Appealing to pragmatics is just to change the subject, in our opinion. It is as if Belnap et al. would have us consider the pragmatics of assertion involved in "*a*-asserts-'The coin will land heads"' while what we should actually be concerned with is the semantics of "The coin will land heads." (Malpass and Wawer 2012, p.124)

This response presupposes something that we should call into question, namely that it is obligatory to give truth values to the propositional contents of predictions, and in particular truth values to those contents at the time at which the predictions are made. This, I contend, is not a datum forced upon us by any commonsense understanding

of the practice of prediction. Rather, the intuitive datum that theorizing in this area must respect is that many predictions eventually are either borne out or not. This, however, is a datum that the Open Future view can accommodate. What is more, when we advert to the conversational role of predictions, we find that our pragmatic characterization of such acts is all we need. Without a settled truth value, predictions can still be entered into conversational common ground. Once that is done, the contents of such predictions can then be treated as true whether or not they currently have a truth value. For instance, my prediction among my fellow parched hikers that we will find water around the hill to the east, can be accepted as true whether or not it in fact is, and once so accepted we may act as if it is true by marching eastward.

This "pragmatic" solution to the assertion problem is compatible with the ascription of determinate content to future-directed speech acts. 'The coin will land heads,' has a determinate set of truth conditions, and as a result is different from 'x is brindle'. So although, in the face of ontic openness, 'The coin will land heads' and 'x is brindle' are alike in lacking truth value, the former still has truth *conditions* that the latter lacks. This is why 'The coin will land heads' is an appropriate vehicle of assertion while 'x is brindle' is not. The point easily generalizes to other members of the assertive family, any of which can be used to make predictions about an ontically open future.

6 The Modal Realism Objection

Mastop (ms) objects to the Open Future view on different grounds from those having to do with future-directed speech acts. Mastop responds to remarks such as those found in Perloff and Belnap (2012) that the notion of indeterminism that they wish to explore is objective.[3] By this they mean that indeterminism is not a matter of our limited knowledge, or due to someone's perspective on the world. Rather, the notion of indeterminism in question pertains to facts of the matter independent of anyone's state of mind, interests, or point of view. In addition, the Open Future approach suggests that each of the possible futures flowing out of an indeterministic moment is ontologically speaking on a par with all the others: unlike what is the case on the TRL approach, no one history is privileged as against the others in any way.

Mastop seems to take these two doctrines as implying, jointly, that the Open Future view is committed to modal realism *sensu* Lewis (2001). According to such a view, each possible future flowing out of an indeterministic moment is concrete but not spatiotemporally related to any other possible future. Mastop takes this modal realist view to be absurd, and infers that because the Open Future implies it, the Open Future view must be absurd as well. Instead, Mastop urges, we should adopt a modal metaphysics such as articulated by Stalnaker (2003), who sees possible worlds as

[3] "As affirmed in FF [Facing the Future], we require a concept of indeterminism that is local, objective, feature-independent, de re, existential, and hard" (Perloff and Belnap 2012, P. 584).

"ways things might be." This view is compatible with a TRL view, and Mastop take this fact to be evidence in favor of the TRL view.

We may remain neutral here on the question of the coherence of modal realism. What is more important is seeing that the Open Future view does not mandate it. Rather, Open Future is compatible with both modal realism and a "ways things might be" conception of possibilia. To see why, observe that the branches that are typically drawn in a tree diagram representing indeterminism are representations of how history might carry on after an indeterministic point. However, such branches need not be taken as representing states of affairs that are in any sense actual, even relative to themselves. By contrast, possible worlds on the modal realist construal are actual relative to themselves. (This is why it is natural for a modal realist to take 'actually' to be an indexical that refers to the possible world at which it is tokened.) Rather, it is compatible with Open Future to hold that such branches represent, "ways history might go." Standing at an indeterministic point, then, we might say of each of the possible future courses of events, "This is a way that history might go; all we claim now is that none of these is what *will* happen."

We have argued that the Open Future can make sense of the ontic status of possible futures, as well as of our thought and talk of the future even in the face of objective indeterminism. If this argument is sound, it will make clear that even if the TRL is a coherent position, it is unwarranted. It posits more than does Open Future, while providing no return for this higher cost. As a result we have no reason to accept the TRL, and every reason to maintain that, at least if our world is indeterministic, there will be moments in time at which the future is truly open.

References

Belnap, N., and M. Green. 1994. Indeterminism and the thin red line. *Philosophical Perspectives* 8: 365–388.

Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future: Agents and choices in our indeterministic world*. New York: Oxford University Press.

Green, M. 2013a. *Speech acts*. The Stanford Online Encyclopedia of Philosophy, ed. by E. Zalta. http://www.plato.stanford.edu/entries/speech-acts/.

Green, M. 2013b. Assertions. In *Handbook of pragmatics, Vol. II: Pragmatics of speech actions*, eds. M. Sbisà, and K. Turner. Berlin: de Gruyter-Mouton.

Green, M. 2009. Speech acts, the handicap principle, and the expression of psychological states. *Mind & Language* 24(2009): 139–163.

Grice, P. 1989. *Studies in the way of words*. Cambridge: Harvard.

Langton, R. 1993. Speech acts and unspeakable acts. *Philosophy and Public Affairs* 22: 293–330.

Lewis, D. 2001. *On the plurality of worlds*. Oxford: Blackwell.

Malpass, A., and J. Wawer. 2012. A future for the thin red line. *Synthese* 188: 117–42.

Mastop, R. (ms). Truths about the future.

Øhrstrøm, P. 2009. In defence of the thin red line: A case for Ockhamism. *Humana Mente* 8: 17–32.

Perloff, M. and N. Belnap. 2012. Future contingents and the battle tomorrow. *Review of Metaphysics* 64: 581–602.
Stalnaker, R. 2003. *Ways a world might be: Metaphysical and anti-metaphysical essays*. Oxford: Oxford University Press.
Stalnaker, R. 1984. *Inquiry*. Cambridge: MIT.

The Intelligibility Question for Free Will: Agency, Choice and Branching Time

Robert Kane

Abstract In their important work, *Facing the Future* (Oxford 2001), Nuel Belnap and his collaborators, Michael Perloff and Ming Xu, say the following (p. 204): "We agree with Kane (1996) that … the question whether a kind of freedom that requires indeterminism can be made intelligible deserves … our most serious attention, and indeed we intend that this book contribute to what Kane calls 'the intelligibility question."' I believe their book does contribute significantly to what I have called "the Intelligibility Question" for free will (which as I understand it is the question of how one might make intelligible a free will requiring indeterminism without reducing such a free will to either mere chance or to mystery and how one might reconcile such a free will with a modern scientific understanding of the cosmos and human beings). The theory of agency and choice in branching time that Belnap has pioneered and which is developed in detail in *Facing the Future* is just what is needed in my view as a logical foundation for an intelligible account of a free will requiring indeterminism, which is usually called libertarian free will. In the first two sections of this article, I explain why I think this to be the case. But the logical framework which Belnap et al. provide, though it is necessary for an intelligible account of an indeterminist or libertarian free will, is nonetheless not sufficient for such an account. In the remaining sections of the article (3–5), I then discuss what further conditions may be needed to fully address "the Intelligibility Question" for free will and I show how I have attempted to meet these further conditions in my own theory of free will, developed over the past four decades.

R. Kane (✉)
Department of Philosophy, The University of Texas at Austin, 2210 Speedway, Stop C3500, Austin, TX 78712-1737, USA
e-mail: rkane@uts.cc.utexas.edu

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_8,

1 The Intelligibility Question: An Introductory Narrative

In their important work, Facing the Future (hereafter FF), Nuel Belnap and his collaborators, Michael Perloff and Ming Xu, say the following (p. 204): "We agree with Kane (1996) that … the question whether a kind of freedom that requires *indeterminism* can be made intelligible deserves, instead of a superficial negative, our most serious attention, and indeed we intend that this book contribute to what Kane calls 'the intelligibility question.'" I believe their book does contribute significantly to what I have called "the Intelligibility Question" for free will. The theory of agency and choice in branching time that Belnap has pioneered and which is developed in the book in detail is just what is needed in my view as a logical foundation for an intelligible account of a kind of free will that requires indeterminism, which is usually called *libertarian* free will. The logical framework they provide, though necessary for an intelligible account of such an indeterminist or libertarian free will, is however not sufficient for such an account. And I want to discuss in this article what further conditions may be needed to adequately address "the Intelligibility Question."

First, I need to say more about what the Intelligibility Question is. Since ancient times philosophers have doubted that one could make sense of a kind of free will that would require indeterminism. Such a free will, it was commonly argued, must reduce freedom of choice either to mere chance or to mystery. When agents face a free choice we assume that different possible pathways (or histories in the language of FF) are open to them; and which possible pathway or history becomes the actual one will depend in part at least on the agents themselves and how they choose. But if a free choice is undetermined then it would appear that which historical future becomes the actual one would be a matter of chance and so not within the control of the agent. An undetermined event, it is often argued, occurs spontaneously and is not controlled by anything, hence not controlled by the agent. If, for example, a choice occurred by virtue of a quantum jump or other undetermined events in an agent's brain it would seem a fluke or accident rather than a responsible choice. Thus it is often argued that indeterminism would not enhance our freedom, but would rather undermine it. For reasons such as these and many others, thinkers have argued for centuries that undetermined free choices would be "arbitrary," "capricious," "random," "irrational," "uncontrolled," "mere matters of luck or chance," and not really free and responsible choices at all. The Epicurean philosophers of old argued that there would be no room in nature for free will if the atoms did not sometimes "swerve" in undetermined ways. But the many ancient critics of their view, including Stoics and skeptics, scoffed at such an idea, arguing that the mere chance swerve of atoms could not amount to freedom of choice.

Defenders of an indeterminist or libertarian free will have had a poor record through the centuries of answering these familiar charges. Realizing that free will could not merely be indeterminism or chance, they have appealed to various obscure or mysterious forms of agency or causation to make up the difference. Immanuel Kant argued that we cannot explain free will in scientific terms, even though we require it for belief in morality. To make sense of it we have to appeal to the agency of

what he called a "noumenal self" outside space and time that could not be studied in scientific terms. Many other philosophers from Descartes onward have believed that only an appeal to a substance dualism of mind and body could make sense of free will. Science might tell us there was some indeterminacy in nature or a place for causal gaps in the brain, but a nonmaterial self would have to fill those causal gaps in the physical world by intervening in the natural order. Nobel physiologist, John Eccles, in the twentieth century, for example, argued that there might be some place for indeterminism in synaptic transmission of neural impulses in the brain (Eccles 1994). But he went on to argue that if we were to make sense of free choice we would have to appeal in dualist fashion to a "transempirical power center" that would intervene in the brain to fill the causal gaps thus left by the indeterminism. And many other philosophers have referred to yet other libertarian strategems to account for free will, such as uncaused causes, prime movers unmoved and special kinds of agent or immanent causation that cannot be explained in terms of ordinary modes of causation in terms of events familiar to the sciences.

In summary, the charge down through the centuries has been that a free will requiring indeterminism was *unintelligible* or incoherent or impossible. Libertarian views of free will must either reduce free will to mere chance or require some appeal to mysterious forms of agency or causation that had no place in the modern scientific picture of the world. As Nietzsche (2002, Sect. 8) summed up the matter in his inimitable prose, freedom of the will in the "superlative metaphysical sense" (as he put it), which requires that free agent somehow be a *causa sui*, is "the best self-contradiction that has been conceived so far" by the mind of man.

The "Intelligibility Question" as I formulated it was a response to this long history of debate and may be stated in this way: *Can one make sense of, or give an intelligible account of, a free will requiring indeterminism without reducing it to either mere chance, on the one hand, or mystery, on the other?*

To explain how I have attempted to answer this question in my own work, a bit of history will be helpful. When I first began thinking about the free will problem in the 1960s, the landscape of the free will debate was much simpler than today. The unstated assumption was that if you had scientific leanings, you would naturally be a *compatibilist* about free will, believing it to be compatible with determinism, unless you denied it all together as did skeptics and hard determinists. And if on the other hand you were a *libertarian* about free will, believing in a free will that was incompatible with determinism, it was assumed that you must invariably appeal to some kind of obscure forms of agency to make sense of it—to uncaused causes, immaterial minds, noumenal selves, prime movers unmoved, or other examples of what P. F. Strawson called the "panicky metaphysics" of libertarianism in his important 1962 essay, "Freedom and Resentment."

If I may add a personal note here, I was a graduate student at Yale University when Strawson's essay first appeared in 1962 and it was there that I first knew Nuel Belnap. He was one of my logic teachers at the time, along with Alan Anderson and Fred Fitch. (Rich Thomason, an important contributor to the branching time logic presupposed in FF, was a fellow graduate student at Yale at the time.) Belnap was not working on the logic of agency and choice in branching time at that point to my

knowledge. That was to come later. As I recall, Belnap was working with Anderson at the time developing a new theory of "relevance logic," another area in which he has made significant contributions.

My own dissertation director and philosophical mentor at this time at Yale was Wilfred Sellars, who soon after was to move to the University of Pittsburgh, along with Belnap and Anderson. Sellars was a compatibilist about free will, like the vast majority of scientists and philosophers of that era, and he did not believe that a libertarian free will requiring indeterminism could be accounted for without appealing to obscure forms of agency of the kinds Strawson had called "panicky metaphysics." Appealing to an influential distinction that Sellars had himself introduced into contemporary philosophical discourse, he granted that free will in some sense was an integral part of what he called *the manifest image* of humans and their world. But he did not believe that a traditional indeterminist or libertarian free will could be reconciled with what he called the *scientific image* of the world; and he challenged me to show otherwise. With the naïveté characteristic of a young graduate student, I suggested that I would return in a few weeks with an answer to this challenge. It has turned out to be a project of somewhat longer duration, still ongoing.

It was a surprise therefore some 40 years later when I received in the mail a complementary copy of Facing the Future, sent to me by Nuel Belnap. It was not sent to me as a former student, but rather as someone who had in the intervening years written extensively on the free will problem, attempting to make sense of the libertarian free will, who might find the book congenial and a significant contribution to that project. (He had in fact forgotten I had ever been a student of his so many years ago and I had to remind him of the fact.) That our intellectual paths should cross this way after so many years was indeed fortuitous. For, as noted above, I do believe that FF provides a logical framework that is congenial to the project of making sense of a free will requiring indeterminism and hence to addressing the Intelligibility Question.

2 Action, Indeterminism, and Facing the Future

I will first give some reasons for thinking this is the case regarding the logical framework of FF before turning to further issues that have to be addressed in order to fully answer the Intelligibility Question. First, there are a number of issues and topics in the philosophy of action related to free will that are made more precise by the stit logic developed in FF, which philosophers who deal with action theory (usually only in informal ways) would do well to take note of. The distinction between the achievement stit and the deliberative stit (pp. 32–40) is particularly important in my view for discussing issues about free will. The achievement stit involves an earlier moment of choice or action that guarantees the later outcome A of an action. The deliberative stit, by contrast, is evaluated at the moment of choice itself, the very moment at which the agent sees to it that the outcome A will occur. The outcome A is guaranteed by the present choice at the moment of choice itself. Both achievement

stits and deliberative stits would play a role I believe in an adequate account of free will. But the idea behind the achievement stit must also be expanded in a certain way to account for free will as I understand it. As I will argue, acts done "of one's own free will," it must be allowed, can also be achievements of multiple choices and actions performed at earlier times which causally influence, even if they do not always guarantee, later choices or actions.

Second, the notion of "settled" truth (pp. 29–32) which is basic to the framework of FF is fundamental to making sense of libertarian free will and indeed to understanding the traditional problem of free will itself. The operative intuition is that when an agent faces a free choice (in particular, a deliberative stit), which choice will be made is not settled true at any time before the choice itself is made. Doctrines of determinism have been thought to be a threat to free will to the extent that they imply that for every choice or action, whether or not it will occur is settled true at some time before it does occur or not. Determinism can be and has been defined in many different ways. But it is this implication of doctrines of determinism in terms of settled truth that has historically been thought to be a threat to free will. The logical framework of FF allows one to express this threat in a clear way.

Third, the framework of FF also helps to resolve a host of controversial issues that have long been discussed in the literature of free will regarding the truth value of future tensed sentences concerning human choices and actions. Since Aristotle, a common assumption has been that if free choices and actions are neither fated nor determined, then future tensed sentences concerning them must be neither true nor false. But this assumption has led to numerous puzzles that are perceptively described and many of which in my view are helpfully resolved in FF (pp. 144–176). To treat future tensed sentences of these kinds as open sentences lacking the assignment of a history parameter seems to me the right way to go to resolve these puzzles. To say that a future tensed sentence concerning a free choice is neither true nor false is not to say that it has some third truth value or a third special status. Given a model and a context, an open sentence about an indeterminate future of this kind will have a truth value, once a suitable value is applied for each of the parameters, including the history parameter. This solution to the assertion problem for such future tensed propositions seems to me quite congenial to libertarian accounts for free will, as is the related solution in FF to the problem of "the thin red line" (pp. 160–174). The solutions of the book to these problems can of course be questioned and its solution to the problem of the thin red line is questioned by other contributors to this volume. I am inclined to agree with its solution to the problem of the thin red line, but will not argue the matter here. I will merely register the general conviction that something like the solutions to these problems about future tensed propositions proposed in FF is what is needed for a coherent conception of free will that requires indeterminism.

Fourth, the logical framework of FF helps to clarify a number of other issues in the philosophy of action and in debates about free will and responsibility. These include its perceptive account of the distinction between "refraining" from an action and simply not performing the action, a distinction which philosophers of mind and action have often puzzled over (pp. 40–45). The interpretation of the distinction in terms of the logic of stit helps one to clearly see how refraining from an action can

be a kind of action even though it also involves not performing an action. Another area where the framework of FF is helpful is in spelling out the different possible meanings of the much discussed expression "could have done otherwise" in the free will literature (pp. 255–270). Belnap at al. show how certain puzzles in the literature concerning the relation of moral responsibility to the ability to do otherwise can be illuminated by distinguishing these different meanings of the ability to do otherwise. Their framework also helps to clarify and formalize the important distinction between so-called "soft facts" and "hard facts" about the past, a distinction that plays a role in many debates about free will and determinism, but is not always carefully defined (pp. 145–174). In these ways and in others, philosophers who deal with the theory of action and free will in more informal ways have much to learn from the formal framework developed by Belnap at al. in this book.

3 From Action to Free Will

While the framework of FF makes a significant contribution to debates about free will in these and other ways, there is at least one point on which I would depart from it—or perhaps better, qualify it to some degree—in giving an account of free will. FF assumes that indeterminism and the logic of branching time presupposed by it are required to account for *action* in general of any kinds, whereas on my view, while indeterminism and branching time are required to explain free will (or more precisely, actions done "of one's own free will"), they are not required to account for action in general. I would find it congenial, to be sure, if it could be shown that all action and agency did require indeterminism, for then, a fortiori, acts of free will would as well. But I am not convinced of this stronger claim and would need to be shown otherwise, for the following reasons.

There seems to be a primordial sense of action and agency that is admittedly presupposed by free will, but leaves open the question of whether determinism or indeterminism is true. According to this primordial sense, *to act is to guide behavior toward a goal or purpose in accordance with a plan* and it involves the capacity to readjust both goal and plan (*ends* and *means,* one might also say) in the light of feedback from the environment. Action in this primordial sense involves a certain kind of *control* of an agent over behavior that we might refer to as *teleological guidance control,* given that the behavior in question is goal-directed and involves guidance. Action in this sense of goal-directed, guided behavior is something other living things are capable of, not merely human beings, though humans have further and more sophisticated higher-order capacities to evaluate and re-evaluate both ends and means. I believe action in this primordial sense can exist in principle in determined worlds. One reason for believing so is that the ability to guide behavior toward a goal does not of itself imply that the agent also has the ability to do otherwise, i. e., to guide behavior to a different goal. Though, importantly, action in this primordial sense is also compatible with some measure of indeterminism. So acknowledging it

as a significant form of action does not settle issues about determinism and indeterminism.

It is when we ask further questions about this primordial conception of action, in my view, that we raise distinctive issues about the freedom of the will. A central question for example is this: W*hence comes the purposes and plans themselves that guide behavior, rendering it action in this primordial sense?* Do the purposes and plans (ends and means) that guide behavior have their *sources* or *originate* in the agents themselves who act, or do these purposes and plans ultimately come entirely from sufficient causes outside the agent and over which the agent does not have control? This is a variant of the free will question; and one can see from it why determinism has been thought by many historically to be a threat to free will. If determinism were true there would be sufficient causes outside the agents and over which the agents did not have control for whatever purposes and plans, ends and means, agents might pursue—sufficient causes going back into the remote past for why they had the purposes and plans they did have rather than some others. Agents might still have the power to control behavior *in accordance with* their purposes and plans (i.e. to act in the primordial sense), but they would not be the ultimate sources of the purposes and plans that guide their behavior. That is, they might be able to do *what* they willed, but they would not be the ultimate creators of what it is that they willed, and in that sense would not be acting "of their own free will" in the sense of "a will of their own free-making."

Yet this notion of freedom of the will as ultimate creation of purposes ("a will of one's own free-making") is itself highly problematic. It immediately conjures up Nietzsche's image, mentioned earlier, of an agent who exercises free will as some kind of ultimate cause of itself, a *causa sui*, the "best self-contradiction conceived so far by the mind of man." The idea of a will of one's own free-making suggests a troubling backtracking regress, since to be the ultimate creator of one's own present will and purposes, one would have to be so by virtue of prior choices and actions which would be motivated by still earlier purposes and plans, which earlier purposes and plans in turn could not have sufficient causes outside the agent and over which the agent did not have control, and so must be created by still earlier choices or actions of the agent, and so on indefinitely.

This regress could be stopped, to be sure, if some choices or actions in the agent's life history did not have sufficient causes at all and so were undetermined. But, while this solution points in the right direction (showing why indeterminism is thought to be important for freedom of will), it brings us back to the dilemma that has historically given rise to the Intelligibility Question: If choices by which we (ultimately) create our purposes and plans were undetermined, it seems that they would not be in our control, since undetermined events occur by chance and are not controlled by anything, hence not by agents. The alternative, as noted, would be to appeal to mysterious forms of agency, to uncaused causes, prime movers, and the like; and in such manner the appeal to ultimate creation of purposes leads us back to the dilemma of chance or mystery once again.

To complicate matters, there is a further problem about indeterminism with regard to free will that is also important for dealing with the Intelligibility Question. Unlike

the previous dilemma, it is a problem that often gets overlooked in historical and contemporary discussions about free will, though as I have argued for several decades, it is crucial for understanding the very notion of the freedom of the will (see, e.g., Kane 1985, 1996, 2002b, 2007).

This problem is that even if one grants that indeterminism is a necessary condition for genuinely free choices and actions, it turns out that it is not a sufficient condition for freedom of will. The reason is that when we wonder about whether the *wills* of agents are free, it is not merely whether they could have done otherwise that concerns us, even if the doing otherwise is undetermined. What interests us is whether they could have done otherwise *voluntarily, intentionally,* and *rationally,* rather than merely by accident or mistake, unintentionally, inadvertently, or irrationally. Or, putting it more generally, we are interested in whether agents could have acted voluntarily (in accordance with their wills), intentionally (on purpose rather than accidentally or inadvertently), and rationally (with good reasons) *in more than one way* rather than in only one way, and in other ways merely by accident or mistake, unintentionally, inadvertently, or irrationally.

I call such conditions–of more-than-one-way voluntariness, intentionality and rationality–"plurality conditions" for free will (Kane 1996, 107–111). And I call the *ability* to choose or act in more than one way voluntarily, intentionally and rationally, i.e. in accordance with these conditions, *plural voluntary control* (PVC). These plurality conditions seem to be deeply embedded in our intuitions about free choice and action. We naturally assume, for example, that freedom and responsibility would be deficient if it were always the case that we could only do otherwise by accident or mistake, unintentionally, involuntarily, or irrationally. It is true that libertarian free will requires that more than one branching pathway (history) into the future be "open" to agents in the manner described in FF (p. 136). But it also requires something about the *way* that agents *select* from among these open pathways: Whichever ones they select, if they are to do so "of their own free will," they must do so voluntarily, intentionally and rationally (*at will,* as we say), rather than merely accidentally, unintentionally or irrationally.

4 Self-forming Actions (SFA's)

We are now in a position to consider what further steps may be necessary to fully address the Intelligibility Question.

The first important step is to note that, as the preceding discussion suggests, indeterminism need not be involved in all acts done "of our own free wills." Often we act from a will (character, motives and purposes) already formed. But it is "our own free will" by virtue of the fact that we formed it to some degree by other choices or actions in the past for which we could have done otherwise and which were undetermined. If this were not so *there is nothing we could have ever done differently in our entire lifetimes to make ourselves and our wills different than they are*—a consequence that I believe is incompatible with our being at least to some degree *ultimately*

responsible for being the way we are, and for the wills we do have, and hence ultimately responsible for the actions that flow from our wills. Compare Aristotle's claim that if a man is responsible for wicked acts that flow from his character and purposes (his will) he must at some time in the past have been responsible for forming the wicked character and purposes from which these acts flow.

I call those choices or actions in agents' life histories by which they formed their present wills and for which they could have done otherwise in a manner that was undetermined, "self-forming actions" or SFAs. (They would be "deliberative stits" in the language of FF.) I believe such self-forming actions occur at those difficult times in life when we are torn between competing visions of what we should do or become; and they are more frequent in everyday life than we may think. We might be torn between doing the moral thing or acting from ambition, or between powerful present desires and long term goals, or faced with difficult tasks for which we have aversions, etc. The uncertainty and inner tension we feel at such soul-searching moments of self-formation, I suggest, would be reflected in some indeterminacy in our neural processes themselves (perhaps chaotically amplified background neural noise) "stirred up," one might say, by the conflicts in our wills. What is experienced personally as uncertainty at such moments would thus correspond physically to the opening of a window of opportunity that temporarily screens off complete determination by influences of the past. (By contrast, when we act from predominant motives and a "settled" will without such inner conflict, the indeterminacy is muted or damped and plays a less significant role.)

In such cases of self-formation, we are faced with competing motivations and whichever choice is made will require an effort of will to overcome the temptation to make the other choice. I thus postulate that, in such cases, multiple goal-directed cognitive processes would be involved in the brain, corresponding to competing efforts, each with a different goal, corresponding to the competing choices that might be made. In short, one might appeal to a form of parallel processing in the free decision-making brain. One of these neural processes has as its goal, the making of one of the competing choices (say, a moral choice), realized by reaching a certain activation threshold, while the other has as its goal the making of the other choice (e.g., a self-interested choice). Likewise, the competing processes have different inputs, moral motives (beliefs, desires, etc.), on the one hand, self-interested motives, on the other. And each of the processes is the realizer of the agent's *effort* or *endeavoring* to bring about *that* particular choice (e.g. the moral choice) *for* those motives (e.g. moral motives), thus taking the input into the corresponding output; and the processes are so connected that if one should succeed, the other will shut down.

Because of the indeterminacy in each of these neural processes stirred up by the conflict in the will, however, for each, it is not certain that it will succeed in reaching its goal, i.e., an activation threshold that amounts to choice. Yet (and here is a further crucial step) if *either* process *does* succeed in reaching its goal (the choice aimed at), despite the indeterminacy involved, one can say that that choice was brought about *by the agent's effort or endeavoring* to bring about *that* choice for *those* motives, because the process itself was the neural realizer of this effort and it succeeded in reaching its goal, despite the indeterminism involved.

Note that, in these circumstances, the choices either way would not be "inadvertent," "accidental," "capricious," or "merely random," because whichever choice is made will be brought about *by* the agent's effort to make that particular choice *for* the reasons motivating that choice, reasons the agent will then and there endorse by making the choice itself. Indeed, the agents will have *plural voluntary control* (PVC) over the choices made, as defined earlier, since whichever choice is made will be made voluntarily (i.e. in accordance with the agent's will, because the prior will is divided and the agent may consequently choose either way at will), intentionally (i.e. on purpose rather than accidentally or inadvertently, since the choice will result from the goal-directed effort to make that choice) and rationally (i.e. because the choice will be made for reasons motivating that choice which are reasons the agent has, and decides to act on then and there).

The idea in sum is to think of the indeterminism involved in free choice, not as a cause *acting on its own,* but as an *ingredient* in larger *goal-directed* or *teleological* activities of the agent, in which the indeterminism functions as a *hindrance* or *interfering* element in the attainment of the goal. The choices that result are then *achievements* brought about by the goal-directed activity (the effort) of the agent, which might have failed since they were undetermined, but one of which succeeds. Moreover, if there are multiple such processes aiming at different goals (as in the conflicted circumstances of an SFA), *whichever choice may be made,* will have been brought about by the agent's effort to bring about *that* particular choice rather than some other, despite the possibility of failure due to the indeterminism.

In such circumstances, as a consequence, the indeterminism, though causally relevant to the choice, would not be the *cause* of the choice because it would have been an interfering element lowering the probability that that choice would be made from what it would have been if there was no interference. The causes of the choice, by contrast, would be those relevant factors that significantly raised the probability that this choice would be made rather than some other, such as the agent's motives for making this choice rather than the other and the agent's deliberative efforts to overcome the temptations to make the contrary choice. Were these factors not present there would be no chance this choice would be made because *there would be no cognitive process of the agent aiming at it*. Moreover, if the choice *was* caused by a deliberative cognitive process of the agent aiming at it, it would also be true to say that the *agent* caused the choice.

A further point is that when indeterminism thus functions as an obstacle to the success of a goal-directed activity of an agent, which succeeds in attaining its goal nonetheless, the indeterminism does not preclude *responsibility*. There are many examples demonstrating this fact (some first suggested by J. L. Austin and Elizabeth Anscombe). Here is one I have previously used. A husband, while arguing with his wife, in anger swings his arm down on her favorite glass-table top in an effort to break it. Imagine that there is some indeterminism in the nerves of his arm making the momentum of his swing indeterminate so that it is literally undetermined whether the table will break right up to the moment when it is struck. Whether the husband breaks the table or not is undetermined; and yet he is clearly responsible if he does break it, because the breaking was caused by his effort to break it by swinging his

arm down forcefully on it. That is why it would be a poor excuse for him to say to his wife "Chance did it (broke the table), not me." Even though chance was causally relevant, because there was chance he would fail, chance didn't do it, *he* did.

But isn't it the case, one might ask, that whether one of these neural processes succeeds (say, in choosing A) rather than the competing process (in choosing B) (i) depends on whether certain neurons involved in the processing fire or do not fire (perhaps within a certain time frame); and isn't it the case that (ii) whether or not these neurons fire is undetermined and hence a matter of chance and hence that (iii) the agent does not have control over whether or not they fire? But if these claims are true, it seems to follow that the choice merely "happened" as a result of these chance firings and so (iv) the agent did not *make* the choice of A rather than B and (v) hence was not *responsible* for making it. As a consequence, it looks like the outcome *must* be merely a matter of chance or luck and not a responsible choice after all.

But those who reason this way do so too hastily. For the surprising thing is that, even if (i)–(iii) are true, (iv) and (v) do not follow *when* the following conditions also hold: (a) the choosing of A rather than B (or B rather than A, whichever occurs) was something the agent was endeavoring or trying to bring about, (b) the indeterminism in the neuron firings was a hindrance or obstacle to the achievement of that goal and (c) the agent nonetheless succeeded in achieving the goal despite the hindering effects of the indeterminism.

For, consider the husband swinging his arm down on the table. It is *also* true in his case that (i) whether or not his endeavoring or trying to break the table succeeds "depends" on whether certain neurons in his arm fire or do not fire; and it is *also* true in his case that (ii) whether these neurons fire or not is undetermined and hence a matter of chance and hence (iii) their firing or not, is not under his control. Yet, even though we *can* say all this, it does not follow that (iv) the husband did not break the table and that (v) he is not responsible for breaking the table, *if* his endeavoring or trying to do so succeeds. Surprising indeed! But this is the kind of significant result one gets when indeterminism or chance plays an interfering or hindering role in larger goal-directed activities of agents that may succeed or fail.

It is well to reflect on this: We tend to reason that if an action (whether an overt action of breaking a table *or* a mental action of making a choice) depends on whether certain neurons fire or not (in the arm *or* in the brain), then the agent must be able to *make* those neurons fire or not, if the agent is to be responsible for the action. In other words, we think we have to crawl down to the place where the indeterminism originates (in the individual neurons) and *make* them go one way or the other. We think we have to become originators at the micro-level and "tip the balance" that chance leaves untipped, if we (and not chance) are to be responsible for the outcome. And we realize, of course, that we can't do that. But we don't have to. It is the wrong place to look. We don't have to micro-manage our individual neurons to perform purposive actions and we do not have such micro-control over our neurons *even when we perform ordinary actions* such as swinging an arm down on a table.

What we need when we perform purposive activities, mental or physical, is macro-control of processes involving many neurons—processes that may succeed in

achieving their goals despite indeterminacies that may be involved in "the naturally noisy processes of sensory transduction." We do not micro-manage our actions by controlling each individual neuron or muscle that might be involved. But that does not prevent us from macro-managing our purposive activities (whether they be mental activities such as practical reasoning, or physical activities, such as arm-swingings) and being responsible when those purposive activities attain their goals, despite the indeterminacies involved. And this would be true in self-forming choices or SFAs, as conceived above, *whichever* of the competing purposive activities succeeds.

5 Further Issues: Efforts, Introspection, Agency, Control, Rationality

Needless to say, there are many further potential objections to the preceding view that need to be addressed, as with any view, and which I have tried to address in many of my writings. In this concluding section I can only briefly respond to a few of these additional objections and refer readers to other writings for discussion of others.[1]

A commonly-made further objection is that it is irrational to make efforts to do incompatible things. I concede that in most ordinary situations it is. But I contend that there are special circumstances in which it is not irrational to make competing efforts: These include circumstances in which (i) we are deliberating between competing options; (ii) we intend to choose one or the other, but cannot choose both; (iii) we have powerful motives for wanting to choose each of the options for different and competing reasons; (iv) there is a consequent resistance in our will to either choice, so that (v) if either choice is to have a chance of being made, effort will have to be made to overcome the temptation to make the other choice; and most importantly, (vi) we want to give each choice a fighting chance of being made because the motives for each choice are important to us. The motives for each choice define in part what sort of person we are; and we would taking them lightly if we did not make an effort in their behalf. And, as it turns out, these are precisely the conditions of "self-forming" actions or SFAs (see e.g., Kane 1996, 128–143, 2002b, 417–124).

It is important to note in this connection that our normal intuitions about efforts are formed in everyday situations in which our will is already "*settled"* on doing something, where obstacles and resistance have to be overcome if we are to succeed in doing it. We want to open a drawer, which is jammed, so we have to make an effort to pull it open. In such everyday situations, it *would* be irrational to make incompatible efforts because our wills are already settled on doing what we are trying or endeavoring to do. But situations of the above kinds involving SFAs are what I call *will-setting* rather than *will-settled.* They are situations in which one's will is not *yet* set on doing either of the things one is trying to do, but where one has strong reasons for doing each (e.g., deciding to A and deciding to B), and neither set

[1] Kane (1985, 1989, 1996, 1999a, b, 2000, 2002a, 2005, 2008, 2009, 2011), Kane (2007).

of reasons is as yet *decisive*. Because most efforts in everyday life are made in will-*settled* situations, we tend to assimilate all effort-making to such situations, thereby failing to consider the uniqueness of will-*setting*, which is of a piece, in my view, with the uniqueness of *free will*.

Another commonly-made objection is that we are not introspectively or consciously aware of making dual efforts and performing multiple cognitive tasks in self-forming choice situations. But I am not claiming that agents are introspectively aware of making dual efforts. What persons are introspectively aware of in SFA situations is that they are trying to decide about which of two (or more) options to choose and that either choice is a difficult one because there are resistant motives pulling them in different directions that will have to be overcome, whichever choice is made. In such introspective conditions, I am theorizing that what is going on underneath is a kind of distributed processing in the brain that involves separate attempts or endeavorings to resolve competing cognitive tasks.

There is a larger point here that I have often emphasized: *Introspective evidence cannot give us the whole story about free will.* Stay on the introspective surface and libertarian free will is likely to appear obscure or mysterious, as it so often has in history. What is needed is a *theory* about what might be going on underneath when we exercise such a free will, not merely a description of what we immediately experience. In this regard, it is my view that new scientific ideas can be a help rather than a hindrance to making sense of free will.

It is now widely believed, for example, that parallel processing takes place in the brain in such cognitive phenomena as visual perception. The theory is that the brain separately processes different features of the visual scene, such as object and background, through distributed and parallel, though interacting, neural pathways or streams.[2] Suppose someone objected that we are not introspectively aware of such distributed processing in ordinary cases of perception. That would hardly be a decisive objection against this new theory of vision. For the claim is that this is what we are doing in visual perception, not necessarily that we are introspectively aware of doing it. And I am making a similar claim about free will. *If parallel distributed processing takes place on the input side of the cognitive ledger (in perception), then why not consider that it also takes place on the output side (in practical reasoning, choice and action)*? That is what I am suggesting we should suppose if we are to make sense of libertarian free will.

Another set of objections involves issues about *control*. Doesn't indeterminism at least *diminish* the control agents exercise over their self-forming choices or SFA's? Indeterminism *does* diminish a certain kind of control that agents may exercise over their self-forming choices, which I have called *antecedent determining control*, the power to guarantee or determine *in advance* that some event will occur. Clearly agents cannot have such control over SFAs (which are deliberative stits) and which must be undetermined at all times before they occur. But from the fact that one does not control which of a set of outcomes is going to occur *before* it occurs, it does not

[2] For an overview of research supporting such views about parallel distributed processing in vision see Bechtel (2001).

follow that one does not control which of them occurs *when* it occurs (Kane 1996, 133–148, 1999a). When the conditions for SFAs are satisfied, agents exercise control over their future lives *then and there* by deciding. Indeed, as argued earlier, they have what I have called "plural voluntary control" over their options in the sense that they are able at the moment of choice to bring about whichever of the options they will, when they will to do so, for the reasons they will to do so, and on purpose rather than by mistake or accident.

And note that it is the diminishment of antecedent determining control over any *one* of the options that makes possible such *plural* voluntary control over each of them. Indeterminism, by being a hindrance to the realization of some of the agent's purposes, opens up the possibility of pursuing other purposes, of doing otherwise, voluntarily and rationally. To be genuinely self-forming agents (creators of ourselves), to have a free will, there must at times in life be such obstacles and hindrances in our wills that must be overcome. Self-formation, as I like to say, is not a gift, but a struggle.

One further remark about control: For an agent to have control generally at a time t over the being or not being (existence or non-existence) of some event (e.g. a choice) is for the agent to have the *ability* or *power* at the time t to *make* that event *be* at t and the ability or power to make it *not be* at t. And in an SFA, one exercises just such control over the choice one makes (e.g. the choice of A rather than B) at the time one makes it. For, one not only has the ability or power at that time to make that choice be, one also has the ability or power at that time to make it not be, *by making the competing choice* (of B rather than A) *be.* One has both these powers because either of the efforts or endeavorings in which one is engaged might succeed in attaining its goal (choosing A or choosing B) at the time. And if either effort does succeed in attaining its goal, the agent can be said to have *brought about* the choice thereby made *by* making that effort to bring it about.

A final objection I will consider here is this: Is there not some truth to the oft-repeated charge that undetermined choices of the kinds required by libertarian free will must be *arbitrary* in a certain sense? A residual arbitrariness seems to remain in all self-forming choices or SFAs since the agents cannot in principle have sufficient or overriding ("conclusive" or "decisive") *prior* reasons for making one option and one set of reasons prevail over the other.

I think there is some truth to this charge, but it is a truth that reveals something important about free will. I have argued elsewhere (Kane 1996, 145–146) that such arbitariness relative to prior reasons tells us that every undetermined self-forming choice or SFA is the creation of novel constraints upon an agent's pathway into the future, constraints that are not fully explained or determined by the agent's past, but are consistent with that past. In making such a choice we say, in effect, "I am opting that these purposes and plans (rather than some others) will be a part of my pathway into the future, my future life. Doing so is not *required* by my past reasons, but is consistent with my past and represents one branching pathway my life can now *meaningfully* take. Whether it is the right choice, only time will tell. Meanwhile, I am willing to take responsibility for it one way or the other."

Of special interest here, as I have often noted, is that the term "arbitrary" comes from the Latin *arbitrium*, which means "judgment"—as in *liberum arbitrium voluntatis*, "free judgment of the will," which is the medieval designation for free will. Imagine a writer in the middle of a novel. The novel's heroine faces a crisis and the writer has not yet developed her character in sufficient detail to say exactly how she will act. The author makes a "judgment" about this that is not determined by the heroine's already formed past which does not give unique direction. In this sense, the judgment (*arbitrium*) of how she will react is "arbitrary," but not entirely so. It had input from the heroine's fictional past and in turn gave input to her projected future.

In a similar way, agents who exercise free will are both authors of and characters in their own stories at once. By virtue of "self-forming" judgments of the will (*arbitria voluntatis*) (SFAs), they are "arbiters" of their own lives, "making themselves" out of past that, if they are truly free, does not limit their future pathways to one. If we should charge them with not having sufficient or *conclusive* prior reasons for choosing as they did, they might reply: "True enough. But I did have *good* reasons for choosing as I did, which I'm willing to endorse and take responsibility for. If they were not sufficient or conclusive reasons, that's because, like the heroine of the novel, I was not a fully formed person before I chose (and still am not, for that matter). Like the author of the novel, I am in the process of writing an unfinished story and forming an unfinished character who, in my case, is myself."

In the logical framework of Belnap et al. Facing the Future, these *libera arbitria voluntatis* or self-forming choices (SFAs) would be *deliberative stits*, or deliberative seeings to it that, of agents. They are represented at moments in the logic of branching time at which there are multiple possible branching future histories; and they determine a particular class of possible future histories within which the future life of the agent must lie. Such self-forming actions or SFA's are not the only kinds of actions that agents can perform "*of their own free wills,"* however, on the above account. As noted earlier, often we act from a will already formed, but it is "our own free will" (a will "of our own free making") to the degree that we formed it by earlier SFAs that were undetermined, and for which we could have done otherwise voluntarily, intentionally and rationally.

Those acts that flow determinately from a will already formed in this manner could be counted as *achievement stits* in the framework of Facing the Future. And they too could be acts done "of our own free wills" to the degree that the wills from which they determinately flow were formed by earlier SFAs. For example, on my way to a class this afternoon on campus, I look up at the clock on the University tower and notice that it is five minutes before the start of the class. Without deliberating about it, I immediately hasten my pace in order to make the class on time. I did not make an explicit choice or decision to hasten my pace at that moment. My doing so was rather guaranteed once I noticed the time (in the manner of an achievement stit) by a prior choice (an SFA) made the day before, when I resolved not to be late for any more classes this semester. I thus hastened my pace "of my own free will" in the sense of a will freely formed in part by a prior self-forming choice

(a deliberative stit) that was undetermined and such that I could have done otherwise when I made it.

In such ways, and in many others, the logical framework pioneered by Nuel Belnap and spelled out by him and his co-authors in Facing the Future provides, in my view, just the right kind of logical framework required to give an account of a traditional (libertarian) free will requiring indeterminism and thereby to answer what I have called the Intelligibility Question.

References

Bechtel, W. (ed.). 2001. *Philosophy and the neurosciences: A reader*. Malden, MA: Blackwell.
Belnap, N., M. Perloff., and M. Xu. 2001. *Facing the future*. Oxford: Oxford University Press.
Eccles, J. 1994. *How the self controls the brain*. Berlin: Springer.
Kane, Robert. 1985. *Free will and values*. Albany, NY: State University of New York Press.
Kane, Robert. 1989. Two kinds of incompatibilism. *Philosophy and Phenomenological Research* 31: 219–254.
Kane, Robert. 1996. *The significance of free will*. New York: Oxford University Press.
Kane, Robert. 1999a. Responsibility, luck, and chance: Reflections on free will and indeterminism. *Journal of Philosophy* 96: 217–240.
Kane, Robert. 1999b. On free will, responsibility and indeterminism: Responses to Clarke, Haji and Mele. *Philosophical Explorations* 2: 105–121.
Kane, Robert. 2000. Précis of *The significance of free will* and Responses to Bernard Berofsky, John Martin Fischer, and Galen Strawson. *Philosophy and Phenomenological Research* 60(129–34): 157–167.
Kane, Robert (ed.). 2002a. *The Oxford handbook of free will*, 1st ed. Oxford: Oxford University Press.
Kane, Robert. 2002b. Free will: New directions for an ancient problem. In R. Kane, ed. *Free Will*, 222–48. Oxford: Blackwell.
Kane, Robert. 2005. *A contemporary introduction to free will*. Oxford: Oxford University Press.
Kane, Robert. 2007. Libertarianism and Responses to Fischer, Pereboom and Vargas. In *Four views of free will*. Fischer, Kane, Pereboom, and Vargas, 5–43 and 166–83. Oxford: Wiley.
Kane, Robert. 2008. Three freedoms, free will, and self-formation: A reply to Levy and other critics. In *Essays on free will and moral responsibility*, ed. N. Trakakis, and D. Cohen, 142–161. Newcastle on Tyne, UK: Cambridge Scholars Press.
Kane, Robert. 2009. Free will and the dialectic of selfhood. *Ideas y Valories* 58: 25–44.
Kane, Robert. 2011. Rethinking free will: New directions for an ancient problem. In *The Oxford handbook of free will*, 2nd ed, ed. R. Kane. Oxford: Oxford University Press.
Nietzsche, Friedrich. 2002. *Beyond good and evil*. Cambridge, UK: Cambridge University Press.
Strawson, P.F. 1962. Freedom and resentment. *Proceedings of the British Academy* 48: 1–25.

What William of Ockham and Luis de Molina Would have said to Nuel Belnap: A Discussion of Some Arguments Against "The Thin Red Line"

Peter Øhrstrøm

Abstract According to A. N. Prior the use of temporal logic makes it possible to obtain a clear understanding of the consequences of accepting the doctrines of indeterminism and free choice. Nuel Belnap is one of the most important writers who have contributed to the further exploration of the tense-logical systems as seen in the tradition after Prior. In some of his early papers Prior suggested the idea of the true future. Obviously, this idea corresponds to an important notion defended by classical writers such as William of Ockham and Luis de Molina. Belnap and others have considered this traditional idea introducing the term, "the thin red line" (TRL), arguing that this idea is rather problematic. In this paper I argue that it is possible to respond to the challenges from Belnap and others in a reasonable manner. It is demonstrated that it is in fact possible to establish a consistent TRL theory. In fact, it turns out that there several such theories which may all be said to support the classical idea of a true future defended by Ockham and Molina.

The Prior Collection at Bodleian Library in Oxford contains a few letters from Nuel Belnap to A. N. Prior and a few letters from Prior in reply—all from the period from 1960 to 1962. From the content of these letters it is evident that the two scholars shared a deep interest in philosophical logic. They both greatly appreciated the beauty of logical structures; in particular, they were interested in modal logic. For a new edition of his *Formal logic* Prior wanted to include some biographical data of some of the logicians he quoted in the book, and in a letter he asked Belnap to help him providing some data for that purpose. Prior received the data from Belnap, and in reply he wrote dated 28 March, 1960, he stated: "1930 seems to have been a good year for modal logic—you, Smiley, Lemmon, Jonathan Bennett …".

Clearly, modal logic attracted several brilliant young logicians during the 1950s and the 1960s. Prior, himself, had worked a lot with modal logic during the 1950s. More and more, these activities came to be combined with his interest in temporal logic

P. Øhrstrøm (✉)
Department of Communication and Psychology, Aalborg University,
Nyhavnsgade 14, 9000 Aalborg, Denmark
e-mail: poe@hum.aau.dk

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_9,

and in the discussions regarding determinism and indeterminism. One of his main interests had to do with the Master Argument of Diodorus Cronus and the search for the so-called Diodorean modality (Prior 1955). It was well-known that Diodorus had formulated his argument about 300 BC in order to demonstrate that the world is deterministic, and to argue for a reductive account of modal notions to temporal notions; specifically that possibility should be conceived as "what is or what is going to be" (Øhrstrøm and Hasle 1995, p. 15 ff; Øhrstrøm and Hasle 2006). To Prior this gave rise to three interesting questions:

1. What is the formal structure of the modal logic in which possibility is defined in the Diodorean way, on the assumption that time is a linear and discrete sequence of instants?
2. How can a formal and valid version the Master Argument of Diodorus be formulated?
3. How can indeterminism be defended (in terms of tense-logical systems consistent with the assumption of free choice) against the valid versions of the Diodorean argument and similar arguments?

Prior worked intensively with these and similar questions from 1953 to his death in 1969. In doing so he found it most useful to study the theories of temporal logic. According to Prior the use of temporal logic would make it possible to obtain a better understanding of the consequences of accepting the idea of free choice. In particular, he also realized that the notion of branching time could be most helpful in this respect.

Question 1 above was fully answered during Prior's lifetime. In fact, Prior dedicated a complete chapter of his *Past, Present and Future* to this problem and its solution (see Prior 1967, p. 20 ff.). As we shall see, the study of this question actually led to the construction of the first branching time models. Prior's work with question 2 led him to the formulation of a reconstruction of the Master Argument (see Prior 1967, p. 32 ff.). Working with question 3, Prior developed some very important systems of temporal logic consistent with the assumption of free choice. In this chapter we shall mainly comment on his Ockhamistic system.

When Prior died in 1969 many additional problems regarding temporal logic and indeterminism had been discovered. Since then several logicians and philosophers have continued Prior's line of thinking. Clearly, Nuel Belnap is one of the most important writers who have contributed to the further exploration of tense-logical approach to the study of indeterminism and free action.

Much of Nuel Belnap's work has been carried out within a Priorean tradition. As we shall see Belnap has elaborated the Priorean view that, although we may formulate a so-called prima facie kind truth of contingent futures, such statements cannot be what Belnap has called "settled true". Belnap has described this inspiration from Arthur Prior in the following way:

> Although I suppose it is unscholarly, I have always thought that what I formulate using "settled" is indeed what he "meant", and what he "would have said" had he been aware of the mischief that could, alas, be caused by not making "settled" explicit. [Personal communication, 31 Oct., 2009].

Branching time

In his book *Time and Modality* (1957), Prior suggested that the modal logic of the Diodorean concept of possibility (and time) is simply the modal system, S4. One of the first readers to react on Prior's book, was Saul Kripke who was only 17 years old when he wrote the following to Prior:

> I have been reading your book *Time and Modality* with considerable interest. The interpretations and discussions of modality contained in your lectures are indeed very fruitful and interesting. There is, however an error in the book which ought to be pointed out, if you have not learned of it already [Letter from Saul Kripke to A. N. Prior, dated Sept. 3, 1958, The Prior Collection, Bodleian Library, Oxford; see Ploug and Øhrstrøm 2011].

Young Saul Kripke then continued his letter by explaining that the formula,

$$\Box\Diamond p \vee \Box\Diamond\sim p$$

can be verified using Prior's representation of Diodorean time as discrete sequences, but that this formula can be shown not to be provable in S4. In this way Kripke made an important contribution to the search for an axiomatic system corresponding to the Diodorean notion of modality. This research engaged several researchers in the late 1950s and the early 1960s. (See Prior 1967, p. 176). Even more important was the following passage from Saul Kripke's letter in which he suggested how the semantics of S4 could be visualized. Kripke's formulation of this very original idea in the letter makes it reasonable to classify the occurrence of this letter as one of the most important events in the history of logic during the twentieth century. Kripke wrote:

> I have in fact obtained this infinite matrix on the basis of my own investigations on semantical completeness theorems for quantified extensions of S4 (with or without the Barcan axiom). However, I shall present it here from the point of view of your "tensed" interpretation. (I myself was working with ordinary modal logic.) The matrix seems related to the "indeterminism" discussed in your last chapters, although it probably cannot be identified with it. Now in an indetermined system, we perhaps should not regard time as a linear series, as you have done. Given the present moment, there are several possibilities for what the next moment may be like—and for each possible next moment, there are several possibilities for the next moment after that. Thus the situation takes the form, not of a linear sequence, but of a "tree" (Fig. 1):

Saul Kripke explains this branching time model in the following way:

> The point 0 (or origin) is the present, and the points 1, 2, and 3 (of rank 2) are the possibilities for the next moment. If the point 1 actually does come to pass, 4, 5, and 6 are *its* possible successors, and so on. The whole tree then represents the entire set of possibilities for present and future; and every point determines a *subtree* consisting of its own present and future. Now if we let a tree sequence attach not three (as above) but a denumerable infinity of points to every point on the tree, we have a characteristic matrix for S4. An element of the matrix is a tree, with either 1 or 3 occupying each point; the designated tree contains only l's. If all points on the proper 'subtree' determined by a point on the tree p are 1's, the corresponding point on Lp is a 1; otherwise, it is a 3. (In other words, a proposition is considered "necessary" if and only if it is and definitely always will be the case.) [Letter from Saul Kripke to A.N. Prior, dated Sept. 3, 1958, The Prior Collection, Bodleian Library, Oxford]; (see Ploug and Øhrstrøm 2011).

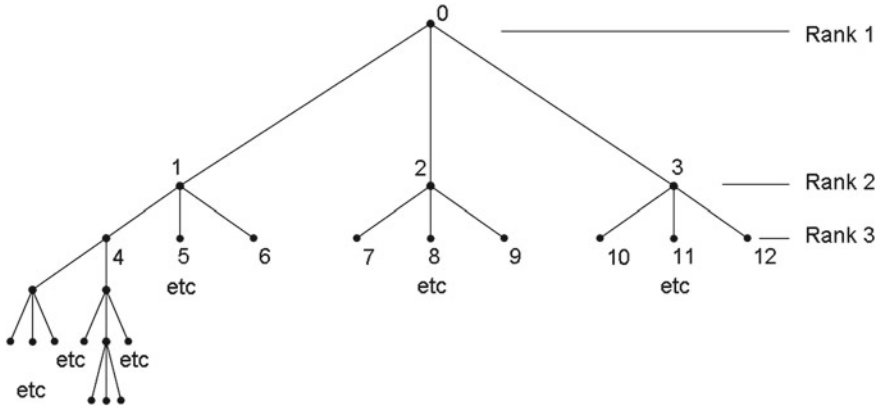


Fig. 1 The branching time model suggested by Saul Kripke

Here '1' stands for 'true' or 'true proposition', and '3' stands for 'false' or 'false proposition'. 'L' stands for the necessity operator.

In this way Saul Kripke argued that S4 corresponds to a branching time system combined with the Diodorean notion of temporal modality. This is the first ever presentation of branching time as a logical system. This was clearly recognised by Prior, who in his book *Past, Present and Future* discussed what he called "Kripke's branching time matrix for S4" (Prior 1967, p. 27). However, there are some obvious shortcomings of Kripke's semantics for predictions, i.e. that 'it will be p' and 'it is possible that it will be that p' are indistinguishable because Kripke keeps the semantic clause from linear time. This observation may have been an important part of Prior's motivation in his further development of branching time models.

Prior seems to have hesitated a bit in embracing the idea of branching time. This probably has to do with the so-called 'B-like' properties of the system (mainly the properties of the before-after relation). Prior clearly wanted a so-called A-theoretic approach to time (i.e. a view of time based on the tenses: past, present and future). On the other hand, he found that the crucial A-theoretical notion of free choice could be represented in terms of branching time in a very clear and convincing manner. In his later further elaboration of branching time Nuel Belnap strongly emphasized the possibility of explaining what indeterminism is using this approach to time. Belnap and Green stated:

> Branching time is not itself an indeterministic theory; instead, it says what indeterminism is, and it says what determinism is, but branching time does not choose between them (Belnap and Green 1994, p. 370).

When it comes to branching time, Belnap takes a clear stand. He argues that what he calls "Our World" can in fact be conceived as a branching time system, (see Belnap and Green 1994, pp. 370, 371 and 386). According to Belnap it is essential that the choices are real, i.e., that the world contains what he calls real possibility. For this reason, he argues that one should reject the idea suggested by David Lewis, according

to which the possibilities should be seen as parallel lines (and not as branching lines). According to Belnap, such a view is misleading because it does not represent the possibilities available to the free agents as belonging to reality (see Belnap 2007).

In his further development of the idea of branching time, Prior found great inspiration in the study of medieval philosophy. In particular, he found the works of William of Ockham (c. 1285–1347) interesting. The central theme in the medieval discussions regarding temporal logic was the apparent conflict between the doctrines of divine foreknowledge and human freedom. Can man be free if God already now knows with certainty what the person in question is going to choose? Ockham wrote a famous book, *Tractatus de praedestinatione et de futuris contingentibus*, on the subject, which exists in a modern translation and edition by Marilyn McCord Adams and Norman Kretzmann [1969]. In the book Ockham asserted that God knows all future contingents, but he also maintained that human beings can freely choose between alternative possibilities. He argued that the doctrines of divine foreknowledge and human freedom are in fact compatible.

Prior's study of Ockham's writings was a great inspiration when he formulated his formal ideas on branching time. Clearly, it should be kept in mind that Ockham himself had no formal language at his disposal. Prior had to transform Ockham's ideas into a modern context. Alex Malpass has edited the hitherto unpublished paper by Prior, *Postulate Sets for Tense Logic* [Forthcoming], which is kept in the Prior Collection at the Bodleian Library in Oxford. This paper was written and circulated in the mid-60s, and is probably a draft of Prior (1966) paper *Postulates for Tense Logic* and chapter VII.4 of *Past, Present and Future* (1967). The paper is the earliest known example of Prior's attempts at formulating a branching theory of his own. In the paper Prior presents what he calls "an Occamist model", which he used to formulate an account of the future tense that was more acceptable to Ockham's philosophical views on future contingents than Kripke's simple semantics. (In his early writing Prior seems to have used the spelling 'Occam', whereas he used 'Ockham' in his later writings.)

> In these models the course of time (in a rather broad sense of this phrase) is represented by a line which, as it moves from left to right (past to future), continually divides into branches, so that from any given point on the diagram there is a unique route backwards (to the left; to the past) but a variety of routes forwards (to the right; to the future). In each model there is a single designated point, representing the actual present moment; and in an Occamist model there is a single designated line (taking one only of the possible forward routes at each fork), which might be picked out in red, representing the actual course of events (Prior 2014).

In his 1966 paper, Prior suggested two versions of the Occamist model, O and O′. In both of them he assumed a designated route. He wrote:

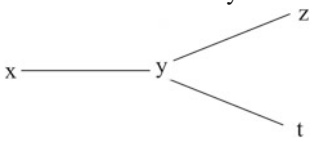
> In each O and O′ model there is a single designated route from left to right, taking one direction only at each fork. This represents the actual course of events (1966, p. 157).

This idea of the true future as a single designated line is an idea which is now seen as rather controversial within the discussion of branching time models. Prior made a formal distinction between A-variables which stand for "those propositions which it is now beyond our power to make true or false" (so-called) and other propositional

variables. Using this distinction and the notion of branching time, Prior showed how actual assignments of truth values at a point in the model and various so-called prima facie assignments could be introduced. He presented the first three steps in the procedure in the following way:

(1) Each A-variable is arbitrarily assigned an actual truth-value at each point, and this is its only prima facie assignment at that point.
(2) A prima facie assignment to Fn**Φ** at a point x will give it the value assigned to **Φ** at the distance n along some path to the right of x (where the diagram forks within this distance, Fn**Φ** will have a number of different prima facie assignments at x).
(3) An actual assignment to Fn**Φ** at x gives it the value of **Φ** at the distance n to the right along the designated line.

In the paper, Prior illustrates his definitions by the following simple model:



It should be noted that Prior in his book *Past, Present and Future* dropped the use of the idea of "an actual assignment" and concentrated on a definition of the Ockhamistic model in terms prima facie assignments only, although no surviving explanation from Prior exists which explains why he dropped the notion. As I have argued in [1981], William of Ockham would not be an Ockhamist in this Priorean sense. However, the theorems of the two Priorean and Ockham-like systems will be the same, and the Ockhamist system defined in *Past, Present and Future* is certainly interesting (see Reynolds 2003).

Prior's Ockhamistic system suggested in Prior (1967, p. 126 ff). may be presented in terms of the following recursive definition (see Øhrstrøm and Hasle 2011):

(a) $Ock(m, c, p) = 1$ iff $TRUE(p, m) = 1$, where p is any propositional constant.
(b) $Ock(m, c, p \wedge q) = 1$ iff both $Ock(m, c, p) = 1$ and $Ock(m, c, q) = 1$
(c) $Ock(m, c, \sim p) = 1$ iff not $Ock(m, c, p) = 1$
(d) $Ock(m, c, Fp) = 1$ iff $Ock(m', c, p) = 1$ for some $m' \in c$ with $m < m'$
(e) $Ock(m, c, Pp) = 1$ iff $Ock(m', c, p) = 1$ for some $m' \in c$ with $m' < m$
(f) $Ock(m, c, \Diamond p) = 1$ iff $Ock(m, c', p) = 1$ for some $c' \in C(m)$

Here *TRUE* is a function, which gives a truth-value (0 or 1) for any propositional constant at any moment m in the branching time structure, $(TIME, \leq)$. What Prior called lines or routes, i.e. the maximal linearly ordered subsets in $(TIME, \leq)$, are often now called chronicles. We shall use this term in the following. $C(m)$ is defined as the set of chronicles through the moment of time m, i.e., $C(m) = \{c \in C | m \in c\}$, where C is the set of all chronicles in $(TIME, \leq)$.

Strictly speaking, (a)–(f) only explain when *Ock* has the value 1 ('true'). It should be added, that the value is 0 ('false'), if it does not follow from the recursive definition above that is 1.

Ock $(m, c, p) = 1$ can be read 'p is true at m in the chronicle c'. A formula p is said to be Ockham-valid if and only if $Ock(m, c, p) = 1$ for any t in any c in any branching time structure, (TIME, $\leq$) and any valuation function *TRUE*. Here C should not be taken as an independent parameter. Furthermore, it should be noted that relative to a single chronicle, (a)–(e) are exactly the same definitions as those used in linear tense-logic (i.e. the tense-logic which follows if (TIME,$\leq$) is a linear structure).

We define the dual operators, H, G, and $\Box$ in the usual manner as $\sim P\sim$, $\sim F\sim$, and $\sim\Diamond\sim$ respectively.

Obviously, there is no designated line (Thin Red Line) in Prior's Ockhamistic system from *Past, Present and Future*, as there were in the two earlier versions of the system mentioned above. If we wish to have such a feature, it has to be added explicitly.

In their 1994 paper, Belnap and Green introduced the term "the Thin Red Line" with reference to an idea very much similar to Prior's "designated line, picked out in red". The term suggested by Belnap and Green was not inspired by Prior's earlier notion. (Belnap apparently never received a copy of Prior's *Postulate Sets for Tense Logic*, and he was not aware of Prior's use of the expression [Personal communication, 25 April, 2012].)

Belnap's and Green's term was inspired by a report from the Crimean War in *The London Times*: "The Russians dashed on towards that thin red-line streak tipped with a line of steel." It has even been suggested that the thin red line should in fact be conceived as infrared indicating "that the Thin Red Line does not imply that mortals are capable of seeing the future" (Belnap et al. 2001, p. 139).

Belnap and his co-workers have presented several arguments against the idea of "the thin red line" and the use of this idea in branching time semantics. In the following, we shall consider some of these arguments and discuss to what extent the idea can be defended. I shall refer to William of Ockham as a main spokesman for the view that the thin red line is important for the proper understanding of temporal reality. In addition I shall refer to the works of Luis de Molina (1535–1600), who much later than Ockham defended an even more elaborated version of the notion of "the thin red line" (see Craig 1988, p. 175). In both cases the notion was presented in terms of the Christian doctrine of divine foreknowledge. It should, however, be pointed out that this view does not have to be linked to a theological framework. Everything which will be said in favour of the idea of the thin red line can be translated into a secular language.

1 There is No Truth Concerning Future Contingents

Nuel Belnap has maintained that "the Thin Red Line" is in no way part of the real world. Before a free choice the alternative possibilities are equally real. There is no designated future if the choice is free. Nobody could know what is going to be freely chosen before the choice has actually been made. In his own words:

> There is no real choice without the reality of alternative possible choices facing the agent. Each of these possibilities is, before the moment of choice, as real as any other. It is true and important that at most one of these possibilities will be realized. It is equally true and equally important that none of these possibilities is a ghostly image of some specially distinguished one among them that some philosopher might label "the actual choice". This form of actualism is a bad idea (Belnap 2001, p. 2).

It seems that Belnap assumes that "a ghostly image" of "the actual choice" is needed in order to make it true that a certain free agent is going to carry out a certain act. However, as Trenton Merricks (2007) has argued the need for truth-makers in order to establish the truth of propositions can certainly be questioned. As Merricks has shown we may alternative hold that being true is a primitive monadic property (2007: 170 ff.) It is, on the other hand, probably true that medieval logicians would have a view closer to what Belnap is criticising as their metaphysical reasoning for believing in "the thin red line".

It is not difficult to imagine how William of Ockham would have replied to Belnap's criticism. He would probably have pointed out that Belnap's position should be accepted as long as we are dealing with human cognition alone. However, there might be a deeper structure in reality which is not directly accessible to the human mind, but which nevertheless is useful for a deeper understanding of natural language and common sense reasoning. As a believer, Ockham stated his view referring to divine foreknowledge. He willingly admitted that this idea is very hard to understand for a human being. However, he attempted to clarify the issue as much as possible. Ockham stated:

> … the divine essence is an intuitive cognition that is so perfect, so clear, that it is an evident cognition of all things past and future, so that it knows which part of a contradiction [involving such things] is true and which part is false (Ockham 1969, p.50).

Ockham had to admit that much of this cannot be stated in a very clear manner. In fact, he maintained that it is impossible to express clearly the way in which God knows future contingents. He also had to conclude that in general the divine knowledge about the contingent future is inaccessible. God is able to communicate the truth about the future to us, but if God reveals the truth about the future by means of unconditional statements, the future statements cannot be contingent anymore. Hence, God's unconditional foreknowledge regarding future contingents is in principle not revealed, whereas conditionals can be communicated to the prophets. Even so, that part of divine foreknowledge about future contingents which is not revealed must also be considered as true according to Ockham.

Ockham was aware that the concept of communication was essential to this discussion—especially, of course, the communication coming from God to human beings. He claimed that God can communicate the truth about the future to us. Nevertheless, according to Ockham divine knowledge regarding future contingents does not imply that they are necessary. As an example Ockham considered the prophecy of Jonah: "Yet forty days, and Nineveh shall be overthrown" [The Book of Jonah ch. 3 v. 4]. This prophecy was a communication from God about the future. Therefore, it might seem to follow that when this prophecy had been proclaimed the future destruction of Nineveh would be necessary. But Ockham did not accept that. Instead,

he made room for human freedom in the face of true prophecies by assuming that "all prophecies about future contingents were conditionals" (Ockham 1969, p. 44). So according to Ockham we must understand the prophecy of Jonah as presupposing the condition "unless the citizens of Nineveh repent". Obviously, this is in fact exactly how the citizens of Nineveh understood the statement of Jonah!

Ockham realised that the revelation of the future by means of an unconditional statement, communicated from God to the prophet, is incompatible with the contingency of the prophecy. If God reveals the future by means of unconditional statements, then the future is inevitable, since the divine revelation must be true. Such possible restrictions on the use of divine communication (revelation) must be taken into consideration, if the belief in divine foreknowledge is to be compatible with the belief in the freedom of human actions.

When translated into a secular language this means that if there is a designated future, which is invisible to human agents, it will not destroy their freedom of choice. In terms of Belnap's notions: If the thin red line is in fact part of reality, then it has to be "infrared" in the sense that it is undetectably to human beings, given that free choice is also part of reality. However, this is not surprising. There are many aspects of reality which we are ready to accept although they are even in principle not verifiable. One such aspect is in fact free choice itself and its rooting the human mind!

2 A Thin Red Line Theory is Insufficient as a Background for a Proper Understanding of the Structure of Tenses in Natural Language

The typical argument given in favour of the assumption of a designated future is that we may in this way deal better with natural language and common sense reasoning. However, it has been argued that this assumption is quite insufficient as a background for a satisfactory model fit for dealing with the logic of tenses in natural language. The Thin Red Line is supposed to help, but perhaps it does not.

Nuel Belnap and Mitchell Green have given a very nice example in support of this criticism of a "Thin Red Line" theory:

> The coin will come up heads. It is possible, though that it will come up tails, and then later it will come up tails again (though at this moment it could come up heads), and then, inevitably, still later it will come up tails yet again (Belnap and Green 1994, p. 379).

Clearly, this example calls for the use of so-called embedded tenses. It is not sufficient to be able to refer to what is actually going to be the case, but we should also be able to discuss what in alternative (counterfactual) situations would have been going to happen. A designated future, it seems, is not enough.

Belnap and Green's statement may be represented in terms of tense logic with τ representing tails and η heads, respectively:

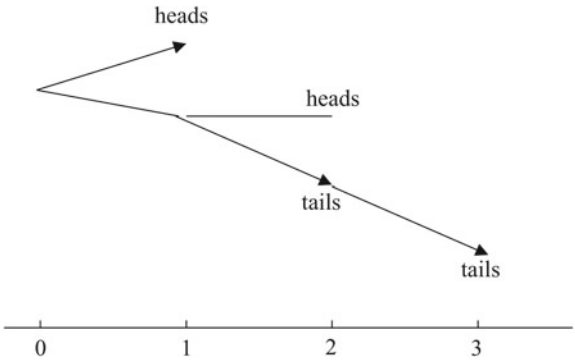


Fig. 2 A branching time model representing an example suggested by Nuel Belnap and Mitchell Green

$$F(1)\eta \wedge \Diamond F(1)(\tau \wedge \Diamond F(1)\eta \wedge F(1)(\tau \wedge \Box F(1)\tau))$$

The problem for a Thin Red Line theory in evaluating this proposition is how to understand the embedded occurrences of the *F*-operator. One way to do this is by using the following branching time structure, which has been enriched with arrows indicating not only a single designated future, but actually a designated future at every branching point in the system (Fig. 2):

The example shows that if the model is taken seriously, then there must be a function TRL, which gives the true future for any moment of time, *m*. More precisely, *TRL*(*m*) yields the linear past as well as the true future of *m*, extended to a maximal set. In this way, *TRL*(*m*) will for any moment of time, *m*, be a chronicle within the branching time system.

It is very likely that William of Ockham would have accepted the points made by Belnap and Green regarding embedded tenses. When analysing the features of the Ockhamistic model, it becomes evident that within the model there must be a true future, not only in every actual situation or instant, but also in every possible situation. This was at least realised by Luis de Molina, who worked some centuries after Ockham, but still very much in the same scholastic tradition. Molina's special contribution is the idea of (God's) middle knowledge, "by which, in virtue of the most profound and inscrutable comprehension of each free will, He saw in His own essence what each such will would do with its innate freedom were it to be placed in this or that or indeed in infinitely many orders of things — even though it would really be able, if it so willed, to do the opposite" (quoted from Craig 1988, p. 175). Craig goes on to explain it as follows: "… whereas by His natural knowledge God knows that, say, Peter when placed in a certain set of circumstances *could* either betray Christ or not betray Christ, being free to do either under identical circumstances, by His middle knowledge God knows what Peter *would* do if placed under those circumstances" (Craig 1988, p. 175). Craig has argued that such counterfactuals of freedom can be true even if there is nothing to make it true and no grounding of such truth. On the contrary, the truth of counterfactuals of freedom might be taken as

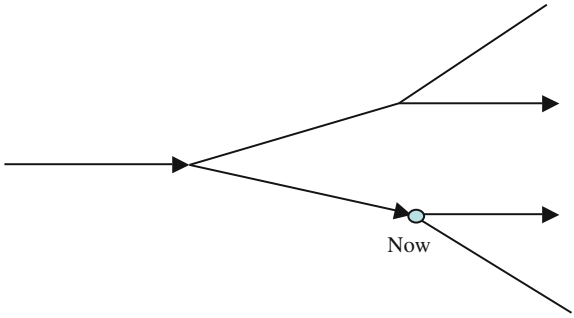


Fig. 3 A representation of the Ockhamistic/Molinistic model in terms of Prior's notion of branching time

indicating the theories of truth-makers and grounding should be rejected. (See Craig 2001, Merricks 2007, 146 ff.)

Using Prior's notion of branching time it might be extended and represented by a diagram such as the following where the idea of the true future (including the idea of 'middle knowledge') is indicated by the use of arrows showing the true or selected courses of events (Fig. 3).

The wisdom obtained from the critical points made by Belnap and Green suggests that a Thin Red Line theory based on a single designated line will be insufficient. If such a theory is possible, it has to include a unique true future at any point in the model although there may be several possible futures at each point in the model. The conclusion is that in the search for a Thin Red Line theory, one should look for a theory based on a TRL-function from temporal moments to histories in the model. We shall call such a theory "a TRL theory".

3 An Obvious Requirement Regarding Iterative Tenses Makes TRL Theories Problematic

Belnap and Green (1994) have argued that any serious TRL theory should imply the validity of the following fundamental relation regarding iterative tenses.

$$(\text{T1})\, P\,P q \supset P q$$
$$(\text{T2})\, F\,F q \supset F q$$

From an intuitive point of view the validity of (T1-2) appears to be rather obvious. T1 says that if it was that it was that q, then it was that q, etc. This understanding of the iterated tenses seems straight forward given the way the tenses are used in natural language and in common sense reasoning. In a similar way, several other basic expressions have to come out as valid in general, if the theory in question is to be accepted. One other obvious proposition which should be valid in general is

$$(\mathrm{M1})\, Fq \supset \Diamond Fq$$

If it will be that q, then it is possible that it will be that q. There can be no doubt that William of Ockham would have understood this type of requirement. After all, he also wanted to formulate a logical theory in accordance with natural language and common sense reasoning.

In their 1994 paper Belnap and Green suggested that the TRL-function in a TRL theory in order to lead to the general validity of expressions like (T1-2) and (M1) satisfy the following conditions:

$$(\mathrm{TRL1})\ m \in TRL(m)$$
$$(\mathrm{TRL2})\ m_1 < m_2 \supset TRL(m_1) = TRL(m_2)$$

However, as Belnap and Green have correctly pointed out the acceptance of the combination of (TRL1) and (TRL2) entails a rejection of the very idea of branching time. The reason is that if (TRL1) and (TRL2) are both accepted, it follows from $m_1 < m_2$ that $m_2 \in TRL(m_1)$, i.e. that all moments of time after m_1 would have to belong to the thin red line through m_1, which means that there will in fact be no branching at all.

This seems to give rise to a problem for the TRL theory. However, it turns out that there is in fact no need to accept (TRL2), which seems to be too strong a requirement. Rather than (TRL2), the weaker condition (TRL2′) can be employed:

$$(\mathrm{TRL2}')(m_1 < m_2 \wedge m_2 \in TRL(m_1)) \supset TRL(m_1) = TRL(m_2)$$

This weaker requirement appears to be much more natural in relation to the basic idea of TRL-theory. Belnap has later accepted that (TRL2′) is a relevant alternative to (TRL2) ([Personal correspondence, 1 Aug. 1996] and Belnap et al. 2001, p. 169).

Following Prior's ideas in *Postulate Sets for Tense Logic* extended to a TRL-model we can formulate the following truth condition for the future operator:

(i) Fq is true a moment m iff there is a moment of time, $m' \in TRL(m)$, such that $m < m'$ and q is true at m'.

In the same way it is possible to define what it means for a proposition, Pq, to be true at the moment m, taking $TRL(m)$ as the designated line.

Given these truth conditions it is easily seen that (T1-2) are valid in general. In addition (M1) will be valid in general, if we accept the following truth condition:

(ii) $\Diamond Fq$ is true a moment m iff there is a moment of time, m', and a chronicle, c, such that $m \in c$, $m' \in c$, $m < m'$ and q is true at m'.

4 TRL Theories Lead to Problematic Evaluations at Counterfactual Moments of Time

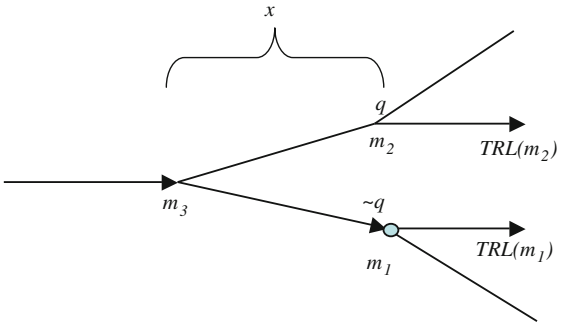
Belnap and Green (1994) have argued that in addition to (T1-2) any serious TRL theory should imply the general validity of expressions like

$$\text{(T3) } q \supset P(x)F(x)q$$

where *P(x)* stands for "it was the case x time units ago" and *F(x)* stands for "it is going to be the case in x time units". (T3) should be true not only at moments belonging to the history which is actually taking place, but also at counterfactual moments.

Again, following a tradition from medieval logic, it seems reasonable to require that statements like (T3) are true even at counterfactual moments of time. Logicians like William of Ockham would be very likely to have accepted that (T3) should be valid in general.

However, this will be difficult to maintain (T3) as valid within a TRL-theory if we assume a rigorous notion of compositionality for the evaluation of truth values. Consider, for instance, a branching time model, which can be illustrated in the following way:



Given this TRL model we may ask whether $q \supset P(x)F(x)q$ is true at m_2. As indicated above q is true at m_2. However, assuming a rigorous notion of compositionality *P(x)F(x)q* is false at m_2, since *F(x)q* appears to false at m_3.

However, alternatively one may insist that any evaluation of a truth value at moment of time, *m*, should be carried out as if *TRL(m)* were the designated line ("The Thin Red Line"). This means that truth of a position, *p*, at a moment of time, *m*, may simply be defined in term of the truth-function in Prior's Ockhamistic system in the following way:

$$\text{true}_{\text{T}}(p, m) = Ock(m, TRL(m), p)$$

If this is accepted no iteration of the tense operators, *P* and *F*, will get us off the designated chronicle when calculating the truth value of a proposition at *m*. Using this approach to the evaluation of counterfactual truth-values, we will in the above case find that the implication, $q \supset P(x)F(x)q$, is in fact true at m_2. This is so, because the evaluation is carried out only referring to $TRL(m_2)$.

Taking this rather simple approach we end up with a logical system with exactly the same theorems as in Priorean Ockhamism, including (T3). This is what we might call the simple Ockhamistic answer to the Belnap-Green challenge.

However, it might be objected that if we were to assume a designated chronicle as a background for the evaluation of the truth-value of a tense-logical proposition it ought to be $TRL(m_1)$ (i.e. the actual history) and not an alternative history such as $TRL(m_2)$. This objection appears to be based on the view, that any counterfactual statement in principle has made as seen from the actual world. When we are claiming that something like (T3) might be true even at a counterfactual moment of time, m_2, what we mean is that at the present moment of time, m_1, it is true for any numbers x and y that

$$P(y)\Diamond F(y)(q \supset P(x)F(x)q)$$

In fact, the claim that (T3) holds in general, means that the implication mentioned in (T3) would have been true no matter what had happened in the past i.e. even if alternative past possibilities had been actualized. This means that at the present time, m_1, the following is true for arbitrary positive numbers z, y and x:

$$P(z)\Box F(y)(q \supset P(x)F(x)q)$$

According to this approach, we suggest that the truth-value of a tense-logical expression at a moment of time, *m*, should be evaluated as in Prior's paper mentioned above using the branching time and taking *TRL(m)* to be the designated (red) line. In order to deal with the modal operators in a precise manner, we need a truth condition for the modal operators which more general than (ii) in Sect. 3. We may consider the following Ockhamistic truth condition:

(iii) $\Diamond p$ is true at the moment of time, *m,* relative to a chronicle c iff there is a chronicle, c', through m, such that p is true at m relative to a c', which is understood as the chronicle that should be used in the further evaluation.

However, it may be objected that in such a model the *TRL*-function has really no role to play in the semantics, in the sense that the properties of the TRL-function does not influence which propositions are valid in general and which are invalid. However, as pointed out in Braüner et al. (2000), Øhrstrøm (2009), it is in fact possible to create an alternative system, in which the TRL-function plays such a role. This may be done using the following Ockhamistic truth condition:

(iv) $\Diamond p$ is true at the moment of time, *m,* relative to a chronicle, *c,* iff there is a chronicle, c', belonging to $C_T(m)$, such that p is true at m relative to a c', which is understood as the chronicle that should be used in the further evaluation,

where

$$C_T(m) = \{c | m \in c \ \& \ TRL(m') = c, \text{ for any } m' \in c \text{ with } m < m'\}$$

Note that $C(m)$ is a subset of all chronicles through m. With this definition any history used in evaluation of the proposition $\Diamond p$ at a moment m, can be conceived as a $TRL(m')$, where m' is a moment immediately after m.

As argued in Braüner et al. (2000), Øhrstrøm (2009) this alternative definition leads to a slightly different semantics, according which e.g. the proposition, $F(x)\Diamond F(y)p \supset \Diamond F(y)F(y)p$, will not be valid in general. This means that in this system something which is not yet possible may become possible, i.e. new possibilities may turn up! However, it should be emphasized there is not absolute need to go for a system like this, but the existence of this alternative system at least shows that simple *TRL*-system mentioned above it not the only possible and that it is possible to define a semantic system in which the *TRL*-function plays a significant role in the semantics.

Conclusion

As argued above, it is possible to respond to the Belnap-Green challenge in a reasonable manner. One solution is the simple Ockhamistic answer. A slightly more sophisticated solution has been suggested in Braüner et al. (2000). There are other interesting solutions such as the one suggested by Malpass and Wawer (2012), where there is a single designated line and a supervaluational account of counterfactual future contingents is given.

Playing with a title of Dummett (*The logical basis of metaphysics*, 1991), Nuel Belnap wrote:

> If you wish to learn the "metaphysical basis of logic" according to some logician, studying the inductive account of the language is useful, but it is crucial to understand his or her explanations of the parameters that are at bottom of the entire enterprise (Belnap 2007, p. 97).

No doubt, William of Ockham would have agreed. He wanted to study the tenses as they are used in natural language and in common sense reasoning. But he certainly wanted to do so based on what he believed to be the fundamental features of our world. A very important feature of the world according to Ockham's view is that exactly one of the many possible ways, in which the world may develop, is the true one. He would insist that we have to develop our logical theories taking this important fact into account. And even more important we have to carry out this task in a logically consistent manner. For this reason William of Ockham would clearly also have appreciated the challenges formulated by Nuel Belnap and his co-workers, since these thoughtful comments have been a great help to anyone who wants to establish a consistent theory of what Belnap and Green have called "the thin red line".

References

Belnap, N. 2001. Double time references: speech-act reports as modalities in an indeterminist setting. In *Advances in Modal Logic*, eds. Wolter, F. et al., vol. 3. Stanford: CSLI Publications.

Belnap, N. 2007. An indeterminist view of the parameters of truth". In *Philosophie der Zeit. Neue Analytische Ansätze*, ed. Müller, T., Frankfurt a.M.: Klostermann.

Belnap, N., and M. Green. 1994. Indeterminism and the thin red line. *Philosophical Perspectives, Logic and Language* 8: 365–388.

Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future. Agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Braüner, T., Hasle, P. and Øhrstrøm, P. 2000. Determinism and the origins of temporal logic. In *Advances in temporal logic*, eds. Barringer, H. et al., 185–206. Dordrecht: Kluwer Academic Publishers.

Craig, W.L. 1988. *The problem of divine foreknowledge and future contingents from Aristotle to Suarez*. New York: E.J. Brill.

Craig, W.L. 2001. Middle-knowledge, truth-makers, and the grounding objection. *Faith and Philosophy* 18: 337–352.

Malpass, A., and J. Wawer. 2012. A future for the thin red line. *Synthese* 188: 117–142.

Merricks, T. 2007. *Truth and ontology*. Oxford: Oxford University Press.

Øhrstrøm, P. and Hasle, P. 1995. *Temporal logic—from ancient ideas to artificial intelligence*. Dordrecht: Kluwer Academic Publishers.

Øhrstrøm, P. and Hasle, P. 2006. Modern temporal logic: the philosophical background, *Handbook of the History of Logic*, 447–498 Bd. 7 Amsterdam: Elsevier.

Øhrstrøm, P. and Hasle, P. 2011. Future contingents. *Stanford Encyclopedia of Philosophy*, Stanford: Stanford University (Summer 2011 Edition).

Øhrstrøm, P. 2009. In defense of the thin red line: a case for Ockhamism. *Humana Mente* 8: 17–32.

Ploug, T. and Øhrstrøm, P. 2011. Branching time, indeterminism and tense logic - unveiling the prior-kripke letters. *Synthese* 188: 367–379 (Nr. 3, 1.10.2012).

Prior, A.N. 1955. Diodoran modalities. *The Philosophical Quarterly* 5: 205–213.

Prior, A.N. 1957. *Time and modality*. Oxford: Clarendon Press.

Prior, A.N. 1966. Postulates for tense-logic. *American Philosophical Quarterly* 3(2): 153–161.

Prior, A.N. 1967. *Present and future*. Oxford: Clarendon Press.

Prior, A.N. Postulate sets for tense logic. (Edited by Alex Malpass.) Forthcoming in *Synthese.*

Reynolds, M. 2003. An axiomatization of Prior's Ockhamist logic of historical necessity. In *Advances in Modal Logic*, eds. P. Balbiani, N-Y. Suzuki, F. Wolter and M. Zakharyaschev, vol. 4, p. 355–370. London: King's College Publications.

William of Ockham. 1969. *Predestination, God's Foreknowledge, and Future Contingents* (trans: Marilyn McCord Adams and Norman Kretzmann). New York.

Branching for General Relativists

Tomasz Placek

Abstract The chapter develops a theory of branching spatiotemporal histories that accommodates indeterminism and the insights of general relativity. A model of this theory can be viewed as a collection of overlapping histories, where histories are defined as maximal consistent subsets of the model's base set. Subsequently, generalized (non-Hausdorff) manifolds are constructed on the theory's models, and the manifold topology is introduced. The set of histories in a model turns out to be identical with the set of maximal subsets of the model's base set with respect to being Hausdorff and downward closed (in the manifold topology). Further postulates ensure that the topology is connected, locally Euclidean, and satisfies the countable sub-cover condition.

1 Introduction

In 1992 Nuel Belnap put forward the branching space-times theory (BST1992) that offered a unified treatment of rudimentary relativistic spacetimes and indeterminism.[1] Building on earlier works on a more frugal theory of branching time (BT), BST1992 represents indeterminism by means of a collection of overlapping histories; in contrast to the linear histories of the former, however, histories are complex objects in BST1992. As a consequence, there are models of BST1992, in which his-

[1] This chapter owes much to Nuel Belnap who spent days listening to my frequently confused arguments, pointing out my mistakes, suggesting corrections, or repairs. I would like to thank Juliusz Doboszewski for reading the proofs of this chapter. I also gratefully acknowledge the support of the research grant 668/N-RNP-ESF/2010/0 of the (Polish) Ministry of Science and Higher Education.

T. Placek (✉)
Department of Philosophy, Jagiellonian University, Grodzka 52 pok. 17,
31-044 Kraków, Poland
e-mail: uzplacek@cyf-kr.edu.pl; Tomasz.Placek@uj.edu.pl

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_10,

tories are isomorphic to the Minkowski spacetime (see Placek and Belnap (2012)). BST1992 can be used to model quantum experiments with non-local correlations (Placek 2010). Furthermore, a branching reading can be given to the consistent histories formulation of quantum mechanics (see Müller (2007)).

This bright picture, however, has been marred by a tension between BST1992 and general relativity (GR). There are serious obstacles to accommodating GR in the branching framework, the most important of which, I believe, is a difference in spirit. The great perception of GR is that coordinalization works by patches: this theory permits the assignments of coordinates (elements of $\mathbb{R}^n$) to subsets (patches) of the totality of events, with the proviso that the patches cover the totality of events. Local coordinalization by patches is to be contrasted with a global coordinalization, as provided by a mapping of a whole spacetime on $\mathbb{R}^n$. Patches, if sufficiently small, have familiar and desirable properties. In essence, they look like subspaces of Minkowski spacetime,[2] which in turn permits a definition of a partial ordering on a patch. Typically these nice properties do not transform to a GR spacetime as a whole, however.

In contrast, BST1992 does not work in terms of local patches. This theory assumes a partial ordering on its base set, and defines history (aka BST spacetime) as a maximal upward directed subset of the base set. With some extra assumptions added, a BST1992 history can be mapped on $\mathbb{R}^n$. Even if one wants to do coordinalization in a piecemeal way, there is no structure in BST1992 that could play the role of patches.

Apart from this difference in spirit, there are technical issues as well: First, the ordering assumed in BST1992 is partial, whereas the natural ordering of a GR spacetime, defined in terms of geodesics, is not necessarily so: it allows for a failure of anti-symmetry. Second, the BST1992 criterion for historicity (or, belonging to one BST spacetime), i.e., being maximally upward directed, flies in the face of some well-studied GR spacetimes, like the Schwarzschild spacetime or the de Sitter cosmological model. The criterion rules out as well some intuitive, although non-physical, candidates for a spacetime since it implies that for two events x and y to belong to some one spacetime, there should be a "later witness", that is, some z such that $x \leqslant z$ and $y \leqslant z$. Consequently, an open square or an open half-plane $\mathbb{R}^- \times \mathbb{R}$, both with Minkowskian ordering, cannot be BST1992 spacetimes.[3] A sought-for generalization of BST1992 should thus modify the criterion for historicity appropriately. (For a discussion as to how one can modify the BST1992 notion of history, see Müller (2013).)

The first attempt to overcome the tensions between GR and BST1992 is Müller (2011). The present chapter continues this work in a somewhat different way, by first generalizing BST1992 appropriately, then defining generalized manifolds on models of generalized BST and, finally, by producing tangent vector spaces.

Although the main aim of this chapter is to offer a GR-friendly generalization of BST1992, I begin by addressing an objection to BST1992. As John Norton once

[2] Strictly speaking, these are properties of tangent spaces rather than of subsets of events.

[3] This ordering $\leqslant_M$ is defined on $\mathbb{R}^n$ by putting $x \leqslant_M y$ iff $x_1 \leqslant y_1$ and $\sum_{i=2}^{n}(x_i - y_i)^2 \leqslant (x_1 - y_1)^2$, where x_1 is the time coordinate and $x_2, \ldots, x_n$ are spatial coordinates.

said, physical theories do not offer the kind of branching that BST1992 assumes.[4] Indeed, the pattern of branching implied by the axioms of BST1992 is particular: If a maximal chain in a base set passes through a maximal element in the overlap of some two histories, then obviously the segment of the chain contained in the overlap has a maximum and, hence, a supremum. But if a maximal chain does not pass through a maximal element in the overlap, the chain's segment contained in the overlap does not have a supremum, but rather two history-relative suprema. Instead of addressing the objection head-on, I argue that a slight modification of BST1992 axioms yields another pattern of branching, which appears to be better suited for a GR-friendly version of BST. In this discussion I introduce choice pairs, a valuable tool for the generalized BST, described in later sections.

The chapter is organized as follows. Section 2 puts forward a version of branching space-times that yields a different pattern of branching histories. Section 3 discusses how BST1992 should be generalized: its basic idea is that topological features of BST1992 should be preserved by the generalization. To this end, this section offers a summary of the topological properties of BST1992 models. Sections 4.1, 4.2, and 4.3 put forward a three-tiered construction of (1) generalized BST models, then (2) generalized manifolds built on these models, and finally, (3) vector spaces of tangent vectors. The next, Sect. 5, addresses some paradoxical issues concerning generalized manifold. Section 6 concludes the chapter with an overview of the chapter's result.

2 BST with a New PCP

Let us recall the basic definitions of BST1992:

A model of BST1992 is a nonempty partial order $\mathcal{W} = \langle W, \leqslant \rangle$ that satisfies the axioms below, with histories in $\mathcal{W}$ defined as maximal upward directed subsets of W. The axioms are as follows:

1. $\mathcal{W}$ has no maximal elements;
2. $\leqslant$ is dense;
3. every lower bounded chain has an infimum in $\mathcal{W}$;
4. every upper bounded chain has a supremum in every history that contains it;
5. for a chain C in $\mathcal{W}$: if $C \subseteq h/h'$, then there is a maximal element in $h \cap h'$ strictly below C (such a maximal element is called a choice point for h and h'; this axiom is called Prior Choice Principle—PCP).

We say that two histories, h, h' are divided at e if e is a maximal element of the intersection $h \cap h'$. And we say that two histories, h, h' are undivided at e if $e \in h \cap h'$ but is not a maximal element of $h \cap h'$. Provably undividedness at e is an equivalence relation on the set of histories containing e. The equivalence classes with respect to this relation are called "elementary possibilities open at e".

[4] After my lunch talk at the Center for the Philosophy of Science of the University of Pittsburgh in February 2008.

A particular pattern of branching mentioned above (aka passive indeterminism or indeterminism without choice—see Placek and Belnap (2012)) is a consequence of PCP. To illustrate, consider a two-history model, with a *single* choice point c, and with histories identified with planes (i.e., $\mathbb{R}^2$), the ordering being Minkowskian. PCP then dictates, first, that the "wings" of the choice point c, that is, the set of events space-like related to c, are in the overlap of the two histories. Second, it prohibits points on the future light cone above c to belong to the overlap; otherwise c would not be maximal in the overlap, i.e., not a choice point.

Our idea is thus to replace PCP by a somewhat different principle, while keeping intact all the other axioms of BST1992.[5] Our new principle postulates the existence of minimal pairs of a particular kind rather than maximal elements in the overlap of histories. As we will see, it enforces a different pattern of branching.

Pairs supreme, hot pairs, and choice pairs. In what follows, we assume tentatively the notion of BST1992 models, with PCP removed.

Definition 1 (pairs supreme) *For $s, s' \in W$, we say that $\{s, s'\}$ is a pair supreme for histories h, h', to be written as $\{s, s'\} \in \mathfrak{S}(h, h')$,* iff $\exists C(C \neq \emptyset \wedge C \subseteq h \cap h' \wedge s = \sup_h(C) \wedge s' = \sup_{h'}(C))$, *where C is an upper bounded chain in $\mathcal{W}$.*

$\{s, s'\}$ is a pair supreme simpliciter, to be written as $\{s, s'\} \in \mathfrak{S}$, iff $\{s, s'\} \in \mathfrak{S}(h, h')$ for some histories h, h'.

Note that the definition allows for a pair supreme $\{s, s'\}$ with identical elements, i.e., $s = s'$, as well as for a pair supreme with distinct elements. To capture the latter case, we define 'hot pairs':

Definition 2 (hot pair) *For $s_1, s_2 \in W$, $\{s_1, s_2\}$ is a hot pair for histories h, h', to be written as $\{s_1, s_2\} \in \mathfrak{H}(h_1, h_2)$,* iff $\{s_1, s_2\} \in \mathfrak{S}(h, h')$ *and $s_1 \neq s_2$. And we say that $\{s, s'\}$ is a hot pair simpliciter, to be written as $\{s, s'\} \in \mathfrak{H}$, iff $\{s, s'\} \in \mathfrak{H}(h, h')$ for some histories h and h'.*

Hot pairs decide between histories in the sense that an event above an element of a hot pair for two histories cannot belong to both these histories.

Fact 3. *If $\{s_1, s_2\} \in \mathfrak{H}(h_1, h_2)$ and $s_i \leqslant e$ for some $i = 1, 2$, then $e \notin h_1 \cap h_2$.*

Proof Obvious. Since histories are downward closed, $e \in h_1 \cap h_2$ and $s_i \leqslant e$ imply $s_i \in h_1 \cap h_2$, which implies $s_1 = s_2$: a contradiction with $\{s_1, s_2\}$ being a hot pair. □

We next define an ordering of pairs supreme (simpliciter):

Definition 4 (ordering of pairs supreme) *Let $s, t \in \mathfrak{S}$, where $s = \{s_1, s_2\}$ and $t = \{t_1, t_2\}$. We define $s \preccurlyeq t$ iff $\exists_{i,j \in \{1,2\}} s_i \leqslant t_j \wedge s_{\tilde{i}} \leqslant t_{\tilde{j}}$, where the tilde function means that $\tilde{n} = 1$ or 2 iff $n = 2$ or 1, resp. $s \prec t$ means that $s \preccurlyeq t$ but $s \neq t$.*

We need to persuade ourselves that $\preccurlyeq$ is a partial ordering.

[5] I learned of the idea to formulate the choice principle in terms of pairs of points rather than of choice points from Nuel Belnap in January 2010, who encouraged me to work it out.

Fact 5. *$\preccurlyeq$ is a reflexive, anti-symmetric, and transitive relation on $\mathfrak{S}(h_1, h_2)$.*

Proof Let $s, t, u \in \mathfrak{S}$, where $s = \{s_1, s_2\}$, $t = \{t_1, t_2\}$, and $u = \{u_1, u_2\}$. It is immediate to see that $s \preccurlyeq s$ (reflexivity). To prove anti-symmetry, let $s \preccurlyeq t$ and $t \preccurlyeq s$, which entails $s_i \leqslant t_j \wedge s_{\tilde{\imath}} \leqslant t_{\tilde{\jmath}}$ and $t_m \leqslant s_n \wedge t_{\tilde{m}} \leqslant s_{\tilde{n}}$, for some $i, j, m, n \in \{1, 2\}$. If $j = m$, then $s_i \leqslant t_j \leqslant s_n$, and since $s_i \leqslant s_n$ implies $s_i = s_n$, we get $s_i = t_j$. We also have $\tilde{\jmath} = \tilde{m}$, which implies, by a similar argument, that $s_{\tilde{\imath}} = t_{\tilde{\jmath}}$. Putting the two together, we get $\{s_1, s_2\} = \{t_1, t_2\}$. If $j \neq m$, then $\tilde{\jmath} = m$, so $s_{\tilde{\imath}} \leqslant t_m \leqslant s_n$, hence $s_{\tilde{\imath}} = s_n$ and then $t_m = s_n$. But also $j = \tilde{m}$, so $s_i \leqslant t_{\tilde{m}} \leqslant s_{\tilde{n}}$, and hence $t_{\tilde{m}} = s_{\tilde{n}}$. Thus $\{s_1, s_2\} = \{t_1, t_2\}$.

Turning to transitivity, let $s \preccurlyeq t$, $t \preccurlyeq u$, and these relations be witnessed by $s_i \leqslant t_j \wedge s_{\tilde{\imath}} \leqslant t_{\tilde{\jmath}}$ and $t_m \leqslant u_n \wedge t_{\tilde{m}} \leqslant u_{\tilde{n}}$, for some $i, j, m, n \in \{1, 2\}$. If $j = m$ (and hence $\tilde{\jmath} = \tilde{m}$), it follows that $s_i \leqslant t_j \leqslant u_n$ and also $s_{\tilde{\imath}} \leqslant t_{\tilde{\jmath}} \leqslant u_{\tilde{n}}$, whence $s \preccurlyeq u$. And, if $j \neq m$ (and hence $\tilde{\jmath} = m$ and $j = \tilde{m}$), we get $s_{\tilde{\imath}} \leqslant t_{\tilde{\jmath}} \leqslant u_n$, and $s_i \leqslant t_{\tilde{\jmath}} \leqslant u_{\tilde{n}}$, so $s_{\tilde{\imath}} \leqslant u_n$ and $s_i \leqslant u_{\tilde{n}}$, whence $s \preccurlyeq u$. □

We next use this ordering to define choice pairs for histories:

Definition 6 (choice pairs) *For $s_1 s_2 \in W$, $\{s_1, s_2\}$ is a choice pair for histories h_1, h_2, to be written as $\{s_1, s_2\} \in \mathfrak{C}(h_1, h_2)$,* iff *$\{s_1, s_2\}$ is a minimal element (wrt $\preccurlyeq$) in the set $\mathfrak{H}(h_1, h_2)$ of hot pairs for h_1 and h_2. We say that $\{s_1, s_2\}$ is a choice pair simpliciter iff there are histories h_1, h_2 such that $\{s_1, s_2\} \in \mathfrak{C}(h_1, h_2)$.*

Having the required notions, we now introduce a substitute for the prior choice principle of BST1992, and we will refer to it by PCP*:

Postulate 7 (PCP*). *If C is a chain in $\mathcal{W}$ and $C \subseteq h_1 \setminus h_2$ for some histories h_1, h_2, then there is a choice pair $\{s_1, s_2\} \in \mathfrak{C}(h_1, h_2)$ such that $s_1 \leqslant C$.*[6]

PCP* postulates choice pairs, where the old PCP postulated choice points. Observe that in contrast to PCP, we need the weak ordering in $s_1 \leqslant C$ above. If C is a one-element chain, i.e, $C = \{e\}$ for some $e \in W$, and $\{e, e'\}$ is a choice pair for h_1 and h_2, there is clearly no choice pair for h_1, h_2 strictly below $\{e, e'\}$.

In the rest of this section we will work with a modified version of BST1992, which results from the definition of models of BST1992, with PCP replaced by PCP*. We call this modified version: BST*1992.

Let us next define in BST*1992 the notions of dividedness and undividedness of histories:

Definition 8 (dividedness and undividedness) *Let $\{s, s'\}$ be a pair supreme (simpliciter). Then histories h_1 and h_2 divide at $\{s, s'\}$, $h_1 \perp_{ss'} h_2$,* iff *$\{s, s'\}$ is a choice pair for h_1, h_2, i.e., $\{s, s'\} \in \mathfrak{C}(h_1, h_2)$.*

[6] Where $s_1 \leqslant C$ means $\forall e \in C\, s_1 \leqslant e$.

Histories h_1 and h_2 are undivided at $\{s, s'\}$, $h_1 \equiv_{ss'} h_2$ iff *$s \in h_1 \cap h_2$ or $s' \in h_1 \cap h_2$ or $\{s, s'\}$ is a hot pair for h_1, h_2, but not a choice pair for h_1, h_2.*

The first line of the above definition decides a category of objects at which histories are divided or undivided: at pairs supreme simpliciter. Note an asymmetry, however: for two histories to be divided at a pair supreme, this pair supreme must be a choice pair for *these* histories. In contrast, two histories may be undivided at a pair supreme, which is *not* a pair supreme for these histories. Clearly, $\perp_{ss'}$ and $\perp_{s's}$ denote the same relation, and this is also true about $\equiv_{ss'}$ and $\equiv_{s's}$. To spell out the definition of $\equiv_{ss'}$, it says that two histories are undivided at a pair supreme $\{s, s'\}$ in exactly three cases: (1) $\{s, s'\}$ is not a pair supreme for these two histories, but one of its elements is shared by the two histories, or (2) $\{s, s'\}$ is a pair supreme for these histories, but not a hot pair for these histories, or (3) it is a hot pair but not a maximal hot pair for the two histories. In case (2), a pair supreme is of the form $\{s, s\}$, so $s \in h_1 \cap h_2$. Case (3) is interesting, as we will see it in a proof below. We prove that $\equiv_{ss'}$ is an equivalence relation on the set $H_{(s)} \cup H_{(s')}$ of histories containing s or s'.

Fact 9. *$\equiv_{ss'}$ is a (1) reflexive, (2) symmetric, and (3) transitive relation on $H_{(s)} \cup H_{(s')}$.*

Proof (1) Pick an $h \in H_{(s)} \cup H_{(s')}$ and assume $s \in h$. (The case with $s' \in h$ is symmetrical). Clearly, $s \in h \cap h$, so $h \equiv_{ss'} h$.

(2) Let $h_1 \equiv_{ss'} h_2$. If s or s' belong to $h_1 \cap h_2$, we immediately get $h_2 \equiv_{ss'} h_1$. Suppose thus that $\{s, s'\} \in \mathfrak{H}(h_1, h_2)$, but it is not a minimal element of $\mathfrak{H}(h_1, h_2)$. By the definitions of pairs supreme and hot pairs, $\{s, s'\} \in \mathfrak{H}(h_1, h_2)$ iff $\{s, s'\} \in \mathfrak{H}(h_2, h_1)$. Accordingly $\{s, s'\} \in \mathfrak{H}(h_2, h_1)$, but it is not a minimal element of $\mathfrak{H}(h_2, h_1)$, and hence $h_2 \equiv_{ss'} h_1$.

(3) For transitivity, let (†) $h_1 \equiv_{s_1 s_2} h_2$ and (‡) $h_2 \equiv_{s_1 s_2} h_3$, and assume the convention that for $i = 1, 2$, $\tilde{i} = 2, 1$, resp. The argument goes by cases, depending on which of the histories: h_1, h_2, h_3, s_i belongs to ($i = 1, 2$):

(a) $s_i \in h_1 \cap h_3$. Then $h_1 \equiv_{s_1 s_2} h_3$.

(b1) $s_i \in h_1 \setminus h_3$ and $s_i \in h_2$. Then by (‡) $s_{\tilde{i}} \in h_3$ and $\{s_1 s_2\} \in \mathfrak{H}(h_2, h_3) \setminus \mathfrak{C}(h_2, h_3)$. It follows that $s_1 \neq s_2$, so $\{s_1 s_2\} \in \mathfrak{H}(h_1, h_3)$. It also follows that there is $\{x_1 x_2\} \in \mathfrak{H}(h_2, h_3)$ such that $\{x_1, x_2\} \prec \{s_1, s_2\}$. Let $x_i < s_i$ and $x_{\tilde{i}} < s_{\tilde{i}}$ (case $x_i < s_{\tilde{i}}$ and $x_{\tilde{i}} < s_i$ is analogous). Since histories are downward closed, $x_i \in h_1$ and $x_{\tilde{i}} \in h_3$, and since $x_i \neq x_{\tilde{i}}$: $\{x_1 x_2\} \in \mathfrak{H}(h_1, h_3)$, so $\{s_1 s_2\} \in \mathfrak{H}(h_1, h_3) \setminus \mathfrak{C}(h_1, h_3)$, whence $h_1 \equiv_{s_1 s_2} h_3$.

(b2) $s_i \in h_1 \setminus h_3$ and $s_i \notin h_2$. By (‡), $s_{\tilde{i}} \in h_2 \cap h_3$. Hence by (†), $\{s_1 s_2\} \in \mathfrak{H}(h_1, h_2) \setminus \mathfrak{C}(h_1, h_2)$, so there is $\{x_1 x_2\} \in \mathfrak{H}(h_1, h_2)$ such that $\{x_1, x_2\} \prec \{s_1, s_2\}$. Let $x_i < s_i$ and $x_{\tilde{i}} < s_{\tilde{i}}$ (the case with $x_i < s_{\tilde{i}}$ and $x_{\tilde{i}} < s_i$ is analogous). Since histories being downward closed, $x_i \in h_1$ and $x_{\tilde{i}} \in h_3$, and since $x_i \neq x_{\tilde{i}}$, we get $\{x_1 x_2\} \in \mathfrak{H}(h_1, h_3)$, and hence $\{s_1 s_2\} \notin \mathfrak{C}(h_1, h_3)$. But since $s_1 \neq s_2$, $\{s_1 s_2\} \in \mathfrak{H}(h_1, h_3)$. Thus, $h_1 \equiv_{s_1 s_2} h_3$.

(c) $s_i \in h_3 \setminus h_1$. As in cases (b1) and (b2) above.

(d) $s_i \notin h_1 \cup h_3$. By (†) $s_{\tilde{i}} \in h_1$ and by (‡): $s_{\tilde{i}} \in h_3$, hence $h_1 \equiv_{s_1 s_2} h_3$. □

With the last result, we define elementary possibilities open at a pair supreme, which is analogous to a BST1992 notion of elementary possibilities open at a point event:

Definition 10 *Let $\{s, s'\}$ be a pair supreme (simpliciter). Then the set $H_{ss'}$ of elementary possibilities open at $\{s, s'\}$ is defined as the set of equivalence classes on $H_{(s)} \cup H_{(s')}$ with respect to the relation $\equiv_{(s),(s')}$ of undividedness at $\{s, s'\}$.*

We next argue that all the action lies at choice pairs, modally speaking:

Fact 11. *Only choice pairs have non-trivial sets of elementary open possibilities.*

Proof Let $\{s, s'\}$ be a pair supreme. If $s = s'$, i.e., $\{s, s'\}$ is not a hot pair, then for any pair $h, h' \in H_{(s)} \cup H_{(s')}$, $s \in h \cap h'$, and hence $h \equiv_{ss'} h'$.

If $s \neq s'$, then $\{s, s'\}$ is a hot pair; let us assume it is not a choice pair, however. Then for some $h, h' \in H_{(s)} \cup H_{(s')}$, there is $\{x, x'\} \in \mathfrak{H}(h, h')$ such that (†) $x < s, x' < s'$. Pick now arbitrary two histories $h_1, h_2 \in H_{(s)} \cup H_{(s')}$. If $h_1, h_2 \in H_{(s)}$ or $h_1, h_2 \in H_{(s')}$, we immediately obtain $h_1 \equiv_{ss'} h_2$. Suppose thus that $h_1 \in H_{(s)} \setminus H_{(s')}$ and $h_2 \in H_{(s')} \setminus H_{(s)}$ (the other case is analogous). Since histories are downward closed, (†) implies $x \in h_1$ and $x' \in h_2$. And, because $x \neq x'$, $\{x, x'\} \in \mathfrak{H}(h_1, h_2)$, which together with (†) entail $\{s, s'\} \in \mathfrak{H}(h_1, h_2) \setminus \mathfrak{C}(h_1, h_2)$. Whence $h_1 \equiv_{ss'} h_2$.

Finally, if $\{s, s'\}$ is a choice pair, there are histories $h, h' \in H_{(s)} \cup H_{(s')}$ such that $h \perp_{ss'} h'$; these two histories determine two elementary possibilities open at the pair. □

Our next fact says that hot pairs abounds:

Fact 12. *Let $\mathcal{W}$ have two histories h_1 and h_2. Let also t be a maximal chain in $\mathcal{W}$ such that $t' := t \cap h_1 \cap h_2 \neq \emptyset$ and $t \cap (h_1 \setminus h_2) \neq \emptyset$. Then (1) t' is upper bounded and (2) $\sup_{h_1}(t') \neq \sup_{h_2}(t')$.*

Proof (1) We claim that any (†) $e \in t'' := t \cap (h_1 \setminus h_2)$ upper bounds t'. Otherwise, since each element of t' and e are comparable, we would have $e < x$ for some $x \in t'$. Since $x \in h_1 \cap h_2$ and histories are downward closed, $e \in h_1 \cap h_2$, contradicting (†).

(2) The above result implies, via the axiom of history-relative suprema, that t' has history-relative suprema. Observe that $\sup_{h_1}(t') = \inf(t'')$. But $t'' \in h_1 \setminus h_2$, so by PCP*, there is (i) $\{s_1, s_2\} \in \mathfrak{C}(h_1, h_2)$ such that (ii) $s_i \leqslant t''$. Thus (iii) $s_i \leqslant \inf(t'') = \sup_{h_1}(t')$. Further, (ii) entails (iv) $s_i \in h_1$. Finally, it follows from (iii), (iv), and Fact 3 that $\sup_{h_1}(t') \notin h_2$, and hence $\sup_{h_1}(t') \neq \sup_{h_2}(t')$. □

Our last fact of this section says the following:

Fact 13. *(1) Every two histories of BST* 1992 overlap and (2) for every two histories, their overlap has no maximal element.*

Proof Ad. (1) For two histories h, h', there must be a chain $C \subseteq h \setminus h'$. By PCP*, there must be a choice pair s, s' for these two histories. By the definition of choice pairs and pairs supreme, there is a chain $C^* \subseteq h \cap h'$. Ad. (2) This is an immediate consequence of Fact 12 (2). □

The last two facts tell us that indeed the new version of BST1992 prescribes a different pattern of branching histories.

A still different pattern of branching is a consequence of a frugal branching framework I worked out with T. Kowalski (Kowalski and Placek 1999). This pattern consists in that every chain contained in the overlap of two histories has a maximum in the overlap. [7]

The upshot of this section is that BST is versatile: if physics tells us how alternative possible courses of events are different, we can modify BST accordingly.

3 How to Generalize BST1992?

In Sect. 1 we argued for a generalization of BST1992 that would accommodate the insights of GR. But how should we do that? We will join a "happy coincidence" as works in different areas point to a similar idea of defining a GR spacetime as a maximal subset of a generalized manifold with respect to being Hausdorff (and perhaps having some additional property as well).

A topology $\mathcal{T}(X)$ is called 'Hausdorff' if for every two distinct $x, y \in X$ there are two non-overlapping open sets containing x and y, respectively. Non-Hausdorff spacetimes were investigated in physics in the 1970s. Importantly, Hájíček (1971) proved the existence theorems for sub-manifolds maximal with respect to being Hausdorff and connected. Nevertheless, in later years a consensus emerged among physicists that a GR spacetime should be Hausdorff. This sentiment is embodied in the dramatic outcry of Penrose (1979, 595): "I must …return firmly to sanity by repeating to myself three times: 'spacetime is a Hausdorff differentiable manifold; spacetime is a Hausdorff …' ". [8] For a survey of the consequences of allowing for non-Hausdorff spacetimes, see Earman (2008).

In a similar spirit, building on Hájíček's results, Müller (2011) defines a history in his generalized BST as a subset of a base set maximal with respect to being Hausdorff and connected. Finally, there is the following result about a natural topology for BST1992, the so-called Bartha topology: given a natural assumption, a BST1992 history is a maximal Hausdorff and downward closed subset of a base set W (see Fact 57).

Thus, our target is to define a candidate for a GR spacetime as a subset of a base set of a generalized BST model maximal with respect to being Hausdorff.

Our second desiderata says that our generalization should be "topologically conservative" with respect to BST1992, that is, the resulting models and histories in these models should have similar topological properties as models and histories of

[7] Here I do not report on this framework any further, since it clashes with the central idea of this chapter that histories are to be identified with maximal subsets of a base set satisfying the Hausdorff property—see Sect. 5.2. The framework's pattern of branching implies that the Hausdorff property is satisfied on an entire base set, a consequence being that every model of this theory has a single generalized history.

[8] This is quoted by Earman (2008).

BST1992. What are then the topological facts about BST1992? BST1992 comes with a natural topology on the entire base set as well as with a natural topology on each history in the model. [9] Both kinds of the topologies are defined by the following condition, known as "the Bartha condition":

Definition 14 (the diamond topology) *Let $\mathcal{W} = \langle W, \leqslant \rangle$ be a BST1992 model and X stand either for W, or for a history h in $\mathcal{W}$.*

Z is an open subset of X, $Z \in \mathcal{T}(X)$, iff *$Z = X$ or for every $e \in Z$ and for every maximal chain t in X containing e there are $e_1, e_2 \in t$ such that $e_1 < e < e_2$ and $\{x \in W \mid e_1 \leqslant x \leqslant e_2\} \subseteq Z$.*

Main topological facts about $\mathcal{T}(W)$ and $\mathcal{T}(h)$, where h is a history in $\mathcal{W}$, are as follows:

1. $\mathcal{T}(h)$ is connected and (given some natural assumptions) Hausdorff [10];
2. $\mathcal{T}(h)$ is maximally Hausdorff in this sense: modulo some natural assumptions, the Bartha condition applied to any proper superset of h yields a non-Hausdorff topology (see Fact 57).
3. for some history h, $\mathcal{T}(h)$ is locally Euclidean, and for some other history h', $\mathcal{T}(h')$ is not locally Euclidean (see Fact 58).
4. $\mathcal{T}(W)$ is connected and non-Hausdorff (unless W contains one history only) [11];
5. $h \notin \mathcal{T}(W)$ (unless $h = W$)—see Placek et al. (2013).
6. $\mathcal{T}(W)$ is not locally Euclidean (unless $W = h$ for some history h and $\mathcal{T}(h)$ is locally Euclidean (see Fact 58)).

In what follows, we will construct a manifold topology on generalized BST, and, in an attempt to be conservative with respect to BST1992, we will see to it that the topology on a generalized history is Hausdorff, and moreover, maximally so. We will also secure that each generalized history is locally Euclidean. In contrast, we will initially allow that the topology on the whole model be not locally Euclidean and non-Hausdorff, and that a history is not open in this topology. In a sequel, we will face a dilemma, however. If we want to construct spaces of tangent vectors (which are needed for the GR equations to make sense), we need to impose a certain restriction on the generalized BST models. The restriction implies that a generalized BST model (as a whole) is locally Euclidean, and that generalized histories are open in the manifold topology. Thus, if we want to have tangent vectors spaces, our resulting construction is not conservative with respect to BST1992, after all.

[9] For an argument that these topologies are natural, see Placek et al. (2013).

[10] The "connected" part is the topic of Fact 53; for a proof of the "Hausdorff" part, see Placek et al. (2013).

[11] The "connected" part is the theme of Fact 54; for a proof of the "non-Hausdorff" part, see Placek et al. (2013).

4 Construction

Our construction proceeds in three steps: First, we will generalize BST1992, second we will construct a generalized differential manifold on a generalized BST model (at this stage we will equip BST models with a topology). Third, we will construct tangent vector spaces, needed for the formulation of GR equations. Our construction is not orthodox in the sense that, in contrast to GR, a base set for a (generalized) differential manifold has some structure: it is assumed to be pre-ordered (i.e., reflexive and transitive, but not necessarily anti-symmetric) and satisfy a few postulates.

4.1 BST Generalized

We take courage from the following theorem of GR.[12] For every event p in an arbitrary GR spacetime there exists a convex normal neighborhood of p, that is, an open set U with $p \in U$ such that for every $q, r \in U$ there is a unique geodesics connecting q and r, and staying entirely in U. Since geodesics fall into three classes, of time-like, space-like, and null-like geodesics, the uniqueness of connectability means that the geodesics can be used to define a partial ordering $\leqslant$ on U: $q \leqslant r$ iff q is connectible to r by a future directed time-like or null-like geodesics. A sufficiently small convex normal set can be charted on an open subset of $\mathbb{R}^n$. In the spirit of this theorem, we will construct a manifold topology such that every element of a base set W has an open neighborhood ("patch"), which is partially ordered. We further postulate that each patch is like a small BST1992 model. As a consequence, in contrast to GR patches, our patches may be modally inconsistent, i.e., containing objects that are not contained in a single spacetime. (So we really "take courage" from the above theorem, it is not a premise of our construction.) Without further ado, let us introduce some terminology and then turn to the definitions:

1. $MC(X)$ is the set of maximal chains in X, where X is a non-empty pre-ordered set;
2. $MC(X; e) = \{t \in MC(X) \mid e \in t\}$;
3. $t^{<x} = \{z \in t \mid z < x\}$, where $t \in MC(X)$ and $x \in X$; $t^{<x}$ is the initial segment of t below x ($t^{\leqslant x}$, $t^{>x}$, and $t^{\geqslant x}$ are similarly defined).

Definition 15 (generalized BST model) *Where $W \neq \emptyset$, $\preccurlyeq$ is a pre-order on W, and $\mathcal{O} \subseteq \mathcal{P}(W)$, a triple $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ is a generalized BST model (genBST model), iff for every $e \in W$ there is a set $\mathcal{O}_e \subseteq \mathcal{O}$ (of patches) around e such that for every $O \in \mathcal{O}_e$:*

1. *$e \in O$;*
2. *$\langle O, \preccurlyeq_{|O}\rangle$ is a nonempty dense partial order satisfying the following:*

   (a) $\forall e' \in O \forall t \in MC(W; e') \exists x, y \in t \cap O \ (x \prec_{|O} e' \prec_{|O} y \ \wedge \ t^{\succ x} \cap t^{\prec y} \subseteq O)$;

[12] See Wald (1984, Thm. 8.1.2).

(b) *every lower bounded chain in* $\langle O, \preccurlyeq_{|O}\rangle$ *has an infimum in* O*;*
(c) *if a chain* C *in* $\langle O, \preccurlyeq_{|O}\rangle$ *is upper bounded by* $b \in O$*, then* $B := \{x \in O \mid C \preccurlyeq_{|O} x \wedge x \preccurlyeq_{|O} b\}$ *has a unique minimum,*
(d) *if* $x, y \in O$ *and* $x \prec z \prec y$*, then* $z \in O$*; and*

3. $\bigcup_{e\in W} \mathcal{O}_e = \mathcal{O}$*;*
4. *If* $x, y \in O \cap O'$*, where* $O, O' \in \mathcal{O}$*, then* $x \preccurlyeq_{|O} y$ *iff* $x \preccurlyeq_{|O'} y$.[13]

Let us next put together some facts about patches:

Fact 16. (about patches). *Let* $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ *be a generalized BST model. Then:*
(i) a subset of O*, where* $O \in \mathcal{O}$*, does not necessarily belong to* $\mathcal{O}$*;*
(ii) the union of $O, O' \in \mathcal{O}$ *does not necessarily belong to* $\mathcal{O}$*, but*
(ii) if $O \cap O' \neq \emptyset$*, where* $O, O' \in \mathcal{O}$*, then* $O \cap O' \in \mathcal{O}$*.*

Proof (i) A subset of $O \in \mathcal{O}$ can fail to satisfy any of the conditions (2a)–(2d). (ii) The ordering $\preccurlyeq_{|O\cup O'}$ on the union of $O, O' \in \mathcal{O}$ may fail to be anti-symmetric; also (2d) can fail on $O \cup O'$. (iii) $\langle O \cap O', \preccurlyeq_{|O\cap O'}\rangle$ is a nonempty dense partial ordering because, by the assumption, $O \cap O' \neq \emptyset$ and each $\preccurlyeq_{|O}$ and $\preccurlyeq_{|O'}$ is a dense partial ordering. It is easy to check that $\langle O \cap O', \preccurlyeq_{|O\cap O'}\rangle$ satisfies (2a) and (2d). To argue for (2b), let C be a chain in $\langle O \cap O', \preccurlyeq_{|O\cap O'}\rangle$, lower bounded by $b \in O \cap O'$. Then C has $\inf_O(C)$ in O and $\inf_{O'}(C)$ in O'. Since $b \preccurlyeq_{|O'} \inf_{O'}(C) \preccurlyeq_{|O'} C$ and $b \preccurlyeq_{|O} \inf_O(C) \preccurlyeq_{|O} C$, by Definition 15 (2d) $\inf_O(C) \in O \cap O'$ and $\inf_{O'}(C) \in O \cap O'$. By the definition of infimum, $\inf_O(C) \preccurlyeq_{|O'} \inf_{O'}(C)$ and $\inf_{O'}(C) \preccurlyeq_{|O} \inf_O(C)$. By Definition 15 (4) $\inf_O(C) = \inf_{O'}(C) := \inf_{O\cap O'}(C)$. To prove (2c), suppose there is a chain $C \subseteq O \cap O'$ upper bounded by $b \in O \cap O'$. Then, by Definition 15 (2d) and (4) $\{x \in O \mid C \preccurlyeq_{|O} x \wedge x \preccurlyeq_{|O} b\}$ and $\{x \in O' \mid C \preccurlyeq_{|O'} x \wedge x \preccurlyeq_{|O'} b\}$ are identical. Thus, a unique minimal element of one must be identical to a unique minimal element of the other, and must belong to $O \cap O'$. □

Generalized BST models allow for causal loops in this sense: $x, y, z \in W$ with $x, y \in O, z \notin O$, $y, z \in O', x \notin O'$ and $x, z \in O''$, $y \notin O''$ and such that $x \preccurlyeq_{|O} y$, $y \preccurlyeq_{|O'} z$, and $z \preccurlyeq_{|O''} x$.

The idea of this chapter is that the Hausdorff property will decide whether a subset of W is contained in a spacetime, or not. We do not have a topology yet, so an appeal to Hausdorffness remains on an intuitive level, to be justified later, when we define a topology. But, in spacetime theories, a bifurcating path, whose trunk has no maximal element indicates a failure of the Hausdorff property. Minimal elements of two upper arms of such a structure will be called "splitting pair".

Definition 17 (splitting pairs) *Let* $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ *be a generalized BST model and* $O \in \mathcal{O}$*. We say that* $e, e' \in O$ *form a splitting pair in* O*,* $\{e, e'\} \in Y_O$*, iff* $e \neq e'$ *and there is a chain* C *in* $\langle O, \preccurlyeq_{|O}\rangle$ *and* $b, b' \in O$ *such that* $C \preccurlyeq_{|O} b$*,* $C \preccurlyeq_{|O} b'$ *and* $e = \min\{x \in O \mid C \preccurlyeq_{|O} x \wedge x \preccurlyeq_{|O} b\}$ *and* $e' = \min\{x \in O \mid C \preccurlyeq_{|O} x \wedge x \preccurlyeq_{|O} b'\}$.
We then define the set Y *of splitting pairs of* $\mathcal{W}$ *as* $Y := \bigcup_{O\in\mathcal{O}} Y_O$.

[13] $e \prec e'$ iff $e \preccurlyeq e'$ but $e \neq e'$.

One may wonder how global pre-ordering $\preccurlyeq$ mesh with splitting pairs. Our postulates do not exclude the following situation:

(*) Events $e, e' \in O$ have a common upper bound with respect to $\preccurlyeq$, but are above a splitting pair $\{x, x'\} \in Y_O$ in the sense that $x \preccurlyeq_{|O} e$ and $x' \preccurlyeq_{|O} e'$.

We would like to prohibit (*): events separated by a splitting pair cannot be connected by causal curves to an event in their (common) future, as they do not have a common future. This intuition goes back to our reading of a splitting pair as a seed of modal inconsistency. Hence this condition:

Condition 18 (Hausdorff separation) *If there is a pair* $\{x, x'\} \in Y$, *then* $\neg\exists z \in W$ $(x \preccurlyeq z \wedge x' \preccurlyeq z)$.

Note the interplay between local and global notions: if x and x' are separated by a splitting pair in some patch O, then x and x' have no common upper bound, no matter how far we go along $\preccurlyeq$, possibly outside O. We next define consistency:

Definition 19 (consistency) $e, e' \in W$ *are consistent* iff *there is no splitting pair* $\{x, x'\} \in Y$ *such that* $x \preccurlyeq e$ *and* $x' \preccurlyeq e'$. $A \subseteq W$ *is consistent iff* $\forall e, e' \in A$: *e and* e' *are consistent.*

Definition 20 (inconsistency) $e, e' \in W$ *are inconsistent* iff *there is a splitting pair* $\{x, x'\} \in Y$ *such* $x \preccurlyeq e \wedge x' \preccurlyeq e'$.

We claim next that there are maximal consistent subsets of W.

Lemma 21 *There is at least one maximal consistent subset of* W.

Proof The proof goes by the Zorn lemma. Observe first that for every $e \in W$, the singleton $\{e\}$ is a consistent set, since $x \preccurlyeq e$, $x' \preccurlyeq e$ and $\{x, x'\} \in Y$ contradict Condition 18. Consider then the set of consistent subsets of W, partially ordered by inclusion. To check if a premise of the Zorn lemma is satisfied, pick a chain $C = A_1, A_2, \ldots, A_\alpha, \ldots$ of consistent subsets of W. Let suppose $\bigcup C$ is not consistent. Then there must be $e, e' \in \bigcup C$ and x, x' such that $\{x, x'\} \in Y$ and $x \preccurlyeq e$ and $x' \preccurlyeq e'$. Thus, for some β, β': $e \in A_\beta$ and $e' \in A_{\beta'}$, where $A_\beta, A_{\beta'} \in C$. Since A_β and $A_{\beta'}$ are comparable by $\subseteq$, for $\beta^* = \max(\beta, \beta')$ we have $e, e' \in A_{\beta^*}$, and hence A_{β^*} is not consistent. Contradiction. □

What are the properties of maximal consistent subsets of W? The fact below list some of them:

Fact 22. (about maximal consistent subsets of W) *Let* A, A' *be maximally consistent subsets of* W, *where* W *is a base set of a gen BST model. Then:*

(1) A *is downward closed.*

(2) Let $e' \in A' \setminus A$. *Then there is a "hot pair"* $\{x, x'\}$ *for* A *and* A', *i.e., there is a a chain* $C \subseteq A \cap A'$, *such that* $x = \sup_A(C)$, $x' = sup_{A'}(C)$, $x \neq x'$, *and* $x' \preccurlyeq e'$.

(3) If $e, e', e^* \in W$ *and* $e \preccurlyeq e^*$ *and* $e' \preccurlyeq e^*$, *then there is a maximally consistent subset* A^* *of* W *such that* $e, e', e^* \in A^*$.

Proof (1) For a reductio, let us assume that A is not downward closed, which means that there are some $e, e' \in W$ such that (i) $e \prec e'$, (ii) $e' \in A$, but (iii) $e \notin A$. Since A is a maximal consistent subset, (iii) implies that e must be inconsistent with some $e^* \in A$, which means that there is a slitting pair $x, x^* \in W$ such that (iv) $x \preccurlyeq e$ and (v) $x^* \preccurlyeq e^*$. By (ii) e' is consistent with e^*, which taken with (v) implies (vi) $\neg(x \preccurlyeq e')$. But by (i) and (iv) we have $x \prec e'$, which contradicts (vi).

(2) Let e', A, and A' be as in the premise. Then e' is inconsistent with some $e \in A$, from which it follows that there is $O \in \mathcal{O}$ and a splitting pair $\{x, x'\} \in Y_O$ such that $x \preccurlyeq e$ and $x' \preccurlyeq e'$. By item (1) of this Fact, $x \in A$ and $x' \in A'$. By Definition 17 of splitting pairs, $x \neq x'$ and there is a chain C in $\langle O, \preccurlyeq_{|O}\rangle$ and $b, b' \in O$ such that $C \preccurlyeq_{|O} b$, $C \preccurlyeq_{|O} b'$ and (†) $x = \min\{y \in O \mid C \preccurlyeq_{|O} y \wedge y \preccurlyeq_{|O} b\}$ and $x' = \min\{y \in O \mid C \preccurlyeq_{|O} y \wedge y \preccurlyeq_{|O} b'\}$. Item (1) of this Fact entails that $C \subseteq A$ and $C \subseteq A'$. To prove that $x = \sup_A(C)$ we argue as follows. Consider the set U of upper bounds of C in A. By condition (2a) of Definition 15, (i) for every upper bound $u \in U$ of C there is $u' \in U \cap O$ such that $C \preccurlyeq_{|O} u' \preccurlyeq u$. (Just connect C with u by a maximal chain in W and apply (2a).) We may thus restrict our attention to the set U' of upper bounds of C in $O \cap A$. Since $U' \subseteq A$, U' is consistent, and hence there are no two upper-bound-relative minima of this kind: $z_1 = \min\{y \in O \mid C \preccurlyeq_{|O} y \wedge y \preccurlyeq_{|O} u_1\}$ and $z_2 = \min\{y \in O \mid C \preccurlyeq_{|O} y \wedge y \preccurlyeq_{|O} u_2\}$, where $u_1, u_2 \in U'$. Otherwise z_1 and z_2 would constitute a splitting pair below u_1 and u_2, respectively, yielding u_1 and u_2 inconsistent, which contradicts $u_1, u_2 \in A$. Thus, there is a unique minimum below (in the sense of $\preccurlyeq_{|O}$) all $u \in U'$, namely x, which, taken together with (i), proves that $x = sup_A(C)$. An argument that $x' = sup_{A'}(C)$ is analogous.

(3) By the Zorn lemma, there is a maximally consistent $A \subseteq W$ such that $e^* \in A$. By item (1) of this Fact, $e, e' \in A$. □

Fact 22 points out to a striking resemblance between histories of BST1992 and maximal consistent subsets of W of a generalized BST model. We take this resemblance to be a good enough justification for calling maximal consistent subsets of W "generalized histories" (or g-histories, for short).

Definition 23 (g-histories) *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model. We say that H is a generalized history (g-history) of $\mathcal{W}$* iff *H is a maximal consistent subset of W. We denote the set of g-histories by $gHist$.*

At this point one may wonder if g-histories extend to the future, as BST1992 histories do. Unfortunately, it is not excluded at this stage that a g-history has a maximal element. This situation will be ruled out, however, in the generalized BST models that admit a manifold structure—see Fact 23. A similar worry concerns PCP. We proved above that there is a hot pair for any two g-histories. A PCP-pair version, however, requires minimal hot pairs for any two histories; we do not know if the latter exist for g-histories.

As a next topic, let us ask what is an intersection of a g-history $H \subseteq W$ with a patch $O \in \mathcal{O}$? The answer is given by this fact:

Fact 24. *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model, H be a g-history of $\mathcal{W}$, and $O \in \mathcal{O}$. Then if $H \cap O \neq \emptyset$, $H \cap O$ is consistent and $\langle H \cap O, \preccurlyeq_{|H\cap O}\rangle$ is a nonempty partial order that satisfies conditions (2b)-(2d) of Definition 15.*

Proof It is left to the reader. □

Note that if a model allows for maximal elements in the intersections of histories, $O \cap H$ does not satisfy clause (2a) of Definition 15, and hence $O \cap H$ is not a patch. This might be a motivation for banning such maximal elements.[14] Observe also that every patch $O \in \mathcal{O}$ is divided between g-histories of $\mathcal{W}$, i.e $\forall x \in O\ \exists A \in gHist\ (x \in A)$. Of course, there might be an element of O shared by a few g-histories; there might also be g-history A and a patch O such that $A \cap O = \emptyset$.

The final question for this section is: does generalized BST extend BST1992 or BST*1992 of Sect. 2, i.e., is genBST worth its name? Since BST1992 and BST*1992 permit models with minimal elements, which generalized BST rules out, the latter does no generalize the former two, strictly speaking. Second, there is a discrepancy between histories of BST1992 and g-histories: the upper fork, extending indefinitely up and down, and with a maximal element in the trunk, is a two-history model of BST1992, but has only one g-history, as there is no splitting pair in it. Still, this fork is a model of generalized BST. Thus, we have the following, qualified, verdict concerning generalization (note that this result does not entail that histories and g-histories are to be identified):

Lemma 25 *Let $\langle W, \leqslant\rangle$ have no minimal element and be a model of either BST1992 or BST*1992. Then $\langle W, \leqslant, \{W\}\rangle$ is a model of generalized BST.*

Sketch of a proof Since a generalized BST model in question has only one patch, W itself, the axioms of BST1992/BST*1992 immediately imply that $\langle W, \leqslant_{|W}\rangle$ is nonempty dense partial order. The axiom of no maximal elements together with the premise of this lemma, no minimal elements, imply clause (2a) of Definition 15. Axioms of infima and history-relative suprema imply clauses (2b) and (2c) of this definition. The remaining clauses, that is, (1), (2d), (3), and (4) are trivially satisfied. □

4.2 Generalized Differential Manifolds and Matters Topological

The aim of this subsection is to set up a (generalized) differential manifold on the base set of a generalized BST model. This is the crux of the construction since, after all, GR spacetimes are differential manifolds of some kind. We do not imply that every generalized BST model can be equipped with the manifold structure—in the sequel we will consider only those that do.

This section generalizes an elegant construction of GR manifolds, due Geroch (1972) and Malament (2012), to modally inconsistent contexts. We will first define

[14] For what we think to be a more serious reason for this move, see Sect. 4.3.

n-dimensional generalized charts on $\mathcal{W}$, in short n-g-charts, and say what it means that such charts are compatible.

Definition 26 (n-g-chart) *An n-g-chart on a generalized BST model $\mathcal{W}=\langle W, \preccurlyeq, \mathcal{O}\rangle$, is a pair $\langle O, \varphi\rangle$, where $O \in \mathcal{O}$ is a patch in $\mathcal{W}$ and $\varphi : O \to \mathbb{R}^n$ satisfies, for every $H \in gHist$:*

If $O \cap H \neq \emptyset$, then

1. *$\varphi_{|O\cap H}$ is injective (i.e., one-to-one),*
2. *$\varphi[O \cap H]$ is an open subset of $\mathbb{R}^n$ (in the standard topology on $\mathbb{R}^n$), and*
3. *$\forall e, e' \in O \cap H$ $e \prec_{|O} e' \Leftrightarrow \varphi(e) <_M \varphi(e')$, where $<_M$ is a (strict) Minkowskian ordering.*

The generalization consists in restricting the chart function to a modally consistent context, that is, to $O \cap H$. Furthermore, the orthodox approach has no analogue of (3).

Definition 27 (compatibility of n-g-charts) *Two n-g-charts on an genBST model $\mathcal{W}$, $\langle O_1, \varphi_1\rangle$ and $\langle O_2, \varphi_2\rangle$, are called compatible iff for every $H \in gHist$ either $O_1 \cap O_2 \cap H = \emptyset$ or $O_1 \cap O_2 \cap H \neq \emptyset$ and these two conditions obtain:*

(1) $\varphi_i[O_1 \cap O_2 \cap H]$ ($i = 1, 2$) are open subsets of $\mathbb{R}^N$, and

(2) $\varphi_2\varphi_1^{-1} : \varphi_1[O_1 \cap O_2 \cap H] \to \mathbb{R}^n$ and $\varphi_1\varphi_2^{-1} : \varphi_2[O_1 \cap O_2 \cap H] \to \mathbb{R}^n$ are both smooth.

A function from $\mathbb{R}^n$ to $\mathbb{R}^n$ is called smooth if it has a continuous derivative of any order. The generalization (with respect to the Geroch-Malament approach) consists in our appeal to histories and considering intersections $O_1 \cap O_2 \cap H$ rather than intersections $O_1 \cap O_2$.[15]

It is easy to see that compatibility is reflexive and symmetric; for an argument that it is not transitive, adapt an argument of Malament (2012) p. 2 appropriately. Following the Geroch-Malament definition of n-dimensional manifold, I define next a smooth n-dimensional *generalized* manifold, n-g-manifold for short.

Definition 28 (n-g-manifold) *An n-g-manifold is a pair $\langle \mathcal{W}, \mathcal{C}\rangle$, where $\mathcal{W}=\langle W, \preccurlyeq, \mathcal{O}\rangle$ is a generalized BST model and $\mathcal{C}$ is a set of n-g-charts on $\mathcal{W}$ satisfying these conditions:*

(M1) Any two n-g-charts in $\mathcal{C}$ are compatible.
(M2) For every $p \in W$ there is $\langle O, \varphi\rangle \in \mathcal{C}$ such that $p \in O$.
(M3) $\mathcal{C}$ is maximal in the sense that every n-g-chart on $\mathcal{W}$ that is compatible with each n-g-chart in $\mathcal{C}$ belongs to $\mathcal{C}$.

The definition mimics Malament's definition, but it drops the requirement of the Hausdorff property. That a maximal collection of n-g-charts (in the sense of (M3))

[15] In their approach, the part beginning with "iff" reads: "iff either $O_1 \cap O_2 = \emptyset$ or if $O_1 \cap O_2 \neq \emptyset$, then (1) $\varphi_i[O_1 \cap O_2]$ ($i = 1, 2$) are open subsets of $\mathbb{R}^N$, and (2) $\varphi_2\varphi_1^{-1} : \varphi_1[O_1 \cap O_2] \to \mathbb{R}^n$ and $\varphi_1\varphi_2^{-1} : \varphi_2[O_1 \cap O_2] \to \mathbb{R}^n$ are both smooth.

exists, can be proved by the Zorn lemma. This would leave open the question of what n-g-manifold s look like. This worry is addressed by the following lemma that gives a simple recipe of how to build n-g-manifold s: find first a collection $\mathcal{C}_0$ of n-g-charts on W satisfying (M1) and (M2), and then add to it the set $\mathcal{C}_1$ of all n-g-charts on W that are compatible with every n-g-chart in $\mathcal{C}_0$.

Lemma 29 *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model and $\mathcal{C}_0$ be a set of n-g-charts on $\mathcal{W}$ satisfying conditions (M1) and (M2). Let $\mathcal{C}_1$ be the set of all n-g-charts on $\mathcal{W}$ that are compatible with every n-g-chart in $\mathcal{C}_0$. Then $\langle \mathcal{W}, \mathcal{C}_0 \cup \mathcal{C}_1\rangle$ is an n-g-manifold .*

Proof Since $\mathcal{C}_0$ satisfies (M2), so does $\mathcal{C}_0 \cup \mathcal{C}_1$. To verify (M1), we need to show that any $\langle O, \varphi\rangle, \langle O', \varphi'\rangle \in \mathcal{C}_1$ are compatible. Pick an arbitrary $H \in gHist$, and since $O \cap O' \cap H = \emptyset$ confirms compatibility of the two charts, assume $O \cap O' \cap H \neq \emptyset$.

We first show that $\varphi[O \cap O' \cap H]$ is open (an argument that $\varphi'[O \cap O' \cap H]$ is open is similar). Pick $p \in O \cap O' \cap H$, so $\varphi(p) \in \varphi[O \cap O' \cap H]$. By (M1) there is $\langle O^*, \varphi^*\rangle \in \mathcal{C}_0$ such that $p \in O^*$, hence $p \in O \cap O' \cap O^* \cap H$ and $\varphi(p) \in \varphi[O \cap O' \cap O^* \cap H]$. Since $\langle O, \varphi\rangle$ is compatible with $\langle O^*, \varphi^*\rangle$ and $\langle O', \varphi'\rangle$ is compatible with $\langle O^*, \varphi^*\rangle$, $\varphi^*[O \cap O^* \cap H]$ and $\varphi^*[O' \cap O^* \cap H]$ are open. Accordingly, their intersection is open, and since φ^* restricted to H is injective, $\varphi^*[O^* \cap O \cap H] \cap \varphi^*[O^* \cap O' \cap H] = \varphi^*[O^* \cap O \cap O' \cap H]$. Observe next that $\varphi[O^* \cap O \cap O' \cap H]$ is open because it is a pre-image of an open set $\varphi^*[O^* \cap O \cap O' \cap H]$ under a continuous (because smooth) map $\varphi^*\varphi^{-1} : \varphi[O \cap O^* \cap H] \to \mathbb{R}^n$. Thus, for any $p \in O \cap O' \cap H$, there is an open set $\varphi[O^* \cap O \cap O' \cap H] \subseteq \varphi[O \cap O' \cap H]$ such that $\varphi(p) \in \varphi[O^* \cap O \cap O' \cap H]$. Thus, $\varphi[O \cap O' \cap H]$ is open.

Second, we verify that (i) $\varphi\varphi'^{-1} : \varphi'[O \cap O' \cap H] \to \mathbb{R}^n$ and (ii) $\varphi'\varphi^{-1} : \varphi[O \cap O' \cap H] \to \mathbb{R}^n$ are smooth. To argue (i), note that for every $x \in \varphi'[O \cap O' \cap H]$, one can find $\langle O^*, \varphi^*\rangle \in \mathcal{C}_0$ such that $\varphi'^{-1}(x) \in O^*$. Then we re-write (i) as the composition $\varphi\varphi^{*\,-1} \circ \varphi^*\varphi'^{-1}$ of two smooth maps, $\varphi^*\varphi'^{-1} : \varphi'[O \cap O' \cap O^* \cap H] \to \varphi^*[O \cap O' \cap O^* \cap H]$ and $\varphi\varphi^{*\,-1} : \varphi^*[O \cap O' \cap O^* \cap H] \to \varphi[O \cap O' \cap O^* \cap H]$. Because a composition of smooth maps is smooth and domains and counter-domains match, the conclusion follows. The argument for (ii) is analogous.

Finally, to prove (M3), note that since a chart not in $\mathcal{C}_1$ must be incompatible with some chart in $\mathcal{C}_0$, $\mathcal{C}_1 \cup \mathcal{C}_0$ is maximal. □

Before we proceed to define topology on W by using n-g-charts, we establish an auxiliary fact:

Fact 30. *Let $\langle \mathcal{W}, \mathcal{C}\rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W}=\langle W, \preccurlyeq, \mathcal{O}\rangle$ and $\langle O, \varphi\rangle \in \mathcal{C}$. Then if $O' \in \mathcal{O}$ and $O' \subseteq O$, then $\langle O', \varphi_{|O'}\rangle \in \mathcal{C}$.*

Proof We need to show that, first, (†) $\langle O', \varphi_{|O'}\rangle$ is an n-g-chart and, second, that (‡) it is compatible with every chart in $\mathcal{C}$. As for (†), observe that a restriction of an injection is an injection. Note also that since φ preserves the ordering on $O \cap H$, it preserves the ordering on $O' \cap H$, for any $H \in gHist$ such that $O' \cap H \neq \emptyset$. It remains to show that $\varphi[O' \cap H]$ is open, if $O' \cap H \neq \emptyset$. Let us pick an arbitrary $\tilde{e} \in \varphi[O' \cap H]$. Our aim is to find an open set in $\varphi[O' \cap H]$ containing $\tilde{e}$. Let us

take a "vertical" maximal chain $\tilde{t} \in MC(\langle \varphi[O \cap H], <_M \rangle, \tilde{e})$[16] and transform it into $t = \varphi^{-1}(\tilde{t})$. Since φ is injective and order preserving on $O \cap H$, t is a maximal chain in $\langle O \cap H, \preccurlyeq_O \rangle$ and $\varphi^{-1}(\tilde{e}) := e \in t$. Recall that $O' \subseteq O$ is a patch as well, so by Definition 15 (2a), t must extend up and down e in O', that is, there are $x, y \in t \cap O'$ such that $x \prec_{|O'} e \prec_{|O'} y$ and $t^{\succ_{|O'}x} \cap t^{\prec_{|O'}y} \subseteq O'$. Since $t \subseteq H$, $t^{\succ_{|O'}x} \cap t^{\prec_{|O'}y} \subseteq O' \cap H$, moreover. Transforming $t^{\succ_{|O'}x} \cap t^{\prec_{|O'}y}$ to $\mathbb{R}^n$, we find $\tilde{t}^{>_M \tilde{x}} \cap \tilde{t}^{<_M \tilde{y}} = \varphi(t^{\succ_{|O'}x} \cap t^{\prec_{|O'}y}) \subseteq \varphi[O' \cap H]$, with $\tilde{x} = \varphi(x)$, $\tilde{y} = \varphi(y)$ such that $\tilde{x}, \tilde{y} \in \tilde{t}$ and $\tilde{x} <_M \tilde{e} <_M \tilde{y}$. Thus, there is a nonempty $\tilde{x}', \tilde{y}' \in \tilde{t}$ such that $\tilde{x} <_M \tilde{x}' <_M \tilde{e} <_M \tilde{y}' <_M \tilde{y}$. Accordingly, $\tilde{x}', \tilde{y}' \in \varphi[O' \cap H]$ and moreover the "diamond" $d = \{\tilde{z} \in \mathbb{R}^n \mid \tilde{x} \leqslant_M \tilde{z} <_M \tilde{y}\}$ contained in $\varphi[O' \cap H]$ (because $z = \varphi^{-1}(\tilde{z})$ is between x and y in $O' \cap H$, thanks to Definition 15 (2d) and histories being downward closed). By removing from d its borders in $\mathbb{R}^n$, we construct the borderless diamond b containing $\tilde{e}$ (because the diamond's vertices $\tilde{x}$ and $\tilde{y}$ belong to the vertical chain $\tilde{t}$ passing through $\tilde{e}$). Clearly, $b \subseteq \varphi[O' \cap H]$ and is open, and hence we proved that $\langle O', \varphi_{|O'} \rangle$ is a chart.

To prove (‡), i.e., compatibility of $\langle O', \varphi_{|O'} \rangle$ with any $\langle O^*, \psi^* \rangle \in \mathcal{C}$, it is enough to consider only such $\langle O^*, \psi^* \rangle$ and $H \in gHist$ that $O' \cap O^* \cap H \neq \emptyset$. As we just showed, $\varphi[O' \cap H]$ is open. Since $\langle O, \varphi \rangle$ is compatible with $\langle O^*, \psi^* \rangle$, $\varphi[O \cap O^* \cap H]$ is open. And $\varphi[O \cap O^* \cap H] \cap \varphi[O' \cap H] = \varphi[O' \cap O^* \cap H]$, because φ_H is an injection. Thus, $\varphi[O' \cap O^* \cap H] = \varphi_{|O'}[O' \cap O^* \cap H]$ is open.

Finally, since $\langle O, \varphi \rangle$ and $\langle O^*, \psi^* \rangle$ are compatible, $\psi^* \varphi^{-1} : \varphi[O \cap O^* \cap H] \to \mathbb{R}^n$ is smooth. And, as shown above, $\varphi_{|O'}[O' \cap O^* \cap H]$ is open. Thus, by making the required restrictions, we see that $\psi^* \varphi^{-1}_{|O'} : \varphi_{|O'}[O' \cap O^* \cap H] \to \mathbb{R}^n$ is smooth. An argument that $\varphi_{|O'} \psi^{*\,-1} : \psi^*[O' \cap O^* \cap H] \to \mathbb{R}^n$ is smooth is analogous. □

Since the intersection of two patches is a patch (Fact 16), the fact above has an immediate corollary, which will be needed to define a topology:

Corollary 31 *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold and $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$ be a generalized BST model. Then:*
if $\langle O, \varphi \rangle, \langle O', \varphi' \rangle \in \mathcal{C}$ and $O \cap O' \neq \emptyset$, then $\langle O \cap O', \varphi_{|O \cap O'} \rangle$ and $\langle O \cap O', \varphi'_{|O \cap O'} \rangle$ belong to $\mathcal{C}$ as well.

Definition 32 (g-manifold topology) *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$. We say that $S \subseteq W$ is open in the g-manifold topology, $S \in \mathcal{T}(W)$, iff*

$$\forall p \in S \; \exists \, \langle O, \varphi \rangle \in \mathcal{C} \; (p \in O \wedge O \subseteq S).$$

We need to check that this definition indeed defines a topology on W.

Fact 33. *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$. Then:*

[16] This means that only the time coordinate of $\tilde{t}$ changes.

(1) $\emptyset \in \mathcal{T}(W)$*;*
(2) $W \in \mathcal{T}(W)$*;*
(3) if $S, S' \in \mathcal{T}(W)$*, then* $S \cap S' \in \mathcal{T}(W)$ *as well;*
(4) if $S_1, S_2, \ldots, S_\alpha, \ldots \in \mathcal{T}(W)$*, then* $\bigcup S_\sigma \in \mathcal{T}(W)$.

Proof It is easy to see that (1) and (2) are true. To prove (3), let $p \in S \cap S'$; since S and S' are open, there are $\langle O, \varphi\rangle, \langle O', \varphi'\rangle \in \mathcal{C}$, such that $p \in O \wedge O \subseteq S$ and $p \in O' \wedge O' \subseteq S'$. Hence $O \cap O' \neq \emptyset$, so by Corollary 31, $\langle O \cap O', \varphi_{|O\cap O'}\rangle \in \mathcal{C}$. Since $p \in O \cap O'$ and $O \cap O' \subseteq S \cap S'$, $S \cap S'$ is open. To verify (4), let us pick $p \in \bigcup_\alpha S_\alpha$. Thus, for some β, $p \in S_\beta \in \mathcal{T}(W)$. Accordingly there is an n-g-chart$\langle O_\beta, \varphi_\beta\rangle$ such that $p \in O_\beta$ and $O_\beta \subseteq S_\beta \subseteq \bigcup_\alpha S_\alpha$. Thus, $\bigcup_\alpha S_\alpha \in \mathcal{T}(W)$. □

We next observe the following fact about a base for this topology.

Fact 34. *Let* $\langle \mathcal{W}, \mathcal{C}\rangle$ *be an n-g-manifold on a generalized BST model* $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$*. Then the base for topology* $T(W)$ *is* $\mathcal{B}_W := \{O \in \mathcal{O} \mid \langle O, \varphi\rangle \in \mathcal{C}$ *for some* $\varphi : O \to \mathbb{R}^n\}$.

Proof It is immediate to see that every element of $\mathbb{B}_W$ is open. From the definition, if $A \in \mathcal{T}(W)$, then $\forall p \in A\ \exists\ \langle O, \varphi\rangle \in \mathcal{C}\ (p \in O \wedge O \subseteq A)$, which implies that $\mathcal{B}_W$ is the basis of this topology. □

By equipping a generalized BST model with a manifold topology, we impose some new properties on histories, not derivable in generalized BST alone.

Fact 35. *Let* $\langle \mathcal{W}, \mathcal{C}\rangle$ *be an n-g-manifold on a generalized BST model* $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ *and H be a g-history in* $\mathcal{W}$*. Then H has no maximal elements.*

Proof Let e be a maximal element of H. There is $\langle O, \varphi\rangle \in \mathcal{C}$ such that $e \in O$. Then $\varphi[O \cap H]$ is an open subset of $\mathbb{R}^n$. Since $\varphi_{|O\cap H}$ respects the ordering, $\varphi(e)$ is a maximal element in $\varphi[O \cap H]$. But then $\varphi[O \cap H]$ is not open, and hence $\langle O, \varphi\rangle \notin \mathcal{C}$. Contradiction. □

At this stage we do not know if g-histories are open, or whether they satisfy PCP. However, as a consequence of the fact above, we have that the openness of g-histories rules out PCP, point-like version:

Lemma 36 *Let* $\langle \mathcal{W}, \mathcal{C}\rangle$ *be an n-g-manifold on a generalized BST model* $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ *and H be a g-history in* $\mathcal{W}$*. Then:*
$H \in \mathcal{T}(W)$ *iff for every* $H' \in gHist$ *there is no maximal element in* $H \cap H'$.

Proof To the right: For reductio, let $H \in \mathcal{T}(W)$ and (†) e^* be a maximal element of $H \cap H'$ for some $H' \in gHist$. Thus, for every $e \in H$, and hence for e^* as well, there is $\langle O, \varphi\rangle \in \mathcal{C}$ such that $e \in O$ and $O \subseteq H$. These last conditions imply that every maximal chain passing through e^* should have some nonempty segment above e^* contained in O, and hence in H. By the Fact above, however, e^* is not a maximal element of H'. Moreover, it is a maximal element in $H \cap H'$. Hence some nonempty chains above e^* are contained in H' rather than H, no matter how short these chains are. Contradiction.

To the left: We need to show that for every $e \in H$ there is $O \in \mathcal{T}(W)$ such that $O \subseteq H$. Let us pick an arbitrary $e \in H$. By Definition 28 there is $\langle O', \varphi \rangle \in \mathcal{C}$ such that $e \in O'$. We claim that the sought-for $O = O' \cap H$. Observe that (‡) if $O \in \mathcal{O}$, we would have by Fact 30 (since $O \subseteq O'$) that $O \in \mathcal{T}(W)$, as required. We thus need to prove that $O \in \mathcal{O}$, which amounts to checking if O satisfies clause (2) of Definition 15.

First, since $e \in O' \cap H$ and $\langle O', \preccurlyeq_{|O'} \rangle$ is a nonempty dense partial order, $\langle O, \preccurlyeq_{|O} \rangle$ is a nonempty dense partial order as well.

Second, we need to prove that for every $e' \in O$ and every $t \in MC(W; e')$ there are $x, y \in t \cap O$ such that $x \prec_{|O} e' \prec_{|O} y$ and $t^{\succ x} \cap t^{\prec y} \subseteq O$. Since $e' \in O' \in \mathcal{O}$, there are x', $y' \in t \cap O'$ such that $x' \prec_{|O'} e' \prec_{|O'} y'$ and (i) $t^{\succ x'} \cap t^{\prec y'} \subseteq O'$. Since histories are downward closed and $e' \in H$, (ii) $t^{\succ x'} \cap t^{\preccurlyeq e'} \subseteq H$. There must also exist $y'' \in t$ such that (iii) $e' \prec_{|O'} y'' \prec_{|O'} y'$ and $y'' \in H$ (hence (iv) $t^{\succ e'} \cap t^{\prec y''} \subseteq H$). Otherwise, for every $z \in t$ such that $e' \prec z$ we would have $z \notin H$. But since $z \in H'$ for some g-history, and hence $e' \in H'$, it would follow that e' is a maximal element in $H \cap H'$, contradicting the Lemma's premise. By (i), (ii), (iii), and (iv): $t^{\succ x'} \cap t^{\prec y''} \subseteq O' \cap H = O$.

Third, every lower bounded chain in $\langle O, \preccurlyeq_{|O} \rangle$ has an infimum in O because it is lower bounded in $\langle O', \preccurlyeq_{|O'} \rangle$, so it has infimum in O', and since histories are downward closed, this infimum is in H as well.

Forth, by a similar argument, if a chain C in $\langle O, \preccurlyeq_{|O} \rangle$ is upper bounded by $b \in O$, then $B := \{x \in O_e \mid C \preccurlyeq_{|O} x \wedge x \preccurlyeq_{|O} b\}$ has a unique minimum. For, since $b \in H$, every $x \preccurlyeq_{|O'} b$ is in H as well.

Finally, since histories are downward closed, if $x, y \in O$ and $x \preccurlyeq z \preccurlyeq y$, then $z \in O$.

These five observations prove that $O = O' \cap H \in \mathcal{O}$, and hence, by (‡), $O \in \mathcal{T}(W)$. Moreover, $e \in O$ and $O \subseteq H$. As this is true for an arbitrary $e \in H$, we showed that $H \in \mathcal{T}(W)$. □

4.2.1 The Hausdorff Property

Before we turn to a discussion of the Hausdorff property in the g-manifold topology defined above, it is helpful to establish an auxiliary fact:

Fact 37. *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$. Then for any $S \in \mathcal{T}(W)$, if $p \in S$, then for any maximal chain $t \in MC(W; p)$, there are $x, y \in t$, $x \prec p \prec y$ such that $t^{\succ x \wedge \prec y} \subseteq S$, where $t^{\succ x \wedge \prec y} := \{z \in t \mid x \prec z \prec y\}$.*

Proof Let $p \in S \in \mathcal{T}(W)$ and let $t \in MC(W; p)$ be an arbitrary maximal chain. There is thus a patch $O \in \mathcal{O}$ such that $p \in O$ and $O \subseteq S$. By Definition 15 (2a), there must be $x, y \in t$, $x \prec p \prec y$ such that $t^{\succ x \wedge \prec y} \subseteq O$. By Definition 15 (2d), for every $z \in t^{\succ x \wedge \prec y}$, $z \in O$ and since $O \subseteq S$, it follows that $t^{\succ x \wedge \prec y} \subseteq S$. □

Theorem 38 (no Hausdorff property) *Let $\langle \mathcal{W}, \mathcal{C}\rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ which has more than one g-history. Then the g-manifold topology on W does not satisfy the Hausdorff property.*

Proof Since $\mathcal{W}$ has more than one g-history, there must be some inconsistent $e, e' \in W$, which is equivalent to the existence of a splitting pair $\{x, x'\}$ such that $x \preccurlyeq e \wedge x' \preccurlyeq e'$. This means that $x \neq x'$ and there is a patch $O \in \mathcal{O}$ and a chain C in $\langle O, \preccurlyeq_{|O}\rangle$ and $b, b' \in O$ such that $C \preccurlyeq_{|O} b$, $C \preccurlyeq_{|O} b'$ and $x = \min\{z \in O \mid C \preccurlyeq_{|O} z \wedge z \preccurlyeq_{|O} b\}$ and $x' = \min\{z \in O \mid C \preccurlyeq_{|O} z \wedge z \preccurlyeq_{|O} b'\}$. Pick next arbitrary $U, U' \in \mathcal{T}(W)$ such that $x \in U$ and $x' \in U'$. Pick also $t \in MC(W; x)$ and $t' \in MC(W; x')$ such that $C \subseteq t \cap t'$ and $x \in t$ and $x' \in t'$. By Fact 37, there are $z \in t$, $z' \in t'$, $z \prec x$, $z' \prec x'$ such that $t^{\succ z \wedge \prec x} \subseteq U$ and $t'^{\succ z' \wedge \prec x'} \subseteq U'$. Accordingly, there is $z^* \in C$ such that $z \prec z^*$ and $z' \prec z^*$. It follows that $z^* \in t^{\succ z \wedge \prec x} \subseteq U$ and $z^* \in t'^{\succ z' \wedge \prec x'} \subseteq U'$, and hence $z^* \in U \cap U'$. Since U and U' are arbitrary, this proves that the Hausdorff property fails in the g-manifold topology on a model with more than one g-history.□

Having established that the topology on a genBST model with more than one g-history is non-Hausdorff, let us now ask if g-histories are Hausdorff. More precisely, we ask if the subspace topology $\mathcal{T}_{\subseteq W}(H)$ has the Hausdorff property, where H is a g-history and the ambient topology is $\mathcal{T}(W)$. To recall the concept of a subspace topology, given (ambient) topology $\mathcal{T}(W)$ and a nonempty subset $A \subseteq W$, the subspace topology on A is $\mathcal{T}_{\subseteq W}(A) = \{A \cap U \mid U \in \mathcal{T}(W)\}$. To proceed, we need an auxiliary fact and a definition, however.

Fact 39. *Let $e_1 \in O \in \mathcal{O}$ and $e_1, e_2 \in H \in gHist$ and suppose that $t^{\preccurlyeq e_2} \neq \emptyset$ and $t^{\preccurlyeq e_2} \preccurlyeq_{|O} e_1$ for some $t \in MC(W; e_1)$. Then $m \preccurlyeq e_2$, where $m = \min\{z \in O \mid t^{\preccurlyeq e_2} \preccurlyeq_{|O} z \wedge z \preccurlyeq_{|O} e_1\}$.*

Proof Clearly, $t^{\preccurlyeq e_2} \preccurlyeq e_2$. By Definition 15 (2c) there is $m' = \min\{z \in O \mid t^{\preccurlyeq e_2} \preccurlyeq_{|O} z \wedge z \preccurlyeq_{|O} e_2\}$. Clearly, $m' \preccurlyeq e_2$. By the same definition, there also exists $m = \min\{z \in O \mid t^{\preccurlyeq e_2} \preccurlyeq_{|O} z \wedge z \preccurlyeq_{|O} e_1\}$. If $m \neq m'$, then the two form a splitting pair and such that $m \preccurlyeq e_1$ and $m' \preccurlyeq e_2$, yielding e_1 and e_2 inconsistent, which contradicts $e_1, e_2 \in H$. Thus, $m = m' \preccurlyeq e_2$. □

Before the next definition, let us introduce some notation. For $e \in W$, we will write $(\succ_e) := \{e' \in W \mid e \prec e'\}$. Also, for $\tilde{x} \in \mathbb{R}^n$, $flc(\tilde{x})$ denote the set of points in $\mathbb{R}^n$ lying on the brim of the future light-cone of $\tilde{x}$.

Definition 40 *Let $\langle \mathcal{W}, \mathcal{C}\rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$, $\langle O, \varphi\rangle \in \mathcal{C}$, and $x \in O$. We define:*

$\bigtriangledown_O(x) := \bigcup_{H \in gHist}\{\varphi^{-1}[\varphi[O \cap H \cap (\succ_x)] \setminus flc(\varphi(x))] \mid O \cap H \neq \emptyset\}$

$\bigwedge(x) := \{z \in W \mid x \not\preccurlyeq z\}$ *and* $\bigwedge_O(x) := \bigwedge(x) \cap O$.

Fact 41. *Let $\langle \mathcal{W}, \mathcal{C}\rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$, $\langle O, \varphi\rangle \in \mathcal{C}$, and $x \in O$. Then*

(1) $\bigtriangledown_O(x) \in \mathcal{O}$ and (2) $\langle \bigtriangledown_O(x), \varphi_{\bigtriangledown_O(x)}\rangle \in \mathcal{C}$.

Moreover, if $\bigwedge_O(x) \neq \emptyset$, then (3) $\bigwedge_O(x) \in O$ and (4) $\langle \bigwedge_O(x), \varphi_{\bigwedge_O(x)}\rangle \in \mathcal{C}$.

Sketch of a proof The proof of (1) and (3) relies on the observation that the image of $\triangledown_O(x) \cap H$ by φ is the inside of the future light cone of $\varphi(x)$ and the image of $\blacktriangle_O(x) \cap H$ is the outside of the future light cone of $\varphi(x)$, where both these images are open in the standard topology on $\mathbb{R}^n$. The argument then relies on noting that properties analogous to those required by Definition 15 (2) obtain in the latter topology, and then transforming these properties, by $\varphi^{-1}_{|O\cap H}$, to generalized BST. Then (2) and (4) follow by Fact 30. □

Fact 42. *Let $\langle \mathcal{W}, \mathcal{C}\rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$, $e_1, e_2 \in H$, $H \in gHist$, and $e_1 \not\preccurlyeq e_2$. Then there is $O \in \mathcal{O}$ and $x \in O$ such that $e_1 \in \triangledown_O(x)$ and $x \not\preccurlyeq e_2$, hence $e_2 \in \blacktriangle(x)$.*

Proof Pick n-g-chart $\langle O, \varphi\rangle \in \mathcal{C}$ such that $e_1 \in O$ so (i) $e_1 \in O \cap H$. Accordingly, $\emptyset \neq \varphi[O \cap H]$ and is open in $\mathbb{R}^n$, so there is a "vertical" maximal chain $\tilde{t} \subseteq \langle \varphi[O \cap H], <_M\rangle$ that contains $\tilde{e}_1 = \varphi(e_1)$ and extends (at least slightly) below and above $\tilde{e}_1$. Clearly, $t = \varphi^{-1}_{|O\cap H}[\tilde{t}] \subseteq O \cap H$ and $e_1 \in t$. Consider next $t^{\succcurlyeq e_2}$. If it is empty, pick any $x \in t^{\succcurlyeq_{|O} e_1}$; then $x \preccurlyeq_{|O} e_1$ and $x \not\preccurlyeq e_2$.

But if $t^{\succcurlyeq e_2} \neq \emptyset$, then $t^{\succcurlyeq e_2}$ is upper bounded by e_1 (because $e_1 \in t$ and $e_1 \not\preccurlyeq e_2$), so by clause (2c) of Definition 15, there is $m = \min\{z \in O \mid t^{\succcurlyeq e_2} \preccurlyeq_{|O} z \wedge z \preccurlyeq_{|O} e_1\}$, so $m \preccurlyeq e_1$. By Fact 39, $m \preccurlyeq e_2$, so $m \neq e_1$, and hence $m \prec e_1$. Pick now $x \in t$ such that $m \prec x \prec e_1$. It follows that $x \not\preccurlyeq e_2$, because otherwise $x \in t^{\succcurlyeq e_2}$, so m would not be an upper bound of $t^{\succcurlyeq e_2}$. Thus, there is $x \in W$ such that (ii) $x \preccurlyeq_{|O} e_1$ and (iii) $x \not\preccurlyeq e_2$. Next, "verticality" of $\tilde{t}$ assures that $\tilde{e}_1 \notin flc(\tilde{x})$, where $\tilde{x} = \varphi(x)$, and this result together with (i) and (ii) implies $e_1 \in \triangledown_O(x)$. On the other hand, (iii) implies $e_2 \in \blacktriangle(x)$. □

Theorem 43 *Let $\langle \mathcal{W}, \mathcal{C}\rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ and $H \in gHist$ of $\mathcal{W}$. Then $\mathcal{T}_{\subseteq W}(H)$ is Hausdorff.*

Proof Let us take distinct $e_1, e_2 \in H$; either $e_1 \not\preccurlyeq e_2$, or $e_2 \not\preccurlyeq e_1$. Suppose the former is true (the latter is proved similarly). By Fact 42, there is $O_1 \in \mathcal{O}$ and $x \in O_1$ such that $e_1 \in \triangledown_{O_1}(x)$ and $e_2 \in \blacktriangle(x)$. Pick next $O_2 \in \mathcal{O}$ such that $e_2 \in O_2$. Accordingly, $e_2 \in \blacktriangle(x) \cap O_2 = \blacktriangle_{O_2}(x)$, so by Fact 41, $\langle \triangledown_{O_1}(x), \varphi_{\triangledown_{O_1}(x)}\rangle \in \mathcal{C}$ and $\langle \blacktriangle_{O_2}(x), \varphi_{\blacktriangle_{O_2}(x)}\rangle \in \mathcal{C}$. It follows that $\triangledown_{O_1}(x)$ and $\blacktriangle_{O_2}(x)$ are open in the manifold topology, yet, by the construction, (†) $\triangledown_{O_1}(x) \cap \blacktriangle_{O_2}(x) = \emptyset$. Moreover, $e_1 \in H \cap \triangledown_{O_1}(x) \in \mathcal{T}_{\subseteq W}(H)$ and $e_2 \in H \cap \blacktriangle_{O_2}(x) \in \mathcal{T}_{\subseteq W}(H)$, which together with (†) show that $\mathcal{T}_{\subseteq W}(H)$ is Hausdorff. □

The next topic of this section is maximality properties. It is a desirable goal that a g-history be not only Hausdorff, but maximally so. Similarly, it is desirable that every subset of base set W maximal with respect to the Hausdorff property be identical to some g-history. The facts below do not fully achieve this goal, as they refer to maximality with respect to the joint property: the Hausdorff property plus being downward closed. This structure is similar to Müller's (2011) maximality results, which refer to the conjunction: Hausdorff plus connectedness.

Let us begin with this observation:

Fact 44. *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$ and $\mathcal{T}(W)$ be its manifold topology. There is a subset of W that is maximal with respect to having the joint property of being Hausdorff and downward closed.*

Proof Left for the reader. Recall that a g-history is downward closed (Fact 22) and has the Hausdorff property (Theorem 43); then apply the Zorn lemma. □

Fact 45. *Let H be a g-history in a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$ and $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on $\mathcal{W}$. Then H is a maximal subset of W with respect to being Hausdorff and downward closed.*

Proof The Fact claims that a subspace topology on any subset $A \subseteq W$ such that $H \subsetneq A$ is either not Hausdorff or A is not downward closed. To prove it, we pick an arbitrary downward closed A such that $A \supsetneq H$ and show that it does not have the Hausdorff property. Since H is maximally consistent, there are $y' \in H$, $y \in A \setminus H$ such that y, y' are not consistent. Accordingly, there is a splitting pair $\langle x, x' \rangle \in Y$ such that $x \preccurlyeq y$ and $x' \preccurlyeq y'$. Since g-histories are downward closed and A is assumed to be downward closed, $x' \in H$ and $x \in A$, and hence $\{x, x'\} \subseteq A$. Accordingly, there is a chain $C \subseteq A$ (because A is downward closed) that has two subsets of upper bounds, with x and x' being their respective minima. Then every open set $U \in \mathcal{T}_{\subseteq W}(A)$ with $x \in U$ contains some nonempty upper segment $C^{\succ z}$ of C, and similarly, every open set $U' \in \mathcal{T}_{\subseteq W}(A)$ with $x' \in U'$ contains some nonempty upper segment $C^{\succ z'}$ of C. Thus, every intersection of such U and U' contains some nonempty segment $C^{\succ z^*}$, $z^* = \max\{z, z'\}$, which shows that the subspace topology $T_{\subseteq W}(A)$ is not Hausdorff. □

Note a striking similarity between the above fact and a property of BST1992 histories (see Fact 57). Next, we have a converse result:

Fact 46. *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$ and $\mathcal{T}(W)$ be its manifold topology. Then if A is a maximal subset of W with respect to being Hausdorff and downward closed, then $A \in gHist$.*

Proof Let us assume that A is as in the premise and, as a reductio hypothesis, that A is not a g-history. Accordingly, either (i) A is not maximally consistent, i.e., $A \subsetneq H$ for some g-history H, or (ii) A is not consistent. If (i), since H has a joint property of being Hausdorff and downward closed, A is not maximal with respect to this property, which contradicts the premise. Turning to (ii), there is a splitting pair $\{x, x'\}$ below some two elements of A, which is generated by some chain C. Since A is assumed to be downward closed, $x, x' \in A$ and $C \subseteq A$. By an argument analogous to that in the last proof, $\mathcal{T}_{\subseteq W}(A)$ is not Hausdorff, which contradicts the Fact's premise. □

4.2.2 The Local Euclidean Property

Let us recall the concept of a locally Euclidean topological space. A topological space is called locally Euclidean if there is $n \in \mathbb{N}$ such that every element of the space has an open neighborhood homeomorphic to an open set of $\mathbb{R}^n$ (in the standard topology of reals). The (standard) definition of differential manifold requires its topology to be locally Euclidean. We should thus learn if our manifold topology $\mathcal{T}(W)$ and the subspace topologies $\mathcal{T}_{\subseteq W}(H)$, where H is a g-history, are locally Euclidean.

Lemma 47 *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$ and $H \in gHist$. Then the subspace topology $\mathcal{T}_{\subseteq W}(H)$ is locally Euclidean.*

Proof We need to show that every $e \in H$ has an open neighborhood $A \in \mathcal{T}_{\subseteq W}(H)$, $e \in A$ such that A is homeomorphic to B, where B is an open subset of $\mathbb{R}^n$. By Definition 28, there is $\langle O, \varphi \rangle \in \mathcal{C}$ such that $e \in O$ and $\varphi[O \cap H] = B$, where B is an open subset of $\mathbb{R}^n$ and $\varphi_{|O \cap H} : O \cap H \to B$ is an injection. By Definition 32, $O \in \mathcal{T}(W)$, so $O \cap H \in \mathcal{T}_{\subseteq W}(H)$. Putting $A = O \cap H$, we need to show that $\varphi : A \to B$ is a homeomorphism.

First, consider an open set $B' \subseteq B$ and ask if $\varphi^{-1}[B']$ is open? Take an arbitrary $e' \in \varphi^{-1}[B']$; then $\tilde{e}' = \varphi(e') \in B'$. Since B' is open, there is a borderless diamond $bd^{\tilde{x}\tilde{y}} \subseteq B'$ such that $\tilde{e}' \in bd^{\tilde{x}\tilde{y}}$. We put next $bd^{xy} := \varphi^{-1}[bd^{\tilde{x}\tilde{y}}]$. Clearly, $bd^{xy} \subseteq \varphi^{-1}[B'] \subseteq A$ and $x = \varphi^{-1}(\tilde{x})$, and $y = \varphi^{-1}(\tilde{y})$. Since φ respects the ordering, bd^{xy} is a borderless diamond in $\langle A, \preccurlyeq_{|O} \rangle$. We next define:

$$z \in O' \text{ iff } z \in O \wedge (z \in H \to z \in bd^{xy}) \wedge (z \notin H \to \exists z' \in bd^{xy} \wedge z' \preccurlyeq_{|O} z)$$

It can be shown (but we leave the proof to the reader) that $O' \in \mathcal{O}$. Then, since $O' \subseteq O$, Fact 30 implies that $O' \in \mathcal{T}(W)$, from which we get $O' \cap H \in \mathcal{T}_{\subseteq W}(H)$. Since $O' \cap H = bd^{xy}$, it follows that $e' \in bd^{xy} \in \mathcal{T}_{\subseteq W}(H)$ and $bd^{xy} \subseteq \varphi^{-1}[B']$. Since this is true about every $e' \in \varphi^{-1}[B']$, we get that $\varphi^{-1}[B'] \in \mathcal{T}_{\subseteq W}(H)$.

Second, pick an arbitrary set $A' \subseteq A$, $A' \in \mathcal{T}_{\subseteq W}(H)$ and ask if $\varphi[A']$ is open. The premise means that $A' = A'' \cap H$ for some $A'' \in \mathcal{T}(W)$. Accordingly, $A'' = \bigcup b_\alpha$, where b_α are elements of the base for $\mathcal{T}(W)$—see Fact 34. Thus, $\varphi[A'] = \varphi[\bigcup(b_\alpha \cap H)]$ which is equal to $\bigcup \varphi[(b_\alpha \cap H)]$ (because φ restricted to A is injective). Since b_α's are domains of the chart maps (see the same Fact), $\varphi[(b_\alpha \cap H)]$ are open subsets of $\mathbb{R}^n$, and hence $\bigcup \varphi[(b_\alpha \cap H)] = \varphi[A']$ is open as well. □

Lemma 48 *Let $\langle \mathcal{W}, \mathcal{C} \rangle$ be an n-g-manifold on a generalized BST model $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O} \rangle$. Then topology $\mathcal{T}(W)$ is not locally Euclidean, if there are g-histories H^1, H^2 in $\mathcal{W}$ whose intersection $H^1 \cap H^2$ has a maximal element.*

Sketch of a proof Let e be a maximal element in $H^1 \cap H^2$ and assume, as a reductio hypothesis, that $\mathcal{T}(W)$ is locally Euclidean. Then there is some b—an element of the base for $\mathcal{T}(W)$ such that $e \in b$, and a homeomorphism $\psi : b \to B$, where B is an open subset of $\mathbb{R}^m$. Clearly, $B \setminus \{\psi(e)\}$ is an open subset of $\mathbb{R}^m$, and hence (since ψ is a homeomorphism), $b \setminus \{e\} \in \mathcal{T}(W)$. Again, since ψ is a homeomorphism, it

preserves a number of maximal connected components. $B \setminus \{\psi(e)\}$ has two maximal connected components if $m = 1$ and one maximal connected component if $m > 1$ (see Munkres 2000, p.165). We have a contradiction since $b \setminus \{e\}$ has at least three maximal connected components [17]: the trunk $\bigtriangledown_b(e) = \{z \in b \mid e \not\prec_{|b} z\}$ and two "rimless futures" $\bigtriangledown_b^1$ and $\bigtriangledown_b^2$ of e, defined as follows (for $i = 1, 2$):

$$\bigtriangledown_b^i = \bigcup_{H^* \in gHist} \{\psi^{-1}[\psi[\{z \in b \cap H^* \mid e \prec_{|b} z\}] \setminus flc(\psi(e)] \mid H^* \approx H^i\},$$

where $H^* \approx H^i$ iff $\exists e' \, (e' \in H^i \cap H^* \wedge e \prec_b e')$,

and $flc(\tilde{x})$ is the set of points in $\mathbb{R}^n$ that lie on the rim of the future light-cone of $\tilde{x}$. □

Lemma 36 and the lemma above show the price that is to be paid for allowing that the intersection of two g-histories has a maximal element (or for assuming PCP, point-like version): g-histories are not open in the topology $\mathcal{T}(W)$ and this topology is not locally Euclidean.

4.2.3 Two Further Postulates

To ensure some desirable topological or differentiability properties in a manifold topology, we need two additional postulates:

Postulate 49 *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model. Then for every g-history H of $\mathcal{W}$ there are no $O_1, O_2 \in \mathcal{O}$ such that $O_1 \cap H \neq \emptyset$, $O_2 \cap H \neq \emptyset$ and $(O_1 \cup O_2) \cap H = H$*

Postulate 50 *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model. Then $\mathcal{O}$ contains a countable sub-cover $\mathcal{O}^*$ of W, i.e., $\mathcal{O}^* \subseteq \mathcal{O}$ and is countable, and $\forall e \in W \exists O \in \mathcal{O}^* \; e \in O$.*

The first postulate ensures that our topologies $\mathcal{T}_{\subseteq W}(H)$ are connected. The second postulate is needed for the existence of affine connections.

4.3 Tangent Vectors

Although we have already constructed a generalized (non-Hausdorff) manifold, whose subsets maximal with respect to being Hausdorff and downward closed are very much like spacetimes of general relativity, we need to equip it with even more structure. GR equations are tensor equations, and tensors need vector spaces to operate. Accordingly, in GR one associates to each element e of a manifold a vector space

[17] It has more if e is a maximal element of the intersection of some other histories, not merely of H^1 and H^2.

of vectors tangent at that point e. We thus need to add vector spaces to our generalized manifolds. That is, for each $e \in W$, where $\mathcal{W}$ is a generalized BST model that admits an n-g-manifold $\langle W, \mathcal{C}\rangle$ (and possibly satisfies Postulates 49 and 50), we will construct the space $V(e)$ of tangent vectors at e.

To recall the GR construction, one begins with the set $S(e) : O \to \mathbb{R}$ of smooth maps, where O is some open set containing e, or another. Since O is generally not a subset of $\mathbb{R}^m$, the concept of smoothness needs an explanation:

A function α from an open set O to $\mathbb{R}$ is said to be smooth iff for every chart $\langle O', \varphi\rangle \in \mathcal{C}$ such that $O \cap O' \neq \emptyset$, $\alpha\varphi^{-1} : \mathbb{R}^n \to \mathbb{R}$ has derivatives of an arbitrary order and is continuous. Finally, a vector in $V(e)$ is defined as a map from $S(e)$ to $\mathbb{R}$ that satisfies some three conditions.[18]

A red light should already blink at this junction since, in the present framework, a chart function φ is not necessarily injective, which makes φ^{-1} undefined. However, each chart function φ of $\langle O, \varphi\rangle \in \mathbb{C}$ is injective if restricted to any g-history H such that $H \cap O \neq \emptyset$. A natural remedy thus is to require that O occurring in the definition of set $S(e)$ should be contained in a g-history.[19] With this remedy, $V(e)$ will not depend on g-histories. Also, if e and e' belong to one g-history, the vector spaces $V(e)$ and $V(e')$ are to be connected in exactly the same way as in GR, that is, by the parallel transport. Finally, if e and e' do not share a g-history, no connection between $V(e)$ and $V(e')$ is postulated.

Unfortunately, the remedy is not going to work if the intersection of some two g-histories H and H' in $\mathcal{W}$ has a maximal element, say m. Each open set in $\mathcal{T}(W)$ containing m must extend upward along every path passing through m, and hence must contain some elements of $H \setminus H'$ as well as some elements of $H' \setminus H$.

We thus are driven to outright prohibit maximal elements in intersections of g-histories by imposing the following postulate on generalized BST models:

Postulate 51 *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model. Then:*

$$\forall e \in W\ \exists H \in gHist\ \exists O \in \mathcal{O}\ (e \in O \wedge O \subseteq H).$$

Postulate 51 has the following consequence:

Fact 52. *Let $\mathcal{W} = \langle W, \preccurlyeq, \mathcal{O}\rangle$ be a generalized BST model that satisfies Postulate 51. Then*

(1) there are no two g-histories in $\mathcal{W}$ whose intersection has a maximal element;
(2) $\forall e \in W\ \exists O \in \mathcal{T}(W)\ \exists H \in gHist\ (e \in O \wedge O \subseteq H)$

Proof A proof of (1) is immediate. As for (2), observe that for every $e \in W$ there is a chart (†) $\langle O', \varphi\rangle \in \mathcal{C}$ such that $e \in O'$ (by Definition 28) and an $O'' \in \mathcal{O}$ such that $(e \in O'' \wedge O'' \subseteq H)$ (by Postulate 51). By Fact 34, (†) implies $O' \in \mathcal{O}$, hence

[18] If $\zeta \in V(e)$, it should satisfy, for arbitrary functions $f_1, f_2 \in S(e)$: (i) $\zeta(f_1+f_2) = \zeta(f_1)+\zeta(f_2)$, (ii) $\zeta(f_1 f_2) = f_1(e)\zeta(f_1) + f_2(e)\zeta(f_1)$ and (iii) if f_1 is constant, $\zeta(f_1) = 0$.

[19] A modified definition will read $S(e) : O \to \mathbb{R}$ is the set of of smooth maps, where O is some open set containing e and $O \subseteq H$ for some g-history H.

$O := O' \cap O'' \in \mathcal{O}$. Since $O \subseteq O'$, by (†) Facts 30 and 34, $O \in \mathcal{T}(W)$. Moreover, $e \in O$ since $e \in O'$ and $e \in O''$ and $O \subseteq H$ since $O \subseteq O'' \subseteq H$. □

Postulate 51 permits a sought-for modification of the construction of tangent vector spaces. The set $S(e)$ is now defined as a set of smooth maps from some $O \in \mathcal{T}(W)$ to $\mathbb{R}$, where O is any open set containing e and contained in some g-history. A vector in $V(e)$ is defined as before, as a map from $S(e)$ to $\mathbb{R}$ that satisfies the three conditions listed in the Footnote 18 above.

Postulate 51 comes at a price: generalized BST does not generalize BST1992 (though it generalizes BST*1992—in the sense of Lemma 25). Nevertheless, the bonuses outweigh the cost: The Postulate assures that there are tangent vector spaces (as required by GR), that g-histories are open in the topology $\mathcal{T}(W)$ (see Lemma 36), and that $\mathcal{T}(W)$ is locally Euclidean (see Lemma 48 and Postulate 51(1)).[20]

5 Discussion

In this sections we address two issues that look troublesome for the generalized BST.

5.1 Hájíček-Müller Quasi-History

Following Hájíček (1971), Müller (2011) discusses an odd subset of a branching model. His tentative definition (which he amends accordingly) takes a history to be a subset of a base set that is maximal with respect to the joint property of being open, connected, and Hausdorff. The subset mentioned above satisfies this definition, but appears to be modally inconsistent (intuitively speaking). The branching model M is the union of two 2-dimensional Minkowski spacetimes M_1 and M_2, each with Minkowskian ordering, and pasted below and in the wings of the origin point $\bar{0} = \langle 0, 0\rangle$, so that the differences of the two Minkowskian spacetimes are the following:

$$M_1 \setminus M_2 = J^+(\bar{0}) \times \{1\}, \quad M_2 \setminus M_1 = J^+(\bar{0}) \times \{2\},$$

where $J^+(\bar{0}) = \{\langle t, x\rangle \mid \bar{0} \leqslant_M \langle t, x\rangle\}$. That is, M_1 and M_2 share neither the point of origin nor its future light cone.

To construct the troublesome subset A of $M_1 \cup M_2$, we subtract from the latter the "left" part of J_1 and the "right" part of J_2, that is,

$$A := M \setminus (J_l \times \{1\} \cup J_r \times \{2\}),$$

[20] It further allows for a simplification of our definitions of charts and of compatibility of charts, Definitions 26 and 27.

where $J_l := \{\langle t, x\rangle \in J^+(\bar{0}) \mid x \leqslant_M 0\}$ and $J_r := \{\langle t, x\rangle \in J^+(\bar{0}) \mid x \geqslant_M 0\}$. Note that A contains no choice pairs, as the "doubled rim" (including $(\bar{0}, 1)$ and $(\bar{0}, 2)$) has been removed from A. For an argument that A is Hausdorff as well as open and connected, see Müller (2011).

From the perspective of the present framework, M with the usual ordering and a single patch, namely M itself, is a model of genBST. However, A turns out to be inconsistent, the witness being any pair $e_1, e_2 \in A$ such that $e_1 \in (J^+ \setminus J_l) \times \{1\}$ and $e_2 \in (J^+ \setminus J_r) \times \{2\}$. Clearly, e_1 is above $(\bar{0}, 1)$ and e_2 is above $(\bar{0}, 2)$, and $(\bar{0}, 1)$, $(\bar{0}, 1)$ constitute a splitting pair. Thus, A is not a g-history (recall that g-history = maximal consistent subset of a base set). This diagnosis agrees with the verdict delivered by Müller's (2011) final definition of histories, which additionally requires, for each subset $C \subseteq h$ of history h that if $\partial C \neq \emptyset$, then $h \cap \partial C \neq \emptyset$ as well.

5.2 Borders in the Overlap

I have already warned against a branching model $\mathcal{W}$ that has more than one maximal upward directed subset (i.e., a BST1992 history) and in which every upper bounded chain has a supremum.[21] Figuratively, in $\mathcal{W}$ the border of the overlap of two BST1992 histories is contained in the overlap. Since a model of this kind does not contain any splitting pair in the sense of Definition 17, from the perspective of the generalized BST $\mathcal{W}$ has a single g-history only, namely, the model itself. As we will now argue, this implies that no generalized manifold in the sense of Definition 28 can be constructed on $\mathcal{W}$. As a reductio hypothesis, let us assume that there is g-manifold constructed on $\mathcal{W}$. Since $\mathcal{W}$ has one g-history only, by Lemma 47 the manifold topology $\mathcal{T}(W)$ must be locally Euclidean. Since upper bounded chains in $\mathcal{W}$ are assumed to have suprema, any nonempty intersection $t \cap h_1 \cap h_2$ of a maximal chain in $\mathcal{W}$ and upward directed subsets h_1, h_2 of $\mathcal{W}$ has a maximal element e'. By an argument analogous to that given in the proof of Lemma 48, e' does not have an open neighborhood homeomorphic to an open subset of $\mathbb{R}^n$ for any natural number n, which contradicts local Euclidicity.

The moral of this argument is that a generalized manifold cannot be constructed on a genBST model that has more than one maximal upper directed subset and in which every upper bounded chain has a supremum.

6 Conclusions

We have developed in this chapter a branching theory that captures the insights of general relativity. To pave the way towards this construction, in Sect. 2 we modified BST1992 by replacing its Prior Choice Principle (stated in terms of maximal points)

[21] Some years ago Tomasz Kowalski and I advocated such a theory, see Kowalski and Placek (1999).

with a pair-like version of this principle. As a consequence, the intersection of any two histories has no maximal element in the resulting theory (termed BST*1992). The construction of the branching theory then proceeded in three stages. In Sect. 4.1 we defined generalized BST models, the underlying idea being that locally, that is, around any element of a base set, the model is similar to BST1992, although the base set is not necessarily partially ordered. Generalized histories are defined as maximally consistent subsets of a base set, where consistency is spelled out in terms of splitting points. In the second stage, in Sect. 4.2 we defined generalized non-Hausdorff manifolds on generalized BST models. The main result of this section is that a generalized history (aka spacetime) turns out to be a subset of a manifold's base set that is maximal with respect to being Hausdorff and downward closed. And, vice versa, every subset of a manifold's base set maximal with respect to being Hausdorff and downward closed is identical to some generalized history. Two postulates (49 and 50) of this section ensure that the manifold topology on a generalized history is connected and that it has a countable sub-cover. We can thus identify a generalized history with a single GR spacetime, and a generalized BST model with a bundle of GR spacetimes. In the third stage (Sect. 4.3), in order to define tangent vector spaces on a generalized history, we had to assume Postulate 51, which comes with significant consequences. First, it prohibits maximal elements in the intersections of generalized histories, making generalized histories similar to histories of BST*1992 rather than to histories of BST1992. On a positive side, it implies that a generalized BST model is (as a whole) locally Euclidean and that a generalized history is open in the manifold topology. We wrapped up this chapter with a discussion (Sect. 5) of two potentially troublesome issues: we showed that the present framework delivers an intuitively adequate verdict concerning an odd structure discussed by Müller (2011) and we argued that generalized manifold cannot be constructed on the branching models advocated by Kowalski and Placek (1999).

Appendix
Topological Facts About BST1992

Let $\mathcal{W} = \langle W, \leqslant \rangle$ be a BST1992 model. To simplify the proofs below, we introduce the concept of "diamond oriented by maximal chain t with vertices e_1 and e_2", to be written as ${d_t}^{e_1 e_2}$:

$$ {d_t}^{e_1 e_2} := \{ y \in W \mid e_1 < e_2 \wedge e_1 \leqslant y \leqslant e_2 \}, $$

where t is a maximal chain in W and $e_1, e_2 \in t$.

Fact 53. *The Bartha topology $\mathcal{T}(h)$ on a history h in a BST1992 model is connected.*

Proof We need to show that the only subsets of history h that are both closed and open, are $\emptyset$ and h itself. To assume to the contrary is to assume that there are open nonempty subsets $A \subsetneq h$ and $B = h \setminus A$. Consider thus $x \in A$ and $y \in B$. Since histories are upward directed there is in h an upper bound z of x and y, and either (i) $z \in A$, or to (ii) $z \in B$. If (i), we consider a maximal chain $t \in MC(h)$ such that $y, z \in t$. (If (ii), consider a maximal chain $t' \in MC(h)$ such that $x, z \in t'$.) By the BST axiom of infima and maximality of t, there is in t an infimum $f = \inf(t \cap A)$. (Analogously, there is in t' an infimum $f' = \inf(t' \cap B)$.) If $f \in A$, then there is no diamond containing f and oriented by t that is a subset of A, so A is not open. But also, if $f \in B = h \setminus A$, then there is no diamond containing f, oriented by t, and a subset of B, so B is not open. We similarly arrive at a contradiction if we ask whether f' is in A, or not. □

Fact 54. *The Bartha topology $\mathcal{T}(W)$ is connected.*

Proof Note that in the proof above, to show that $\mathcal{T}(h)$ is connected, we used a maximal chain $t \in MC(h)$ that intersects both A and $h \setminus A$. Now, if we only know that there is at least one $t \in MC(W)$ that intersects $A \subseteq W$ and $B := W \setminus A$, where each A and B is open and nonempty, we could use the same trick as above to prove that $\mathcal{T}(W)$ is connected. Thus, let us assume for an arbitrary pair of A, B of the sort described above that (†) $\forall t \in MC(W)\ t \subseteq A \vee t \subseteq B$. Let us then pick some $t \subseteq A$ nonempty $t \subseteq A$ (the case with $t \subseteq B$ proceeds analogously). Clearly, for some history h, $t \in MC(h)$. Suppose now that (i) there is some $x \in h \cap B$. Then we pick some $y \in t$, produce an upper bound z of x and y. If $z \in A$, there is a maximal chain containing z and x, and if $z \in B$ there is a maximal chain containing z and y, where each of these chains intersects A and B—this contradicts (†). Let us thus suppose that (ii) $h \cap B = \emptyset$, which entails $h \subseteq A$. Then, for any $x \in B$, we must have $x \notin h$, but $x \in h'$ for some history h'. By PCP, there is a choice point c such that $c < x$ and $h \perp_c h'$. It follows that any maximal chain containing c and x intersects with A and B since $x \in B$ and $c \in h \subseteq A$, which again contradicts (†). □

Fact 55. *For every $A \subseteq W$, the Bartha condition applied to A yields topology $\mathcal{T}(A)$.*

Proof Rearrange Facts 8 and 9 of Placek et al. (2013) by replacing h by A. □

Our next fact appeals to continuous branching, which is defined as below:

Definition 56 (continuous branching surface) *Histories h and h' branch along a continuous branching surface* iff *there is $x \in h \setminus h'$ such that for every chain $t \in h \cap h'$ upper bounded by x:* $\sup_h(t) = \sup_{h'}(t)$.

Note that $x \in h \setminus h'$ entails (by PCP) that there is some $x' \in h \cap h'$ and below x, which in turn ensures that some chains containing x pass through this intersection.

Fact 57. *Let A be a proper superset of some history h of W (i.e., $h \subsetneq A$). Let also A be downward closed and there is no continuous branching surface for any two histories in W. Then $\mathcal{T}(A)$ does not satisfy the Hausdorff property.*

Proof Let (i) $h \subsetneq A$. Pick some $x \in A \setminus h$; hence $x \in h'$ for some $h' \in Hist$. Since h and h' do not branch along a continuous branching surface, there is a chain (ii) $t^* \subseteq h \cap h'$ such that (iii) $t^* < x$ and $\sup_h(t^*) \neq \sup_{h'}(t^*)$. By (iii) and downward closure of A, $\sup_h(t^*), \sup_{h'}(t^*) \in A$. Consider then an arbitrary pair of open sets $O, O' \in \mathcal{T}(A)$ containing $s = \sup_h(t^*)$ and $s' = \sup_{h'}(t^*)$, respectively. This means that for every pair of maximal chains t, t' such that $s \in t, s' \in t'$ and $y < s < z$, $y' < s' < z'$, there are oriented diamonds $d_t^{yz} \subseteq O$ and $d_{t'}^{y'z'} \subseteq O'$. Picking t and t' such that $t^* \subseteq t \cap t'$, we obtain that $\max\{y, y'\} \in d_t^{yz} \cap d_{t'}^{y'z'} \neq \emptyset$. Accordingly, any $O, O' \in \mathcal{T}(A)$ containing s, s', respectively, must overlap.

Lemma 58 *(1) There are BST histories such that $\mathcal{T}(h)$ is not locally Euclidean (in the Bartha topology).*

(2) $\mathcal{T}(W)$ is not locally Euclidean (unless $W = h$ for some history h);

(3) There are BST models such that, for every history h of such a model, $\mathcal{T}(h)$ is locally Euclidean (again, in the Bartha topology).

Proof As an example for (1), consider a downward fork, with its upper arm having a minimal element —this a one-history BST model. For reductio, suppose there is homeomorphism f between a neighborhood u of the vertex e and an open ball $b \subseteq \mathbb{R}^n$, for some $n \in \mathbb{N}$. Clearly, $b \setminus \{f(e)\}$ is open in standard topology on $\mathbb{R}^n$, so $u \setminus e$ must be open in the Bartha topology. However, $u \setminus \{e\}$ has three connected components (two lower arms and the top arm), whereas $b \setminus \{f(e)\}$ has two if $n = 1$, or one (itself) if $n > 1$. Thus, f cannot be a homeomorphism.[22]

As for (2), the above construction shows that any W containing a choice point (that is, having more than one history) is not locally Euclidean;

For (3), take a history in a Minkowskian Branching Structure[23]—it is locally (and globally) Euclidean since it is isomorphic to $\mathbb{R}^n$. □

References

Belnap, N. 1992. Branching space-time. *Synthese*, 92:385–434. Postprint archived at PhilSci Archive, http://philsci-archive.pitt.edu/archive/00001003

Earman, J. 2008. Pruning some branches from branching spacetimes. In *The ontology of spacetime II*, ed. D. Dieks, 187–206. Amsterdam: Elsevier.

Geroch, R. 1972. Differential geometry. http://home.uchicago.edu/geroch/

Hájíček, P. 1971. Causality in non-Hausdorff space-times. *Communications in Mathematical Physics* 21: 75–84.

Kowalski, T., and T. Placek. 1999. Outcomes in branching space-time and GHZ-Bell theorems. *British Journal for the Philosophy of Science* 50: 349–375.

[22] See Munkres (2000, p.150, 159). Some other examples of locally non-Euclidean histories involve a dimension change, like Müller's (2005) history, one part of which is homeomorphic to a half-line, and the other part — to the half-plane.

[23] For a theory of Minkowskian Branching Structures, see Placek and Belnap (2012).

Malament, D. 2012. *Topics in the foundations of general relativity and Newtonian gravitation theory*. Chicago: University of Chicago Press.
Müller, T. 2005. Probability theory and causation: A branching space-times analysis. *British Journal for the Philosophy of Science* 56(3): 487–520.
Müller, T. 2007. Branch dependence in the 'consistent histories' approach to quantum mechanics. *Foundations of Physics* 37(2): 253–276.
Müller, T. 2011. Branching space-times, general relativity, the Hausdorff property, and modal consistency. Pittsburgh PhilSci archive. http://philsci-archive.pitt.edu/8577/.
Müller, T. 2013. *Alternatives to histories? Employing a local notion of modal consistency in branching theories* (Forthcoming in Erkenntnis). doi:10.1007/s10670-013-9453-4.
Munkres, J.R. 2000. *Topology*. London: Prentice-Hall.
Penrose, R. 1979. Singularities and time-asymmetry. In *General relativity: An Einstein centenary survey*, ed. S. Hawking, and W. Israel, 581–638. Cambridge: Cambridge University Press.
Placek, T. 2010. On propensity-frequentist models for stochastic phenomena with applications to Bell's theorem. In *The analytical way*, ed. T. Czarnecki, K. Kijania-Placek, O. Poller, and J. Woleński, 105–144. London: College Publications.
Placek, T., and N. Belnap. 2012. Indeterminism is a modal notion: Branching spacetimes and Earman's pruning. *Synthese* 187(2): 441–469. doi:10.1007/s11229-010-9846-8.
Placek, T., N. Belnap, and K. Kishida. 2013. On topological issues of indeterminism. *Forthcoming in Erkenntnis*. doi:10.1007/s10670-013-9455-2.
Wald, R.M. 1984. *General relativity*. Chicago: University of Chicago Press.

Some Examples Formulated in a 'Seeing to It That' Logic: Illustrations, Observations, Problems

Marek Sergot

Abstract The chapter presents a series of small examples and discusses how they might be formulated in a 'seeing to it that' logic. The aim is to identify some of the strengths and weaknesses of this approach to the treatment of action. The examples have a very simple temporal structure. An element of indeterminism is introduced by uncertainty in the environment and by the actions of other agents. The formalism chosen combines a logic of agency with a transition-based account of action: the semantical framework is a labelled transition system extended with a component that picks out the contribution of a particular agent in a given transition. Although this is not a species of the *stit* logics associated with Nuel Belnap and colleagues, it does have many features in common. Most of the points that arise apply equally to *stit* logics. They are, in summary: whether explicit names for actions can be avoided, the need for weaker forms of responsibility or 'bringing it about' than are captured by *stit* and similar logics, some common patterns in which one agent's actions constrain or determine the actions of another, and some comments on the effects that level of detail, or 'granularity', of a representation can have on the properties we wish to examine.

1 Introduction

Logics of 'seeing to it that' or 'bringing it about that' have a long tradition in the analytical study of agency, ability, and action. The best known examples are perhaps the *stit* ('seeing to it that') family associated with Nuel Belnap and colleagues. (See e.g. Belnap and Perloff 1988; Horty and Belnap 1995; Horty 2001; Belnap et al. 2001 and some of the other chapters in this volume). Segerberg (1992) provides a summary

M. Sergot (✉)
Department of Computing, Imperial College London, 180 Queen's Gate,
London SW7 2BZ, UK
e-mail: m.sergot@imperial.ac.uk

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_11,

of early work in this area, and Hilpinen (1997) an overview of the main semantical devices that have been used, in *stit* and other approaches. With some exceptions, notably (Pörn 1977), the semantics is based on a branching-time structure of some kind.

In recent years logics of this kind have also been attracting attention in computer science. They have been seen as a potentially valuable tool in the formal modelling of agent interaction (human or artificial), in distributed computer systems and in the field of multi-agent systems. Works in this area have tended to be quite technical, focussing on various extensions, usually to the *stit* framework, or on connections to other formalisms used in computer science. There are however very few examples to my knowledge of any actual applications and so the usefulness of these formalisms in practice remains something of an open question. Forms of *stit* and Pörn's 'brings it about' have also been used as a kind of semi-formal device in representation languages for regulations and norms and in discussions of the logical form of normative and legal constructs.

In this chapter I want to look at a series of simple examples and how they might be formulated in a *stit*-like logic. An element of indeterminism is introduced by the environment—in some examples it may be raining, in others a fragile object might or might not break when it falls—and by the actions of other agents. The aim is, first, to explore something of the expressive power of this framework. An important feature of *stit* is that actions themselves are never referred to explicitly. The semantics abstracts away these details. *stit* thereby sidesteps what remains one of the most contentious questions in the philosophy of action, which is the question of what is action itself. If a man raises his arm, the arm goes up. But what is the *action* of raising the arm? Opinions are divided on this point. In *stit*, actions are not referred to directly and do not have to be named. On the other hand, there is sometimes a price to be paid for this abstraction since it is difficult to do without names for actions in all circumstances. Some of the examples are intended to explore this question. Second, I want to comment on some common patterns that arise, particularly when one agent's actions constrain, or possibly even determine, the actions of another. Relying on informal readings of these patterns can be misleading. And third, I want to identify some of the limitations and inadequacies of the framework as a representational device. These concern the treatment of causality, and questions regarding the effects of granularity, or level of detail, of a representation. I am making no claims of completeness. The treatment of temporal features is rudimentary, I will not touch on topics such as voluntary, deliberative, intentional, purposeful action, and even in these simple examples there are many issues that will not be addressed.

I will not formulate the examples in any form of *stit*-logic exactly, but using a different formalism (Sergot 2008a, b) that nevertheless has much in common. It combines a logic of 'brings it about' with a transition-based account of action: the semantical framework is a form of labelled transition system extended with an extra component that picks out the contribution—intentional, deliberative but perhaps also unwitting—of a particular agent in a given transition. Although the development was influenced by the constructions used in (Pörn 1977), it turned out (unexpectedly) to have much greater similarity with *stit*. Indeed, as explained later, it can be seen as

a special case of the deliberative *stit*, with a different informal reading and some additional features. Although some aspects of the representations will be specific to the use of my preferred formalism, nearly all the points I want to make will apply equally to *stit*-logics.

2 Syntax and Semantics

2.1 Preliminaries: Transition Systems

Transition systems A labelled transition system (LTS) is usually defined as a structure $\langle S, A, R\rangle$ where

- S is a (non-empty) set of *states*;
- A is a set of *transition labels*, also called *events*;
- R is a (non-empty) set of labelled *transitions*, $R \subseteq S \times A \times S$.

When (s, ε, s') is a transition in R, s is the initial state and s' is the resulting state, or end state, of the transition. ε is *executable* in a state s when there is a transition (s, ε, s') in R, and *non-deterministic* in s when there are transitions (s, ε, s') and (s, ε, s'') in R with $s' \neq s''$. A *path* or *run* of length m of the labelled transition system $\langle S, A, R\rangle$ is a sequence $s_0\, \varepsilon_0\, s_1 \cdots s_{m-1}\, \varepsilon_{m-1}\, s_m \quad (m \geq 0)$ such that $(s_{i-1}, \varepsilon_{i-1}, s_i) \in R$ for $i \in 1 \ldots m$. Some authors prefer to deal with structures $\langle S, \{R_a\}_{a\in A}\rangle$ where each R_a is a binary relation on S.

It is helpful in what follows to take a slightly more general and abstract view of transition systems. A transition system is a structure $\langle S, R, \mathit{prev}, \mathit{post}\rangle$ where

- S and R are disjoint, non-empty sets of *states* and *transitions* respectively;
- *prev* and *post* are functions from R to S: $\mathit{prev}(\tau)$ denotes the initial state of a transition τ, and $\mathit{post}(\tau)$ its resulting state.

A *path* or *run* of length m of the transition system $\langle S, R, \mathit{prev}, \mathit{post}\rangle$ is a sequence $\tau_1 \cdots \tau_{m-1}\, \tau_m \quad (m \geq 0)$ such that $\tau_i \in R$ for every $i \in 1 \ldots m$, and $\mathit{post}(\tau_i) = \mathit{prev}(\tau_{i+1})$ for every $i \in 1 \ldots m-1$.

Two-sorted language Given a labelled transition system, it is usual to define a language of propositional atoms or 'state variables' in order to express properties of states. We employ a *two-sorted* language. We have a set $\mathcal{P}_\mathrm{f}$ of propositional atoms for expressing properties of states, and a disjoint set $\mathcal{P}_\mathrm{a}$ of propositional atoms for expressing properties of transitions. Models are structures

$$\mathcal{M} = \langle S, R, \mathit{prev}, \mathit{post}, h^\mathrm{f}, h^\mathrm{a}\rangle$$

where h^f is a valuation function for atomic propositions $\mathcal{P}_\mathrm{f}$ in states S and h^a is a valuation function for atomic propositions $\mathcal{P}_\mathrm{a}$ in transitions R.

Transition atoms are used to represent events and attributes of events, and properties of transitions as a whole. For example, atoms $x{:}move{=}l$ and $x{:}move{=}r$ might be used to represent that agent x moves in direction l and r, respectively. The atom $falls(vase)$ might be used to represent transitions in which the object *vase* falls. Transition atoms are also used to express properties of a transition as whole: for instance, whether it is desirable or undesirable, timely or untimely, permitted or not permitted, and so on. So, for example, the formula

$$a{:}lifts \wedge \neg\, b{:}lifts \wedge c{:}move{=}l \wedge \neg\, d{:}move{=}l \wedge falls(vase) \wedge trans{=}red$$

might represent an event in which a lifts its end of the table and b does not while c moves in direction l, d does not move in direction l, and the vase falls. The atom $trans{=}red$ might represent that this event is illegal (say), or undesirable, or not permitted.

When a transition satisfies a transition formula φ we say it is a transition of type φ. So, for example, all transitions of type $a{:}lifts \wedge \neg b{:}lifts$ are also transitions of type $a{:}lifts$, and also transitions of type $\neg b{:}lifts$.

Formulas We extend this two-sorted propositional language with (modal) operators for converting state formulas to transition formulas, and transition formulas to state formulas.

Formulas are *state formulas* and *transition formulas*. State formulas are:

$$F ::= \text{ any atom } p \text{ of } \mathcal{P}_{\mathrm{f}} \mid \neg F \mid F \wedge F \mid \underset{\rightarrow}{\Box}\varphi$$

where φ is any transition formula. Transition formulas are

$$\varphi ::= \text{ any atom } \alpha \text{ of } \mathcal{P}_{\mathrm{a}} \mid \neg\varphi \mid \varphi \wedge \varphi \mid 0{:}F \mid 1{:}F$$

where F is any state formula.

We have the usual truth-functional abbreviations. $\underset{\rightarrow}{\Diamond}$ is the dual of $\underset{\rightarrow}{\Box}$: $\underset{\rightarrow}{\Diamond}\varphi =_{\mathrm{def}} \neg\underset{\rightarrow}{\Box}\neg\varphi$.

Semantics Models are structures

$$\mathcal{M} = \langle S, R, prev, post, h^{\mathrm{f}}, h^{\mathrm{a}}\rangle$$

where h^{f} and h^{a} are the valuation functions for state atoms and transition atoms respectively. Truth-functional connectives have the usual interpretations. The satisfaction definitions for the other operators are as follows, for any state formula F and any transition formula φ.

State formulas:

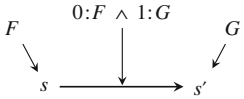
$$\mathcal{M}, s \models \underset{\rightarrow}{\Box}\varphi \quad \text{iff} \quad \mathcal{M}, \tau \models \varphi \text{ for every } \tau \in \mathrm{R} \text{ such that } prev(\tau) = \mathrm{s}$$

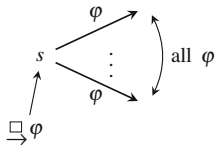
$\underset{\rightarrow}{\Box}\varphi$ is true at a state s when every transition from state s satisfies φ. $\underset{\rightarrow}{\Diamond}\varphi$ says that there is a transition of type φ from the current state.

Transition formulas:

$$\mathcal{M}, \tau \models 0{:}F \quad \text{iff} \quad \mathcal{M}, \mathit{prev}(\tau) \models F$$
$$\mathcal{M}, \tau \models 1{:}F \quad \text{iff} \quad \mathcal{M}, \mathit{post}(\tau) \models F$$

A transition is of type $0{:}F$ when its initial state satisfies the state formula F, and of type $1{:}F$ when its resulting state satisfies F.





As usual, we say a state formula F is *valid* in a model $\mathcal{M}$, written $\mathcal{M} \models F$, when $\mathcal{M}, s \models F$ for every state s in S, and a transition formula φ is *valid* in a model $\mathcal{M}$, written $\mathcal{M} \models \varphi$, when $\mathcal{M}, \tau \models \varphi$ for every transition τ in R. A formula is *valid* if it is valid in every model (written $\models F$ and $\models \varphi$, respectively).

We use the following notation for 'truth sets':

$$\|F\|^{\mathcal{M}} =_{\text{def}} \{s \in S \mid \mathcal{M}, s \models F\}; \quad \|\varphi\|^{\mathcal{M}} =_{\text{def}} \{\tau \in R \mid \mathcal{M}, \tau \models \varphi\}.$$

$\mathcal{M}$ is omitted when it is obvious from context.

Examples: transition formulas The following represents a transition from a state where (state atom) p holds to a state where it does not:

$$0{:}p \wedge 1{:}\neg p$$

von Wright (1963) uses the notation $p\,\mathrm{T}\,q$ to represent a transition from a state where p holds to one where q holds. It would be expressed here in the more general notation as the transition formula:

$$0{:}p \wedge 1{:}q$$

Let the state atom *on-table*(*vase*) represent that a certain vase is on the table. A transition of type $0{:}\mathit{on\text{-}table}(\mathit{vase}) \wedge 1{:}\neg \mathit{on\text{-}table}(\mathit{vase})$, equivalently, of type $0{:}\mathit{on\text{-}table}(\mathit{vase}) \wedge \neg 1{:}\mathit{on\text{-}table}(\mathit{vase})$ is one from a state in which the vase is on the table to one in which it is not on the table. Let the transition atom *falls*(*vase*) represent the falling of the vase from the table. Any model $\mathcal{M}$ modelling this system will have the property:

$$\mathcal{M} \models \mathit{falls}(\mathit{vase}) \rightarrow (0{:}\mathit{on\text{-}table}(\mathit{vase}) \wedge 1{:}\neg \mathit{on\text{-}table}(\mathit{vase}))$$

There may be other ways that the vase can get from the table to the ground. Some agent might move it, for example. That would also be a transition of type $0{:}\mathit{on\text{-}table}(\mathit{vase}) \wedge 1{:}\neg\mathit{on\text{-}table}(\mathit{vase})$ but not a transition of type $\mathit{falls}(\mathit{vase})$.

The operators $0{:}$ and $1{:}$ are not normal in the usual sense because formulas F and $0{:}F$ (and $1{:}F$) are of different sorts. However, they behave like normal operators in the sense that, for all $n \geq 0$, if $F_1 \wedge \cdots \wedge F_n \rightarrow F$ is valid then so are $0{:}F_1 \wedge \cdots \wedge 0{:}F_n \rightarrow 0{:}F$ and $1{:}F_1 \wedge \cdots \wedge 1{:}F_n \rightarrow 1{:}F$. Since *prev* and *post* are (total) functions on R, we have

$$\models 0{:}F \leftrightarrow \neg 0{:}\neg F \qquad \text{and} \qquad \models 1{:}F \leftrightarrow \neg 1{:}\neg F$$

(and $0{:}$ and $1{:}$ distribute over all truth-functional connectives).

Examples: state formulas $\underset{\rightarrow}{\Diamond}\varphi$ says that there is a transition of type φ from the current state, or in the terminology of transition systems, that φ is 'executable'. $\underset{\rightarrow}{\Diamond}\, 1{:}F$ expresses that there is a transition from the current state to a state where F is true. $\underset{\rightarrow}{\Box}(\varphi \rightarrow 1{:}F)$ says that all transitions of type φ from the current state result in a state where F is true.

There are various relationships between state formulas and transition formulas. For example, the state formula $F \rightarrow \underset{\rightarrow}{\Box}\, 0{:}F$ is valid (true in all states, in all models). Further details are given in the next section.

2.2 Agency Modalities

We now extend the language with operators to talk about the actions of agents and sets of agents in a transition. Ag is a finite set of (names of) agents. The account can be generalised to deal with (countably) infinite sets of agents but we will not do so here.

Language Transition formulas are extended with the operators $\Box$, $[x]$ and $[G]$ for every agent x in Ag and every non-empty subset G of Ag. State formulas are unchanged. $\Box\varphi$, $[x]\varphi$ and $[G]\varphi$ are transition formulas when φ is a transition formula. $\Diamond$, $\langle x\rangle$ and $\langle G\rangle$ are the respective duals.

Semantics Models are relational structures of the form

$$\langle S, R, \mathit{prev}, \mathit{post}, \sim, \{\sim_x\}_{x\in Ag}, h^{\mathrm{f}}, h^{\mathrm{a}}\rangle$$

where $\langle S, R, \mathit{prev}, \mathit{post}, h^{\mathrm{f}}, h^{\mathrm{a}}\rangle$ is a labelled transition model of the type discussed above, and $\sim$ and every $\sim_x$ are equivalence relations on R.

$$\sim =_{\mathrm{def}} \{(\tau, \tau') \mid \mathit{prev}(\tau) = \mathit{prev}(\tau')\}$$

and, for every $x \in Ag$: $\sim_x \subseteq \sim$.

Informally, for any transitions τ, τ' in R, $\tau \sim \tau'$ represents that τ and τ' are transitions from the same initial state, and $\tau \sim_x \tau'$ that τ and τ' are transitions from the same initial state ($\sim_x \subseteq \sim$) in which agent x performs the same action in τ' as it does in τ.

The truth conditions are

$$\mathcal{M}, \tau \models \Box\varphi \quad \text{iff} \quad \mathcal{M}, \tau' \models \varphi \text{ for every } \tau' \text{ such that } \tau \sim \tau'$$
$$\mathcal{M}, \tau \models [x]\varphi \quad \text{iff} \quad \mathcal{M}, \tau' \models \varphi \text{ for every } \tau' \text{ such that } \tau \sim_x \tau'$$

$[x]$ is what some authors (e.g. Horty 2001) call the 'Chellas *stit*'. However, it is important to stress that $[x]\varphi$ is a *transition formula* expressing a property of transitions and that φ is also a transition formula. When $[x]\varphi$ is true at a transition τ, we will say that φ is necessary for how x acts in τ. $\Box$ and each $[x]$ are normal modal operators of type S5. The schema

$$\Box\varphi \rightarrow [x]\varphi$$

is valid for all agents x in Ag.

We also have the following relationships between state formulas and transition formulas. All instances of the transition formula $0{:}\underset{\rightarrow}{\Box}\varphi \leftrightarrow \Box\varphi$ are valid, as are the state formulas $F \rightarrow \underset{\rightarrow}{\Box}\, 0{:}F$ and $(\underset{\rightarrow}{\Diamond}\top \wedge \underset{\rightarrow}{\Box}\, 0{:}F) \rightarrow F$, i.e., $\underset{\rightarrow}{\Diamond}\, 0{:}F \leftrightarrow (\underset{\rightarrow}{\Diamond}\top \wedge F)$.

In what follows it is convenient to employ a functional notation. Let:

$$alt(\tau) =_{\text{def}} \{\tau' \mid \tau \sim \tau'\}$$
$$alt_x(\tau) =_{\text{def}} \{\tau' \mid \tau \sim_x \tau'\}$$

alt is for 'alternative'. ($alt(\tau)$ and $alt_x(\tau)$ are thus the equivalence classes $[\tau]^{\sim}$ and $[\tau]^{\sim_x}$ respectively. The alt_x notation is slightly easier to read).

For every $x \in Ag$ and every $\tau \in R$, we have $alt_x(\tau) \subseteq alt(\tau)$. The truth conditions can be expressed as:

$$\mathcal{M}, \tau \models \Box\varphi \quad \text{iff} \quad alt(\tau) \subseteq \|\varphi\|^{\mathcal{M}}$$
$$\mathcal{M}, \tau \models [x]\varphi \quad \text{iff} \quad alt_x(\tau) \subseteq \|\varphi\|^{\mathcal{M}}$$

$alt(\tau)$ is the set of transitions from the same initial state as τ, and $alt_x(\tau)$ is the set of transitions from the same initial state as τ in which x performs the same action as it does in τ: these are the possible alternative actions that could be performed by x (deliberatively, intentionally, but possibly also unwittingly). $alt_x(\tau)$ is the equivalence class that contains τ, and so, just as in the *stit* framework, it can be regarded as the action performed by x in the transition τ.

For readers familiar with *stit* models, and models for the deliberative *stit* in particular, the set of transitions from any given state s can be seen (some technical details

aside) as the set of histories passing through a moment s. (It would be better to speak of mappings from moments to states but I do not want to dwell on technical details here.) Since every transition τ has a unique initial state $prev(\tau)$, every transition can also be thought of as a moment-history pair m/h where the moment m is the initial state $prev(\tau)$ and the history h is the transition τ. Putting aside technical details, one can think of transition system models as the special case of a (deliberative) *stit* model in which there is a single moment-history pair for every history. Evaluating formulas on transitions, as we do, is then like evaluating formulas on moment-history pairs in *stit*-models. Evaluating formulas on states, as we also do, would be like evaluating formulas on moments in *stit*-models. (Mark Brown in his chapter in this volume raises the question of whether points of valuation should be moments or moment/history pairs. We want both, which is why we employ a two-sorted language.) Put in these terms, $\tau \sim \tau'$ represents two moment-history pairs $\tau = m/h$ and $\tau' = m/h'$ through the same moment m. The equivalence relations $\sim_x$ determine what in *stit* would be the agent x's choice function. When $\tau = m/h$, $alt(\tau)$ is the set H_m of histories passing through m, and $alt_x(\tau)$ is $Choice_x^m(h)$, i.e., the action performed by x at moment m in history h, or equivalently, the subset of histories H_m in which x performs the same action at moment m as it does at moment m in history h.

Indeed, if we ignore states (or formulas on moments) and look only at transitions (or formulas on moment-history pairs), then models are of the form

$$\langle R, \sim, \{\sim_x\}_{x \in Ag}, h^{\mathrm{a}} \rangle$$

These are exactly the abstract models of the deliberative *stit* discussed in (Balbiani et al. 2008) *except that* there the models have a slightly different, but equivalent, form because they incorporate an extra, very strong 'independence of agents' assumption characteristic of *stit*.

stit-independence says (Horty 2001, p. 30) that 'at each moment, each agent must be able to perform any of his available actions, no matter which actions are performed at that moment by the other agents' or (Belnap and Perloff 1993, p. 26) 'any combination of choices made by distinct agents at exactly the same moment is consistent'.

Expressed as a condition on alt_x, *stit*-independence would require that, for all pairs of agents x and y in Ag, for all τ_x and τ_y such that $\tau_x \sim \tau_y$,

$$alt_x(\tau_x) \cap alt_y(\tau_y) \neq \emptyset$$

and more generally that, for all transitions τ and all mappings $s'_\tau : Ag \to alt(\tau)$:

$$\bigcap_{x \in Ag} alt_x(s'_\tau(x)) \neq \emptyset$$

We will not need the more general form in this chapter since none of the examples have more than two agents.

I do not understand what the 'independence of agents' assumption is for and why it is adopted without question in works on *stit*. I have not been able to find

any convincing justification for it in the literature. (Belnap and Perloff 1993, p. 26) remark that '… we do not consider the evident fact that agents interact in space-time' but do not say why. Why *not* consider the evident fact that agents interact in space-time? It is only a matter of dropping the *stit*-independence condition. What purpose does it serve? It is sometimes suggested that *stit*-independence is needed in order to ensure that some combination of actions by individual agents always exists. But that is not so. In the *stit* framework some combination of actions by agents always exists, without the *stit*-independence assumption. The *stit*-independence condition insists that *every* combination of actions always exists, which is much stronger. Further discussion is for another occasion. In what follows, some of the models will satisfy the *stit*-independence condition and some will not.

Group actions Just as in *stit*, the account generalises naturally to dealing with the joint actions of groups (sets) of agents. Let G be a non-empty subset of Ag. $alt_x(\tau)$ represents the action performed by x in the transition τ, which is the set of transitions in $alt(\tau)$ in which x performs the same action as it does in τ. $\bigcap_{x\in G} alt_x(\tau)$ is the set of transitions in $alt(\tau)$ in which every agent in G performs the same action as it does in τ, and is thus a representation of the joint action performed by the group G in the transition τ.

The truth conditions are:

$$\mathcal{M}, \tau \models [G]\varphi \quad \text{iff} \quad alt_G(\tau) \subseteq \|\varphi\|^{\mathcal{M}}$$

where

$$\begin{aligned} alt_G(\tau) &=_{\text{def}} \bigcap_{x\in G} alt_x(\tau) \\ \sim_G &=_{\text{def}} \bigcap_{x\in G} \sim_x \end{aligned}$$

That is, expressed in the relational notation:

$$\begin{aligned} \mathcal{M}, \tau \models [G]\varphi \quad \text{iff} \quad & \mathcal{M}, \tau' \models \varphi \text{ for every } \tau' \in \bigcap_{x\in G} alt_x(\tau) \\ \text{iff} \quad & \mathcal{M}, \tau' \models \varphi \text{ for every } \tau' \in alt_G(\tau) \\ & \text{where } alt_G(\tau) =_{\text{def}} \bigcap_{x\in G} alt_x(\tau) \\ \text{iff} \quad & \mathcal{M}, \tau' \models \varphi \text{ for every } \tau' \text{such that} \tau \sim_G \tau' \\ & \text{where } \sim_G =_{\text{def}} \bigcap_{x\in G} \sim_x \end{aligned}$$

When $[G]\varphi$ is true at τ we will say that φ is necessary for how the agents G collectively act in τ. (Which is not the same as saying that they act together, i.e., as a kind of coalition or collective agent. We are not discussing genuine collective agency in this chapter.) Clearly $\models [\{x\}]\varphi \leftrightarrow [x]\varphi$ for every x in Ag.

Axiomatisation $\Box$ and every $[x]$ and every $[G]$ are normal modal operators of type S5. The logic is the smallest normal logic containing all instances of the following axiom schemas, for all non-empty subsets G and G' of Ag:

$\Box$ type S5
$[G]$ type S5
$\Box\varphi \rightarrow [G]\varphi$
$[G]\varphi \rightarrow [G']\varphi \quad (G \subseteq G')$

2.3 Acts Differently

We also want to be able speak about alternative transitions from the same initial state in which an agent x, or set of agents G, acts *differently* from the way it acts in a transition τ. We further extend the language of transition formulas with operators $[\overline{x}]$ and $[\overline{G}]$ for every agent x in Ag and every non-empty subset G of Ag: $[\overline{x}]\varphi$ and $[\overline{G}]\varphi$ are transition formulas when φ is a transition formula. $\langle\overline{x}\rangle$ and $\langle\overline{G}\rangle$ are the respective duals.

The truth conditions are:

$$\mathcal{M}, \tau \models [\overline{x}]\varphi \quad \text{iff} \quad (alt(\tau) - alt_x(\tau)) \subseteq \|\varphi\|^{\mathcal{M}}$$
$$\mathcal{M}, \tau \models [\overline{G}]\varphi \quad \text{iff} \quad (alt(\tau) - alt_G(\tau)) \subseteq \|\varphi\|^{\mathcal{M}}$$

Note that $\models [\overline{x}]\varphi \leftrightarrow [\overline{\{x\}}]\varphi$, and that:

$$\models \langle\overline{G}\rangle\varphi \leftrightarrow \bigvee_{x\in G}\langle\overline{x}\rangle\varphi$$
$$\models [\overline{G}]\varphi \leftrightarrow \bigwedge_{x\in G}[\overline{x}]\varphi$$

2.4 'Brings It About' Modalities

In logics of agency, expressions of the form 'agent x brings it about that' or 'sees to it that' are typically constructed from two components. The first is a 'necessity condition': φ must be necessary for how agent x acts. The second component is used to capture the fundamental idea that φ is, in some sense, caused by or is the result of actions by x. Most accounts of agency introduce a negative counterfactual or 'counteraction' condition for this purpose, to express that had x not acted in the way that it did then the world would, or might, have been different.

Let $\mathrm{E}_x\varphi$ represent that agent x brings it about, perhaps unwittingly, that (a transition has) a certain property φ. $\mathrm{E}_x\varphi$ is satisfied by a transition τ in a model $\mathcal{M}$ when:

(1) (necessity) $\mathcal{M}, \tau \models [x]\varphi$, that is, all transitions from the same initial state as τ in which x acts in the same way as it does in τ are of type φ, or as we also say, φ is necessary for how x acts in τ;

(2) (counteraction) had x acted differently than it did in τ then the transition might have been different: there exists a transition τ' in $\mathcal{M}$ such that $\tau \sim \tau'$ and $\tau \not\sim_x \tau'$ and $\mathcal{M}, \tau' \models \neg\varphi$.

$\mathrm{E}_x\varphi$ is then defined as $\mathrm{E}_x\varphi =_{\text{def}} [x]\varphi \wedge \langle\overline{x}\rangle\neg\varphi$, or equivalently:

$$\mathrm{E}_x\varphi =_{\text{def}} [x]\varphi \wedge \neg[\overline{x}]\varphi$$

The difference modalities $[\overline{x}]$ are useful in their own right (see Sect. 6), but in order to avoid introducing further technical machinery, we note that if our purpose is only to construct the E_x modalities, then we can simplify. The counterfactual condition (2) can be simplified because of the necessity condition (1): if there is a transition τ' in $\mathcal{M}$ such that $\tau \sim \tau'$ and $\mathcal{M}, \tau' \models \neg\varphi$ but where $\tau \not\sim_x \tau'$, then the necessity condition (1) does not hold: $\mathcal{M}, \tau \not\models \varphi$. In other words, the following schema is valid, for all x in Ag:

$$\models ([x]\varphi \wedge [\overline{x}]\varphi) \leftrightarrow \Box\varphi$$

So instead of (2) for the counteraction condition we can take simply:

(2′) there exists a transition τ' in $\mathcal{M}$ such that $\tau \sim \tau'$ and $\mathcal{M}, \tau' \models \neg\varphi$.

This is just $\mathcal{M}, \tau \models \Diamond\neg\varphi$, or equivalently, $\mathcal{M}, \tau \models \neg\Box\varphi$.

The following simpler definition is thus equivalent to the original:

$$\mathrm{E}_x\varphi =_{\text{def}} [x]\varphi \wedge \neg\Box\varphi$$

This is exactly the construction used in the definition of the 'deliberative *stit*' (Horty and Belnap 1995)

$$[x \textit{ dstit}\colon \varphi] =_{\text{def}} [x]\varphi \wedge \neg\Box\varphi$$

except of course that we are reading φ as expressing a property of a transition.

The notation $\mathrm{E}_x\varphi$ is from (Pörn 1977) (though the semantics are different). It is chosen in preference to the *dstit* notation because it is more concise, and in order to emphasise that we do not want to incorporate the very strong *stit*-independence assumption that is built into *dstit*.

Notice that $\mathrm{E}_x\varphi \wedge \mathrm{E}_y\varphi$ is satisfiable even when $x \neq y$. Indeed

$$\models \mathrm{E}_x\varphi \wedge \mathrm{E}_y\varphi \leftrightarrow [x]\varphi \wedge [y]\varphi \wedge \neg\Box\varphi$$

It is possible to define a stronger kind of 'brings it about' modality which represents a sense in which it is agent *x and x alone* who brings it about that φ. We will not need that stronger form in this chapter since none of the examples has more than two agents. See (Sergot 2008a, b) for details and for discussion of some forms of collective action by groups (sets) of agents.

Note that adding the *stit*-independence condition validates, among other things, the following schema, for all distinct x and y in Ag:

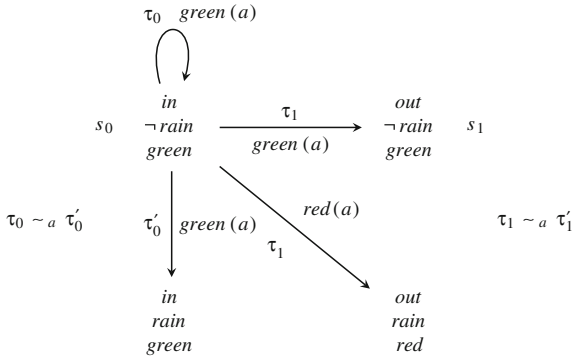


Fig. 1 Transitions from state s_0 ($in \wedge \neg rain$)

$$\neg \mathrm{E}_x \mathrm{E}_y \varphi \qquad (x \neq y)$$

Finally, in many of the examples that follow we will be interested in expressions of the form $\mathrm{E}_x(0{:}F \wedge 1{:}G)$. We note for future reference that:

$$\models \mathrm{E}_x(0{:}F \wedge 1{:}G) \leftrightarrow (0{:}F \wedge \mathrm{E}_x 1{:}G)$$

3 Example: Vase (One Agent)

We begin with a very simple example containing just a single agent a. Agent a can move a certain (precious) vase between indoors and outdoors. An element of indeterminism is introduced by allowing that it might be raining or not raining in any state, which is something that is outside the control of the agent a. Further, for the sake of an example, suppose it is forbidden, illegal, wrong for the vase to be outside in the rain.

Let state atoms *in* represent that the vase is indoors, *rain* that it is raining, and *red* that the state is forbidden/illegal. *out* is shorthand for $\neg in$; *green* is shorthand for $\neg red$.

Figure 1 shows a fragment of a transition system modelling this example, depicting the transitions from state s_0 ($in \wedge \neg rain$). The labels $green(a)$ and $red(a)$ on transitions will be explained presently. Figure 2 shows the transitions from state s_1 ($out \wedge \neg rain$). They are shown in a separate diagram simply to reduce clutter. Not shown in the diagrams are the transitions from the other two states in the model, where it is raining.

I have deliberately not included any transition atoms to name the actions by a. A perceived advantage of the *stit* treatment of action is that we are not forced to say exactly what action is performed by a when the vase is moved or left where it is. We need only say (in the example as I am thinking of it) that, whatever these actions are, the actions by a are the same in the two transitions τ_0 and τ'_0 ($\tau_0 \sim_a$

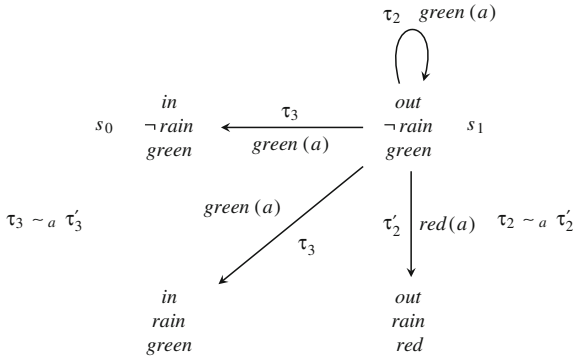


Fig. 2 Transitions from s_1 ($out \wedge \neg rain$)

τ_0'); they differ only in whether it is raining or not in the resulting state and not in what agent a does when the vase stays in place. And similarly, $\tau_1 \sim_a \tau_1'$. The possible actions by a in state s_0 are thus $\{\{\tau_0, \tau_0'\}, \{\tau_1, \tau_1'\}\}$, and those in state s_1 are $\{\{\tau_2, \tau_2'\}, \{\tau_3, \tau_3'\}\}$. From the diagram, one can see that they can be characterised in various ways, including:

$$\begin{aligned}
\{\tau_0, \tau_0'\} &= \|\neg 0{:}rain \wedge 0{:}in \wedge 1{:}in\| \\
&= \|\neg 0{:}rain \wedge \mathrm{E}_a\,(0{:}in \wedge 1{:}in)\| \\
&= \|\neg 0{:}rain \wedge 0{:}in \wedge \mathrm{E}_a\,1{:}in\| \\
\{\tau_1, \tau_1'\} &= \|\neg 0{:}rain \wedge 0{:}in \wedge 1{:}out\| \\
&= \|\neg 0{:}rain \wedge \mathrm{E}_a\,(0{:}in \wedge 1{:}out)\| \\
&= \|\neg 0{:}rain \wedge 0{:}in \wedge \mathrm{E}_a\,1{:}out\|
\end{aligned}$$

And similarly for a's possible actions in state s_1. $\{\tau_2, \tau_2'\} = \|\neg 0{:}rain \wedge 0{:}out \wedge 1{:}out\|$ and $\{\tau_3, \tau_3'\} = \|\neg 0{:}rain \wedge 0{:}out \wedge 1{:}in\|$, and so on.

Not shown in Figs. 1 and 2 are the transitions from the two states where it is raining. It is for that reason that the actions by a in state s_0 are not just $\|0{:}in \wedge 1{:}in\|$ and $\|0{:}in \wedge 1{:}out\|$ but $\|\neg 0{:}rain \wedge 0{:}in \wedge 1{:}in\|$ and $\|\neg 0{:}rain \wedge 0{:}in \wedge 1{:}out\|$. The example as formulated leaves open the possibility that moving-when-it-is-raining-now is not the same action as moving-when-it-is-not-raining-now, and not-moving-when-it-is-raining-now is not the same action as not-moving-when-it-is-not-raining-now.

Suppose however that we *do* want to say that the actions performed by a are the same irrespective of whether it is raining or not in the initial or final states: suppose the actions performed by a are the same in all transitions $\|0{:}in \wedge 1{:}out\|$, the same in all transitions $\|0{:}out \wedge 1{:}in\|$, and the same in all transitions $\|(0{:}in \wedge 1{:}in) \vee (0{:}out \wedge 1{:}out)\|$ where the vase stays where it is.

That would require an adjustment to the model structures. We could add a relation $=_x$ for every agent x in Ag, using $\tau =_x \tau'$ to represent that the action performed by x is the same in any transitions τ and τ' not just those that have the same initial state. We would then have:

$$\sim_x \ =_{\text{def}} \ \sim \cap =_x$$

A strong argument could be made that, for modelling purposes, this would be a useful and natural extension. It is easy to accommodate but I will not do so in the rest of this chapter. It would not fit so well in the *stit*-framework since that would require relating actions/choices across moments in different, incompatible histories which does not seem so natural.

One final remark: I am thinking here of 'moving' as a basic, simple kind of act, such as moving an arm while it grasps the vase or pushing the vase in one movement from one location to another. I am not thinking of 'moving' as an extended process of some kind requiring the vase to be packed up, transported somehow to the new location, and unpacked (say). In the latter case, the transitions in the diagrams would correspond to executions of this more elaborate 'moving' process. In that case we might well *not* want to say that $\tau_1 \sim_a \tau_1'$, since the moving process might be different if it happens to be raining as the vase reaches the *out* location. Indeed there might be many different 'moving' transitions between *in* and *out*, each corresponding to a different combination of actions by a. We will return to this point later under discussions of granularity of representations.

Example: Obligations

There is an obligation on a that the vase is not outside in the rain. Let the transition atom $red(a)$ represent a transition in which a fails to comply with this obligation. $green(a)$ is shorthand for $\neg red(a)$ and so is satisfied by transitions in which a does comply. Figures 1 and 2 show these labels on transitions. (It is an open question whether the transitions from a *red* state where the vase is already out in the rain should be $green(a)$ or $red(a)$ transitions. We will ignore that question here).

One sense of 'it is obligatory for agent x to 'do' φ' in a state s can be defined as follows:

$$\mathrm{O}_x\varphi \ =_{\text{def}} \ \underset{\rightarrow}{\Box}(green(x) \rightarrow \varphi)$$

or equivalently $\mathrm{O}_x\varphi =_{\text{def}} \underset{\rightarrow}{\Box}(\neg\varphi \rightarrow red(x))$. It follows that $\models \mathrm{O}_x\, green(x)$.

But *can* x comply with its obligations?

One sense of agent ability is that discussed by Brown (1988); it is expressed in the *stit* framework by the formula $\Diamond[x]\varphi$. In the present framework where we distinguish between state formulas and transition formulas, that sense of x can 'do' φ in state s would be expressed:

$$Can_x\varphi \ =_{\text{def}} \ \underset{\rightarrow}{\Diamond}[x]\varphi$$

In the example:

$$s_1 \models Can_a\, green(a) \qquad (\underset{\rightarrow}{\Diamond}[a]green(a))$$

But *what* should a do to ensure $green(a)$? We are looking for transitions from s_1 of type $[a]green(a)$. There are such transitions: those in which the vase is moved from *out* to *in*. a might also comply with its obligation by leaving the vase outdoors but compliance then is a matter of chance, outside a's control.

'Absence of moral luck' (Craven and Sergot 2008) is an (optional) rationality constraint that we might often want to check for when considering sets of regulations or specifications for computer systems. It reflects the idea that, for practical purposes, whether actions of agent x are in accordance with the norms directed at x should depend only on x's actions, not on the actions of other agents, nor actions in the environment, nor other extraneous factors. It can be expressed

- 'absence of moral luck' (in a model $\mathcal{M}$)

$$\mathcal{M} \models green(x) \rightarrow [x]green(x)$$

- 'absence of moral luck' (locally, in a state s)

$$\mathcal{M}, s \models \underset{\rightarrow}{\Box}(green(x) \rightarrow [x]green(x))$$

In the example, if agent a leaves the vase outside, it is a matter of luck whether it complies with its obligation or not, for this will depend on whether it rains, and that is an extraneous factor outside of a's control. Thus:

$$\tau_2 \models green(a) \rightarrow [a]green(a)$$
$$s_1 \models \underset{\rightarrow}{\Box}(green(a) \rightarrow [a]green(a)) \qquad \text{(no 'absence of moral luck')}$$

'Absence of moral luck' is a rather strong form of 'Ought implies Can'. Other, weaker forms can also be expressed. For instance, 'Ought implies Can' (1) (at state s in model $\mathcal{M}$)

$$\mathcal{M}, s \models \mathrm{O}_x\varphi \rightarrow \underset{\rightarrow}{\Diamond}\varphi \qquad \text{for all } \varphi$$

This is equivalent (it turns out) to $\mathcal{M}, s \models \underset{\rightarrow}{\Diamond} green(x)$ and to $\mathcal{M}, s \models \neg \mathrm{O}_x \bot$.

Compare 'Ought implies Can' (2) (at state s in model $\mathcal{M}$):

$$\mathcal{M}, s \models \mathrm{O}_x\varphi \rightarrow Can_x\varphi \qquad \text{for all } \varphi$$

This is equivalent (it turns out) to $\mathcal{M}, s \models Can_x\, green(x)$. It is easy to check that 'Ought implies Can' (2) is stronger than (implies) 'Ought implies Can' (1). 'Absence of moral luck' is stronger still: it implies 'Ought implies Can' (2).

In the example,

$$s_1 \models Can_a\, green(a)$$

and so 'Ought implies Can' (2) at s_1. But there is no 'absence of moral luck' at s_1, as demonstrated earlier.

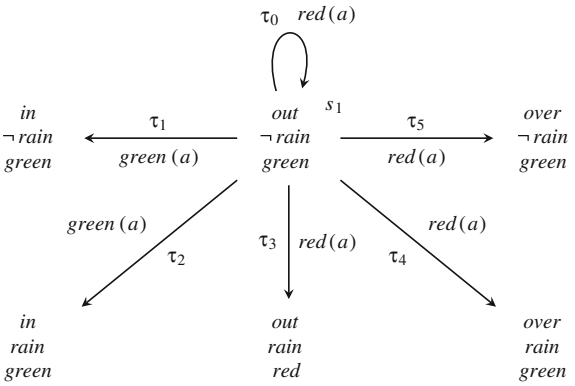


Fig. 3 Transitions from state s_1 ($out \wedge \neg rain$)

States where the vase is *in* are similar. Refer to Fig. 1. Here again

$$s_0 \not\models \underset{\rightarrow}{\Box}(green(a) \rightarrow [a]green(a)) \qquad \text{(no 'absence of moral luck')}$$
$$s_0 \models Can_a\, green(a) \qquad (\underset{\rightarrow}{\Diamond}[a]green(a))$$

In contrast, suppose that the obligation on a is not to ensure that the vase is not outdoors in the rain but instead that the vase is to be *moved* indoors if it is outdoors. In that case, transition τ_2 in Fig. 2, which was labelled $green(a)$ would be labelled $red(a)$. In that modified form of the example we have:

$$s_1 \models \mathrm{O}_a(0{:}out \wedge 1{:}in)$$
$$s_1 \models \underset{\rightarrow}{\Box}(green(a) \rightarrow [a]green(a)) \qquad \text{('absence of moral luck')}$$
$$s_1 \models Can_a\, green(a) \qquad (\underset{\rightarrow}{\Diamond}[a]green(a)) \qquad \text{(which follows from the above)}$$

4 Example: Vase (Two Agents)

Let us now introduce another agent, b. Suppose that the vase can be in one of three possible, mutually exclusive, locations, *in*, *out*, and *over*, say. Agent a can move the vase between *in* and *out*, and b can move it between *out* and *over* (but not simultaneously). There is an obligation on a to move the vase to *in* if it is *out*. There is no obligation on b to move the vase.

Figure 3 shows the transitions from the state s_1 where the vase is *out* and it is not raining.

The possible actions by a in state s_1 are (as we conceive the example) $\{\{\tau_0, \tau_3, \tau_4, \tau_5\}, \{\tau_1, \tau_2\}\}$. From the diagram:

$$\{\tau_0, \tau_3, \tau_4, \tau_5\} = \|0{:}\neg rain \wedge 0{:}out \wedge 1{:}\neg in\|$$
$$\{\tau_1, \tau_2\} = \|0{:}\neg rain \wedge 0{:}out \wedge 1{:}in\|$$

The possible actions by b in state s_1 are $\{\{\tau_0, \tau_1, \tau_2, \tau_3\}, \{\tau_4, \tau_5\}\}$.

$$\begin{aligned} \{\tau_0, \tau_1, \tau_2, \tau_3\} &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge 1{:}\neg \mathit{over}\| \\ \{\tau_4, \tau_5\} &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge 1{:}\mathit{over}\| \end{aligned}$$

This model does not satisfy *stit*-independence:

$$\{\tau_1, \tau_2\} \cap \{\tau_4, \tau_5\} = \emptyset$$

That is as it should be: the actions of a and b are not independent. If a moves the vase to *in* then b cannot simultaneously move it to *over*, and vice-versa.

a still has an obligation to move the vase *out* from *in*: the transitions in the diagram are labelled $\mathit{green}(a)$ and $\mathit{red}(a)$ accordingly.

$$\begin{aligned} &s_1 \models \mathrm{O}_a\,(0{:}\mathit{out} \wedge 1{:}\mathit{in}) \\ &s_1 \models \underset{\rightarrow}{\Box}\,(\mathit{green}(a) \rightarrow [a]\mathit{green}(a)) \qquad \text{('no moral luck')} \\ &s_1 \models \mathit{Can}_a\,\mathit{green}(a) \qquad (\underset{\rightarrow}{\Diamond}\,[a]\mathit{green}(a)) \end{aligned}$$

That is as expected. But note also that:

$$s_1 \models \underset{\rightarrow}{\Diamond}\,[b]\mathit{red}(a) \qquad (\mathit{Can}_b\,\mathit{red}(a))$$

The last says that in state s_1, b can act in such a way that a necessarily fails to fulfill its obligation. And that is also surely right: if b moves the vase from *out* to *over*, a could not simultaneously move it from *out* to *in*, which is a's obligation.

One can see from the diagram that all transitions where b moves the vase to *over* are $\mathit{red}(a)$. Thus:

$$\begin{aligned} &s_1 \models \underset{\rightarrow}{\Box}\,(\,\mathrm{E}_b\,1{:}\mathit{over} \rightarrow \mathit{red}(a)\,) \\ &s_1 \models \underset{\rightarrow}{\Box}\,(\,\mathit{red}(a) \rightarrow [a]\mathit{red}(a)\,) \qquad \text{('no moral luck')} \\ &s_1 \models \underset{\rightarrow}{\Box}\,(\,\mathrm{E}_b\,1{:}\mathit{over} \rightarrow [a]\mathit{red}(a)\,) \end{aligned}$$

and moreover:

$$\begin{aligned} &s_1 \models \underset{\rightarrow}{\Box}\,(\,\mathrm{E}_b\,1{:}\mathit{over} \rightarrow \mathrm{E}_a\,\mathit{red}(a)\,) \\ &s_1 \models \underset{\rightarrow}{\Box}\,(\,\mathrm{E}_b\,1{:}\mathit{over} \rightarrow \mathrm{E}_b\mathrm{E}_a\,\mathit{red}(a)\,) \end{aligned}$$

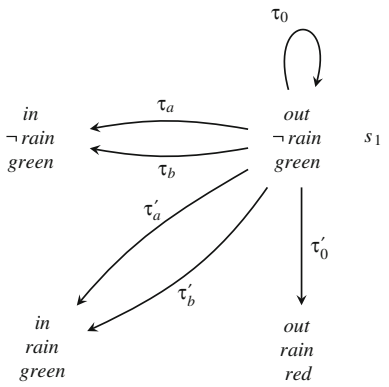


Fig. 4 Transitions from state s_1

The last two formulas in particular may seem counterintuitive on an informal reading. The first seems to say that if b brings about or is responsible for the vase's moving to *over* then a brings about or is responsible for violating its obligation; the second that b thereby brings about or is responsible for a's violating its obligation. The question of how these formulas may be read informally as *stit* statements does not arise because the example is not a *stit*-model. It does not satisfy the *stit*-independence condition.

5 Example: Vase, Minor Variation

The following minor variation of the example is intended to make some further observations about the representation of actions.

Let us suppose there are just two (mutually exclusive) locations *in* and *out* for the vase, and that agents a and b can both move the vase between them (but not simultaneously).

Informally, in Fig. 4 τ_a and τ'_a are transitions where a moves the vase, and τ_b and τ'_b are transitions where b moves it.

$$\begin{aligned} &\text{Actions by } a \text{ in state } s_1: \quad \{\{\tau_0, \tau'_0, \tau_b, \tau'_b\},\ \{\tau_a, \tau'_a\}\} \\ &\text{Actions by } b \text{ in state } s_1: \quad \{\{\tau_0, \tau'_0, \tau_a, \tau'_a\},\ \{\tau_b, \tau'_b\}\} \end{aligned}$$

There is no *stit*-independence in this model: a and b cannot both move the vase simultaneously.

$$\{\tau_a, \tau'_a\} \cap \{\tau_b, \tau'_b\} = \emptyset$$

Suppose for the sake of an example that a and b *both* have an obligation to move the vase *in* if it is *out*: the transitions τ_a and τ'_a are $green(a)$, τ_b and τ'_b are $green(b)$, and all other transitions from state s_1 are $red(a)$ and $red(b)$.

In this example there are different transitions between the same pairs of states and we cannot identify the actions of a and b by reference only to what holds in initial and final states.

$$\{\tau_a, \tau_a'\} \neq \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge 1{:}\mathit{in}\|$$
$$\{\tau_0, \tau_0', \tau_b, \tau_b'\} \neq \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge 1{:}\neg \mathit{out}\|$$

(And likewise for b.)

It seems that in order to refer to a and b's actions we are forced to introduce some new (transition) atoms, which is something we were trying to avoid. But it happens that in this example the actions by a in state s_1 can be picked out as follows:

$$\begin{aligned}\{\tau_a, \tau_a'\} &= \|0{:}\neg \mathit{rain} \wedge \mathrm{E}_a(0{:}\mathit{out} \wedge 1{:}\mathit{in})\| \\ &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge \mathrm{E}_a 1{:}\mathit{in}\| \\ \{\tau_0, \tau_0', \tau_b, \tau_b'\} &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out}\| - \{\tau_a, \tau_a'\} \\ &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge \neg \mathrm{E}_a(0{:}\mathit{out} \wedge 1{:}\mathit{in})\| \\ &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge \neg \mathrm{E}_a 1{:}\mathit{in}\|\end{aligned}$$

(And likewise for b.)

So, for convenience only, in this example we could define two new transition atoms $a{:}\mathit{moves}(\mathit{out}, \mathit{in})$ and $b{:}\mathit{moves}(\mathit{out}, \mathit{in})$ as follows:

$$\begin{aligned}a{:}\mathit{moves}(\mathit{out}, \mathit{in}) &=_{\text{def}} \mathrm{E}_a(0{:}\mathit{out} \wedge 1{:}\mathit{in}) \\ b{:}\mathit{moves}(\mathit{out}, \mathit{in}) &=_{\text{def}} \mathrm{E}_b(0{:}\mathit{out} \wedge 1{:}\mathit{in})\end{aligned}$$

The possible actions by a in state s_1 are thus:

$$\begin{aligned}\{\tau_0, \tau_0', \tau_b, \tau_b'\} &= \|0{:}\neg \mathit{rain} \wedge 0{:}\mathit{out} \wedge \neg a{:}\mathit{moves}(\mathit{out}, \mathit{in})\| \\ \{\tau_a, \tau_a'\} &= \|0{:}\neg \mathit{rain} \wedge a{:}\mathit{moves}(\mathit{out}, \mathit{in})\|\end{aligned}$$

(And likewise for b.)

I am not suggesting there is a general principle at work here. This is a very simple example where there are just two agents, and where each agent has just two possible actions in any state. In more complicated examples it is very far from obvious how to characterise possible actions by means of 'brings it about' formulas in this way. In bigger examples it very rarely works out so neatly.

It is perhaps worth reiterating that what seems natural in this framework is to say that the action performed by x in transition τ is not $[\tau]^{\sim x}$ but $[\tau]^{=x}$. Then a's possible actions in state s_1 would be simply $\|0{:}\mathit{out} \wedge \mathrm{E}_a 1{:}\mathit{in}\|$ and $\|0{:}\mathit{out} \wedge \neg \mathrm{E}_a 1{:}\mathit{in}\|$, i.e., $\|a{:}\mathit{moves}(\mathit{out}, \mathit{in})\|$ and $\|0{:}\mathit{out} \wedge \neg a{:}\mathit{moves}(\mathit{out}, \mathit{in})\|$.

In this example we have, among other things:

$$s_1 \models \mathrm{O}_a\, a{:}\mathit{moves}(\mathit{out}, \mathit{in}) \wedge \mathrm{O}_b\, b{:}\mathit{moves}(\mathit{out}, \mathit{in})$$
$$s_1 \models \mathrm{O}_a \mathrm{E}_a\, (0{:}\mathit{out} \wedge 1{:}\mathit{in}) \wedge \mathrm{O}_b \mathrm{E}_b\, (0{:}\mathit{out} \wedge 1{:}\mathit{in})$$
$$s_1 \models \mathit{Can}_a\, a{:}\mathit{moves}(\mathit{out}, \mathit{in}) \wedge \mathit{Can}_b\, b{:}\mathit{moves}(\mathit{out}, \mathit{in})$$
$$s_1 \models \neg \underset{\rightarrow}{\Diamond}\, (a{:}\mathit{moves}(\mathit{out}, \mathit{in}) \wedge b{:}\mathit{moves}(\mathit{out}, \mathit{in}))$$
$$s_1 \models \underset{\rightarrow}{\Box}\, (\, \mathit{green}(a) \leftrightarrow \mathit{red}(b)\,)$$
$$s_1 \models \underset{\rightarrow}{\Box}\, (\, a{:}\mathit{moves}(\mathit{out}, \mathit{in}) \rightarrow \mathrm{E}_b\, \mathit{red}(b)\,)$$
$$s_1 \models \mathit{Can}_a \mathrm{E}_b\, \mathit{red}(b)$$
$$s_1 \models \underset{\rightarrow}{\Box}\, (\, a{:}\mathit{moves}(\mathit{out}, \mathit{in}) \rightarrow \mathrm{E}_a \mathrm{E}_b\, \mathit{red}(b)\,)$$

6 Example: Table

This example is intended to raise some questions about the treatment of agency, and in particular about the 'necessity' condition.

Suppose there is an agent a who can lift or lower its end of a table, or do neither. On the table stands a vase. If the table tilts, the vase might fall or it might not. If the vase falls, it might break or it might not. If the table does not tilt then the vase does not fall; if it does not fall, it does not break.

Figure 5 shows transitions from the state in which the table is level and the vase stands on it. State atoms *level*, *on-table* and *broken* have the obvious intended readings. There are other transitions not shown in the diagram and two more states, those in which the table is level (*level*) but the vase is not on it ($\neg$*on-table*); in one of these the vase is broken, in the other it is not.

For convenience, let the transition atom *falls* be defined as follows:

$$\mathit{falls} \;=_{\text{def}}\; 0{:}\mathit{on\text{-}table} \wedge 1{:}\neg \mathit{on\text{-}table}$$

The possible actions by a in this state are $\{\{\tau_0\},\ \{\tau_1, \tau_1', \tau_1''\}\}$. By reference to previous examples, there are various ways we can describe them, e.g.

$$\begin{aligned} \{\tau_1, \tau_1', \tau''\} &= \|0{:}\mathit{on\text{-}table} \wedge \mathrm{E}_a\, (0{:}\mathit{level} \wedge 1{:}\neg \mathit{level})\| \\ &= \|0{:}\mathit{on\text{-}table} \wedge 0{:}\mathit{level} \wedge \mathrm{E}_a\, 1{:}\neg \mathit{level}\| \\ \{\tau_0\} &= \|0{:}\mathit{on\text{-}table} \wedge \neg \mathrm{E}_a\, (0{:}\mathit{level} \wedge 1{:}\neg \mathit{level})\| \\ &= \|0{:}\mathit{on\text{-}table} \wedge 0{:}\mathit{level} \wedge \neg \mathrm{E}_a\, 1{:}\neg \mathit{level}\| \end{aligned}$$

(Here, there is just a single agent a in the example and so the operator E_a could be omitted from all of the above.) The simpler expressions $\|\mathrm{E}_a\, (0{:}\mathit{level} \wedge 1{:}\neg \mathit{level})\|$ and $\|\neg \mathrm{E}_a\, (0{:}\mathit{level} \wedge 1{:}\neg \mathit{level})\|$ are not sufficient to pick out a's actions: there are other transitions not shown in the diagram where a lifts or lowers its end of the table

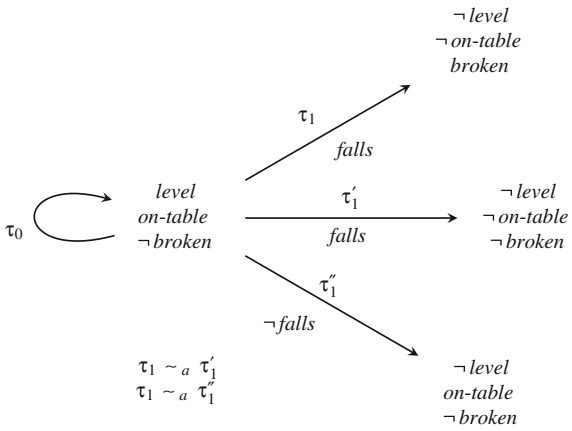


Fig. 5 Transitions from the state in which the table is level and the vase stands on it

when the vase is not on it. On the other hand, as observed earlier, we might well want to say that the actions of a's lifting its end of the table or not lifting are the same whether the vase stands on it or not. That would identify actions with equivalence classes of $=_a$ rather than $\sim_a$.

But here is the main point. Suppose that a tilts the table and the vase falls and breaks:

$$\tau_1 \models \mathit{falls} \wedge 0{:}\neg\mathit{broken} \wedge 1{:}\mathit{broken}$$

Had a not tilted the table the vase would not have fallen. But a is not responsible for, does not bring about, the breaking of the vase:

$$\tau_1 \not\models \mathrm{E}_a\mathit{falls} \qquad \text{(because } \tau_1 \not\models [a]\mathit{falls}\text{)}$$
$$\tau_1 \not\models \mathrm{E}_a\, 1{:}\mathit{broken} \qquad \text{(because } \tau_1 \not\models [a]\, 1{:}\mathit{broken}\text{)}$$

It is not necessary for what a does in τ_1 that the vase falls, and it is not necessary for what a does in τ_1 that the vase breaks.

Examples such as this, and many others, suggest that there is a weaker sense in which a 'brings about' or is responsible for the falling and breaking of the vase when a tilts the table. What is this weaker form?

There are two obvious candidates:

(1) $\varphi \wedge [\overline{x}]\neg\varphi$

(2) $\varphi \wedge \langle\overline{x}\rangle\neg\varphi$

The first says that x is responsible for φ because φ is true and had x acted differently, φ would not have been true. (1) is too strong (demands too much). The second is more plausible and is mentioned briefly in (Pörn 1977): x is responsible for φ because φ is true, and had x acted differently, φ might not have been true. But (2) is too weak.

In this particular example, both are plausible at first sight: in transition τ_1, the vase fell and broke but had *a* acted differently and not tilted the table, the vase would not have fallen and would not have broken.

(1) $\varphi \wedge [\overline{x}]\neg\varphi$ is too strong (demands too much). For suppose there were another way in which agent *a* could cause the vase to fall: suppose that *a* could dislodge the vase by jolting the table (say). Now, suppose that *a* lifts its end of the table and the vase falls. That would be a transition of type *falls*; but $falls \wedge [\overline{a}]\neg falls$ is false in that transition since there is another transition, where *a* jolts instead of lifting, which also has *falls* true. So on that reading, *a* is not responsible for the vase's falling.

The candidate form (2) $\varphi \wedge \langle\overline{x}\rangle\neg\varphi$ is more plausible but is too weak. Consider a version of the earlier vase example in which agent *a* moves the vase between *in* and *out*. Consider a transition in which *a* moves the vase to *out* and it rains, that is, a transition of type $1{:}(out \wedge rain)$. It is *a* who moves the vase, no-one else. In that transition, $\mathrm{E}_a\, 1{:}(out \wedge rain)$ is false because it is not necessary for what *a* does that $1{:}(out \wedge rain)$: $[a]1{:}(out \wedge rain)$ is false because it might not have rained. However, had *a* acted differently (by not moving the vase) it might have been otherwise: $1{:}(out \wedge rain) \wedge \langle\overline{a}\rangle\neg 1{:}(out \wedge rain)$ is true.

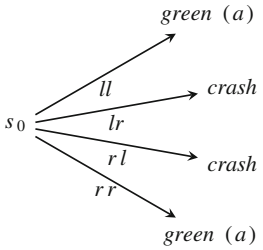
But that is too weak. By exactly the same argument, $1{:}rain \wedge \langle\overline{a}\rangle\neg 1{:}rain$ is also true in that transition: it rains, and had *x* acted differently, it might not have rained. Yet we would not want to say that agent *a* is reponsible for, or the cause of, or the one who brings about that it is raining.

It is far from clear whether this weaker sense of 'brings it about' or 'responsible for' can be articulated using the available resources. The problem is that nearly everything we want to say about agency in practice is of this weaker form. If a man walks into a room, puts a loaded revolver in his mouth and blows his brains out, we would surely want to say that he killed himself, that he was responsible for his death, that it was his actions that caused it. Yet he did not see to it or bring it about: it was not necessary for what he did that he died. The gun might have jammed, the bullet might have hit a thick part of the skull, the resulting injury might not have been fatal for any number of reasons. And this has nothing to do with probabilities. If a man walks in a room, picks a bullet at random from a barrel containing live and blank ammunition, loads his revolver, spins the chamber, then pulls the trigger and blows his brains out, we would say that he killed himself, even though the likelihood that those actions result in death is very small.

7 Example: Avoidance (Fixed)

The next series of examples illustrates some common patterns in which the actions of one agent constrain, or possibly even determine, the actions of another.

Two agents *a* and *b* (cyclists, say) approach each other on a path. If both swerve left or both swerve right they avoid the crash; otherwise they crash. There is an obligation on *a* that there is no crash.



$$
\begin{array}{lcl}
ll & \models & green(a) \wedge \neg[a]green(a) \\
lr & \models & red(a) \wedge \neg[a]red(a) \\
rl & \models & red(a) \wedge \neg[a]red(a) \\
rr & \models & green(a) \wedge \neg[a]green(a)
\end{array}
$$

Fig. 6 *a* and *b* can both swerve to *left* or *right*

Figure 6 shows the possible transitions as the agents approach each other. The labels $ll, lr, \ldots$ on transitions are just mnemonics: ll indicates that a and b both swerve left, lr that a swerves left and b swerves right, and so on. *crash* is a transition atom with the obvious intended reading. The transition atom $green(a)$ represents transitions in which a complies with its obligation. $red(a)$ is shorthand for $\neg green(a)$. In this model, $green(a) \leftrightarrow \neg crash$ is valid, or at least true in all transitions from the state s_0 depicted in the diagram.

One can see that in the case of a crash, agents a and b collectively bring it about that there is a crash, though neither individually does so. And similarly in the case where both swerve and there is no crash. We will not discuss possible forms of collective agency in this chapter.

a has an obligation to avoid the crash but cannot guarantee that its actions will comply: 'Ought implies Can' (2) fails for this obligation:

$$s_0 \not\models Can_a\, green(a) \qquad (s_0 \not\models \underset{\rightarrow}{\Diamond}[a]green(a))$$

And there is no 'absence of moral luck' (which follows from the above)

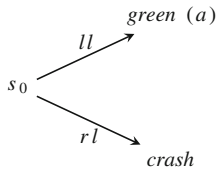
$$s_0 \not\models \underset{\rightarrow}{\Box}(green(a) \rightarrow [a]green(a))$$

Agent as Automaton

Consider the same example but suppose now that b has a fixed behaviour in this situation—a reflex or a deliberative decision procedure of some kind that always chooses the same action by b in these circumstances—b always swerves left (say). The obligation is still on a that there is no crash ($green(a) \leftrightarrow \neg crash$).

At one level of detail, the possible transitions are as shown in Fig. 7. Note first that there is 'absence of moral luck': $s_0 \models \underset{\rightarrow}{\Box}(green(a) \rightarrow [a]green(a))$. Moreover:

Fig. 7 *b* always swerves *left*



$$s_0 \models \underset{\rightarrow}{\Diamond}[a]\neg crash \qquad (Can_a\,\neg crash)$$

(though a might not know this, or know how to avoid the crash).

But who is responsible in the case of a crash?

$$ll \models [a]\neg crash \wedge \neg[b]\neg crash \wedge \mathrm{E}_a\,\neg crash$$
$$rl \models [a]crash \wedge \neg[b]crash \wedge \mathrm{E}_a\,crash$$

Because b's actions are fixed, b never brings about crash or no-crash: a is always solely responsible.

$$s_0 \models \underset{\rightarrow}{\Box}(crash \leftrightarrow \mathrm{E}_a\,crash) \;\wedge\; \underset{\rightarrow}{\Box}(\neg crash \leftrightarrow \mathrm{E}_a\,\neg crash)$$

Perhaps this seems odd. Perhaps not—after all, this transition system models how b will *actually* behave. b's behaviour is treated here as if it were just part of the environment in which a operates, like a gate operated by a sensor or a traffic light. This seems perfectly reasonable if b is an automaton or a mechanical device of some kind. But what if b is not an automaton? What if b makes deliberate decisions about other actions but reacts automatically when faced by an oncoming a as here? b behaves like an automaton *in this respect* but not in every other.

Here is an alternative way of modelling this scenario. Let transition atom $prog_b$ represent that b acts in accordance with its protocol/decision procedure. (Here, to swerve left whatever a does.) We can assume $\mathcal{M} \models prog_b \leftrightarrow [b]prog_b$.

We need some way of referring to b's actions. Unlike in previous examples, there seems to be no recourse but to introduce a transition atom for this purpose. Let transition atom $b{:}l$ represent that b swerves left. b's protocol requires that $s_0 \models \underset{\rightarrow}{\Box}(prog_b \leftrightarrow b{:}l)$.

Figure 8 depicts the model. In this version:

$$s_0 \models \underset{\rightarrow}{\Box}(prog_b \rightarrow \mathrm{E}_b prog_b) \wedge \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_b\, b{:}l)$$
$$s_0 \not\models Can_b\,\neg crash$$
$$s_0 \models Can_a(prog_b \rightarrow \neg crash)$$
$$s_0 \not\models \underset{\rightarrow}{\Box}(crash \rightarrow \mathrm{E}_a(prog_b \rightarrow crash))$$

The last is because $\mathrm{E}_a(prog_b \rightarrow crash)$ is false in the transition lr. Moreover:

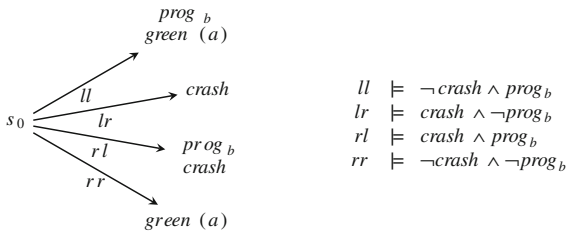


Fig. 8 *b* swerves *left* (explicit protocol)

$$s_0 \not\models \Box\!\!\!\!\rightarrow (prog_b \rightarrow (crash \rightarrow \mathrm{E}_a\, crash))$$

Of course it is a matter of *choice* how we model the example. It is not that one is right and the other is wrong. They model different things. Let us call the models in Figs. 7 and 8 *actual* and *explicit protocol*, respectively.

In both models, *b* cannot avoid the crash, in the sense that:

$$s_0 \not\models Can_b \neg crash$$

And in both models *a* can avoid the crash (though *a* might not know this, nor know how). In the 'actual' model (Fig. 7):

$$s_0 \models Can_a \neg crash$$

In the 'explicit protocol' model (Fig. 8):

$$s_0 \models Can_a\,(prog_b \rightarrow \neg crash\,)$$

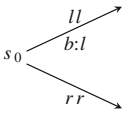
What differs is who is responsible in the case of a crash. In the 'actual' model (Fig. 7) it is *a*:

$$s_0 \models \Box\!\!\!\!\rightarrow (\,crash \rightarrow \mathrm{E}_a\, crash\,)$$

However in the 'explicit protocol' model (Fig. 8):

$$s_0 \not\models \Box\!\!\!\!\rightarrow (\,crash \rightarrow \mathrm{E}_a\,(prog_b \rightarrow crash))$$
$$s_0 \not\models \Box\!\!\!\!\rightarrow (prog_b \rightarrow (crash \rightarrow \mathrm{E}_a\, crash))$$

I find this slightly disturbing. I cannot see any general principles for choosing one of these models over the other. Both seem reasonable formalisations of the example, in their own way. And if one model has it that *a* is responsible for the crash, then it seems the other should have something comparable. But what? The two obvious candidates (the last two formulas above) do not work. It is not immediately obvious



$ll \models [a]\neg crash \wedge [b]\neg crash$
$rr \models [a]\neg crash \wedge [b]\neg crash$

Fig. 9 *b* reacts to *a* (atemporal, 'actual')

whether a sense of responsibility for crashing in the second model could be expressed and related neatly to the first.

8 Example: Avoidance (Reaction)

Suppose now that *b*'s fixed reflex, program, deliberative procedure is to *react* to *a*—if *a* goes left, so does *b*; if *a* goes right, so does *b*. (The obligation on *a* that there is no crash will play no role in this example).

As a first shot, let us ignore the temporal structure implicit in the term 'reacts to' and represent the possible behaviours in the example as atomic transitions.

We begin with 'actual' behaviour, as depicted in Fig. 9.

From the diagram:

$$s_0 \models Can_b \neg crash \qquad \text{(trivially, since } \underset{\rightarrow}{\Box} \neg crash\text{)}$$
$$s_0 \models Can_a \neg crash$$
$$s_0 \models \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_b b{:}l)$$

But also:

$$s_0 \models \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_a b{:}l)$$

For consider:

$$ll \models [a]b{:}l \wedge \neg\Box b{:}l, \quad \text{and hence } ll \models \mathrm{E}_a b{:}l$$
$$ll \models \mathrm{E}_b b{:}l \qquad \text{(similarly)}$$

Furthermore:

$$ll \models [a]\mathrm{E}_b b{:}l \wedge \neg\Box \mathrm{E}_b b{:}l$$
$$ll \models \mathrm{E}_a \mathrm{E}_b b{:}l$$
$$ll \models \mathrm{E}_b \mathrm{E}_a b{:}l \qquad \text{(similarly)}$$

So then:

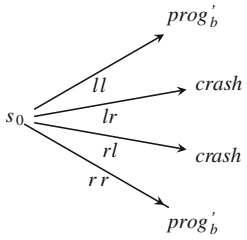


Fig. 10 b reacts to a (atemporal, 'explicit protocol')

$$\begin{aligned} ll &\models \neg crash \wedge prog'_b \\ lr &\models crash \wedge \neg prog'_b \\ rl &\models crash \wedge \neg prog'_b \\ rr &\models \neg crash \wedge prog'_b \end{aligned}$$

$$s_0 \models \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_a\, b{:}l)$$
$$s_0 \models \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_a \mathrm{E}_b\, b{:}l)$$
$$s_0 \models \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_b \mathrm{E}_a \mathrm{E}_b\, b{:}l)$$

There is obviously no *stit*-independence in this model. If there were then $\mathrm{E}_a \mathrm{E}_b \varphi$ would be false for every formula φ. That is a property validated by *stit*-independence.

Perhaps some of these formulas seem counterintuitive? What if we represent the temporal structure implicit in 'reacts to'? We will turn to that in a moment. Before that, for the sake of completeness, let us consider the 'explicit protocol' formulation of the atemporal model.

Let the transition atom $prog'_b$ represent that b acts in accordance with its reaction procedure. We can assume $\mathcal{M} \models prog'_b \rightarrow [b]prog'_b$. See Fig. 10.

Obviously in this example: $s_0 \models \underset{\rightarrow}{\Box}(crash \leftrightarrow \neg prog'_b)$. But suppose that b fails to react correctly, that is, that $prog'_b$ is false. Is b then responsible for the crash? No:

$$s_0 \not\models \underset{\rightarrow}{\Box}(\neg prog'_b \rightarrow [b]crash)$$
$$s_0 \not\models \underset{\rightarrow}{\Box}(\neg prog'_b \rightarrow \mathrm{E}_b crash)$$

b's protocol is to react to a: if b goes left and by doing so abides by its protocol, does it follow that a brings this about? No:

$$s_0 \not\models \underset{\rightarrow}{\Box}(b{:}l \rightarrow (prog'_b \rightarrow \mathrm{E}_a\, b{:}l))$$

Though a does bring it about in the following sense:

$$s_0 \models \underset{\rightarrow}{\Box}(b{:}l \rightarrow \mathrm{E}_a(prog'_b \rightarrow b{:}l))$$

And if transition atom $a{:}l$ represents that a swerves left, then:

$$s_0 \models \underset{\rightarrow}{\Box}(a{:}l \rightarrow \mathrm{E}_a(prog'_b \rightarrow b{:}l))$$

Furthermore:

$$s_0 \not\models \mathit{Can}_b\,\neg\mathit{crash}, \quad \text{but } s_0 \models \mathit{Can}_b\,(\mathit{prog}'_b \rightarrow \neg\mathit{crash}))$$
$$s_0 \not\models \mathit{Can}_a\,\neg\mathit{crash}, \quad \text{but } s_0 \models \mathit{Can}_a\,(\mathit{prog}'_b \rightarrow \neg\mathit{crash}))$$
$$s_0 \not\models \underset{\rightarrow}{\Box}\,(\neg\mathit{crash} \rightarrow \mathrm{E}_a\,(\mathit{prog}'_b \rightarrow \neg\mathit{crash}))$$
$$s_0 \not\models \underset{\rightarrow}{\Box}\,(\neg\mathit{crash} \rightarrow \mathrm{E}_b\,(\mathit{prog}'_b \rightarrow \neg\mathit{crash}))$$

Notice that in this example we have had to rely on transition atoms to refer to the actions of a and b. I cannot see how we could do without them.

Temporal Structure

Let us now compare a model at a finer level of detail, by making explicit the temporal structure implicit in the term 'reacts to'. We will consider the 'actual behaviour' model. The 'explicit protocol' version can be constructed in similar fashion but adds little new so we leave it out.

In transition τ_2 of Fig. 11, b reacts by swerving left after a swerves left in transition τ_1. We have

$$\tau_2 \models [b]b{:}l \wedge \neg\mathrm{E}_b\,b{:}l$$

and hence at τ_1

$$\tau_1 \models \mathrm{E}_a\,1{:}\underset{\rightarrow}{\Box}\,b{:}l$$

So:

$$s_0 \models \underset{\rightarrow}{\Box}\,(a{:}l \rightarrow \mathrm{E}_a\,1{:}\underset{\rightarrow}{\Box}\,b{:}l\,)$$
$$s_0 \not\models \underset{\rightarrow}{\Box}\,(a{:}l \rightarrow \mathrm{E}_a\,1{:}\underset{\rightarrow}{\Box}\,\mathrm{E}_b\,b{:}l\,)$$

Indeed, in general the following are validities:

$$\models \neg\mathrm{E}_x\,1{:}\underset{\rightarrow}{\Box}\,\mathrm{E}_y\,\varphi \qquad (\text{any } x, y, \text{ including } x = y)$$
$$\models \neg[x]1{:}\underset{\rightarrow}{\Box}\,\mathrm{E}_y\,\varphi$$

This is because:

$$\models \neg\underset{\rightarrow}{\Box}\,\mathrm{E}_x\,\varphi$$

It is straightforward to derive this in the logic, or one can argue informally as follows. Suppose $s \models \underset{\rightarrow}{\Box}\,\mathrm{E}_x\,\varphi$. Then all transitions from s must have $\mathrm{E}_x\,\varphi$ true and hence also φ true and $\Diamond\neg\varphi$ true (by definition of $\mathrm{E}_x\,\varphi$). But if any transition from s has

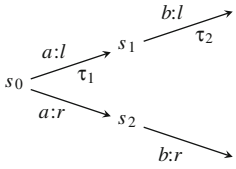


Fig. 11 *b* reacts to *a* (temporal structure, 'actual behaviour')

$\Diamond\neg\varphi$ true then it cannot be that all transitions from s have φ true, which contradicts the assumption.

So to recap: in the atemporal representation where the behaviours of a then b are modelled as atomic transitions

$$s_0 \models \underset{\rightarrow}{\Box}(a{:}l \rightarrow \mathrm{E}_a\, b{:}l)$$
$$s_0 \models \underset{\rightarrow}{\Box}(a{:}l \rightarrow \mathrm{E}_a \mathrm{E}_b\, b{:}l)$$

At this level of detail a brings it about that b brings it about that b turns left. But at a finer level of detail where we make the temporal structure explicit

$$s_0 \not\models \underset{\rightarrow}{\Box}(a{:}l \rightarrow \mathrm{E}_a\, 1{:}\underset{\rightarrow}{\Box}\mathrm{E}_b\, b{:}l)$$

Instead a's actions force b's reaction, in the following sense:

$$s_0 \models \underset{\rightarrow}{\Box}(a{:}l \rightarrow \mathrm{E}_a\, 1{:}\underset{\rightarrow}{\Box}\, b{:}l)$$

In the temporal model then, $b{:}l \leftrightarrow \mathrm{E}_b\, b{:}l$ is not valid. On a casual reading one might think it should be.

My point is that I can see no general principle why we should always insist on picking the most detailed model. Indeed, why should we think that there is a most detailed model? What looks like an atomic transition at one level of detail can always be decomposed into something with finer structure if we look closely enough.

9 Example: Granularity

This last example is to illustrate that granularity of a model does not always depend on temporal structure.

Suppose there are two agents a and b. Both can be in one of two rooms, left and right, separated by a doorway. The agents can stay where they are or pass from one room to the other, but not simultaneously (the doorway is too narrow).

The diagram on the left of Fig. 12 shows the possible transitions from the state where both agents are in the room on the left.

The possible actions by a and b in this state are as follows:

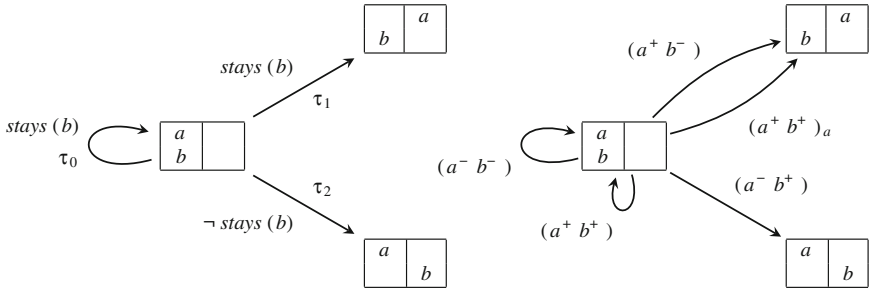


Fig. 12 The same example at two different levels of detail

$$\text{Actions by } a: \ \{\{\tau_0, \tau_2\}, \{\tau_1\}\}$$
$$\text{Actions by } b: \ \{\{\tau_0, \tau_1\}, \{\tau_2\}\}$$

There is no *stit*-independence in this model (a and b cannot both pass through the doorway at the same time):

$$\{\tau_1\} \cap \{\tau_2\} = \emptyset$$

Let transition atom *stays*(b) be true in transitions where b remains in the room on the left, as shown in the diagram.

Consider the transition τ_1 where a moves from left to right:

$$\tau_1 \models \mathrm{E}_b \mathit{stays}(b)$$

But also:

$$\tau_1 \models \mathrm{E}_a \mathit{stays}(b)$$
$$\tau_1 \models \mathrm{E}_a \mathrm{E}_b \mathit{stays}(b)$$

Indeed, if transition atom *moves*(a) represents transitions where the location of a changes from left to right, then (amongst other things):

$$\mathcal{M} \models \mathit{moves}(a) \rightarrow \mathrm{E}_a \mathrm{E}_b \mathit{stays}(b)$$

Let us now consider the same example, but at a greater level of precision. a and b cannot both pass through the doorway at the same time. Why? For the sake of an example, suppose that if both try then one of two things can happen: either both fail and stay in the room on the left, or just a succeeds in moving through, because a is a little stronger or faster than b, say. b can never get through ahead of a. (Many other versions of the example are possible).

The diagram on the right of Fig. 12 depicts a model at this finer level of detail. The labels on the transitions are just mnemonics. In (a^+b^-), a tries to move to the room on the right (and succeeds) while b does not try to move. In (a^+b^+) both a and b try to get through the door but neither succeeds. In $(a^+b^+)_a$ both try to get through

the door; a succeeds but b does not. In (a^-b^-) neither try. The possible actions of a and b in this state are therefore:

$$\text{Actions by } a: \quad \{\{(a^-b^-), (a^-b^+)\}, \{(a^+b^-), (a^+b^+), (a^+b^+)_a\}\}$$
$$\text{Actions by } b: \quad \{\{(a^-b^-), (a^+b^-)\}, \{(a^-b^+), (a^+b^+), (a^+b^+)_a\}\}$$

At this level of detail there is *stit*-independence in the model. (That does not always happen. Adding detail does not always produce *stit*-independence. It happens in this example).

Let the transition atom $stays(b)$ again represent those transitions where b stays on the left. *stit*-independence validates $\neg \mathrm{E}_a \mathrm{E}_b \varphi$ for all transition formulas φ. In particular, in this more detailed model of the example

$$\mathcal{M} \not\models moves(a) \rightarrow \mathrm{E}_a \mathrm{E}_b stays(b)$$

We still have:

$$\mathcal{M} \models moves(a) \rightarrow \mathrm{E}_a stays(b)$$

Evidently in this more detailed version of the model

$$\mathcal{M} \not\models stays(b) \leftrightarrow \mathrm{E}_b stays(b)$$

My point is that again important properties of the example change as detail is added. And it is not as though there is some most detailed model for which we should always aim. In the present example, a can sometimes get through the doorway ahead of b but not the other way round. We could also build a more detailed representation that models how that happens. So again: the models are different in some essential respects. We look to see which agent is responsible for, say, bringing it about that b stays where it is. At one level of detail, it is both a and b; at another level of detail it is only a. Indeed, it could be that at this level of detail, a brings it about that b does not bring it about that b stays where it is.

10 Conclusion

The purpose of the chapter was to explore how easy or difficult it would be to formulate some simple examples in a *stit*-like framework. I deliberately picked examples with a simple temporal structure. An element of indeterminism is present, either because of the uncertainty of the environment or because of the actions of other agents (for simplicity in these examples, at most one other). Here is a brief summary of the main points.

(1) An essential feature of the *stit* framework is that it does not refer explicitly to the actions performed by an agent but only to the way an agent's choices (intentional,

deliberative but also possibly automatic or unwitting) shape the course of future histories. The result is a very elegant and appealing abstraction which gives a natural denotation for actions whilst doing away with the need to identify and name them. The examples were intended in part to explore how easy it would be to exploit this abstract treatment. In the first few it worked out quite neatly. Here it was possible to identify and describe an agent's actions in terms of transitions of certain kinds between observable states, such as the location of a vase or whether a certain table was level or not. In other examples that does not work out so well. Often it is necessary to refer to the occurrence of a specific kind of action—jolting a table, swerving to the left, kicking an opponent—where the action cannot be picked out by reference to its effects on states. Perhaps dislodging a vase by lifting one end of a table is forbidden but causing it to fall by jolting the table is not. In these cases there seems to be no alternative but to introduce propositional atoms to name specific actions.

(2) We very often want to say that the actions of a particular agent are responsible for or the cause of such-and-such in a much weaker sense than is captured by typical *stit* or 'brings it about' constructions. Here it is the 'necessity' condition that is too strong. When an agent lifts a table and a vase standing on it falls and breaks, we want to say that the agent 'broke the vase': it was his actions that were responsible for the falling and the breaking, even though the vase might not have fallen when he lifted the table, and might not have broken when it fell. I looked briefly at two natural candidates for expressing a weaker sense of 'brings it about', which refer to what would, or might, have happened had the agent acted differently. One of these candidates is clearly much too strong (too demanding); the other is much too weak. It is far from clear that there is a way of expressing the required causal relationships using the available resources. I believe this is an important and urgent question because in practice it is precisely these weaker senses of responsibility and 'brings it about' that dominate.

(3) Sometimes an agent (human or artificial) behaves in some respects like an automaton, in that in some circumstances it follows a fixed protocol or decision procedure to select its course of action. It might do this unwittingly, as in the case of a reflex, or as a result of a long process of deliberation. Either way it seems very unsatisfactory to model this form of behaviour as if it were a fixed part of the environment in which other agents act. I suggested a simple device for distinguishing between modelling what I called 'actual' and 'explicit protocol' behaviours. I am sure there is much more that can be said about these matters, and about the formal relationships between models of these respective kinds.

(4) Finally, some of the key properties of the examples seem to depend critically on the level of detail that is being modelled. For some purposes it is perfectly reasonable to model, say, the moving of a vase from one place to another as an atomic transition with no further structure. For other purposes we might want to look more closely, and model in more detail how the vase is picked up, transported, and set down. For some purposes, we choose to model the movements of agents, physical robotic devices, say, as atomic transitions where unarticulated spatio-temporal constraints make certain combinations of movings impossible. At another level of detail, we model something of what these spatio-temporal constraints are: a doorway is too narrow to allow two

agents to pass through simultaneously, there is a single power source which some of the agents have to share, some of the agents are connected by inextensible physical wires, and so on. What we find is that at one level of detail, agent b sees to it that φ, and agent a sees to it that b sees to it that φ. When more detail is added, the same example says that b does not see to it that φ, and perhaps even that a sees to it that b does not see to it that φ. This is disturbing because these are precisely the kinds of properties that we want to examine. Of course there is nothing surprising about the fact that models at different levels of detail have different properties. Some properties are preserved as detail is added and some are not. There is a great deal of work on these matters, for example in the current literature on abstraction methods in model checking. However, the 'stit' and 'brings it about' patterns seem unusually sensitive to choice of detail. I would like to understand better how different models of the same example at different levels of detail relate to one another.

References

Balbiani, Philippe, Andreas Herzig, and Nicolas Troquard. 2008. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic* 37(4): 387–406.

Belnap, N., and M. Perloff. 1988. Seeing to it that: A canonical form for agentives. *Theoria*, 54:175–199. Corrected version in (Belnap and Perloff 1990).

Belnap, N., and M. Perloff. 1990. Seeing to it that: a canonical form for agentives. In *Knowledge representation and defeasible reasoing: Studies in cognitive systems*, vol. 5, eds. H.E. Kyburg, Jr., R.P. Loui, and G.N. Carlson, 167–190. Dordrecht: Kluwer.

Belnap, Nuel, and Michael Perloff. 1993. In the realm of agents. *Annals of Mathematics and Artificial Intelligence* 9(1–2): 25–48.

Belnap, Nuel, Michael Perloff, and Ming Xu 2001. *Facing the future: Agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Brown, Mark A. 1988. On the logic of ability. *Journal of Philosophical Logic* 17: 1–26.

Craven, Robert, and Marek Sergot. June 2008. Agent strands in the action language nC+. *Journal of Applied Logic* 6(2): 172–191.

Hilpinen, R. 1997. On action and agency. In *Logic, action and cognition—essays in philosophical logic: Trends in logic, Studia Logica library,* vol. 2, eds. E. Ejerhed, and S. Lindström, 3–27. Dordrecht: Kluwer Academic Publishers.

Horty, J.F. 2001. *Agency and deontic logic*. Oxford: Oxford University Press.

Horty, J.F., and N. Belnap. 1995. The deliberative stit: A study of action, omission, ability, and obligation. *Journal of Philosophical Logic* 24(6): 583–644.

Pörn, Ingmar. 1977. *Action theory and social science: Some formal models: Synthese library,* vol. 120. Dordrecht: D. Reidel.

Segerberg, K. 1992. Getting started: Beginnings in the logic of action. *Studia Logica* 51(3–4): 347–378.

Sergot, Marek. 2008a. Action and agency in norm-governed multi-agent systems. In *Engineering societies in the agents world VIII. 8th annual international workshop, ESAW 2007, Athens, Oct 2007, Revised selected papers*, LNCS 4995, eds. Artikis, A., G.M.P. O'Hare, K. Stathis, and G. Vouros, 1–54. Berlin: Springer.

Sergot, Marek. 2008b. The logic of unwitting collective agency. Technical report 2008/6, Department of Computing, Imperial College London.

von Wright, Georg Henrik. 1963. *Norm and action—a logical enquiry*. London: Routledge and Kegan Paul.

In Retrospect: Can BST Models be Reinterpreted for What Decisions, Speciation Events and Ontogeny Might Have in Common?

Niko Strobach

Abstract This chapter addresses two interrelated topics: (1) a formal theory of biological ancestry (FTA); (2) ontological retrospect. The point of departure is a reinterpretation of Nuel Belnap's work on branching spacetime (BST) in terms of biological ancestry. Thus, Belnap's prior choice principle reappears as a principle of the genealogical unity of all life. While the modal dimension of BST gets lost under reinterpretation, a modal dimension is added again in the course of defining an indeterministic FTA where possible worlds are alternatives in terms of offspring. Indeterministic FTA allows to model important aspects of ontological retrospect. Not only is ontological retrospect a plausible account for the perspectival character of Thomason-style supervaluations, but it is shown to be a pervasive ontological feature of a world in development, since it is relevant for cases as diverse as speciation, the individual ontogeny of organisms and decisions of agents. One consequence of an indeterministic FTA which includes the idea of retrospect is that, contrary to what Kripke famously claims, species membership is not always an essential feature, but may depend on the way the world develops. The chapter is followed by a postscript by Martin Pleitz and Niko Strobach which provides a version of indeterministic FTA that is technically even closer to Belnap's BST than the one in this chapter and which allows for a discussion of further philosophical details.

1 Introduction

This chapter is about a subclass of the structures which Nuel Belnap defined in his epoch-making 1992 article "Branching space-time"[1] (*BST*) and which have, therefore, been called BST structures ever since. BST structures have triggered an

[1] Belnap (1992).

N. Strobach (✉)
Westfälische Wilhelms-Universität Philosophisches Seminar, Domplatz 6,
48143 Münster, Germany
e-mail: nstro_01@uni-muenster.de

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_12,

impressive amount of interesting research. Still, as far as I'm aware, this chapter provides a novel interpretation of them. The reinterpretation is in terms of biological ancestry and links BST with the philosophy of biology. The present chapter is followed by a postscript, which was co-authored by Martin Pleitz and Niko Strobach. The postscript ties the modal version of the theory of ancestry even more closely to BST structures than the present chapter does and thus draws attention to a number of important features of a modal formal theory of biological ancestry which are not discussed in the present chapter.

In what follows, the suggested reinterpretation of BST structures will be expanded in certain respects. Thus, it is possible to link two relatively independent topics. One topic is a formal theory of biological ancestry, the other is ontological retrospect.

I shall proceed in two steps. In a first step, certain kind of BST structures will be reinterpreted as a certain kind of structures of the formal theory of ancestry (in what follows: FTA). I shall then briefly explain what can be done with FTA structures. The most characteristic feature of BST structures is the so-called prior choice principle. It will turn out that the prior choice principle corresponds precisely to a particularly important and intuitively controversial feature that may be added to FTA: the postulate of the unity of life. The fact that maximal directed subsets (MDSs) of BST structures may "branch" is crucial to the original space-time interpretation, because it is interpreted as *modal* branching of spatio-temporal histories. This modal character of MDSs is lost when those BST structures, which are suitable for FTA interpretation, receive a biological interpretation.

So, in a second step, I suggest adding the modal dimension again. Roughly, I will have structures of FTA play the role of histories in the *original* interpretation of BST structures.

This allows me to address the topic of retrospect. My main claim is that, at least sometimes, retrospect is not an epistemological, but an important *ontological* feature of reality which has, so far, been neglected. Finding this plausible presupposes pretty strong intuitions in favour of the temporal A-series.[2] If ontological retrospect appears plausible in itself, this will, by contraposition, provide some further support for an A-theoretical view of time.

I shall point out that ontological retrospect calls into doubt Kripke's thesis that belonging to the biological species to which one belongs is an essential property. I shall then point out how one might transfer the idea of retrospect to the very beginning of life. I shall consider retrospect in connection with speciation and in connection with the ontogeny of individual living beings (which might have implications for moral philosophy).

I conclude by indicating how *decisions* fit into the picture, a topic on which Nuel Belnap has made such an important contribution by developing STIT-models and by co-authoring *Facing the Future*.[3]

---

[2] Cf. McTaggart (1908).

[3] Belnap et al. (2001).

2 First Step: BST Structures and Structures of FTA

2.1 BST Structures

As is familiar to readers of Belnap's *BST*, a BST structure is an ordered pair which consists of a nonempty domain D, usually interpreted as a set of possible point events, and an accessibility relation $\leq$, usually interpreted as possible causal-influence-or-identity, which satisfies a number of postulates. In order to conveniently formulate the postulates, the following definitions are needed[4]:

Definition $<$	$x < y$ iff $x \leq y \wedge \sim x = y$;
Definition directed subset	m is a directed subset over $\langle D, \leq \rangle$ iff for any x, y from D in m there is a z from D in m such that $x \leq z$ & $y \leq z$ (*BST* D4);
Definition history / MDS	h is a history over $\langle D, \leq \rangle$ iff h is a *maximal* directed subset over $\langle D, \leq \rangle$ (*BST* D5);
Definition obviously undivided	histories h and h' (over $\langle D, \leq \rangle$) are obviously undivided at x (from D) iff there is a y from D such that $y \in h$ & $y \in h'$ & $x \leq y$ (*BST* D18);
Definition c[hoice] point	x (from D) is a choice point between h and h' iff $\{\{h, ...\}, ... , \{h',...\}\}$ is the finest partition of histories which contain x such that any h'', h''' from any element of the partition are obviously undivided at x (*BST* D19–21, 24).

The BST postulates are[5]:

BST postulate 1a	$\forall x\ (x \leq x)$	[reflexivity]
BST postulate 1b	$\forall xyz\ (x \leq y \wedge y \leq z \supset x \leq z)$	[transitivity]
BST postulate 1c	$\forall xy\ (x \leq y \wedge y \leq x \supset x = y)$	[antisymmetry]
BST postulate 2	for all x from D, all histories h, h' over $\langle D, \leq \rangle$: If $x \in h - h'$, then there is a y from D such that $y < x$ and y is a choice point for h and h',	[prior choice principle/PCP]

2.2 BTA Structures

Now let us isolate some core of a formal theory of ancestry (FTA). Let us call it the *basic* theory of ancestry: BTA. BTA structures are based on a two-place relation $<$.

[4] Belnap (1992), 390, 409.

[5] Postulates 1a to 1c are called postulate 1, postulate 2 is called postulate 28 in Belnap (1992), D is called OW.

Just in order to be able to read the formulae let us say that "$<$" is read "is an ancestor of". Let us add the following definitions[6]:

Definition $>$	$x > y$ iff $y < x$	[descendant]
Definition $\geq$	$x \geq y$ iff $y \leq x$	[descendant or identical]
Definition $\gg$	$x \gg y$ iff $x > y \wedge \sim \exists z\,(x > z \wedge z > y)$	[direct descendant]

Definition "BTA structure":
A BTA structure is an ordered pair which consists of a nonempty and finite domain D and an accessibility relation $<$ that satifies the following postulates[7]:

BTA postulate 1	$\forall xy\,(x < y \supset \sim y < x)$	[asymmetry, thus irreflexivity]
BTA postulate 2	$\forall xyz\,(x<y \wedge y<z \supset x<z)$	[transitivity]

Postulates 1 and 2 postulate a partial strong order on D with respect to $<$. Postulating a *finite* domain has at least the following consequences[8]:

BTA C1	$\forall xy\,(x < y \supset \ldots$ $\ldots\exists z\,(x < z \wedge z \leq y \wedge \forall w\,(x < w \wedge \leq y \supset z \leq w)) \wedge \ldots$ $\ldots\ \exists z'(x \leq z' \wedge z' < y \wedge \forall w\,(x \leq w \wedge w< y \supset w \leq z'))$	[discreteness]
BTA C2	For every x there are only finitely many y such that $y \gg x$.	[no infinity of direct ancestors]
BTA C3	For every x there are only finitely many x such that $x \gg y$.	[no infinity of direct descendants]
BTA C4	$\forall x\,(\sim \exists y\ x < y \vee \exists y\,(x < y \wedge \sim \exists z\ y < z))$	[endpoint(s)]
BTA C5	$\forall x\,(\sim \exists y\ x > y \vee \exists y\,(x > y \wedge \sim \exists z\ y > z))$	[starting point(s)]
BTA C6	There are only finitely many x such that $\sim \exists y\ y > x$.	[no infinity of endpoint(s)]
BTA C7	There are only finitely many x such that $\sim \exists y\ y < x$.	[no infinity of starting point(s)]

[6] Martin Pleitz has pointed out to me that, alternatively, it should be possible to base BTA structures on a primitive relation of direct ancestry or direct descent and to introduce a more general relation by definition. However, the results of both approaches do not to seem to be interdefinable in any simple and obvious way. One reason is the case of Antigone: On the alternative approach, Antigone would clearly have two direct ancestors. Possibly, the alternative approach is closer to branching space-time models which are based on local transitions (cf. Müller 2011) than to the models of Belnap (1992).

[7] Strobach (2010, 2011) contain the same postulates, except postulate 3, which is there presented in a weaker version which I now consider slightly too weak.

[8] Regarding BTA C1, cf. Goranko et al. (2004), 15. The formula means: To each ancestor x of y there is (1) some descendant z such that any w after (ancestor-wise) x and up to y *is* z at the earliest, and (2) there is some descendant z' such that any w after x from x on and before y *is* z' at the latest.

In what follows, BTA C1 to C7 will also be called the finiteness *postulates*. Note, however, that even all of them together will not *guarantee* a finite domain.[9] Although I think that BTA is basic, C1 to C7 are, in a way, arranged in diminishing degrees of intuitive basicness: C1 to C3 are absolutely essential to a biological interpretation, giving up C3 and C4 is very hard to imagine on a biological interpretation. Intuitively, C4 and C5 look somewhat less basic, perhaps C4 even less than C5 (C4 and C5 are independent of each other). Given C4, it is very hard not to accept C6, and given C5, it is very hard not to accept C7 on a biological interpretation.

A BTA structure enriched by the postulate of the unity of all life (BTA+U) is a BTA structure that satisfies the following additional postulate which forbids isolated substructures:

Postulate U $\forall$ xy (x $\neq$ y $\supset$ (x $<$ y $\vee$ y $<$ x $\vee$ ... [unity of life]
... $\exists$z (z $<$ x $\wedge$ z $<$ y) $\vee$ ...
... $\exists$z (x $<$ z $\wedge$ y $<$ z)).

2.3 BTA+U Structures are BST Structures

It is easily possible to establish the following purely formal result: The class of BTA+U structures is a subclass of the class of BST structures. The proof uses finiteness, BTA C4 (endpoints) and postulate U (unity of life). Hurried readers might like to skip it and continue with Sect. 2.4.

First, it is a standard result that the BST postulates 1a to 1c are equivalent to the postulates 1 and 2 of BTA. So in order to establish the mentioned result it is enough to show that the postulates for BTA+U imply the PCP.

Clearly, the BTA finiteness postulates could be added to the postulates of BST in order to single out a certain subclass of BST structures whose elements satisfy some extra constraints. BST structures need not be dense and may well contain "endpoints", i.e. some element x of D may satisfy $\sim\exists$y x $<$ y or $\sim\exists$y y $<$ x. It is true that, according to the original intended interpretation of BST structures, both alternatives seem rather strange, the first even more so than the second, but they are not excluded. Let us note some facts first:

Fact 1: There is only one way for an MDS of BST/BTA to end: in a single endpoint (according to the space-time interpretation of BST: a single big crunch event). If a subset of a BST or a BTA structure has a spliced end it will not be a directed subset, because splicing precludes a common upper bound. So either an MDS of BST/BTA

[9] Take the positive integers in their usual order (1 being the first element) and the negative integers in reverse order (-1 being the last element) and define that every positive integer precedes every negative one. This structure satisfies C1 to C7, but is infinite. It seems that finiteness is guaranteed by postulating finite chains.

does not terminate at all, or it terminates in a single element of D. So each MDS of BTA terminates in exactly one element due to BTA C4.

Fact 2: Postulate U implies that any two MDSs intersect. For due to fact 1, every MDS terminates in exactly one element of D. But the endpoints of *two different* MDSs have no common upper bound. So in order to avoid a violation of postulate U they must intersect somewhere else below.

Fact 3: If some item e belongs to an MDS, so must any predecessor of e in terms of $<$. If e is a member of some MDS h and if $e' < e$, then e' is a member of h, too. For if e is a member of h, but e' isn't, then there is a proper superset of h, i.e. $h \cup \{e'\}$, which is a directed subset of D. For, by transitivity, every common upper bound of e and some e'' from h is a common upper bound of and e' and e'', too. So h is no MDS, but was supposed to be one.

Now let $\langle D, < \rangle$ be a BTA+U structure, let e_1 be an element of D, h_1 and h_2 maximal directed subsets (MDSs) of D with respect to $\leq$. Assume the antecedent of the PCP, i.e. that $e_1 \in h_1 - h_2$. It is easy to see that h_1 and h_2 must be different from each other, and that $h_2 - h_1$ is nonempty. So there is some element of $h_2 - h_1$, which we may call e_2, which is different from e_1 and which does not belong to h_1.

Any two MDSs intersect. So h_1 and h_2 do. Clearly, neither e_1 nor e_2 is a common member of both of them. Can any successor of e_1 be a common member of h_1 and h_2? No, because if so e_1 would have to be a member of h_2, too. But we know e_1 isn't a member of h_2. Analogously for h_1 and e_2. So there must be a common member of h_1 and h_2 which precedes both e_1 and e_2, call it e_3.

Now, because of finiteness, there is a(t least one) *last* common $<$-predecessor of both e_1 and e_2, either identical with or after e_3, say e_c.

Now consider the finest partition of all MDSs which contain e_c such that it bundles obviously undivided MDSs at e_c. It must contain h_1 in one of the bundles and h_2 in another. So e_c is a c-point for h_1 and h_2. Also, e_c precedes e_1. So there is a c-point for h_1 and h_2 which precedes e_1. So the consequent of the PCP is satisfied on the assumption that its antecedent is satisfied. So the class of BTA+U structures is a subclass of the class of BST structures.

2.4 What Does it all Mean?

Why reinterpret (some) BST structures in terms of living beings? What is the intended interpretation of FTA structures? Certain notions in contemporary biology cry out for formal modeling, in particular cladistic notions like "most recent common ancestor" (concestor) or "last universal common ancestor" (LUCA).[10] In contemporary

[10] Some formal modeling of biology was attempted by the logically-minded biologist Joseph Henry Woodger between the 1930s and 1950s. Woodger makes some natural assumptions which are built into BTA, too. However, the FTA presented here (with BTA as its basic version) is without any reference to Woodger. Cf. Woodger (1937). Some summary of Woodger is contained in Carnap (1958). I am grateful to Barry Smith for drawing my attention to Woodger.

bioinformatics, phylogenetic trees are usually reconstructed by using algorithms which include pretty specific constraints on branching structures, e.g. that every parent node has exactly two daughter nodes.[11] That is fine for cladograms, but one should have something much more basic and more general which should yield the usual trees only after adding quite a lot of constraints.

The notions of concestor and LUCA presuppose that one species may be called the ancestor of another species. What this means is not quite as clear as one might wish. It is appropriate to start from the ancestor-relation between individual living beings. This is a relation we are well-acquainted with. The intended interpretation is quite broad, though: Not only are parents ancestors of their children, but also, literally, parent cells of daughter cells.[12] We are well-acquainted with individual living beings, whose existence is beyond doubt (the same cannot be said of species). So our domain D is interpreted as containing living beings. They may be multicellular or unicellular, the size of a bacterium or the size of a whale, plant or animal, reproducing sexually or asexually. I shall not try to define what a living being is. My approach will, however, show some affinity[13] towards a recursive definition with an ostensive base: "We are living beings, all ancestors of living beings are living beings and all descendants of living beings are living beings, and nothing else is a living being." The tricky bit is the final clause of the recursive definition "...and nothing else is a living being". I am sympathetic with it. No angels. And, as will become clear later on, no Martians either.

BTA contains no postulates which preclude forward or backward branching. Individual biological ancestry is a network. It possesses nothing like the maximally fine twigs of Prior's tempo-modal trees, even though the network of some BTA structure may, by and large, be tree-shaped, if you look at it from afar. Not only does a living being often have several direct descendants, but also several direct ancestors, in the case of sexual reproduction: usually two (but beware of Antigone!).[14]

BTA C4 and C5 deserve a little extra comment. C5 says that every living being has either no ancestor or is a descendant of some living being that had no ancestor. One might motivate this by saying that BTA structures are supposed to be local and that the primordial soup is out of focus. But although one might do so, I am rather up to some large-scale modeling of all the life that there has been so far. If life had been going on forever, as Aristotle thought, and, thus, the domain is infinite, BTA C5 would be false even on the largest scale (although even Aristotle would have allowed for starting points in the structure due to spontaneous generation).[15] Even today,

[11] Cf. Gusfield (2007).

[12] Even ancestors of endosymbionts may be called ancestors of what, in later generations, they will be endosymbionts in. That depends on what turns out to be the best description of the fusion of host cells and endosymbionts.

[13] I am not claiming that there is a one-to-one correspondence between this definition and postulate U. There isn't.

[14] Antigone has only one direct ancestor: Oedipus. For her mother, Iokaste, has a descendant, her son Oedipus, who is an ancestor of Antigone, i.e. her father.

[15] For a detailed comparison of Aristotelian and contemporary biology in terms of structural postulates cf. Strobach (forthcoming).

we might wonder: Must there really have been ancestor-less living beings? Yes, wherever in the vague morning haze of evolution the horizon of life may be hidden. For suppose, BTA C5 were not true. Then there would be at least one topologically infinite lineage of living beings, which, since every reproductive step takes some finite time and there is some lower bound to the length of such a time,[16] would also be metrically infinite. So there would have been living beings before big bang, which is absurd.

BTA C5 comes naturally along with BTA C7: There is no infinity of ancestor-less living beings. They neither presuppose nor preclude one single ancestor-less living being, the one and only primordial cell. It is probable that the primordial soup was boiling on more than one stove. Life grew together.[17]

The least obvious principle is BTA C4: Every living being has either no descendants or is an ancestor of some living being that has no descendants. This makes BTA structures topologically finite towards the reproductive future. One might motivate this by the fateful certainty of the sun running out of fuel some day in the future, or by big crunch or big chill scenarios. Large-scale modeling can be depressing. My reason for BTA C4 is rather pragmatic: let us model life up to now. Or up to any time in the past we choose (as long as it contains life). Remember, by the way, that the overwhelming majority of all living beings that ever existed never reproduced, which already makes for lots of endpoints in a BTA structure. MDSs of BTA structures, on the biological interpretation, are just the set of all ancestors of some descendant-less living being.

No temporal sequence is modeled directly. No life-spans are modeled. There is no way to express the fact that individual lives are finite. It is true that remote ancestors will not coexist with their remote descendants. But there isn't even so much simultaneity modeled in a BTA structure to express this.

2.5 The Unity of Life

Postulate U had better be separated from the basic parcel of axioms which have been explained. "U" is supposed to abbreviate "unity of life", and that is what it is about on the intended interpretation. It says: Any two living beings are related by the ancestor relation (one way or the other) or have a common ancestor or have a common descendant.

Note that the class of BTA structures without restrictions is *not* a subclass of the class of BST structures. The class of BTA structures contains structures in which some maximal directed subsets do not intersect. This cannot happen with a BST structure. On the usual BST interpretation this would mean that there are several causally unrelated bundles of histories, several parallel universes with all their modal

[16] The last clause precludes, as Martin Pleitz put it in conversation, "inverse Thomson lamps".

[17] There is some tension between BTA C5 and postulate U, if there is no single primordial cell. It will be resolved later on, in the section on retrospect.

branching, which are part of the same structure. On the intended large-scale BTA interpretation this would mean that there are several independent trees or networks of life.

My own point of view on the matter is a bit complicated. I am not even entirely happy with the prior choice principle on the original BST interpretation.[18] Although I am against postulating the prior choice principle in the manner of Belnap (1992) if one interprets BST structures the usual way, I prefer BTA+U to BTA without U. However, in discussions the following points have been raised against postulate U:

(1) Postulates of a supposed FTA should be conceptual truths, uncontroversial truths about the use of the word "life". But it does not belong to the meaning of the word "life" that all life is genealogically coherent.
(2) If we met Martians pretty similar to us, but genealogically completely unrelated to us, how could we possibly deny that they are living beings?

So it seems that the largest plausible scale of an intended interpretation of a BTA+U structure can be life-on-earth.

However, as to (1), I am not sure if there are such things as conceptual truths independent of any background theory. If not, why not take the best background theory we have today? Of course, this does not yet settle the point. Did 20th century biology discover the unity of all life? Is this anything a science can empirically discover? If so, how about the Martians?

I want the Martians out. I tend to deny that they are living beings. Both objections to postulate U might be symptoms of a profound misunderstanding of how the word "life" should be understood. Defining the word "life" by a set of criteria has never really worked. It rather seems that if any term works like a natural kind term the way Kripke thinks, then "life" is a good candidate for a natural kind term. In fact, "life" might even work better as a natural kind term than the species terms of traditional biology or folk biology. It is a very amazing fact that what we have been pointing to as life all the time has indeed turned out to be one kind of thing. "Life" might even be a proper name that refers to a single object which is coherent in four-dimensional space-time. That does not necessarily contradict its being a natural kind term. Even natural kind terms like "gold" might best be interpreted as proper names, not just as

[18] Strobach (2007a), 219ff. Recent work by Müller (2011) and by Tomasz Placek (in the present volume) on BST has refrained from postulating choice points in the manner of Belnap (1992) in order to make BST more "GR-friendly" (GR = general relativity), for first points of divergence suit GR topology better than last points of coincidence. I think that there are independent metaphysical reasons for preferring first points of divergence: branching is nothing that takes place, but world history develops by zillions of local decisions and thus continuously excludes possible alternative developments it might have had by the course it takes. Decisions do not take place at instants/events but by events occurring and thus not failing to occur. That picture suggests first points of divergence. So I welcome the result BST research has reached for different reasons than the ones I gave in Strobach (2007a). I suspect that if the PCP is abandoned, also the inclusion of the "wings" as a necessary feature of BSTs is gone (I would welcome this, too). At least the proof of fact 31 given in Belnap (1992), 411, seems to rely on the PCP.

something close to proper names.[19] We might owe the Martians some respect if they are capable of suffering, though. Still, I do not think they would be living beings if "life" is the kind of term I think it is.

There are, however, at least two serious counter-objections to postulate U.

(1) Why not fix the reference of the term "life" by pointing to life on planet earth and find more of it on Mars? Maybe the Martians and us would be like H_2O and XYZ if they did not share the same microchemistry with us. But what if the Martians do not just look alive, but even share their microchemistry with us, while it is beyond doubt that we have no common ancestors with them? So at most, postulate U can be is a risky and contingent assumption.
(2) How can "early" sections of a BTA+U model without a primordial cell, when life has not yet grown together, be models of *life*? But if a BTA+U structure is supposed to be a model of life, how can an early, incoherent, section of it fail to be one?[20]

The second objection can only be dealt with in the section on retrospect. As to the first objection: Wouldn't we say that the situation in which the Martians even have DNA etc. would be one where life on earth clearly isn't all the life there is? That is not so clear. One might even consider dismissing the scenario as just too silly. However, dismissing any scenario which structurally resembles the one with Martian DNA as just too silly, would be too easy a way out. The progress of so-called synthetic biology might soon cause a situation which is, in principle, not too dissimilar from the scenario. As long as existing cells are reprogrammed, postulate U remains plausible. Once cells with the same microchemistry as life can be built from the scratch, postulate U might have to be reconsidered. My opinion is that one should be on the cautious side when it comes to calling them instances of life. At least we should be clear about the fact that we might be facing a fundamental conceptual decision in a few years and that subsuming artificial cells under the term "life" is not a matter of course. Furthermore, clearly, if there is some essential property which

[19] The relevant passage is Kripke (1980), 127: "[T]erms for natural kinds are much closer to proper names than is ordinarily supposed. The old term 'common name' is thus quite appropriate for predicates marking out species or natural kinds, such as 'cow' or 'tiger'. My considerations apply also, however, to certain mass terms for natural kinds, such as 'gold', 'water', and the like." But neither "gold" nor "tiger" is a predicate. "…is a portion of gold" and "…is a tiger" are. "Gold" and "tiger" are singular terms. They are proper names for natural kinds, not just something close to names, while natural kinds are individual elements of the universe of discourse.

[20] As Martin Pleitz has remarked to me in conversation, there are even extreme cases in which a structure may "lose" the property of satisfying postulate U again by acquiring an additional descendant. Think of expanding an N-shaped structure with ancestry downward into an M-shaped structure. This shows that postulate U is very strong. Still, I do not think it is unrealistically strong. Why is it so strong anyway? It is not only related to the PCP, but it is the minimal condition you need for, once an object language has been defined, being able to highlight the whole structure from a certain context by using sequences of quantifier-like modal operators without the help of actuality operators. In fact, postulate U came to my mind as a constraint on models of modal logic in connection with Crossley and Humberstone (1977) investigation of "actually" and Kienzle (2007) investigation of non-isolated structures of modal logic. This does, of course, not help its philosophical motivation.

a and b share, a and b need not be numerically identical. Any two different human beings in 2013 might serve as a counter-example. So even if being DNA(etc.)-based were an essential property of both life (i.e. life on earth) and of the Martians' way of being, that would not force somebody to admit that there was life on Mars, too. Life might be very special in a combination of a number of respects, describable or not even describable; or it might be special just due to its very continuous history, which might be termed the maximal biography.

Suffice it to say that the question of the status and the acceptability of postulate U is not easily settled and involves fundamental conceptual issues concerning the word "life". Interestingly, of all the conceivable postulates of FTA, the one that raises the most difficult questions is the one that formally corresponds to Nuel Belnap's prior choice principle.

2.6 What Else Can be Done with BTA?

Here is some very brief impression of what else can be done on the basis of BTA.[21]

(1) Independently of U, BTA structures may be expanded to BTA+CS by adding a second primitive relation, the relation of being conspecific, i.e. of being of the same species. Conspecifity should be postulated as symmetric, but *not* as reflexive (if mules don't belong to any species, no mule is conspecific with itself). It is also plausible to postulate that if a living being is conspecific with itself then there is some *other* living being with which it is conspecific (i.e. that nothing is *sui generis*). The transitivity of conspecifity is a tricky issue. It ensures that no living being belongs to more than one species in a BTA+CS model. However, a transitivity postulate might cause trouble in connection with ring species (like the sea-gulls around the arctic) or with historical borderline cases.

(2) Somehow the members of a species extension cohere genealogically. This would even be the case if life as a whole did not. So, quite independently of postulate U, an analogue of postulate U, and thus, again, of the prior choice principle, should be postulated. However, the simple analogue to postulate U does not suffice for the kind of genealogical coherence one has in mind for the members of a species. For it does not preclude alien intermediate generations and is, thus, not tight enough to conform to our intuitions. However, some additional postulate does the job.[22]

(3) It is possible to give a clear account of what it means that a species is an ancestor of some other species in the context of a BTA+CS structure. Take the following definition:

---

[21] Points 1 to 5 are discussed in detail in Strobach (2011), point 6 is the topic of Strobach (2010).

[22] I suggest: "If x is a conspecific ancestor of y, then x has a direct descendant that is conspecific with both x and y and which is an ancestor of or identical with y." Cf. Strobach (2011).

A biological species s is a species ancestor of some biological species s‘ iff

(1) every organism that belongs to s‘ is a descendant of some organism that belongs to s and
(2) no organism belonging to s is a descendant of any organism that belongs to s‘.

BTA+CS provides the resources to explain why the relation which is thus defined is an ancestor relation: It satisfies the conditions which were postulated for the individual ancestor relation when stating the definition of a BTA structure, including the finiteness postulates. An analogue to postulate U cannot be deduced, and plausibly so: the beginning of life may have been species-less for a long time, and completely independent species trees may have grown out of the same origin of life.

(4) It is remarkable that the conditions for species-ancestry can all be expressed as, albeit very long and convoluted, statements about conspecific individuals. This establishes the possibility of being a nominalist about biological species. Thus, a bit of homework from Quine's "On What There Is" could finally be done.[23] Although nominalism about species is possible, the fact that the reconstruction is so complicated might itself rather be an argument for accepting species.
(5) Sometimes species fuse. However, if this is deliberately disregarded as a rare phenomenon, it can be shown that no more backward branching of the ancestor relation between species is possible, but that the ancestor relation between species is semi-linear like the accessibility relation of a Prior-style modal tree.
(6) It is possible to define a non-reductionist multi-layer ontological structure with species on top, living beings in the middle and cells on the ground-floor, all of them being admitted to the domain of discourse. Now the postulates must be sorted. There are bridge principles between the different levels like: "x is a species iff it has members" (while members must be living beings, or "x is a living being iff it at least one y is a cell of x"[24]). There is a far-reaching analogy between the relations between living beings and species on the one hand and the relations between cells and living beings on the other. Just as there is the extension of a species there is the cell-extension of a living being: the set of all cells that ever belonged to it. There is an ancestor relation for cells which intuitively satisfies the BTA postulates. Just as life as a whole may be imagined as a huge ancestral network of living beings, life may be imagined as an ancestral network of cells. Cell-extensions of organisms are themselves coherent subnets of this structure (often huge, though minimal in the case of unicellular living beings, whose cell extensions are their singletons). It is now possible to *define* what BTA took as basic, i.e. the ancestor relation between living beings, in terms of the ancestor relation between cells, and to do so completely analogously to

[23] Quine (1951), 13: "When we say that some zoological species are cross-fertile we are committing ourselves to recognizing as entities the several species themselves […]. We remain so committed at least until we devise some way of so paraphrasing the statement as to show that the seeming reference to species on the part of our bound variable was an avoidable manner of speaking."

[24] Viruses are tricky in this respect.

the way the ancestor relation between species is defined in terms of the ancestor relation between living beings. There are, however, good independent reasons for not being a nominalist about living beings in spite of this result. Cell-extensions which originate in one single cell are particularly interesting. For we are among those living beings whose reproductive cycle typically goes through single-cell bottlenecks. However, this is far from being a universal feature of life.

Future work might involve the following points:

(1) Adding a gene layer (which might be a difficult task).
(2) Knitting BTA structures or even multi-layered FTA structures onto histories of BST structures in their original space-time interpretation. A living being would then correspond to a small worm-shaped subset of point events of a history.[25]
(3) Enriching FTA structures by some explicit modeling of temporal order.

3 Retrospect

3.1 The Story so Far

Let us turn to the topic of retrospect. Here are some examples of it:

(1) There is no such thing as a photograph taken at an instant: Opening the shutter for zero seconds would be just too short to take a picture. Although cameras and human beings differ in that human beings are (self-)consciuos and cameras are not, I do not think that they differ in this respect: sensations, thoughts, feelings of awareness are time-consuming. If that is so, we experience changes *in retrospect* with our backs turned towards our future.[26]
(2) As to the problem of future contingents, Thomason-style supervaluations[27] are to be preferred over all other solutions that have been proposed. Statements about future contingents lack a truth-value, while future necessities are already true; claiming that a future contingent is not only true, but even settled in advance turns out false. This is already quite a nice combination of attractive results which is not at all easy to achieve (in the 1950s Quine famously called it a fantasy).[28] But there is even more to supervaluations: In the case that a certain contingent event did take place, *in retrospect*, it was the case that this former future contingent event was going to happen, and it is then settled that it was going to happen, although it was not necessarily going to happen. Formally, this result is achieved because if you evaluate the statement "It was going to be the case that p" *post*

[25] These subset would probably resemble the "Vorkommnisse" in Kienzle (2007), which are, regrettably, restricted to one dimension.
[26] Strobach (1998), 201–234.
[27] Thomason (1970).
[28] Quine (1953).

festum you do so at a position in a branching tempo-modal structure à la Prior[29] where the event in question has occurred, so only such branches on which it has occurred are taken into account for the evaluation of the statement. This sensitivity to positions is a marvelous feature of supervaluations.

(3) This feature can be transferred to branching relativistic space-time. I have argued that, as long as it is in the space-like of a given event, even what happens at a "spatially" remote position from my position in space-time is ontologically undetermined with respect to my position, because positional necessity and contingency should be generalised to spacetime. *In retrospect*, however, once the position in question has entered the past light-cone of my world-line it is true to say that the event occurred. It occurred without ever having been occurring.[30] Putnam, for instance, finds this absurd[31]; but it is plausible.

Might ontological retrospect play a role in connection with our formal theory of biological ancestry, FTA? I think so. Can it be modeled starting off from BTA structures by reintroducing a modal dimension? Here is how it might be done.

3.2 Theory of Possible Ancestry (TPA)

A TPA structure is a nonempty set of BTA structures $\{\langle D_{P1}, <_{P1}\rangle, ...\langle D_{Pn}, <_{Pn}\rangle\}$, which, as components of a TPA structure, will be called Possibilities (with a capital "P"), such that the following condition is satisfied:

$$\text{TPA 1} \quad \forall x\, \forall y\, \forall P\, \forall P'(x <_P y \wedge y \in D_{P'} \rightarrow x <_{P'} y)$$

Note that the following is clearly equivalent to TPA 1 (swop "<" and ">", then "x" and "y"):

$$\text{TPA C1} \quad \forall x\, \forall y\, \forall P\, \forall P'(x >_P y \wedge x \in D_{P'} \rightarrow x >_{P'} y))$$

Note furthermore that, of course, if some x stands to some y in the relation $<_P$, then both x and y have to belong to D_P. So it follows from TPA 1 that

$$\text{TPA C2} \quad \forall x\, \forall y\, \forall P\, \forall P'(x <_P y \wedge y \in D_{P'} \rightarrow x \in D_{P'})$$

Finally, note that TPA C2 may be rewritten, using the import / export law of propositional logic, as

$$\text{TPA C3} \quad \forall x\, \forall y\, \forall P\, \forall P'(x <_P y \rightarrow (y \in D_{P'} \rightarrow x \in D_{P'})),$$

which yields, by contraposition,

---

[29] Prior (1967), chap. 7.

[30] Strobach (2007a). Summary: Strobach (2007b), cf. also Müller and Strobach (2011).

[31] Putnam (1967).

$$\text{TPA C4} \quad \forall x\, \forall y\, \forall P\, \forall P'(x <_P y \rightarrow (x \notin D_{P'} \rightarrow y \notin D_{P'})),$$

which yields, again by import / export,

$$\text{TPA C5} \quad \forall x\, \forall y\, \forall P\, \forall P'(x <_P y \wedge x \notin D_{P'} \rightarrow y \notin D_{P'}).$$

To underline the similarity between BST and TPA in spirit, if not in technical detail, one might say that a is a choice individual between P and P′ iff

$$a \in D_P \wedge a \in D_{P'} \wedge \sim \forall y(y \gg_P a \equiv y \gg_{P'} a)$$

TPA might be strengthened in the following way: Use BTA+U structures instead of BTA structures and add

$$\text{TPA 2} \quad \exists x\, \forall P\, x \in D_P$$

Call the result a TPA+U structure. A TPA*+U structure is an TPA+U structure which, in addition to TPA 2, even satisfies the following, slightly stronger condition:

$$\text{TPA 2}^* \quad \exists x\, \forall P(x \in D_P \wedge \forall y(\sim \exists z\, z <_P y \rightarrow x \geq_P y))$$

The very same individual may be a member of the domains of different Possibilities of a TPA structure. Condition TPA 2 even ensures that this is the case for at least one individual. Possibilities are not maximal directed subsets. There is no single ancestor relation across Possibilities, but there is a whole family of them, one per Possibility, which satisfies the usual BTA+U conditions. They are, however, closely related.

Roughly speaking, Possibilities of TPA structures are possible worlds. If you have a close look at them, though, they turn out to be less fine-grained than possible worlds, for there are more properties of living beings than just having such and such descendants. However, TPA structures focus entirely on this property. So different Possibilities in a TPA structure are alternatives in terms of offspring and in terms of nothing else.

The user of TPA structures should be willing to confess to a certain naïvité concerning future and/or possible individuals, at least while using these structures. Anyway, in different Possibilities, different things happen, and thus different individuals exist: In P_1 a has children b and c with d and no children with anyone else; in P_2 a stays single and never has children; in P_3 a marries e instead of d and has children f and g with e; in P_4 a marries and has children b and c with d plus another child with h; in P_5 a has a child c with d, but no b makes it to existence; in P_6 a, instead of having children b and c with d, has children j and k with d. TPA structures allow for all that. However, they respect the Kripkean idea of the necessity of origin,[32] because it is highly plausible: If a is b's ancestor, a will not just exist in any Possibility in which b exists, but will also be b's *ancestor* in any such Possibility (TPA 1). TPA C1 says: If b is a descendant of a in P_1 and b exists in P_2, b must also be a descendant of a in

[32] Kripke (1980), 112f.

P_2. It may, however, happen, that b is a descendant of a in P_1 and b does not exist in P_2. In that case, b is, of course, not a descendant of a in P_2. In fact, the minimal deviation of two Possibilities would be that some individual just has one descendant less, *everything else being equal*. If a exists in P′, so must all of a's ancestors (TPA C3). For anyone with different ancestors could not be *a*. If a doesn't exist in P, neither will any of a's descendants (TPA C4). For *they* could not fail to be descendants of a.

TPA 2 postulates that there is at least one individual that exists in all Possibilities of the structure. Because of TPA 1 it cannot do so on its own if it has any ancestors, but they will exist in all Possibilities, too. So, as the prior choice principle in BST guarantees coherence of histories (according to the original interpretation), TPA 2 guarantees coherence of Possibilities in TPA+U structures.

Can we carve alternatives out of the structure rather than investing them? Answering this question is the topic of the postscript to the present chapter.

3.3 The Growth of Life Itself

TPA 2* implies TPA 2, but is stronger. Both postulates do not differ if there is a primordial cell. But if there are several ancestor-less individuals, TPA 2* can be false while TPA 2 is true. For TPA 2* postulates that all Possibilities have an individual in common which comes so late that all ancestor-less individuals are among its ancestors: a first common descendant of all origins of life.

This takes us back to a curious consequence of postulate U, the postulate of the unity of life. Consider a BTA+U structure with several ancestor-less individuals. Consider an "early" substructure of it, which does not yet contain a common descendant of *all* of them. This substructure will be a BTA structure, but not a BTA+U structure.

How are we to interpret this result? Here is a suggestion (maybe controversial): Before the occurrence of a first universal descendant, life did not exist. But neither did it come into being when a first universal descendant occurred. Rather, once the first universal descendant occurred, it happened that, *in retrospect* each of its ancestor-less ancestors became an origin of life, and life started with the earliest of them.

3.4 Speciation

How about adding conspecifity to TPA structures? Clearly, an TPA+CS structure would have to be a set of BTA+CS structures which satisfy at least the same constraints as the components of TPA structures. A natural question to ask is: Should there be a bridge principle which makes species membership an essential property, just as suggested by some famous examples in Kripke's *Naming and Necessity*?[33]

[33] Kripke (1980), 125f., 147.

Given the rest of TPA+CS, such a principle is easy to state. Let us call it the principle of species membership as an essential property, SMEP:

$$(\text{SMEP})\forall x\ \forall y\ \forall P\ \forall P'(x\ CS_P\ y \wedge x \in D_{P'} \wedge y \in D_{P'} \rightarrow x\ CS_{P'}\ y)$$

If x and y belong to the same species in P and both exist in P′, they also belong to the same species in P′. That this renders the intuition that species membership is essential becomes particularly clear in cases where x = y. But should SMEP be added as a postulate? Kripke is right in that I could not possibly be a lion. Still, there is some reason for rejecting the principle that has just been stated.[34]

Take a couple of birds that makes it to a remote island. They are the beginning of a founder population which flourishes and, after a while, diverges so considerably from their mainland cousins that they could not mate with them any longer, so a new species was born. Are there any first members of the new species? According to the story just told, I should say that the first two birds on the island are good candidates. But they do not differ in any way from their direct ancestors on the mainland, so must they not belong to the same species as their ancestors do? My favorite account of the situation is this: If they had not founded the new population they would have belonged to the same species as their parents. But since they did, they don't. That is, in P_1 they don't, which is supposed to be an alternative in which they were successful founders. They become the first members of a new species *in retrospect*, once the new species has developed. But once it has, it is true that the new species started with them and not any later. But take P_2 in which they starve on the island without leaving any descendants. Clearly, in P_2 they belong to the same species as their parents. So species membership is not an essential property, but may vary from possible world to possible world. So SMEP should not be a postulate of TPA+CS. Should we even say in retrospect that the founders changed species membership during their lifetime? Probably we should.

3.5 Individual Ontogeny

Let us switch to the cell level. BTA can be extended to a multi-layer FTA that includes a cell level. What would its modal version, a TPA with cells, look like? Again, a natural question is if there is any cross-Possibility bridge principle. My proposal is that there is at least one such a principle, but that it is weaker than the one that first comes to mind. According to Kripke, there is a microscopic version of the principle of the necessity of origin.[35] I could not have originated from a different sperm and egg. Let us focus on living beings like us whose reproductive cycle includes single-cell bottlenecks. Let us define what a first cell is. The definition presupposes the relation

[34] The story ignores vague boundaries. That might be a mistake. Perhaps species talk functions pretty differently at the end of the day.

[35] Kripke (1980), 112f.

CellOf, which, in turn, holds only between a cell and a living being and is structurally similar to the relation of species membership:

$$\text{x FCO y iff x CellOf y} \wedge \sim \exists\, \text{z}(\text{z} < \text{x} \wedge \text{z CellOf y})$$

Once we move on to the modal version, all the relations have Possibility parameters. Now consider the following principle:

$$(\text{origin1})\ \forall \text{x}\, \forall \text{y}\, \forall \text{P}\, \forall \text{P}'\ (\text{x FCO}_{\text{P}}\ \text{y} \rightarrow ((\text{x} \in \text{D}_{\text{P}'} \equiv \text{y} \in \text{D}_{\text{P}'}) \wedge (\text{x} \in \text{D}_{\text{P}'} \rightarrow \text{x FCO}_{\text{P}'}\ \text{y})))$$

If x is a first cell of y in P, then x and y either coexist or both fail to exist in any P′, and if x exists in P′, x is a first cell of y in P′, too. So clearly, in every alternative in which y exists, x is its first cell; and in every alternative which x exists, y exists already because x exists, being y's first cell.

While it is nice that this principle can be stated within the framework that has been presented so far, I think it is too strong to merit acceptance. Like in the cases of life or speciation, something might have to reach a certain size before y exists. If y comes into existence, nothing speaks against some cell's being y's first cell *in retrospect* which would otherwise not have been a cell of *any* living being. I started from a zygote, which is the first cell of the set of all cells that were, are or will be cells of my body. After things turned out fine it is even true to say that *I* started of *as* a zygote. But only in retrospect. Had the blastula into which the same zygote turned by cell division been destroyed I would never have existed. This is no contradiction. The point may have implications for moral philosophy, which might concern PID or the very difficult issue of abortion. I shall not pursue them here. But let me state a weaker principle than the one above, which I do think is plausible at least in connection with single-cell bottlenecks:

$$(\text{origin 2})\ \forall \text{x}\, \forall \text{y}\, \forall \text{P}\, \forall \text{P}'(\text{x FCO}_{\text{P}}\ \text{y} \wedge \text{y} \in \text{D}_{\text{P}'} \rightarrow \text{x} \in \text{D}_{\text{P}'} \wedge \text{x FCO}_{\text{P}'}\ \text{y})$$

If y exists in P′, then so must x, and x must be first cell of y in P′, too. But this does not rule out a Possibility in which y never exists and x exists but isn't the first cell of anything.[36]

4 Afterthought: Resuscitation and Decisions

To conclude, let me mention two more possible applications of the idea of ontological retrospect.

[36] The same point is argued in Strobach (2010) without explicit modal modeling. Possibilities are quite useful to clearly state the difference between the two positions, which can only be done using a version of TPA.

(1) The first example is related to biology. Perhaps it is outdated in the days of brain death. The idea goes like this: It is a world-relative matter (in terms of temporally structured possible worlds) whether or not a certain event is the death of a certain living being. If a patient is successfully resuscitated the very same event will not be the patient's death which, if resuscitation fails, will *in retrospect* be the patient's death in the sense that *that was when* the patient died.
(2) The last example is about decisions. At least, it belongs to the range of the common use of the term "decision", although it is probably not about anything which is called a decision in connection with STIT models. It is about a mental state which will, later on, be called the decision or perhaps, more cautiously, "what I felt the moment when I knew I had made up my mind". It is not completely far-fetched to call a decision what I remember as one. I claim that, in this sense of the word "decision", decisions are retrospective events, i.e. events which acquire the status of being decisions only in retrospect and contingently. One might think that a decision in this sense necessitates the action. But this is actually not true: "Only the execution of the intention provides it with the stamp 'decision' ", says Schopenhauer[37]. Never mind that Schopenhauer was wrong in that he was a determinist.[38] This is a point he got right: As long as nothing has been done, someone or something might interfere, or I might interfere by hesitating and suddenly starting to reconsider and reevaluate. If nobody and nothing interferes and I act, my "decisive feeling" becomes the decision *in retrospect*. Then that was indeed when I decided. If something interferes the very same mental event never made it to be a decision. So the same event may be a decision in one possible world history and not be one in another.

5 Summary

To sum up, I have argued (1) that a certain subclass of FTA structures is identical with a certain subclass of BST structures; (2) that the feature of FTA which formally corresponds to the prior choice principle of BST is a fundamental principle of the unity of life; (3) that the basic theory of ancestry BTA may modalised in such a way that it incorporates the principle of the necessity of origin; (4) that an account of speciation along the lines I have suggested calls into doubt the idea that species-membership is always an essential property; (5) that retrospect is an ontological feature of reality; (6) that the beginning of life on earth, speciation, individual ontogeny, death and decisions involve retrospect.

[37] Schopenhauer (1818), 152 [WWV I 1 §18]: "Nur die Ausführung stämpelt den Entschluss, der bis dahin noch immer veränderlicher Vorsatz ist".

[38] Schopenhauer (1839), 372–383 [section II].

References

Belnap, N. 1992. Branching space-time. *Synthese* 92: 385–434.
Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the Future*. Oxford: Oxford University Press.
Carnap, R. 1958. *Introduction to symbolic logic and its applications*. New York: Dover (translation of *Einführung in die symbolische Logik*. Wien: Julius Springer 1954).
Crossley, J.L., and I.M. Humberstone. 1977. The logic of 'Actually'. *Reports on Mathematical Logic* 8: 11–29.
Goranko, V., A. Montanari, and G. Sciavicco. 2004. A road map of interval temporal logics and duration calculi. *Journal of Applied Non-Classical Logics* 14(1/2004): 1–48.
Gusfield, D. 2007. *Algorithms on strings, trees, and sequences*, 10th ed. Cambridge: Cambridge University Press.
Kienzle, B. 2007. *Die Bestimmung des Janus*. Tübingen: Mohr Siebeck.
Kripke, S. 1980. *Naming and necessity*. Oxford: Blackwell.
McTaggart, J.M.E. 1908. The unreality of time. *Mind* 18: 457–484.
Müller, T., and Strobach, N. (2011). A letter on the present state of affairs. Prior, indeterminism and relativity 40 years later. *Synthese*. Preprint online at Springerlink: doi:10.1007/s11229-011-9939-z.
Müller, T. (2011). Branching space-times, general relativity, the Hausdorff property, and modal consistency. Preprint: http://philsci-archive.pitt.edu/8577/1/bst_hausdorff20apr11.pdf.
Prior, A. 1967. *Past, Present and Future*. Oxford: Oxford University Press.
Putnam, H. 1967. Time and physical geometry. *Journal of Philosophy* 64: 240–247.
Quine, W.V.O. 1951. On what there is. In: *From a logical point of view*, 1–19. Cambridge/Mass: Harvard University Press.
Quine, W.V.O. 1953. On a so-called paradox. *Mind* 62: 65–67.
Schopenhauer, A. 1818. Die Welt als Wille und Vorstellung I [The World as Will and Representation volume I]. In: *Arthur Schopenhauers Werke in fünf Bänden*, ed. Ludger Lütkehaus, Zürich: Haffman 1988.
Schopenhauer, A. 1839. Preisschrift ueber die Freiheit des menschlichen Willens. Volume 3 of the same edition, 359–458.
Strobach, N. 1998. *The moment of change - A systematic history in the philosophy of space and time*. Dordrecht: Kluwer.
Strobach, N. 2007a. *Alternativen in der Raumzeit - eine Studie zur philosophischen Anwendung multimodaler Aussagenlogiken*. Berlin: Logos.
Strobach, N. 2007b. Fooling around with tenses. *Studies in the History and Philosophy of Modern Physics* 38(3): 653–672.
Strobach, N. 2010. Zellen in der Logik des Lebens. In: *Logos, Freie Zeitschrift für wissenschaftliche Philosophie*. 2 (2010), 2–51. http://fzwp.de/0002/urn:nbn:de:0265--00022.
Strobach, N. 2011. Die Persistenz der biologischen Arten. Provisorische Überlegungen zu einer biologischen Interpretation zeitlogischer Strukturen. In: *Persistenz - Indexikalität - Zeiterfahrung*, ed. P. Schmechtig and G. Schönrich, 371–403. Heusenstamm: Ontos.
Strobach, N. (forthcoming). 2013. Aristoteles und die Konstanz der Arten. Forthcoming in the proceedings volume to a conference on Aristotle's biology at Kassel University in 2009, ed. G. Heinemann and R. Timme. Freiburg: Alber.
Thomason, R. 1970. Indeterminist time and truth-value gaps. *Theoria* 36: 264–281.
Woodger, J.H. 1937. *The Axiomatic Method in Biology*. With Appendices by Alfred Tarski and W. F. Floyd. Cambridge: Cambridge University Press.

A Theory of Possible Ancestry in the Style of Nuel Belnap's Branching Space-Time

Martin Pleitz and Niko Strobach

Abstract We present a general theory of possible ancestry that is a case of modal ersatzism because we do not take possibilities in terms of offspring as given, but construct them from objects of another kind. Our construction resembles Nuel Belnap's theory of branching space-time insofar as we also carve all possibilities from a single pre-existing structure. According to the basic theory of possible ancestry, there is a discrete partially ordered set called a *structure of possibilia*, any subset of which is called *admissible* iff it is downward closed under the ordering relation. A structure of possibilia is meant to model possible living beings standing in the relation of possible ancestry, and the admissible sets are meant to model possible scenarios. Thus the Kripkean intuition of the necessity of (ancestral) origin is incorporated at the very core of our theory. In order to obtain a more general formulation of our theory which allows numerous specifications that might be useful in concrete biological modeling, we single out two places in our framework where further requirements can be implemented: *Global requirements* will put further constraints on the ordering relation; *local requirements* will put further constraints on admissibility. To make our theory applicable in an indeterminist world, we use admissible sets to construct

This postscript and the family of theories it describes grew from discussions between Niko Strobach and Martin Pleitz about earlier drafts of Strobach's paper "In Retrospect". It was written jointly but with somewhat diverging motivations, with Strobach having a special interest in the question of ontological competition and the clue it provides for a general relativity friendly variant of BST and Pleitz having a special interest in the most general form of a theory of possible ancestry and the constructions that allow to answer the question of embeddability. We are grateful to the participants of the discussion during the WIRP II workshop at GAP 8 in Konstanz in September 2012, in particular to Peter Fritz and Thomas Müller.

M. Pleitz (✉) · N. Strobach
Philosophisches Seminar, Westfälische Wilhelms-Universität, Domplatz 6,
48143 Münster, Germany
e-mail: martinpleitz@web.de

N. Strobach
e-mail: pslogik@uni-muenster.de

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_13,

the (possible) moments and (possible) histories of a branching time structure. We then show how the problem of ontological competition can be solved by adding an incompatibility partition to a structure of possibilia, and conclude with some remarks about how this addition might provide a clue for developing a variant of the theory of branching space-time that can account for the trousers worlds of general relativity.

1 Ersatzism of Belnapian Elegance

The aim of this postscript is to present a theory of possible ancestry that emulates the elegance of Nuel Belnap's theory of branching space-time (BST), and in particular of the modal side of BST [1]. To bring out what is particularly elegant about it, let us have a look at how other theories of modality model possibilities. It is most common to model a possibility as a possible world, viewing the collection of all possible worlds as modal space, which is structured by the relation of accessibility that holds between worlds. *Modal primitivism* takes a possible world as a *given* object, irreducible to anything else. *Modal ersatzism* reduces each possible world to a construction from objects of a different kind, typically to a maximally consistent set of sentences or to a maximally coherent collection of states of affairs.[2] On this basis, both primitivism and ersatzism of the typical kind form modal space by knitting together their respective modal components by adding the relation of accessibility to the collection of possible worlds. BST, in contrast, gives a picture of much more cohesion and unity. Far from constructing modal space by knitting together possibilities, which (for typical ersatzism) are themselves the result of pasting together some of a plurality of modal atoms, it *carves* possibilities out from a single pre-existing structure, Our World. This is so because what corresponds naturally to possible worlds in the BST framework are *histories*, and these are just subsets of Our World that are defined by recourse to the inner structure of Our World (its ordering relation) alone. So BST, though of course also a case of modal ersatzism, is ersatzism of an untypically elegant kind.

The theory of possible ancestry sketched in Strobach's "In Retrospect" is like primitivism and like typical ersatzism insofar it also knits together possibilities to form modal space. Hence we will here dub it "TPA^{knit}". What we want to achieve in this sequel to Strobach's paper is to find a theory of possible ancestry such that there is some obvious one-to-one correspondence of some of its elements to the possibilities of TPA^{knit}, but which is closer to Belnap's BST insofar as it is based on carving out possibilities rather than knitting them together. (To make this contrast explicit, we will sometimes call our theory of possible ancestry "$\text{TPA}^{\text{carve}}$", but usually we will stick to the shorter "TPA".) The results will not quite be Belnap-style BST structures, but nearly so, and they will give a flexible framework for biological modeling.

---

[1] Belnap (1992).

[2] In our use, the term "ersatzism" is meant only as a neutral description of one kind of metaphysical theory. It was probably introduced with derogatory overtones, though. Cf. Lewis (1986), 142–165, for a highly valuable discussion of ersatzism given by one of its staunchest opponents.

2 The Basic Theory of Possible Ancestry

We start out by giving a basic variant of a theory of possible ancestry, which will suffice to explain some core notions and our main idea. The basic theory is as follows.

A *structure of possibilia* is an ordered pair $\langle D, \prec \rangle$, where D is a non-empty set of objects and $\prec$ is a relation on D. Nothing is required of D; in particular, D may be infinite. The relation $\prec$ is required to be irreflexive, anti-symmetric (and hence will be asymmetric),[3] and transitive, so that it is a partial strong order on D (hence the notation, "$\prec$"), and to be discrete. Nothing else is required of $\prec$; in particular, $\prec$ need not be connected and $\prec$ (when viewed as a graph) may contain both upward and downward branches and may thus be unlike the tree structure of branching time. Furthermore, some subsets of D are singled out as *admissible*, a set being admissible just in case it is downward closed under the relation $\prec$.

Formally[4]: (Irreflexivity) $\neg(x \prec x)$.
(Anti-Symmetry) $x \neq y \wedge x \prec y \rightarrow \neg(y \prec x)$.
(Transitivity) $x \prec y \wedge y \prec z \rightarrow x \prec z$.
(Discreteness)[5] $x \prec y \rightarrow \exists z(x \preccurlyeq z \wedge z \prec y \wedge \forall w(x \preccurlyeq w \wedge w \prec y \rightarrow w \preccurlyeq z))$
$\wedge \exists z'(x \prec z' \wedge z' \preccurlyeq y \wedge \forall w(x \prec w \wedge w \preccurlyeq y \rightarrow z' \preccurlyeq w))$.
(Def. Closure) A set $M \subseteq D$ is *downward closed* iff $x \in M \wedge y \prec x \rightarrow y \in M$.
(Def. Admissibility) A set $M \subseteq D$ is *admissible* iff M is downward closed.

The intended material interpretation of our formal ontology of a structure of possibilia and its admissible sets is as follows. The elements of the domain D represent possible living beings. The relation $\prec$ on D is represents possible ancestry, i.e., $x \prec y$ if and only if x is a possible ancestor of y (or, equivalently, y is a possible descendant of x). The admissible sets represent ancestral possibilities—alternatives in terms of offspring.[6]

In the light of this interpretation, we can explain our choice of requirements. Partly, the reasons for the requirements on the relation of possible ancestry mirror those for the corresponding requirements of (actual) ancestry made in Strobach's "In Retrospect". This is so for (Irreflexivity), (Transitivity), and (Discreteness). No being can be its own ancestor. If a first being can be an ancestor of a second, and the second being can be an ancestor of a third, then the first can be an ancestor of the third. And in view of our understanding of the ordering relation as one of possible

[3] We split up the requirement of asymmetry because it will turn out that the motivations for irreflexivity and for anti-symmetry belong to different levels.

[4] We suppress initial universal quantifiers, which are understood to range over the domain D.

[5] The weak order $\preccurlyeq$ is defined by recourse to the strong order $\prec$ in the usual way, with $x \preccurlyeq y$ iff $x \prec y \vee x = y$.

[6] We postpone the decision whether an admissible set is to model a possible state, a possible moment, or a possible history until Sect. 4.

ancestry, it just makes no sense to allow dense patches in the structure of possibilia, to say nothing of continuous ones.

The motivation of (Closure) and, as it will turn out, also of (Anti-Symmetry), is more substantial. We want our theory of possible ancestry to respect the Kripkean claim of the necessity of (ancestral) origin: *Any possible being has each one of its possible ancestors of necessity*—a being with distinct ancestors just would be a distinct being. Therefore any possibility must be downward closed, i.e., for any being that it contains it must also contain each one of the possible ancestors of that being. In other words, possibly being an ancestor entails being an ancestor, so that in many situations we may abridge talk about possible ancestry to talk about ancestry.[7]

It turns out that the Kripkean claim also motivates (Anti-Symmetry). If ancestral origin were *contingent*, we might well have two distinct possible living beings A and B such that in one possibility being A is an ancestor of being B and in another possibility being B is an ancestor of being A. But because of the explication of the Kripkean claim in terms of (Closure), and in view of (Transitivity), A and B would be their own ancestors in each one of the two possibilities of this scenario, which obviously conflicts with (Irreflexivity). So, reflecting on the inadmissibility of such *circular* relations of ancestry as those between A and B lets us note that incorporating the relation of possible ancestry *on the level of possibilia* already does quite much to commit us to a Kripkean doctrine of the necessity of ancestral origin. Or, what probably amounts to the same thing, an incorporation of the relation of possible ancestry on the level of possibilia and a restriction of ancestral relations within each possibility like (Closure) make sense only when they are implemented *together*.

Note that, according to the above definition, the empty set is admissible. This will be technically convenient later on,[8] and it can also be motivated intuitively. For is it not possible that there is no (and there never has been any) living being at all? A similar claim that involved a truly unrestricted quantifier (i.e., that it is possible that there is *nothing at all*) might well be contentious. But in the case of the present framework, the intuitive background story has other entities—e.g., atoms, chemical compounds, water, air, and the planet Earth—besides the possible living beings modeled by the elements of the domain D. Clearly an entirely uninhabited Earth is possible relative to some moments in time (especially in the far past), and we can even imagine entire possible histories in which live never evolves.[9]

---

[7] This is not so in all situations, because the converse claim does not hold: It is not the case that possibly being a descendant entails being a descendant!

[8] Cf. the role played by the empty state in Sect. 4.

[9] The above argument presupposes that to be in an admissible set intuitively is *to exist or have existed* relative to the possibility modeled by it. (The temporal aspects of the intuitive interpretation of our formal ontology will get clearer in Sect. 4.) In terms of quantified modal logic, we thus understand each admissible set as the variable domain of the possibility modeled by that very set— but rather than use the variable domain as the extension a contingent existence predicate has relative to that possibility, we understand it as containing all objects that are *identifiable* relative to it. For more on this way of singling out local *identifiabilia* from a set of global *possibilia* in an application of quantified modal logic to an indeterminist world, cf. Pleitz (forthcoming).

With our basic theory, we have already achieved some similarity to Belnap's BST. This becomes evident by contrasting the present $\text{TPA}^{\text{carve}}$ to the TPA^{knit} of Strobach's paper. While the latter starts out with given possibilities, each with its own relation of ancestry, which have to be made to match each other by certain requirements on those relations to enable the next step of knitting them together, the former needs no corresponding requirements because possibilities are carved out of a single pre-existing object, the structure of possibilia.

However, we yet have no natural one-to-one correspondence between the possibilities of $\text{TPA}^{\text{carve}}$ and the possibilities of TPA^{knit}. This is so because our basic theory leaves out many of the requirements, especially of cardinality and connectedness, which are implemented in TPA^{knit} (which it in turn had inherited from the non-modal formal theory of ancestry, FTA). So, although we already have captured a few basic metaphysical intuitions, there is still some way to go for our $\text{TPA}^{\text{carve}}$ to model the biological realm in a satisfactory way.

3 The General Form of a Theory of Possible Ancestry and Some Specific Theories

In order to do some biological modeling, we will now leave behind the *basic theory* of possible ancestry and move on to a plurality of *specific theories* of possible ancestry. We will start with what is common to them all, with the *general form* of a theory of possible ancestry. Going specific means adding details to the structure of possibilia and the possibilities it contains. We do this by adding requirements to the basic theory. The general form of a theory of possible ancestry tells us that there are two different places in the theory where we can implement the extra requirements.

The general formulation of a theory $\text{TPA}^{\text{carve}}$ is as follows. A *structure of possibilia* is a domain ordered by an asymmetrical, transitive, and discrete relation *such that the domain and the relation satisfy some additional theory-specific requirement* Γ (gamma for "*global*"), e.g., of cardinality or connectedness. An *admissible set* is a downward closed subset of the domain *such that the set together with the ordering relation satisfy some additional theory-specific requirement* Λ (lambda for "*local*"), e.g., again, of cardinality or connectedness.

Formally:

(Irreflexivity), (Anti-Symmetry), (Transitivity), & (Discreteness)[10]
(Gamma) $\langle D, \prec \rangle$ satisfies the additional requirement Γ.
(Def. Admissibility) A set $M \subseteq D$ is *admissible* iff M is downward closed and $\langle M, \prec \rangle$ satisfies the additional requirement Λ.

To see how this general framework can be used we will look at some examples of specific theories that can be obtained by choosing particular sentences Γ and Λ.

[10] For these four postulates and the definition of downward closure, cf. Sect. 2.

First, the trivial example. The basic theory is that special case where for Γ and Λ we insert some tautologies into the general form, i.e., where further constraints are put neither on the structure of possibilia nor on the admissible sets.

Next, the example of a direct counterpart of the theory $\mathrm{TPA}^{\mathrm{knit}}$, called simply "TPA" in Sect. 3.2 of "In Retrospect". Its axiom TPA 1 corresponds to the framework delivered already by our basic theory, but it does its work on a more specific structure of possibilia, which can be obtained by putting the conjunction of all the BTA postulates from Sect. 2.2 of "In Retrospect" in the place of Γ. The possible strengthened theories discussed by Strobach in Sect. 3.2 can be obtained by adding U as a further conjunct of Γ and adding TPA 2 or TPA 2* as a further conjunct of Λ. Thus we have found a natural one-to-one correspondence between the possibilities of $\mathrm{TPA}^{\mathrm{carve}}$ and the possibilities of $\mathrm{TPA}^{\mathrm{knit}}$ (which our basic theory could not yet deliver): Each one of the pairs $\langle D_P, <_P \rangle$ of $\mathrm{TPA}^{\mathrm{knit}}$ corresponds to an admissible set of the present specific variant of $\mathrm{TPA}^{\mathrm{carve}}$, because, when the domain of $\mathrm{TPA}^{\mathrm{carve}}$ is taken to be a superset of the union of all the D_P, each relation $<_P$ need only be taken as the restriction of the relation $\prec$ of $\mathrm{TPA}^{\mathrm{carve}}$ to the admissible set corresponding to D_P, and everything will fall into place nicely.

As our next family of examples, we have some specific theories that share the following characteristic with the above direct counterpart of the theory $\mathrm{TPA}^{\mathrm{knit}}$: Each one of the possibilities they deliver satisfies all the postulates of the non-modal basic theory of ancestry (BTA) of Sect. 2.2 of "In Retrospect". We have constructed the direct counterpart by putting all the BTA postulates into the slot held open by "Γ" in the general form of a theory of possible ancestry. It is interesting to see what happens when we move some of them over to the slot "Λ", that is, when we understand them not as *global* but as *local* requirements. The results are impressive in the case of constraints of cardinality, where it arguably makes much biological sense, too.

Using the cardinality constraints[11] as conjuncts not of Γ but of Λ will allow the structure of possibilia to have an infinite domain because it moves the requirements of finiteness into the admissible sets. For example, any possible being in each possibility has only finitely many direct descendants, but it may nevertheless stand in the relation of direct possible ancestry to infinitely many possible beings. Here is a reason why we should not demand that some ancestor has only finitely many possible descendants. While a living being cannot actually reproduce infinitely often within a given amount of time and cannot actually leave infinitely many direct descendants, it may well do so *possibly* in the following sense: While no infinite branching within the same alternative is possible, the same parents may have an infinity of different *possible* children. (The more strongly we understand the metaphysical principle of the necessity of origin, the more plausible this gets. For according to a very strict reading of that principle, even offspring from the same sperm and egg would be a different individual if both had met a second earlier, or half a second, or a quarter of a second etc. Maybe this reading of the principle is too strict to be credible.[12] Still,

[11] BTA postulates C2, C3, C6, and C7.

[12] So thinks Pleitz; Strobach likes the strict reading. So here the two authors disagree.

we do not take it to be to the disadvantage of our framework that it can model the consequences of the strict reading.)

Another class of examples shows again how sensitive the general form of a theory is to choices whether a certain constraint is implemented in a global or a local way. When we put a constraint of connectedness (like postulate U) into the place of Γ, we get a much more severe restriction on our frame of possibilia than when we put it into the place of Λ. In the former case, all possible living beings are connected, whereas in the latter case, only the living beings *of each possibility* are connected, which would allow for the possibility of an entirely disjoint alternative to the actual development of life even in the face of the intuition of the unity of life and its deictic component that motivates postulate U.[13] Something similar can be said about the postulates about starting points and endpoints.[14]

There are also some specific theories that do *not* result from recombining the BTA postulates, but can be obtained easily from the general form by some small additions to the language of TPA. If we add species predicates ("… is a horse", "… is a dog"), we can formulate postulates that state the impossibility of interbreeding, and put species-specific upper limits on the number of direct ancestors a being can have (the number being *two* for mammals and *one* for cells that reproduce by fission), and so on. Thus a whole wealth of distinctions becomes available. For instance, although prima facie it is natural to give a rigid interpretation to all species predicates, they arguably can also be construed as flaccid.[15] But then it may come about that some given individuals that belong to distinct species in *some* possibilities belong to the same species in *other* possibilities.

The preceding examples should be enough to show that in its general form our theory of possible ancestry allows to do some realistic biological modeling. We want to close this section with a remark about the way in which we have split up general principles and specific constraints by using some postulates, like the transitivity of possible ancestry and the downward closure of ancestral possibilities, to formulate the general form and others, like those of cardinality and connectedness, to formulate specific requirements of a global or local sort. What is behind this way of splitting up postulates is a conviction about how to draw the line between metaphysical and empirical inquiry. The postulates enshrined in the general form of all our theories of possible ancestry are motivated by metaphysical principles, first and foremost Kripke's claim about the necessity of ancestral origin. In contrast, all the optional specific postulates—of cardinality, connectedness, the existence of starting points and endpoints, and all species-relative constraints—are in principle open to empirical revision.[16] To see the metaphysical character of Kripke's claim, just try to devise an experiment (or, more generally, try to come up with any empirical consideration) that would make it possible to falsify it! It nevertheless is fitting that Kripke's claim

[13] Cf. Sect. 2.5 of "In Retrospect".

[14] BTA postulates C4 and C5.

[15] Cf. Sect. 3.4 of "In Retrospect".

[16] Of course, some postulates might be of an unclear status concerning the metaphysical/ empirical divide. Strobach's intuition behind postulate U, for example, seems to be of a purely conceptual

and the other motivations behind the general form are incorporated at the core of a *biological* theory because these are metaphysical facts *about the subject matter of biology*.

Now, after we have given our theory a lot of flexibility, which will allow it to model quite a large part of biological reality, we should investigate how it relates to the picture of branching time, which after all has provided the intuitive background all along.

4 The Question of Embeddability: States, Moments, and Histories

Belnap developed BST as a formal framework to model an *indeterminist* world, making a significant step toward accommodating modern physics by adding resources to model spatial variation to the branching time framework of Kripke and Prior. The intuitive interpretation of our theory of possible ancestry presupposes a similar intuition of indeterminism.[17] Hence the question arises of how its ontology can be embedded in a branching time structure.

The task is not trivial, because typically neither a structure of possibilia nor any of its restrictions to an admissible set will have the requisite property of having no downward branches. A typical family tree is not a tree in the branching-time sense, and the same goes for any typical structure of possibilia. (If fission were the only means of reproduction, a structure of possibilia need indeed not have downward branches. But they will appear as soon as there can be reproductive acts that require more than one participant.)

So, how can we construct a branching time structure from the means at our disposal? Here, the admissible sets clearly play a central role. We will motivate our construction by a look at how the possibilities they model correspond to elements of the branching time structure that is implicit in our intended interpretation. Facing the future, it is obvious that from the possibility modeled by an admissible set there typically will sprout many different historical paths towards later possible situations, depending on which possible children (modeled by direct descendants in terms of $\prec$) of some of the inhabitants of that possibility come into existence. Facing the past, we encounter a small surprise because there may also be a plurality of paths branches leading *up* to the possibility modeled by an admissible set. E.g., towards the admissible set containing Eve, Adam, and their two sons Abel and Cain, there is one path via the admissible set {Eve, Adam, Abel} and a second path via the admissible

(Footnote 16 continued)
nature (cf. Sect. 2.5 of "In Retrospect") so that there might be reason to understand postulate U in a metaphysical way. But in this special case there remains a pragmatic reason to group it with other *specific* requirements, namely its high degree of contentiousness.

[17] This will be obvious in the examples of Sect. 5, where we admit, for some given possibility, that *from then on* things may take one of many different courses: Elizabeth and Peter have these children, but they might have had others, and so on.

set {Eve, Adam, Cain}, because the corresponding structure of possibilia does not determine whether Cain or Abel was born first. As the structure of admissible sets as ordered by inclusion may have downward branches, chains of admissible sets are unfit to model (*possible*) *histories* or parts thereof, and our admissible sets turn out to be too coarse-grained to correspond to the nodes of a branching time tree, which we might call (*possible*) *moments*.

Nonetheless, it should also be evident from the biblical examples that there is a close connection between admissible sets and moments. We can understand an admissible set as the (*possible*) *state* the world is in at some a moment, a state that in many cases will be shared by a plurality of distinct moments. This observation puts us in a position to construct (possible) moments and (possible) histories. We will say that a moment is individuated not only by its state, but also by the chain of states that it is reached by. In our example we can thus pry apart the moment where Eve, Adam, Abel, and Cain belong to the state and Abel is firstborn and the distinct moment where the same family of four belongs to the state but Cain is firstborn. Here is our construction:

A set of elements from D is a (*possible*) *state* iff it is an admissible set.

Note that {} is a state; we call it the *empty state*. The set of all states is partially ordered by inclusion, with the empty state being smaller than all other states.

A set of states is a (*possible*) *moment* iff it is a maximal chain[18] in the set of all states as ordered by inclusion from the empty state {} to some state.

Note that {{}} is a moment. The set of all moments is partially ordered by inclusion, which plays the role that the relation of accessibility has in a branching time structure. As it precedes all other moments in terms of this relation, {{}} may aptly be called the *first moment*.

A set of moments is a *(possible) history* iff it is a maximal chain in the set of all moments as ordered by inclusion from the first moment {{}} to some moment.

Note that {{{}}} is a history—it accounts for the possibility that live evolves never. All histories overlap—in fact their union forms a tree-like structure with a single root in the first moment. Thus we have found a way of embedding our ancestral alternatives in the tree structure of indeterminist time.

***

By now we have come as near in our analogy to the structure of BST as we will get. In the next section we will deal with a phenomenon that resists treatment in this framework.

[18] A subset of an ordered set $\langle M, < \rangle$ is a *chain* iff for all elements x and y of M either $x < y$ or $x = y$ or $y < x$.

5 Ontological Competition

In the past sections we have looked at structures of possibilia mainly under the aspect of *co-possibility*, investigating possible collections of possible beings that *can* (or even *must*) be grouped together. We now turn to the contradictory aspect of *incompatibility*.

Let us have a look at three examples from the realm that is to provide the material interpretation to our formal theory, which we will consider as test cases for its power to model phenomena of incompatibility.

> (Example 1) A human couple, Elizabeth and Peter, actually have a lot of children and they could have had even more children, or a distinct lot of possible children, or another lot, or … However, there clearly is an upper limit to the number of children they could have had, determined (roughly speaking) by the minimum length of a pregnancy and the maximum duration a woman can bear children. So, the number of the possible children of Elizabeth and Peter exceeds the upper limit of children they could have had by far. Any collection of possible children of a number that exceeds this upper limit is not compatible.
>
> (Example 2) Two possible mammals, A and B, are such that they not only have the same ancestors but result from the very same sperm and egg, while their dates of conception are a few seconds apart. For the scope of this example we understand beings of the kind of A and B to be individuated by their time of conception (among other things). Hence we must see A and B as incompatible individuals.
>
> (Example 3) A cell of a kind that reproduces by fission actually splits into the two daughter cells D_1 and D_2, but it could have split in a different way (e.g., distributing its material in a different way) that would have led to the two possible daughter cells E_1 and E_2. But every cell can split only once—though its daughter cells may split in turn, it is no longer there to split up again. Hence D_1 is incompatible with E_1, D_1 is incompatible with E_2, and so on. In fact, as there plausibly are many possibilities for a cell to split up, there will be a large plurality of incompatible pairs of possible daughter cells for each cell.

These examples illustrate three general observations that are relevant to the task of modeling phenomena of incompatibility. Firstly, incompatibility need not be a two-place relation. It may well be that each two of the many possible children of Elizabeth and Peter of (Example 1) are compatible, but clearly no collection of a hundred of them is a population of any possibility.[19] Robert Brandom makes a corresponding observation with respect to sentences or claims: "the claim that the piece of fruit in my hand is a blackberry is incompatible with the *two* claims that it is red and that it

[19] Here we are bracketing the intuitions behind (Example 2) and (Example 3). But even taking into account complications due to time of conception and possible monozygotic twins, there will remain many collections of possible children that are jointly impossible even though each two of them are compatible.

is ripe, though not with either of them individually—in keeping with the childhood slogan that blackberries are red when they're green."[20]

Secondly, incompatibility comes in two flavors, extrinsic and intrinsic. The decisive question is this: Is the incompatibility due to some external factor or is it founded in the possible objects themselves? A clear case of *extrinsic* incompatibility is provided by (Example 1), because there is nothing in the possible children themselves that explains their joint incompatibility, which rather rests on those external factors determining the maximum number of children a woman can bear. A clear case of *intrinsic* incompatibility can be found in (Example 3), because it is due to the very nature of some possible daughter cells that they are not possible siblings and hence are incompatible. (Example 2) might be less easy to classify, but the strict reading of the principle of the necessity of origin we adopt for its scope would tip the scales in favor of construing the time of conception as an internal factor, because of its individuating force.

Thirdly, all three examples taken together provide ample evidence for the importance of the metaphysical phenomenon of *ontological competition*: Some possible individuals can come to be only by cutting off the chance of existence for others (of course, they do not literally *struggle*). We think that this phenomenon has not received the attention it deserves. So, let us get on with modeling it in our formal ontology!

Extrinsic incompatibility is easily accommodated in the general framework laid out in Sect. 3, at least if we decide to invest species membership into a model. We need only add to Λ a species-relative cardinality constraint on the number of direct descendants of any pair of possible parents, and there will be only possibilities that conform to the intuitions behind (Example 1).

With intrinsic incompatibility, however, we reach the limits of what our theory of possible ancestry in its current state can achieve. There just is no natural way to model those cases of incompatibility that are due to the nature of the incompatible possibilia themselves by any general statements that act as constraints on either the global or the local level.[21] In the scenario of (Example 3), what leads to the incompatibility of the possible cells D_1 or E_1 is no property that they might share with some other possible cells, but their individual nature.

How are we to enhance our theory of possible ancestry to give it the power to model intrinsic incompatibility?

There is the somewhat brutal way of adding an arbitrary filter that admits only some of our admissible sets: To the structure of possibilia $\langle D, \prec \rangle$ and the admissible sets defined in terms of it we add **P**, which is a subset of the set of admissible sets. No element of **P** may contain any collection of possibilia that according to our intended interpretation are intrinsically incompatible. We obtain the theory $\mathrm{TPA}^{\mathrm{carve+filter}}$ from

[20] Brandom (2008), 123.

[21] We could model intrinsic incompatibilities by adding a large number of *singular* statements as further conjuncts to Λ. In the case of (Example 3), those would be of the kind "No admissible set contains both D_1 and E_1", "No admissible set contains both D_1 and E_2", and so on. This way strikes us as quite inelegant and unnatural.

$\text{TPA}^{\text{carve}}$ by adding to the requirement Λ the postulate that a set is only admissible if it is an element of **P**. In the terms of Sect. 1, any such theory would be a hybrid between elegant ersatzism ("carve") and primitivism ("filter"). By taking the brutal way we thus would be in danger of losing something of what we have gained so far.

We are optimistic that there also is a somewhat more subtle way.[22] Inspired by the work of Brandom in incompatibility semantics, we add to our framework an incompatibility partition INC. It is a subset of the powerset of the domain D and satisfies the single postulate of *persistence*: Any superset of a set in INC is also in INC. The intuitive reason for the property of persistence is that you just cannot remove an incompatibility between some objects by adding further objects.[23] Now we add the local requirement that no subset of an admissible set may be in INC. What we thus acquire is a tool to model intrinsic incompatibility— but a tool that had to be added to our structure of possibilia because we have found no way of carving it from it.

6 Back To Branching Space-Time: General Relativity

We have moved some distance away from the structures of BST that inspired our theory of possible ancestry in the beginning. But in fact, apart from its intended literal interpretation, this postscript might be read as a little parable on a certain aspect of BST. Investing a suitable incompatibility partition for possible point events may be one way to solve the problem posed to BST by the so-called *trousers worlds* of general relativity theory (GTR).[24]

Belnap's theory BST reduces the incompatibility of two possible point events to their not having a common upper bound. But there is a price to be paid for this elegance, because due to this reduction BST cannot distinguish between the basic scenario of indeterminism, where a history after some time branches into two histories, and what happens in the single history of a trousers world of GTR, where (roughly speaking) a connected space after some time splits up into two disconnected spaces.[25]

---

[22] The question whether the INC approach suffices or whether we have to take the brutal way after all hinges on the following objection: Might there not be ontological co-dependence in addition to ontological competition? An example would be cellular fission as in the scenario of (Example 3): Does not the emergence of one of the cells necessitate the emergence of its sibling, too, such that either both belong to a certain possibility or none does? This could not be modeled by the INC approach. Our preferred answer is to deny ontological co-dependence. The emergence of one cell does not necessitate the emergence of its sibling. There is always so much that can go wrong and make the other half crumble before it is a cell; and that would establish possibilities with one of the allegedly co-dependent entities, but not the other.

[23] Brandom (2008), 117ff.

[24] Müller (2011) suggests a different solution, which is based on local transitions.

[25] Earman (2008). Cf. Müller (2011), specifically Sect. 2.1.

So one might add another primitive to the theory, an incompatibility partition INC. Now the required distinctions can be made:

A collection of possible point events is *incompatible* iff the set formed by them is in INC.

A collection of possible point events is *merely spatially disconnected* (towards the future) iff the set formed by them has no upper bound but is not in INC.[26]

We have, however, to tread very carefully. For note that Belnap's reduction of incompatibility to being without an upper bound is reached via his definition of a history: In BST, first a history is defined as a maximally directed set—a set such that each two of its elements have an upper bound in it—and then compatibility is defined as belonging to a single history. To implement our idea for a solution of the trousers world problem we thus have to change the definition of a history. Upward directedness will no longer do because a history that is a trousers world for some of its points does not include an upper bound. On this approach, a new definition of "history" would resemble the one we have given for admissible sets in the context of the theory of possible ancestry:

A subset of Our World is a *history* iff it is downward closed with respect to the causal order and none of its subsets is in INC.

We conjecture that with an incompatibility partition as a new primitive and this revised definition of a history, a theory can be constructed in a way similar to BST that has a broader range of application when it comes to GTR. This would be a comparatively simple approach, though it admittedly lacks the elegance of the original BST style of dealing with histories. Perhaps, one day, the approach might turn out useful for discrete models of space-time which take quantum gravity into account. But as our proposal for a solution to the trousers world problem does not consist merely in an addition to Belnap's theory, but in changing one of its most basic ingredients, there remains much work to be done in the future.

References

Belnap, N. 1992. Branching space-time. *Synthese* 92: 385–434.

Brandom, R. 2008. *Between Saying and Doing. Towards an Analytic Pragmatism*. Oxford: Oxford University Press.

Earman, J. 2008. Pruning some branches from branching space-times. In *The Ontology of Spacetime II*, ed. D. Dieks, 187–206. Amsterdam: Elsevier.

[26] Note that due to the character of BST we here employ not a *topological*, but an *order theoretic* notion of spatial connectedness, which in terms of the theory of relativity corresponds to *not having a common causal future*.

Lewis, D. 1986. *On the Plurality of Worlds*. Oxford and New York: Blackwell.
Müller, T. 2011. Branching space-times, general relativity, the Hausdorff property, and modal consistency. http://philsci-archive.pitt.edu/8577/1/bst_hausdorff20apr11.pdf
Pleitz, M. (forthcoming). "*This Sentence*": *The Liar Paradox from the Viewpoint of Logic, Semantics, and Metaphysics*.

Connecting Logics of Choice and Change

Johan van Benthem and Eric Pacuit

Abstract This chapter is an attempt at clarifying the current scene of sometimes competing action logics, looking for compatibilities and convergences. Current paradigms for deliberate action fall into two broad families: dynamic logics of events, and STIT logics of achieving specified effects. We compare the two frameworks, and show how they can be related technically by embedding basic STIT into a modal logic of matrix games. Amongst various things, this analysis shows how the attractive principle of independence of agents' actions in STIT might actually be a source of high complexity in the total action logic. Our main point, however, is the compatibility of dynamic logics with explicit events and STIT logics based on a notion that we call 'control'—and we present a new system of dynamic-epistemic logic with control that has both. Finally, we discuss how dynamic logic and STIT face similar issues when including further crucial aspects of agency such as knowledge, preference, strategic behavior, and explicit acts of choice and deliberation.

1 Introduction: Logical Frameworks for Agency

The STIT logic of Belnap et al. (2001) and its variants have proven fruitful tools to help philosophers and computer scientists explore their intuitions about agency and social interaction. These logics provide a framework to reason about choices,

We thank Roberto Ciuni and Thomas Müller for helpful comments on this chaper.

J. van Benthem (✉)
Institute for Logic, Language and Computation (ILLC), University of Amsterdam,
P. O. Box 94242, 1090 GE Amsterdam, The Netherlands
e-mail: J.vanBenthem@uva.nl

E. Pacuit
Department of Philosophy, Skinner Building, University of Maryland,
College Park, MD 20742-7505, USA
e-mail: epacuit@umd.edu

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_14,

abilities and actions of agents, all placed in a temporal setting. And further issues lie just below the surface: what agents know or believe at the time of choice, how they act based on preferences, and engage in deliberate strategic interaction (cf. Horty 2001).

But STIT is not the only game in town. Many logical paradigms are active in the above territory, and they often show clear similarities. This calls for analysis and reflection. For instance, van Benthem and Pacuit (2006) relate the major varieties of epistemic temporal logics, coming from mathematical logic, computational logic, and studies of agency. Continuing in this line, van Benthem et al. (2009) prove representation theorems linking dynamic-epistemic models with epistemic-temporal ones, making it possible to enlist ideas from one logic in the service of the other. In the case of STIT, too, much has been done to clarify its connections with other frameworks. In fact, Belnap et al. (2001) already pointed out links with earlier work of Chellas (1992), to which one can add the neighborhood logics of ability in (Brown 1988, 1992). Moreover, connections have been found with coalition logic (Broersen et al. 2006b) and alternating-time temporal logic (Broersen et al. 2006a), while Lorini and Schwarzentruber (2010) relates STIT to logics for strategic and extensive games—a line that we will continue in this chapter (cf. also Herzig and Lorini 2010). Finally, Ciuni and Zanardo (2010) shows how STIT extends well-known logics of branching time.

Our aim in this chapter is to continue in the latter vein, and connect STIT models further with modal models for action from the realm of propositional dynamic logic (PDL), modal game logics (see van Benthem 2014), and dynamic-epistemic logic DEL (van Benthem 2011). We start by addressing an initial barrier to making any comparison between these different logical frameworks.

STIT logics are primarily intended as logics of *ontic* freedom and indeterminacy while the logical systems we discuss in this chapter are focused on *epistemic* uncertainty (i.e., knowledge about what will happen next). The heart of our comparison is the simple observation that the basic STIT modality turns out to be precisely the "knowledge" modality found in many epistemically-oriented logical systems. Importantly, however, we are *not* suggesting that all discussion about "agency" and agents making choices in an indeterministic world can or should be replaced with an analysis of what the agents know about their own choices and the consequences of their actions in an indeterministic world, or vice versa. Our point is simply that similar logical frameworks are open to different interpretations. The goal is not to argue for the primacy of any single interpretation, but rather to demonstrate how two different perspectives on modeling rational agency can lead to similar insights. This is in line with a broader goal. The arena of logics for agency appears to be moving from an initial stage of a "Battle of the Sects" to a more detached understanding of both similarities and relative advantages of different paradigms, leading to a more unified sense of purpose and methodology.

2 Preliminaries: The STIT Framework

In this section, we introduce the basic STIT framework. We will be very brief, only touching on the key notions we need later in this chapter. For more information, the reader is invited to consult (Horty 2001; Belnap et al. 2001; Horty and Belnap 1995; Balbiani et al. 2008).

STIT structures STIT models are based on *branching-time frames*, structures $\langle T, < \rangle$ where T is a nonempty set of "moments", and $<$ is a strict partial order on T without backwards branching: for all m, m', m'', if $m' < m$ and $m'' < m$, then either $m' \leq m''$ or $m'' \leq m'$ (where $x \leq y$ iff $x < y$ or $x = y$). A *history* is a maximal linearly ordered subset of T. Let Hist denote the set of all histories and for $t \in T$, $H_t = \{h \in \mathsf{Hist} \mid t \in h\}$ is the set of histories containing moment t.

At each moment, there is a choice available to the agent. Let $\mathcal{A}$ be the set of agents. Formally, the choices available to agent i at moment t are represented by a partition $Choice_i^t$ on the set H_t of histories containing t. Let $Choice_i^t(h)$ denote the cell containing h. Since $Choice_i^t$ is a partition, we have for each $i \in \mathcal{A}$ and $t \in T$, $Choice_i^t \neq \emptyset$ and $\emptyset \notin Choice_i^t$. In addition, the choice partitions of the agents must satisfy one additional condition:

Independence For all $t \in T$ and all $s_t : \mathcal{A} \to \wp(H_t)$ with $s_t(i) \in Choice_i^t$, $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$.

Now we define a *STIT model* as a tuple $\langle T, <, \mathcal{A}, Choice, V \rangle$, where $\langle T, < \rangle$ is a branching-time frame, $\mathcal{A}$ is a finite set of agents, $Choice$ is a function assigning to each $i \in \mathcal{A}$ and $t \in T$ a partition on H_t satisfying $\mathsf{Independence}$, and V is a function assigning to each atomic proposition a set of history/moment pairs ($V : \mathsf{At} \to \wp(T \times \mathsf{Hist})$).

STIT language Let At be a set of atomic propositions. The STIT language is the smallest set of formulas generated by the following grammar

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i \text{ stit}]\varphi \mid \Box\varphi$$

where $p \in \mathsf{At}$ and $i \in \mathcal{A}$. Additional boolean connectives ($\vee, \rightarrow, \leftrightarrow$) are defined as usual. Further, $\langle i \text{ stit}\rangle\varphi$ is the dual modality $\neg[i \text{ stit}]\neg\varphi$ and $\Diamond$ the dual $\neg\Box\neg\varphi$. The interpretation of $[i \text{ stit}]\varphi$ is that "agent i *sees to it that* φ is true" and the historic necessity $\Box\varphi$ means that "φ is true at all alternative histories".

STIT Semantics Let $\mathcal{M} = \langle T, <, \mathcal{A}, Choice, V \rangle$ be a STIT model. Truth of a STIT formula φ is defined inductively as follows, at pairs t/h of histories h and moments t on them:

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i \text{ stit}]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$

In addition, one sometimes defines an additional STIT operator (the so-called "deliberative STIT"):

- $\mathcal{M}, t/h \models [i \text{ dstit}]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$ and there is a $h'' \in H_t$ such that $\mathcal{M}, t/h'' \models \neg\varphi$

This modality is definable in the basic language: $[i \text{ dstit}]\varphi := [i \text{ stit}]\varphi \wedge \Diamond\neg\varphi$. A number of other STIT-operators can be found in the literature. For example, the "achievement STIT operator" (see Horty and Belnap 1995, Sect. 2.2 for a definition and discussion) and the "next time STIT operator" (Broersen 2011) both make use of the underlying past and future time structure.

Logic and axiomatics The models and language are one major aspect of current uses of STIT, as a style of representing action semantically. However, there is also the issue of syntactic proof rules for reasoning about action. The following axiomatization was proven sound and complete for the class of all STIT models in (Xu 1995; Balbiani et al. 2008):

- The **S5** axioms for $\Box$ and $[i \text{ stit}]$: $\bigcirc(\varphi \to \psi) \to (\bigcirc\varphi \to \bigcirc\psi)$, $\bigcirc\varphi \to \varphi$, $\bigcirc\varphi \to \bigcirc\bigcirc\varphi$, $\neg\bigcirc\varphi \to \bigcirc\neg\bigcirc\varphi$, for $\bigcirc \in \{\Box, [i \text{ stit}]\}$
- $\Box\varphi \to [i \text{ stit}]\varphi$
- $\left(\bigwedge_{i\in\mathcal{A}} \Diamond[i \text{ stit}]\varphi_i\right) \to \Diamond\left(\bigwedge_{i\in\mathcal{A}}[i \text{ stit}]\varphi_i\right)$
- Modus Ponens and Necessitation for $\Box$.

It will be clear that these axioms do not reflect, let alone enforce, any particular view of time, whether branching or linear. This is no accident. The basic ideas of STIT seem compatible with about every major temporal logic that is on the market.

Now that we have all major components of STIT on the table, we will discuss its semantics and axiomatics in relation to other approaches for studying agency coming from the "dynamic logic family". We will not define these other frameworks in any detail, but refer the reader to the literature on dynamic logic, game logics, and dynamic-epistemic logics cited in this chapter.

3 Modeling Choice Situations

3.1 The Modal Heart of Choice

Abstracting from the temporal component that could come from any existing framework, the heart of STIT-style choice is a very simple **S5** logic.[1] A **STIT choice scenario** for a set of agents $\mathcal{A}$ is a tuple $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$, where W is a nonempty set, for each $i \in \mathcal{A}$, $\sim_i$ is an equivalence relation on W (we write $[w]_i$

[1] An earlier modal analysis of STIT scenarios can be found in Herzig and Schwarzentruber (2010), Balbiani et al. (2008) and follow-up literature—but in this chapter, we will eventually choose a path of our own.

for the equivalence class of w under $\sim_i$) and V is a valuation function. We focus on two agents ($\mathcal{A} = \{1, 2\}$) for convenience in what follows. STIT choice scenarios are standard multi-agent **S5** models, and so a simple modal language describes them: for each $i \in \mathcal{A}$, use '$[\,i\,]$' for the modality matching the relation $\sim_i$ and 'E' for the existential modality.[2] The Independence assumption above corresponds to the validity of the following *product axiom*:

$$(E[\,1\,]\varphi \wedge E[\,2\,]\psi) \rightarrow E(\varphi \wedge \psi)$$

By standard frame correspondence, this says that any pair of choices for the two agents overlap.

The key idea of STIT in these models may be called *control*: the equivalence relations represent the extent to which agents control outcomes by their choices. The product axiom says that no agent can prevent any other agent from making any of her choices. There is more to this condition than meets the eye. For instance, assume that agent 1 has a singleton choice somewhere. Since 2's choices must always overlap with this singleton, and different choices are disjoint, it follows that 2 has only one choice set.

The logic of these models is many-agent **S5** plus the product axiom. In this basic system, we can derive interesting facts, such as

$$[\,1\,][\,2\,]\varphi \leftrightarrow [\,2\,][\,1\,]\varphi \leftrightarrow U\varphi$$

where U is the universal modality dual to E.[3] In slightly extended modal languages, more can be proved. For instance, the previous comment about singleton choices amounts to the validity of $([\,1\,](\varphi \wedge \neg D\varphi) \wedge E[\,2\,]\varphi) \rightarrow U\varphi$, where $D\varphi$ is the *difference modality* true at a world w if there is a $v \neq w$ such that $\mathcal{M}, v \models \varphi$. Thus, the product axiom packs a lot of punch.

So, basic STIT logic is a nice simple multi-**S5**-extension. This first natural connection with modal logic shows that we are at least generally in the same world as modal logics of action.[4]

---

[2] Truth for these operators is defined as usual: $\mathcal{M}, w \models [\,i\,]\varphi$ iff for all $v \in W$, if $w \sim_i v$ then $\mathcal{M}, v \models \varphi$, and $\mathcal{M}, w \models E\varphi$ iff for some $v \in W$, $\mathcal{M}, v \models \varphi$.

[3] As observed in Balbiani et al. (2008), this principle can also function as a product axiom by itself. Also inter-derivable with our version of the product axiom is the stronger-looking $(E[\,1\,]\varphi \wedge E[\,2\,]\psi) \rightarrow E([\,1\,]\varphi \wedge [\,2\,]\psi)$., for which Roberto Ciuni has proposed an interesting epistemic interpretation.

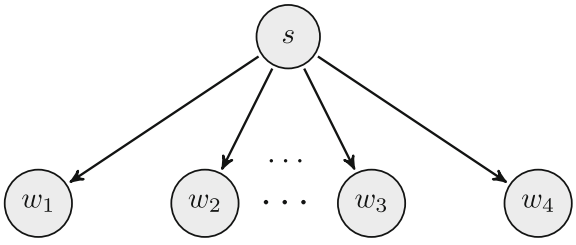
[4] These simple modal equivalence models show up when studying many aspects of rational agency: they work for specifying ranges of knowledge, issues in the logic of questions, etc.

3.2 An Initial Comparison with Modal Logics of Action

Broadly speaking, there are two general views about how to model the actions available to an agent. The first is the view found in STIT as presented in Sects. 2 and 3.1 above. Let us now consider the second view, that of modal and dynamic logics of actions (see Harel et al. 2000 for a discussion), which is also the main model of action in Situation Calculus (Reiter 2001), Automata Theory, Decision Theory and Game Theory. Its general idea is to think of actions as *transitions* moving between different "states of the system". This happens again in standard modal models $\mathcal{M} = \langle W, \{R_a\}_{a \in \mathsf{Act}}, V, s \rangle$, with worlds in W viewed as states of some process (s is the initial state), and labeled transition relations $R_a \subseteq W \times W$ for each action label $a \in \mathsf{Act}$. Each relation R_a indicates the possible executions of the basic action a. Modal languages over these models then describe possible effects of actions, while real dynamic logics also have an explicit language for speaking about complex actions defined by means of sequential composition, conditional choice, or iteration.[5] We use the phrase *PDL scenarios* for this family of paradigms.

At a first glance, these are very different views. While both perspectives acknowledge variety in possible outcomes of actions, they also have structure that the other lacks. In action-labeled approaches, the primary emphasis is on actions or events themselves and their properties, of which a description of outcome states seems only one. For instance, dancing a tango involves many features in addition to its end state: we would trivialize the process by just having an end state of 'having danced a tango'. On the other hand, many daily actions expressed in natural language are largely defined by just post-conditions on their outcomes, witness 'opening the door' or 'posting a letter'. In that sense, STIT's approach to describing actions is very natural.

We now proceed to a more technical comparison of the two styles. But to do so, we need some further touches. For a start, our simple modal picture of STIT choice situations takes out all of temporal structure. However, for a comparison with PDL, it seems more concrete to view the above 'worlds' as steps emanating from some root toward next states in a tree, a snap-shot of an ongoing decision process. The actual world is then the actual transition from the root to some next state:



[5] This is just a first intuitive pass. We will have occasion to spell out things further later on.

What this suggests is introducing a richer modal language for basic STIT, referring also to the two stages: 'now' and 'next'. This motivates the NEXT-STIT of Broersen (2011), and we will also encounter this setting in the DEL-style logic of Sect. 6. But right now, we continue in a semantic mode with worlds viewed intuitively as transitions.

Likewise, in order to compare STIT with PDL, we must also clarify the intuitive interpretation of PDL-style models. In particular, there are two broad views in the literature. One is that of transition models as abstract processes or machines, the other as unraveled temporal executions. On the *process view*, worlds are states in a process, and the relations indicate possible transitions. On this view, the model is a sort of automaton, perhaps in a very compact form, where many different transition relations can go from one state to the same next state. By contrast, the second view of PDL-models is one of *unraveled temporal execution*. Intuitively, once a process starts working, it produces a temporal universe of *executions*, being histories of successive admissible actions (cf. Clarke et al. 2000; Clarke and Emerson 1981 for this view). For the usual modal languages of action, the difference between the two views does not matter, since the execution tree is just a bisimilar unraveling of the process. And vice versa, we can think of a process as a sort of bisimulation-contracted essence of what can happen in the execution tree. But in our present setting, comparing with STIT seems to favor the temporal execution view.[6]

We therefore continue with the temporal view, where for simplicity, all event labels are taken to be unique.[7] Like with the above basic STIT, we will not take the full temporal models here, but just the snapshots of a one-step action. A **PDL action scenario** is a set of labeled transitions from some initial state s, each leading to a different successor state. This can be viewed as an obvious special "one-shot" case of the earlier-mentioned transition models.

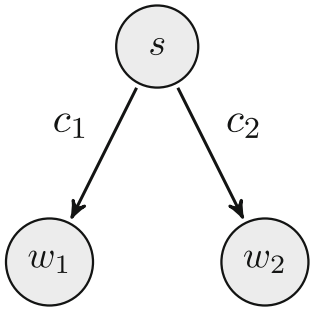
3.3 Merging the Two Perspectives on Action

Our goal in this chapter is not to reduce STIT models to PDL models, or vice versa. We find it more rewarding to show connections between the two perspectives leading to merged systems.

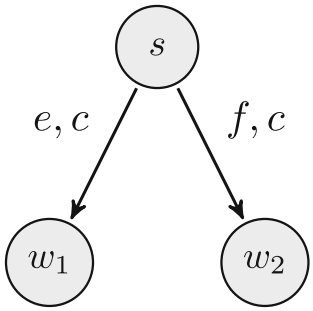
For better focus, we start with the single-agent case. Consider a simple STIT choice situation with two states $W = \{w_1, w_2\}$ and two equivalence classes: $[w_1] = \{w_1\}$ and $[w_2] = \{w_2\}$. Thus, there are two choices for the agent, which we label c_1 and c_2, respectively. A simple corresponding PDL action scenario has two transitions from the root state s labeled by c_1 and c_2, respectively:

[6] However, the process view of PDL may be closer to the dynamics of agents making choices and performing actions. We do not claim that our take in this chapter is the only way to go.

[7] This uniqueness is standard modeling practice in many temporal formalisms: if histories differ at a point, then there should be a difference in the next event.



This seems straightforward, but we have not yet found the real structure that we need. To get at this, consider a STIT choice situation with the same two states, but now only one equivalence class $[w_1] = [w_2] = \{w_1, w_2\}$. Now the agent only has one choice c. We cannot label the two transitions by c now, since that gives a PDL model with the same event, and it is unclear how this would fit the scenario. Here the difficulty is not that we cannot label the transitions: We can introduce different events for them, say e and f. In fact, this makes sense even in STIT, since histories consist of events, and as we said earlier, if two histories are different, this is because different events take place on them. But this still does not address the matter of the choice structure, and crucially related to this: how we interpret the branching in our PDL model.[8] What emerges here is an ambiguity in the usual talk about PDL models. In particular, *what do branchings mean*? Sometimes, people talk as if these are conscious choices a process or an agent can make, sometimes as if they are variations that cannot be predicted. What we need to distinguish the two senses is precisely the notion provided by STIT, that of *control*. In our first scenario, the two labels c_1 and c_2, when added to events, divide them into two control equivalence classes. In the second scenario, adding the label c to both e and f indicates how the events belong to the same control class. The agent cannot choose between the events.



[8] We could view the branching as "non-determinism" in PDL, but this does not clarify the issues very much. Non-determinism usually means that a process has several options, 'ways of doing c', but that is not the situation in the STIT model: it is not up to the agent to non-deterministically chose one or the other transition.

Action as 'events under control' To us, the preceding discussion suggests that it make sense to pool ideas. PDL has labeled events, and this makes sense, if we want to describe what happens on histories regardless of agents' choices. But STIT adds the notion of control, which also makes sense as a key feature of agency, and this helps remove a potential ambiguity in thinking about PDL models. The resulting view of actions is this:

$$Action = events + control$$

In line with this, it makes sense to merge the basic ideas of PDL and STIT into a logic with both features. Its models can have pair labels: $(event, choice)$ for transitions, thinking of equivalence relations of control on either whole transition relations, or on concrete state transitions. Such structures support a joint language with PDL event modalities $[\,e\,]$ and STIT modalities $\langle [i] \rangle$. We will not pursue technical details here, since we will discuss concrete systems of this kind later in this chapter in the modal game logics of Sect. 4, and the dynamic epistemic logic of Sect. 6. For the moment, it suffices to note that it is quite possible to have the best of both worlds, in combined logics that might be called "eventful STIT" or "controlled PDL".

We end with two more general comments about this encounter.

More on interpretations of PDL The confrontation with STIT leads to some useful clarification. We already mentioned the two main views of models as representing 'process' structure versus 'execution space'. We also discussed a major ambiguity in how one interprets branching. As a final point, we mention the issue of *events versus actions*. There is a lot of loose talk in the PDL literature about action and choice. For instance, back-and-forth clauses in bisimulation are justified by looking at 'internal choices' that a process has, and one often talks about events in PDL models as actions performed by agents.[9] But really, PDL talks about arbitrary events, all further meaning for these has to be supplied additionally in different settings. In particular, actions by agents are events with special further structure, and if it matters, these need to be made explicit. The above case of 'control' is one clear instance.[10]

A caveat about framework comparison In this section, we have engaged in high-level framework comparison. But rarefied air can exaggerate ideological differences, and it is important to also think of applied experience. In modeling practice, framework differences often prove much less dramatic than expected, as is well-known from the fact that the same real process can often be specified very happily in quite different computational paradigms. For instance, in our setting, dealing with concrete scenarios of choices and actions requires an explicit *modeler's decision* as to individuating states and actions: formal frameworks themselves do not tell us how

[9] The same is true in the dynamic epistemic logic literature: notice the terminology 'action models' versus 'event models' for its core update rules.

[10] By itself our point is not new. Adding internal structure is crucial when modeling *simultaneous* action, where one endows PDL events with internal vector structure, as in the 'interpreted systems' of Fagin et al. (1995) or the parallel games of van Benthem et al. (2008).

to do that. But then, differences between STIT and PDL tools may just amount to different legitimate decisions on how one individuates actions.[11] The problem of individuating actions has been discussed extensively in philosophy (see footnote 3 on pg. 588 of Horty and Belnap (1995) for a concise explanation), and it also shows in modeling practice in computer science.

This brings us to an important philosophical issue which we have thus far swept under the rug. In PDL models, actions are labels of transitions and this basic sorting of transitions by their labels seems to suggest a particular ontology of actions and events. In STIT models, there is no such sorting and, indeed, the only way to characterize an action is by reference to the outcomes. This raises an important question for the philosophical logician: Does adopting PDL as a logic of *actions* force one to take sides in philosophical debates about the ontology of events and actions? Our response is to bracket this question since we feel that both STIT and PDL models are open to a wide range of philosophical interpretations, regardless of the original intended interpretation of these logical frameworks. However, we certainly admit that this rather mathematical "formal modeling" view is itself controversial and we welcome (and enjoy) debates on this issue. Nonetheless, we hope that the comparative points we are making in this chapter still make sense.

4 A Merged System: Matrix Game Logic

Now, we want to make our comparisons and merges more concrete by looking at a concrete modal logic that already existed independently, and that turns out to shed some additional light on the semantic and axiomatic aspects of STIT meeting PDL.

Choices and pair events Let us return to the STIT choice situation for two agents. There is an actual world with the choices that were actually made. It makes sense to think of the worlds here as pairs of actions chosen. Note that each world w can be mapped to a unique pair of equivalence classes containing it, one for each agent, and by the product axiom, this map to pairs of equivalence classes is surjective. What we do not know is whether the map is injective, and indeed it may not be, unless we modify the product axiom to require that *different choices for all the agents have singleton intersections*. The latter constraint says that all slack in choices has been explained by introducing enough agents—perhaps including the 'environment' to take up all remaining slack. There is some simple arithmetic involved here. Assume that our model is finite. The product axiom with the singleton clause forces all

[11] As a concrete example, suppose there are two histories h, h' where an agent refrains from choosing either. Presumably, refraining means she could have made a choice for h or for h'. One way of viewing this involves three actions: choosing h, choosing h', or 'leaving things be': h, h'. This would violate the disjointness constraint of STIT. But we can also individuate events differently, with four histories: one where h is chosen, one where h' is chosen, and two copies of these except for the fact that no choice was made.

equivalence classes for agent 1 to have the same size n, as they need room for representatives of all choices of 2. The total size will be $n \times k$, with k the fixed size for 2 that exists similarly. But this suggests a viewpoint in terms of "matrix models" for joint actions that is well-known from logics of games in strategic form (cf. Osborne and Rubinstein 1994). We will develop this analogy here, using a logic proposed in (van Benthem 2007) that provides a particularly apt comparison for STIT, while also doing full justice to the PDL perspective.[12]

4.1 Modal logic of matrix games

Games induce natural models for epistemic, doxastic and preference logics, as well as conditional logics and temporal logics of action. See van der Hoek and Pauly (2006) for an overview of many such systems. Our discussion just takes a small slice.

Recall the definition of a *strategic game* for a set of players N: (1) a set A_i of actions for each $i \in N$, and (2) a utility function or preference ordering on the set of outcomes. For simplicity, one often identifies the outcomes with the set $S = \Pi_{i \in N} A_i$ of *strategy profiles*. Given a strategy profile $\sigma \in S$ with $\sigma = (a_1, \ldots, a_n)$, σ_i is the ith projection (*i.e.*, $\sigma_i = a_i$) and σ_{-i} lists the choices of all agents except agent i: $\sigma_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$.

Now, from a logical perspective, it is natural to treat the set S of strategy profiles as a universe of "possible worlds".[13] Following (van Benthem et al. 2011) for the rest of this subsection, two natural relations can be defined on these worlds. For each $\sigma, \sigma' \in S$, set for each player $i \in N$:

- $\sigma \sim_i \sigma'$ iff $\sigma_i = \sigma'_i$: this epistemic relation represents player i's "view of the game" at the *ex interim* stage where i's choice is fixed but the choices of the other players' are unknown,
- $\sigma \approx_i \sigma'$ iff $\sigma_{-i} = \sigma'_{-i}$: this relation of "action freedom" (a term taken from Seligman (2010)) gives the alternative choices for player i when the other players' choices are fixed.

Control can be freedom Our earlier discussion of STIT was in terms of *control*, including the lack of it inside players' equivalence classes. But in a multi-agent perspective, one person's lack of control is another person's freedom, and labels can switch easily.

This can all be packaged in a standard relational structure

$$\mathcal{M} = \langle S, \{\sim_i\}_{i \in N}, \{\approx_i\}_{i \in N} \rangle$$

[12] What follows here has strong resemblances to earlier work by a number of authors, including (Herzig and Lorini 2010; Balbiani et al. 2008; Lorini 2010; Lorini et al. 2009).

[13] One can also have more abstract worlds in so-called 'models of games', as is usual in epistemic game theory, see (Aumann 1999)—but this generality is not needed in what follows.

with S the set of strategy profiles and the relations just defined. Adding a valuation function interpreting a set At of atomic propositions that represent basic facts about strategy profiles (physical, or game-internal), we get standard multi-modal models.[14]

Such game models support many logical languages, from simple modal formalisms to 'hybrid modal logics', first-order logic, or even non-first-order fixed-point logics. Cf. van Benthem (2010) and Blackburn et al. (2002) on the balance of expressive power and computational complexity that arises in such design choices, a topic that will return below. However, the simplest system will do for us here. In particular, here are the key modalities for a modal logic of strategic games:

- $\sigma \models [\sim_i]\varphi$ iff for all σ', if $\sigma \sim_i \sigma'$ then $\sigma' \models \varphi$.
- $\sigma \models [\approx_i]\varphi$ iff for all σ', if $\sigma \approx_i \sigma'$ then $\sigma' \models \varphi$.

The first modality expresses the knowledge a player has once her choice is made, and given her uncertainty about what others will do, the second modality refers to her freedom of choice. As is well-known, combining the two modalities makes φ true in each world of a matrix game model: $[\sim_i][\approx_i]\varphi$ acts as a universal modality U.[15] This reflects an earlier observation about STIT—and that is no coincidence, witness the observations in Sect. 4.2 below.

What is the deductive power of the basic modal logic of strategic games? As before, we restrict attention to two-player games. First, given the nature of our relations, the separate logics are standard modal **S5** for epistemic outlook and action freedom. In addition, the interaction of these modalities validates further laws. In particular, the above fact about the universal modality is reflected in the following law:

the equivalence $[\sim_i][\approx_i]\varphi \leftrightarrow [\approx_i][\sim_i]\varphi$ is valid in all matrix game models.

This validity depends on, and in fact it expresses, the geometrical "grid property" of game matrices that, if one can go on a path $x \sim_i y \approx_i z$, then there also exists a point u with $x \approx_i u \sim_i z$. We will discuss what this feature means in some more detail in Sect. 4.3.

This concludes our brief introduction to the modal logic of matrix games. For details and further issues, the reader is referred to (van Benthem 2014).

4.2 STIT in Modal Matrix Logic

Given our discussion in Sect. 3, it will be evident how to translate the basic STIT operators into our modal language of matrix games:

---

[14] For example, a proposition p_i^a might say "agent i plays action a in the current profile"—but atomic propositions could also encode utility values for players.

[15] As noted in (van Benthem 2007), another interesting feature of our models is that 'distributed knowledge' $D_G\varphi$ for a group of players accesses those profiles where only players outside the group still have options.

$$[i \text{ stit}]\varphi := [\sim_i]\varphi, \quad \Box\varphi := [\sim_i][\approx_i]\varphi$$

This connection gives just the right combination of what we have called freedom plus knowledge.

Fact 4.1. *Our translation embeds STIT logic faithfully into the modal logic of full matrix games.*

Proof. First consider the direction from STIT theoremhood to modal game logic. Our translation validates the earlier STIT axioms, where the action modality refers to all consequences of the choice actually made, while the freedom modality looks at all alternative histories passing through the current profile. In particular, the quantifier combination employed in the Freedom axiom now becomes derivable through the theorems that are derivable for the STIT modality plus the existential modality E defined as $\langle\approx_1\rangle\langle\approx_2\rangle$:

Fact 4.2. *The formula* $(E[\sim_1]\varphi \wedge E[\sim_2]\psi) \rightarrow E(\varphi \wedge \psi)$ *is derivable in multi-***S5** *plus the commutation law for the two modalities.*

Conversely, to prove that the embedding is faithful, we need to refute each non-valid STIT law in our matrix models. To do so, take any STIT temporal counter-model in the sense of Sect. 2, and note that it suffices to look at the current moment and the next moments only (recall, that our STIT language does not contain temporal modalities). Furthermore, without loss of generality, we assume that this model is finite. More precisely, as in Sect. 3.1, we can abstract a finite two-agent basic STIT **S5**-model out of the temporal structure by letting histories be worlds, and defining agent's equivalence relations respecting their choice partitions. Now, the historic necessity operator is the universal modality while the two STIT modalities are the modalities for the equivalence relations. The last step is to show that we can transform this model into a matrix model.

If the intersections of the equivalence classes, one from each agent, are singletons, then we are done. Otherwise, we proceed as follows. A **cell** is an intersection of the agents' equivalence classes (i.e., $C = [w]_1 \cap [v]_2$ for some states w and v). Since the model is finite, there are finitely many cells and each cell has only finitely many states in them. Furthermore, by the independence assumption, each cell is non-empty. Let m be the number of elements in the largest cell. Without loss of generality, we can assume that all cells contain exactly m states (this may require adding copies of states to the model).

Organize the cells so that they form a matrix where each row contains all the cells making up a 1-equivalence class and each column contains all cells making up a 2-equivalence class. Label each cell by its position in the matrix (so, the pair (x, y) corresponds to the cell in row x and column y). There may be more than one way to organize the cells so that the rows correspond to a 1-equivalence class and the columns correspond to a 2-equivalence class. Our construction does not depend on the choice of labeling. For the remainder of the proof, fix such a $r \times c$ matrix.

Now, construct an $m \times m$ matrix for each cell. Fix a cell C labelled with (x, y) containing states $w_1, \ldots, w_m$. Worlds in the new model with be 4-tuples (i, j, x, y) where (i, j) denotes the position in the matrix and (x, y) denotes the cell containing the world. Formally, let (i, j, x, y) be a copy of $w_{i+j-1 \mod m}$. So, for example, if $m = 3$, then the world $(2, 3, x, y)$ is a copy of w_1. Note that each row and each column contains a copy of all the worlds in C.

The model is $\mathcal{M} = \langle W', \sim'_1, \sim'_2, V' \rangle$ where $W' = \{(i, j, x, y) \mid i, j \leq m, x \leq r, y \leq c\}$ (where r and c are the number of rows and columns respectively in the outer matrix). We define the uncertainty relations for the agents as follows:

- $(i, j, x, y) \sim^0_1 (i, j', x, y)$ for all $j, j' \leq m$
- $(i, j, x, y) \sim^0_2 (i', j, x, y)$ for all $i, i' \leq m$

So $\sim'_1$ runs along the rows of each inner matrix, and $\sim'_2$ runs along the columns. We extend this relation as follows:

- $(i, j, x, y) \sim^0_1 (i, 0, x, y + 1)$, where the addition is taken modulo m
- $(i, j, x, y) \sim^0_2 (0, j, x + 1, y)$, where the addition is taken modulo m

Let $\sim'_1$ and $\sim'_2$ be the reflexive and transitive closure of $\sim^0_1$ and $\sim^0_2$, respectively. Finally, the valuation V' is copied from the original valuation in the obvious way. We note the following two facts about the construction:

1. If $(i, j, x, y) \sim'_1 (i', j', x', y')$, then $i' = i$ and $x' = x$. If $y' = y$, then $w_{i+(j-1) \mod m}$ and $w_{i+(j'-1) \mod m}$ are both in the cell labeled by (x, y), and so $w_{i+(j-1) \mod m} \sim_1 w_{i+(j'-1) \mod m}$. If $y' \neq y$, then $w_{i+(j-1) \mod m}$ and $v_{i+(j'-1) \mod m}$ are in different cells. However, we still have $w_{i+(j-1) \mod m} \sim_1 v_{i+(j'-1) \mod m}$ since we assume that cells in the same row are in the same 1-equivalence class.
2. If $(i, j, x, y) \sim'_2 (i', j', x', y')$ then $j' = j$ and $y' = y$. If $x' = x$, then $w_{i+(j-1) \mod m}$ and $w_{i+(j'-1) \mod m}$ are both in the cell labeled by (x, y), and so $w_{i+(j-1) \mod m} \sim_2 w_{i+(j'-1) \mod m}$. If $x' \neq x$, then $w_{i+(j-1) \mod m}$ and $v_{i+(j'-1) \mod m}$ are in different cells. However, we still have $w_{i+(j-1) \mod m} \sim_2 v_{i+(j'-1) \mod m}$ since we assume that cells in the same column are in the same 2-equivalence class.

These observations show immediately that the newly constructed model is bisimilar to the original STIT model. Hence, they satisfy the same formulas in our language.

The last thing we need to check is that the intersection of agents' equivalence classes are singletons. Suppose that $(i_0, j_0, x_0, y_0) \sim'_1 (i', j', x', y')$, $(i_0, j_0, x_0, y_0) \sim'_1 (i'', j'', x'', y'')$, $(i_1, j_1, x_1, y_1) \sim'_2 (i', j', x', y')$ and $(i_1, j_1, x_1, y_1) \sim'_2 (i'', j'', x'', y'')$. Then, by construction, $i' = i'' = i_0$, $x' = x'' = x_0$ and $j' = j'' = j_1$ and $y' = y'' = y_1$. Hence, $(i', j', x', y') = (i'', j'', x'', y'')$, as desired.

We have shown that our translation is both correct and faithful.[16] QED

[16] Note that the construction given in this proof is only needed because the *singleton intersection property* (the intersection of all the agents equivalence classes are singletons) is not definable in our

This proof exploits the fact that matrix game models are close to the multi-**S5** models for basic STIT defined in Sect. 3.1. Still, the geometrical matrix perspective is useful, since it links up with a body of existing results. We will see a number of examples as we proceed.

4.3 Complexity and Correlation

While the preceding embedding makes sense, it does embed STIT in a system whose behavior is potentially complex. Richer modal logics of matrix games may well be unaxiomatizable and worse. The reason is the above commutation law for the two equivalence relations. While this may look like a pleasant structural feature of matrices, its logical effects are delicate. It is well-known that the general logic of bi-modal languages plus a universal modality on 'grid models' with two immediate successor relations is not decidable, and not even axiomatizable: indeed, it is "Π^1_1-complete" (cf. Halpern and Vardi 1989; Marx 2007; Gabbay et al. 2003; van Benthem and Pacuit 2006). The reason is that grid structure can be exploited to encode computations of Turing machines on successive rows, or geometrical "tiling problems" of known high complexity.

Now, it is not clear whether our most basic modal game logic falls into this trap, since our models only have two *equivalence relations*, one horizontal and one vertical. Indeed, its closeness to STIT may suggest that it remains decidable–even though this does not follow from our earlier embedding result, that went in the opposite direction. Still, Halpern and Vardi (1989) and Spaan (1990) show high complexity of modal logics on grid models with reflexive transitive relations, using an encoding trick with alternating proposition letters.[17]

This potential high complexity, while not directly threatening to STIT, does raise an interesting issue in modeling action. A standard way of defusing high complexity results is by allowing *more models*. In the present setting, the resulting structures are *general game models* where certain strategy profiles may be absent. Then general modal game logic becomes much simpler, being just multi-agent modal **S5** without any connecting axioms (van Benthem 1997).[18]

(Footnote 16 continued)
language. However, if the language contains *group* STIT operators $[G\,\mathsf{stit}]\varphi$ meaning that the group G can see to it that φ is true, then the singleton intersection property is definable via the formula $[\mathcal{A}\,\mathsf{stit}](\varphi \vee \psi) \rightarrow [\mathcal{A}\,\mathsf{stit}]\varphi \vee [\mathcal{A}\,\mathsf{stit}]\psi$, where $\mathcal{A}$ is the set of all agents. Furthermore, the argument would be very different once we consider STIT formulas with temporal operators shifting moments along histories, as is suggested in Sect. 7.

[17] Such encodings also work with two equivalence relations and common knowledge in one dimension of the grid model, while time provides the other dimension. See van Benthem and Pacuit (2006) for an extensive survey.

[18] For a concrete counter-example, note that the formula in Fact 4.2 is not valid on such models. Suppose that $\approx_i$ and $\sim_i$ are arbitrary equivalence relations for each i. Consider a model where $w \approx_1 v$ and $w \approx_2 v'$ with $v \neq v'$, and both v and v' are dead-end states (i.e., we only have $v \sim_1 v$ and $v' \sim_2 v'$). Suppose that φ is true at v only

Now this is not just a technical move: "profile gaps" encode something interesting, namely *correlations* between behavior of agents. In a general game model, if player i changes her move, then the only available profiles for this may now be ones where some other player j has changed his move as well. Game theorists have studied correlations extensively: cf. (Aumann 1987; Brandenburger and Friedenberg 2008). But the same notion has come up in logic, since correlations provide "information channels" where the behavior of one agent can carry information about that of another (Barwise and Seligman 1997). And more recently, generalized forms of such dependencies have become the focus of attention in "dependence logics" (Väänänen 2007). In other words, independence may be costly, and the Product Axiom that seemed the pride of STIT may eventually stand in the way, being just an extreme case of a more sophisticated theory of agent behavior.[19]

In the rest of this chapter, we look at extensions of the current framework with features that seem essential to rational agency, and that have been the subject of study in dynamic logics.

5 The Roles of Knowledge

Our connection between STIT and matrix games introduced a notion of knowledge, of agents that have decided, but do not know yet what the others have chosen. Knowledge is not mentioned explicitly in STIT framework, but it seems to be lurking behind the scenes here. In fact, it is present in more than one way: choice and action naturally come with *varieties of knowledge*. Here is how this can happen, even in the simple setting that we have considered.

Consider a one-step action. Before I have made my choice, I only know that one of the available future histories will occur: and in that sense, the STIT tree modality already acts as a form of knowledge about how the whole future can unfold. This knowledge can be significant, since the tree encodes the "protocol" of all possible runs of the current process.

Next, right after I have chosen my action, I know what I am going to do, but I still do not know what the others will do, and this was the sort of knowledge based on personal decisions that was made explicit in the matrix models for games of Sect. 4.

Finally, once both our actions have actually taken place, agents do know what was chosen, if we assume that they observe these actions publicly. Knowledge from observation of events is a major source of information in a temporal world. It is is often encoded in epistemic uncertainty relations between moments of time that are used to model information-driven processes, such as games with imperfect information (cf. Binmore 2009; Parikh and Ramanujam 2003; Fagin et al. 1995). As for its driving

(Footnote 18 continued)
and ψ is true at v' only. Then the antecedent is true, but the consequence is not.

[19] Also relevant to the issue of generalized "profile models" is recent work by Roberto Ciuni on connections between generalized STIT models and notions of effectivity in games, and on actions whose effects are only given probabilistically: Ciuni (2013).

forces, updating knowledge from public observation or more private sources is the key topic in dynamic-epistemic logics (van Benthem 2011).

It is natural to add epistemic operators of all these sorts to logics of decision and action, and in fact, this is happening in logics of games (cf. van Benthem 2014). Many kinds of knowledge relevant to action scenarios are local, having to do with what agents know temporarily as they make a choice. But more global "procedural knowledge" about the future of the process is essential, too, and then the trees of STIT may lose their grip. If I know something about your space of possible strategies, the informational situation will need "STIT forests" rather than trees to distinguish the alternatives (cf. van Benthem et al. 2009). The same complication arises in genuine multi-agent scenarios. One cannot assume that agents know everything about others, and to cope with this variation, again, models have to be complicated beyond the basic STIT format.

Pursuing these matters is beyond the scope of this chaper, but explicit modeling of knowledge seems inescapable in a serious theory of choice and action. For a discussion along these lines, see Pacuit and Simon (2011) for a logical system that merges ideas from STIT and PDL while explicitly representing the agents' knowledge. We see it as one virtue of our linking up STIT and PDL that experiences in the latter area can then be enlisted for the former. Our next section will present a case study, of one particular dynamic epistemic logic with added STIT features.

6 Dynamic Epistemic Logic Meets STIT

Temporal trees with epistemic features may be viewed as a record of actions unfolding over time, while marking local uncertainties (or information) that agents had. If we want to understand the dynamics that gives rise to such a record, we need an account of information update in a temporal universe. A typical system where PDL-style events and knowledge come together is *dynamic epistemic logic* (DEL). We assume the reader is familiar with its basics, and so, we only give the key definitions here (see van Benthem 2011 for more details and motivation).

The basics of DEL update The basic structures are **epistemic models**, tuples $\langle W, \{R_i\}_{i \in I}, V \rangle$ with W a (finite) set of worlds, $R_i \subseteq W \times W$ an equivalence relation, and $V : \mathsf{At} \to \wp(W)$ a valuation function marking at which worlds the atomic propositions in At are true. Over these models the basic language of epistemic logic $\mathcal{L}_{EL}$ can be interpreted, including universal modalities $K_i\varphi$ for "agent i knows that φ. This much is completely standard.

The central idea of *dynamic* epistemic logic is now to describe social interaction, including agents' uncertainty about the events they witness, in so-called **event models**. These are tuples $\mathbf{E} = \langle E, \{S_i\}_{i \in I}, \mathsf{pre} \rangle$ with E a (finite) set of basic events, $S_i \subseteq E \times E$ is an uncertainty relation, and $\mathsf{pre} : E \to \mathcal{L}_{EL}$ assigns to each event $e \in E$ a formula that serves as a **precondition** for that event.

Now, dynamic changes in agents' information can be described by means of *product update* transforming a current pointed[20] epistemic model $\mathcal{M}$ using the event model $\mathbf{E}$. The **product model** $\mathcal{M} \oplus \mathbf{E} = \langle W', \{R'_i\}_{i \in I}, V' \rangle$ is defined as follows:

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \models \mathsf{pre}(e)\}$;
- $(w, e) R'_i (w', e')$ iff $w R_i w'$ and $e S_i e'$; and
- $(w, e) \in V'(p)$ iff $w \in V(p)$

More precisely, the understanding is that $\mathcal{M}$ has an actual world w, while $\mathbf{E}$ has an actual event e. Product update works for many epistemic scenarios, while it has also been extended to deal with belief and preference change. The language of DEL then adds dynamic modalities $\langle \mathbf{E}, e \rangle \varphi$ that describe at worlds w in $\mathcal{M}$ what is true one step later in the product model with $\mathbf{E}$ and actual event e. The resulting logic of informational events can be axiomatized completely by a compositional technique of 'recursion axioms' analyzing compounds $\langle \mathbf{E}, e \rangle K_i \varphi$ in terms of conditional knowledge that agents had before the update. The details of this are beyond our needs here, but see van Ditmarsch et al. (2007), van Benthem (2011) for more extensive analysis.

Our aim in this section is just to show how, in line with our analysis of Sect. 3, STIT ideas of control fit quite well in this PDL stronghold.

Extending DEL with control A first easy task is adding the earlier control relations for different agents to event models, which just requires adding equivalence relations.[21] Now we can set up a calculus of reasoning. Our dynamic-epistemic language still has its basic event modalities $\langle \mathbf{E}, e \rangle \varphi$, but now we can also introduce a STIT operator

$$\langle \mathbf{E}, e, i \rangle \varphi$$

saying in $\mathcal{M}, w$ that φ is true in all product models $(\mathcal{M}, w) \oplus (\mathbf{E}, f)$ for all events f that are control equivalent to e for agent i. This is formally quite similar to an operator that would already make sense in DEL as it stands, namely, stating the 'observational knowledge' that an agent has acquired after product update with the current event model $\mathbf{E}$.

The complete dynamic logic of this expanded system lies embedded in the base logic of DEL in an obvious manner. Its laws for the new control operator will be essentially those of STIT. But what we obtain in this way is a much richer logic of one-step information flow plus an explicit account of agents' choices of actions where relevant. However, it should be noted that this logic still runs on the usual analysis of DEL's standard dynamic modality.

One crucial feature is that, unlike standard DEL logics, this new system does not have a modality reflecting its dynamic control relations in the static epistemic

[20] A pointed epistemic model is a model with a distinguished state intended to represent the "actual" state of the world.

[21] This may have to be modified when we want some events to just happen without agency. Also, there are problems of intuitive interpretation for control in private-information scenarios, but we ignore these here.

base models $\mathcal{M}$.[22] One might think of this negatively as limiting the logical status of control, reflecting its ephemeral nature. Our own more positive view is that this feature makes event models really come into their own, as carrying crucial information that is *sui generis*.

While our proposal for merging enriches DEL with STIT ideas, what good does it do in the opposite direction? One effect is that we now have a logic that describes both steps in the fork models of Sect. 3, before and after the choice. Thus, it is a logic of choosing and moving ahead, like the NEXT-STIT system of (Broersen, 2011). But the main virtue is that, given the long experience in DEL, our merged system plugs STIT into the world of private versus public information, imperfect information games, and much more.

Dynamifying STIT But the DEL perspective also suggests a more radical move, affecting our view of the scenarios that motivated STIT in the first place. STIT is a logic of deliberate choice and action, but remarkably, it does not analyze any of these activities explicitly, recording only their outcomes.[23] By contrast, the DEL methodology follows a main principle of Logical Dynamics:

Where there is a change, there is an event.

Taking this line, can we 'dynamify' STIT in DEL style? What are the main events that take place in a choice scenario? Here are the main stages as we see them:

deliberation, decision, action, and observation

In a first *deliberation stage*, we analyze our options, and find optimal choices. Next, at the *decision stage*, we make up our mind and choose an action of our own. Then at the *action stage*, everyone acts publicly, and this gets observed, something that we can also model as a separate *observation stage*, though things happen simultaneously.

All these stages can be analyzed using DEL-style models. Perhaps the easiest is the final stage, where an event model will do with all possible events, marking the actual one, and giving agents the right amount of observational powers: totally public in STIT, perhaps more mixed in other settings. But the intermediate stage, too, invites event models. We can have pair events with control relations, as we just introduced in Sect. 6, and then get the matrix models of Sect. 4 as an output. Finally, modeling the initial deliberation stage is more complex, since many factors can weigh in here that are not represented in basic STIT models, such as agents' *preferences* over outcomes. Still, there is a growing body of work on deliberation analyzed in terms of DEL-style updates (van Benthem 2007; Pacuit and Roy 2011), and this might inform an account of deliberation that seems a natural companion to any logic of "deliberate action".

[22] This makes our DEL system with control different, e.g., from DEL logics of questions with issue relations: see van Benthem and Minică (2012).

[23] This output orientation on choice is of course precisely the official STIT view of actions.

7 Further Directions

There are many follow-up topics to our analysis, of which we mention three.

The first is the addition of agents' *preferences*. Clearly, this further structure is crucial to the game logics of Sect. 4, and existing modal systems do incorporate preferences in order to define and reason about notions like 'best response', Nash equilibrium, and rational behavior generally (cf. van Benthem 2011, 2014 for such notions analyzed in DEL). In particular, the interplay of actions and preferences has already been studied in the matrix logics of Sect. 4, using techniques from (Liu 2012; van Benthem et al. 2009).[24]

Adding preferences seems a necessity for STIT as well, since rational agency is about *best actions* rather than just any actions, and agents may also prefer ensuring φ rather than ψ for other reasons, including deontic norms. This need not be a simple matter, since best action is not just a matter of, say, finding Pareto-optimal simultaneous choices for all agents. As we know from game theory, more complex deliberation methods are needed, such as iterated removal of dominated actions.[25] All this seems a happy marriage with STIT, and indeed, many of the relevant issues are addressed in Horty's book (Horty 2001).

The next extension would be the study of long-term *temporal evolution*. Our logics so far described single steps in a larger process, but it has long been acknowledged that the proper stage for studying agency is that of a linear- or branching-time temporal logic (Fagin et al. 1995; Parikh and Ramanujam 2003). The same is true for STIT, and one question that seems of interest is whether our one-step event models with control relations can be related systematically to epistemic temporal universes via representation theorems extending those of van Benthem et al. (2009).

Our final topic is *strategic interactive behavior*. We started our presentation of STIT with its basic properties for agents' choices: for each agent, these formed a *partition* of all possible outcomes (call this the Partition property), and also, any two choices for different agents have to overlap. This level of stating constraints is similar to that of representation theorems for games characterizing players' strategic *powers*, forcing the game to end in certain sets of outcomes by playing one of their strategies against any counterplay of the opponent. The latter type of result, however, usually refers to powers in a longer extensive game that can take many individual steps. For instance, van Benthem (2001) characterizes players' powers in finite determined two-player games in terms of three constraints: *Monotonicity* (powers are upward closed), *Consistency* (any two powers of different players overlap), and *Determinacy* (if a set of outcomes is not a power for one of the players, then its complement is a power for the other player). Of these three, Determinacy is typically lost in the

[24] Many interesting new problems arise in this area. One is finding a formalization of basic game-theoretic reasoning that makes sense for rational action generally: as initiated in (van Benthem 2007). Another unresolved issue is whether introducing preference structure increases the computational complexity of the modal logic of action, an issue known as the "price of rationality".

[25] Thus one might first iteratively prune a given choice situation in this way, and only follow the standard STIT-style format once an equilibrium has been reached.

STIT setting of simultaneous action. Nevertheless, it seems significant that there are extended representation results for players' powers in extensive games with *imperfect information* that require only Monotonicity and the typical STIT constraint of Consistency (cf. again van Benthem 2001).[26]

We end with just one simple observation. What happens to the key STIT constraints when we consider iterated simultaneous action? Most importantly, the crucial property of Partition disappears, and the reason is very instructive. When we make consecutive choices, our available strategies get enriched. In a one-step scenario, agents could only choose one of their actions *ab initio*. But now, they can have strategies letting their next action depend on the observed behavior of the other agents. A standard example of this is the famous strategy *Tit for Tat* in evolutionary game theory: one copies the opponent's preceding move. Hence, the strategies available at the second level do not just consist of choosing an action uniformly, they can depend on the behavior of others. It is easy to see that the disjointness property for sets of outcomes (i.e., the powers matching these strategies) are no longer disjoint.[27] On the other hand, this richer set of strategies does depend crucially on a special feature of the STIT scenario, namely the public observation of everyone's moves. If there were no such observation, then players' could not make their choices dependent on what others have done, and we would get a simple product model of two consecutive actions that does satisfy the Partition condition. Put differently, one-step simultaneous action does not allow for sequential *dependence* of actions, though it may allow for *correlation* as we saw in Sect. 4. But it is precisely the observation feature built into STIT that does make more sophisticated dependent behavior possible as actions get repeated.

8 Conclusion

In this chapter, we have lightly compared the STIT approach to choice and action with that offered by dynamic logic, broadly conceived (including dynamic-epistemic logics). We found that, despite differences in style and presentation, these frameworks are much more congenial than is often thought. Indeed, key ideas from STIT about actions and control merged well with modal logics of games, and in particular, they led to natural dynamic-epistemic logics of information and events that incorporate the crucial STIT notion of *control*. We have only proposed a few such bridges here, without any sustained development, suggesting how ideas might flow across, and further directions pursued. Even so, we hope to have put to rest some views about vast chasms separating STIT and PDL that are sometimes found in the literature.

---

[26] There is also a literature with more sophisticated representation results that are significant here, of which we mention Bonanno (1992), Pauly (2001) and Goranko et al. (2013).

[27] It is an interesting problem whether some special properties remain for STIT powers. In particular, the temporal logic of Ciuni and Zanardo (2010) seems relevant to analyzing these matters, including the special constraints imposed on STIT models if we do insists on the above properties of powers for single agents, or groups of these: cf. Zanardo (2013).

We are by no means the first to have observed the compatibility of STIT and ideas from the world of PDL and DEL. Notably, Horty articulates many of the idea sketched in this chapter in his important book (Horty 2001). Also, Xu (2010, 2012) are interesting examples of STIT systems that have borrowed notions of action and strategy from the PDL tradition to form richer frameworks for strategic agency. We see our analysis as making a small push in the same direction.

Finally, we recall an earlier point made at the start of our analysis. A paradigm is not just a set of definitions of structures and axioms for reasoning. It is also a belt of applications, in the terminology of Kuhn (1962), a growing family of successful "exemplars". This makes frameworks harder to compare and merge, since their success does not just depend on their formal backbone, but also on the "art of modeling" that has been invested by skilled practitioners. In a practical setting, choices between paradigms may just be choices of taste and life-style, and these of course will not be affected much by theoretical analysis. Still, tastes can at least be diversified—and we hope to have contributed at least to what is on the menu in the logical study of deliberate action.

References

Aumann, R.J. 1987. Correlated equilibrium as an expression of bayesian rationality. *Econometrica* 55: 1–18.

Aumann, R. 1999. Interactive epistemology I: Knowledge. *International Journal of Game Theory* 28: 263–300.

Balbiani, P., A. Herzig, and N. Troquard. 2008. Alternative axiomatics and complexity of deliberative STIT theories. *Journal of Philosophical Logic* 37(4): 387–406.

Barwise, J. and J. Seligman. 1997. *Information flow: The logic of distributed systems*. Cambridge: Cambridge University Press.

Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future: Agents and choice in our indeterminist world*. Oxford: Oxford University Press.

Binmore, K. 2009. *Rational decisions*. Princeton: Princeton University Press.

Blackburn, P., M. de Rijke, and Y. Venema. 2002. *Modal logic*. Cambridge: Cambridge University Press.

Bonanno, G. 1992. Set-theoretic equivalence of extensive-form games. *International Journal of Game Theory* 20: 429–447.

Brandenburger, A., and A. Friedenberg. 2008. Intrinsic correlations in games. *Journal of Economic Theory* 141(1): 28–67.

Broersen, J., A. Herzig, and N. Troquard. 2006a. Embedding alternating-time temporal logic in strategic STIT logic of agency. *Journal of Logic and Computations* 16(5): 559–578.

Broersen, J., A. Herzig, and N. Troquard. 2006b. From coalition logic to stit. *Electronic Notes Theoretical Computer Science* 157(4): 23–35.

Broersen, J. 2011. Deontic epistemic STIT logic distinguishing modes of mens rea. *Journal of Applied Logic* 9(2): 127–152.

Brown, M.A. 1988. On the logic of ability. *Journal of Philosophical Logic* 17(1): 1–26.

Brown, M.A. 1992. Normal bimodal logics of ability and action. *Studia Logica* 51(3/4): 519–532.
Chellas, B.F. 1992. Time and modality in the logic of agency. *Studia Logica* 51(3/4): 485–518.
Ciuni, R. 2013. Ought implies can, omission, and probabilistic deliberative STIT Lecture at the Institute for Logic, Language and Computation, LIRa Session, 30 March 2013.
Ciuni, R., and A. Zanardo. 2010. Completeness of a branching-time logic with possible choices. *Studia Logica* 96(3): 393–420.
Clarke, E.M., and E.A. Emerson. 1981. Design and synthesis of synchronization skeletons using branching-time temporal logic. In *Logic of programs*, ed. D. Kozen, 52–71. Berlin: Springer.
Clarke, E.M., O. Grumberg, and D.A. Peled. 2000. *Model checking*. Cambridge: The MIT Press.
Fagin, R., J. Halpern, Y. Moses, and M. Vardi. 1995. *Reasoning about knowledge*. Cambridge: The MIT Press.
Gabbay, D., A. Kurucz, F. Wolter, and M. Zakharyaschev. 2003. *Many-dimensional modal logics: Theory and applications*. Amsterdam: Elsevier.
Goranko, V., W. Jamroga, and P. Turrini. 2013. Strategic games and truly playable effectivity functions. *Autonomous Agents and Multi-Agent Systems* 26(2): 288–314.
Halpern, J., and M. Vardi. 1989. The complexity of reasoning about knowledge and time. *Journal of Computer and System Sciences* 38: 195–237.
Harel, D., D. Kozen, and J. Tiuryn. 2000. *Dynamic logic*. Cambridge: The MIT Press.
Herzig, A., and F. Schwarzentruber. 2010. Properties of logics of individual and group agency. In *Proceedings of advances in modal logic*, vol. 7, eds. C. Areces, and R. Goldblatt, 133–149. London: College Publications.
Herzig, A., and E. Lorini. 2010. A dynamic logic of agency I: STIT, capabilities and powers. *Journal of Logic, Language and Information* 19(1): 89–121.
Horty, J. 2001. *Agency and deontic logic*. Oxford: Oxford University Press.
Horty, J.F., and N. Belnap. 1995. The deliberative stit: A study of action, omission, ability, and obligation. *Journal of Philosophical Logic* 24(6): 583–644.
Kuhn, T. 1962. *The structure of scientific revolution*. IL: The University of Chicago Press.
Liu, F. 2012. *Reasoning about preference dynamics*. Synthese library. Berlin: Springer.
Lorini, E., F. Schwarzentruber, and A. Herzig. 2009. Epistemic games in modal logic: Joint actions, knowledge and preferences all together. In *LORI-II Workshop on Logic, Rationality and Interaction, Chongqing, China*, eds. X. He, J. Horty, and E. Pacuit, 212–226. Berlin: Springer.
Lorini, E. 2010. A dynamic logic of agency II: Deterministic $\mathcal{DLA}$, coalition logic, and game theory. *Journal of Logic, Language and Information* 19(3): 327–351.
Lorini, E., and F. Schwarzentruber. 2010. A modal logic of epistemic games. *Games* 1(4): 478–526.
Marx, M. 2007. Complexity of modal logic. In *Handbook of modal logic*, eds. P. Blackburn, J. van Benthem, and F. Wolter, 139–189. New York: Elsevier.
Osborne, M., and A. Rubinstein. 1994. *A course in game theory*. Cambridge: The MIT Press.
Pacuit, E., and O. Roy. 2011. A dynamic analysis of interactive rationality. In *LORI, Volume 6953 of lecture notes in computer science*, eds. H.P. van Ditmarsch, J. Lang, and S. Ju, 244–257. Berlin: Springer.
Pacuit, E., and S. Simon. 2011. Reasoning with protocols under imperfect information. *The Review of Symbolic Logic* 4(3): 412–444.
Parikh, R., and R. Ramanujam. 2003. A knowledge based semantics of messages. *Journal of Logic, Language and Information* 12: 453–467.
Pauly, M. 2001. *Logic for social software*. Ph. D. thesis, ILLC University of Amsterdam (Dissertation Series 2001–10).
Reiter, R. 2001. *Knowledge in action: Logical foundations for specifying and implementing dynamic systems*. Cambridge: The MIT Press.
Seligman, J. 2010. Hybrid logic for analyzing games. Lecture at the Workshop Door to Logic, Beijing, May 2010.
Spaan, E. 1990. Nexttime is not necessary. In *Proceedings of TARK*, ed. R. Parikh, 241–256.
Väänänen, J. 2007. *Dependence logic*. Cambridge: Cambridge University Press.

van Benthem, J., and Ştefan Minică. 2012. Toward a dynamic logic of questions. *Journal of Philosophical Logic* 41(4): 633–669.
van Benthem, J., and E. Pacuit. 2006. The tree of knowledge in action: Towards a common perspective. In *Proceedings of Advances in Modal Logic*, vol. 6, eds. G. Governatori, I. Hodkinson, and Y. Venema, 87–106. London: King's College Press.
van Benthem, J., J. Gerbrandy, T. Hoshi, and E. Pacuit. 2009a. Merging frameworks for interaction. *Journal of Philosophical Logic* 38(5): 491–526.
van Benthem, J., S. Ghosh, and F. Liu. 2008. Modelling simultaneous games in dynamic logic. *Synthese* 165(2): 247–268.
van Benthem, J., P. Girard, and O. Roy. 2009b. Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic* 38: 83–125.
van Benthem, J., E. Pacuit, and O. Board. 2011. Toward a theory of play: A logical perspective on games and interaction. *Games* 2(1): 52–86.
van Benthem, J. 1997. Modal foundations for predicate logic. *Logic Journal of the IGPL* 5(2): 259–286.
van Benthem, J. 2001. Games in dynamic epistemic logic. *Bulletin of Economic Research* 53(4): 219–248.
van Benthem, J. 2007. Rational dynamics and epistemic logic in games. *International Journal of Game Theory Review* 9(1): 13–45.
van Benthem, J. 2010. *Modal logic for open minds*. Stanford: CSLI Publications.
van Benthem, J. 2011. *Logical dynamics of information and interaction*. Cambridge: Cambridge University Press.
van Benthem, J. 2014. *Logic and games*, manuscript, ILLC, University of Amsterdam. To appear with The MIT Press, Cambridge MA.
van der Hoek, W., and M. Pauly. 2006. Modal logic for games and information. In *Handbook of modal logic: Studies in logic*, vol. 3, eds. P. Blackburn, J. van Benthem, and F. Wolter, 1077–1148. New york: Elsevier.
van Ditmarsch, H., W. van der Hoek, and B. Kooi. 2007. *Dynamic Epistemic Logic*. Synthese library. Berlin: Springer.
Xu, M. 1995. On the basic logic of STIT with a single agent. *Journal of Symbolic Logic* 60(2): 459–483.
Xu, M. 2010. Combinations of stit and actions. *Journal of Logic, Language and Information* 19(4): 485–503.
Xu, M. 2012. Events and actions. *Journal of Philosophical Logic* 41(4): 765–809.
Zanardo, A. 2013. Indistinguishability, choices, and logics of agency. *Studia Logica on-line* 1–22.

Intentionality and Minimal Rationality in the Logic of Action

Daniel Vanderveken

Abstract Philosophers have overall studied intentional actions that agents attempt to perform in the world. However the pioneers of the logic of action, Belnap and Perloff, and their followers have tended to neglect the intentionality proper to human action. My primary goal is to formulate here a more general logic of action where intentional actions are primary as in contemporary philosophy of mind. In my view, any action that an agent performs involuntarily could in principle be intentional. Moreover any involuntary action of an agent is an effect of intentional actions of that agent. However, not all unintended effects of intentional actions are the contents of unintentional actions, but only those that are historically contingent and that the agent could have attempted to perform. So many events which happen to us in our life are not really actions. My logic of action contains a theory of attempt, success and action generation. Human agents are or at least feel free to act. Moreover their actions are not determined. As Belnap pointed out, we need branching time and historic modalities in the logic of action in order to account for indeterminism and the freedom of action. Propositions with the same truth conditions are identified in standard logic. However they are not the contents of the same attitudes of human agents. I will exploit the resources of a non classical predicative propositional logic which analyzes adequately the contents of attitudes. In order to explicate the nature of intentional actions one must deal with the beliefs, desires and intentions of agents. According to the current logical analysis of propositional attitudes based on Hintikka's epistemic logic, human agents are either perfectly rational or completely irrational. I will criticize Hintikka's approach and present a general logic of all cognitive and volitive propositional attitudes that accounts for the imperfect but minimal rationality of human agents. I will consider subjective as well as objective possibilities and explicate formally possession and satisfaction conditions of propositional attitudes. Contrary to Belnap, I will take into account the intentionality of human agents and

D. Vanderveken (✉)
Department of Philosophy, University of Quebec at Trois-Rivières, Trois-Rivières, QC G9A 5H7, Canada
e-mail: daniel.vanderveken@gmail.com

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_15,

explicate success as well as satisfaction conditions of attempts and the various forms of action generation. This chapter is a contribution to the logic of practical reason. I will formulate at the end many fundamental laws of rationality in thought and action.

I will only consider here *individual actions* and *attitudes* of single agents at one moment. Examples of such individual actions are intended body movements like voluntarily raising one's arm, some effects of these movements like touching something and saluting someone, mental actions like judgements and elementary illocutionary acts such as assertions and requests. Whoever performs an action at a moment has individual beliefs, desires and intentions at that very moment. Individual actions (and attitudes) of agents at a single moment are the simplest kinds of action (and attitudes) from a logical point of view. They part of other kinds of individual or collective actions (and attitudes) which last during several moments of time.

In order to contribute to the foundations of the logic of action I will attempt to answer general philosophical questions: What is the logical form of proper intentional actions? Which attitudes do they contain? In my view, attempts are constitutive of intentional actions. Attempted actions have success conditions: either agents succeed or fail in performing them? How can we define success and failure? We need an account of agents' *reasons* in our logic of action and attitudes. Indeed agents have *theoretical reasons* for believing propositions and they make their attempts *for practical reasons*. Their intentional actions can both create reasons and be subject to demands for reasons i.e. for justifications. Moreover voluntary actions are related by the relation of being means to achieve ends (Aristotle). Agents make their attempts in order to perform other actions. How can we account for their objectives? Our intentional actions have involuntary effects in the world. In walking intentionally on the snow an agent might unintentionally slip and fall. What are the logical relations that exist between our intentional and unintentional actions? Some types of action *strongly commit the agent to* performing other types of action. Whoever shouts produces sounds. Any instance of an action of the first type contains an action of the second type. Moreover certain action tokens *generate* others in certain circumstances. Whoever expresses an attitude that he does not have, is lying. But he could be sincere at another moment. What are the basic laws governing agentive commitment and action generation? In particular, how can agents perform certain actions by way of performing others? Are all actions performed by an agent at a moment generated by a single basic intentional action of that agent at that moment? If yes, what is the nature of basic actions?

As Brentano (1993) pointed out, agents of propositional attitudes and intentional actions have *intentionality*: they are *directed at* objects and facts of the world that they represent. From a logical point of view, propositional attitudes have logically related *conditions* of *possession* and of *satisfaction*. Whoever *possesses* a propositional attitude is in a certain mental state: he or she represents what has to happen in the world in order that his or her attitude is satisfied. Beliefs are satisfied whenever they are true, desires whenever they are realized and intentions whenever they are executed. So agents having beliefs represent how things are in the world according to them. Agents having desires represent how they would prefer things to be in the world and agents having intentions represent how they should act in order to execute

their intentions. Propositional attitudes consist of a *psychological mode M* with a *propositional content P*. They are the simplest kinds of individual attitudes directed at facts. My first objective here is to explicate adequately *possession and satisfaction conditions of all propositional attitudes*. My second objective is to explicate the nature of intentional actions and the different kinds of action generation whether voluntary or not. According to standard epistemic logic human agents are either perfectly rational or totally irrational. I will advocate an intermediate position compatible with contemporary philosophy of mind according to which human agents are not perfectly but minimally rational. In my logical approach, one can formulate adequate laws of psychological commitment and avoid current epistemic and volitive paradoxes. In order to account for minimal rationality[1] I will exploit the resources of a non classical propositional *predicative logic* that distinguishes propositions with the same truth conditions that do not have the same cognitive or volitive value.

The structure of this chapter is the following. I will explain in the first section my predicative analysis of propositional contents. Next I will explicate components of psychological modes and define possession and satisfaction conditions of propositional attitudes. I will explain in the third section the principles[2] of my logic of action where intentional actions are primary as in contemporary philosophy.[3] Because *intentional actions* are actions that agents *attempt* to perform in the world, the basic individual actions of each agent are in my logic his or her *primary attempts* (usually attempts of body movement). My ideographical object-language has richer expressive capacities than that of Belnap. It expresses in addition to modalities, time and individual actions of agents, their attempts and their cognitive and volitive propositional attitudes. I will give a formal account of intentionality and explicate the nature of attempts and forms of action generation in the fourth section.[4] In the last section I will enumerate a few valid laws of my logic after having criticized Searle's skepticism against the logic of practical reason. I will also explain why the logic of action is so important for the purposes of illocutionary logic.

1 Analysis of Propositional Contents of Attitudes

Propositions with the same truth conditions are not the contents of the same attitudes and intentional actions. One can believe and assert that Rome is a capital without believing and asserting that it is a capital and not an erythrocyte. Moreover human agents do not know *a priori* by virtue of competence the necessary truth of many propositions. We have to learn a lot of essential properties of objects. By *essential property* of an object I mean here a property that it *really* possesses in any possible

[1] The notion of minimal rationality was first discussed by Cherniak (1986).

[2] These principles were first stated in my paper "Attitudes, tentatives et actions" (Vanderveken 2008a).

[3] See Bratman (1987), Davidson (1980), Goldman (1970), Searle (1982).

[4] I define a model-theoretical semantics for my object-language in my next book *Truth, Thought and Action*.

circumstance. Each human agent has the essential property to have certain parents. But some do not know their parents. Others are wrong about their identity; in that case they have necessary false beliefs. However when agents are inconsistent, they remain paraconsistent: as the Greek philosophers pointed out, they never believe nor desire everything.

According to standard logic of attitudes (Hintikka 1971), relations of psychological compatibility with the truth of beliefs and the realization of desires are modal relations of accessibility between agents and moments, on one hand, and possible circumstances, on the other hand. Possible circumstances are compatible with the truth of agents' beliefs at each moment of time. To each agent *a* and moment *m* there corresponds in each model a unique set *Belief*(*a,m*) of possible circumstances that are compatible with the truth of all beliefs of that agent at that moment. On Hintikka's view, an agent *believes a proposition at a moment* when that proposition is true in all possible circumstances that are compatible with what that agent then believes. Given such a formal approach, human agents are *logically omniscient*. They believe all necessarily true propositions and their beliefs are closed under logical implication. Moreover, human agents are either *perfectly rational* or *totally irrational*. They are perfectly rational when at least one possible circumstance is compatible with what they believe. Otherwise, they are totally irrational. Whoever believes a necessary falsehood believes all propositions according to the standard approach. But this conclusion is clearly false.

One could introduce in logic so-called *impossible circumstances* where necessarily false propositions would be true. But this move is very *ad hoc* and neither necessary nor sufficient. In my approach, all *circumstances* remain *possible*. So objects keep their essential properties (each of us keeps his real parents) and necessarily false propositions remain false in all circumstances. In order to account for human inconsistency, we have to consider *subjective* in addition to *objective possibilities*. Many subjective possibilities are not objective. So we need a non classical propositional logic. My logic is *predicative* in the general sense that it takes into account acts of predication that agents make in expressing and understanding propositions.[5]

In my view, each proposition has a finite *structure of constituents*. It predicates *attributes* (properties or relations) of *objects subsumed under concepts*. We understand a proposition when we understand which attributes objects of reference must possess in a possible circumstance in order that this proposition be true in that circumstance. As Frege (1977) pointed out, we always *refer to* objects by subsuming them under senses. We cannot directly have in mind *individual objects* like material bodies and persons. We rather have in mind *concepts* of individuals and we *indirectly* refer to them and predicate attributes of them through these concepts. So our attitudes are directed towards *individuals under a concept* (called an *individual concept*) rather than towards pure individuals. By recognizing the indispensable role of concepts in reference and predication, predicative logic accounts for attitudes directed towards

[5] See my papers "Propositional Identity, Truth According to Predication and Strong Implication" (Vanderveken 2005a) and "Aspects cognitifs en logique intensionnelle et théorie de la vérité" (Vanderveken 2009a).

inexistent and even impossible objects. It also explains why attitudes and intentional actions directed towards an individual under a concept are often not directed towards the same individual under other concepts. Jocasta, the queen of Thebes, is Oedipus' mother. In marrying Jocasta, Oedipus has then married his own mother. However he believed at the time of his wedding that he had another mother. So he did not then intend to marry his mother.

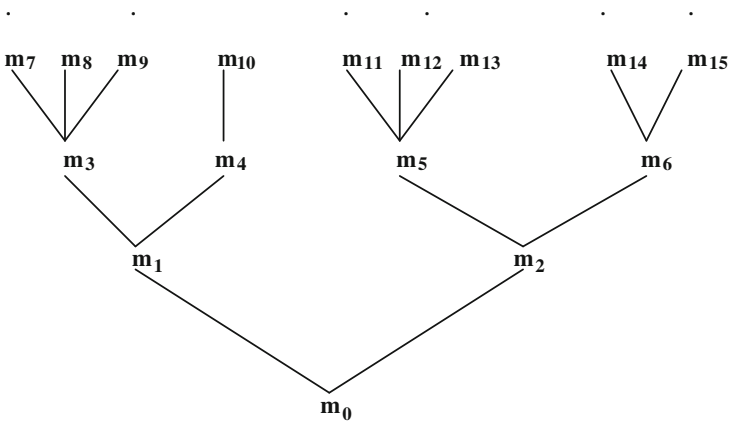
The logic of attitudes needs more than an analysis of the structure of constituents of propositions; it requires a *better explication of their truth conditions*. Because we ignore *real denotations* of most attributes and concepts in many circumstances, we understand most propositions without knowing in which possible circumstances they are true. One can refer to a friend's wife without knowing who she is. However we can always in principle think of persons who could be his wife. In my view most possible uses and interpretations of a natural language, let us say for short, most models for that language, consider a lot of *possible denotation assignments to attributes and concepts* in addition to the standard *real denotation assignment* of classical logic which associates with each propositional constituent its actual denotation in every possible circumstance. All possible denotation assignments to attributes and concepts of each model are functions of the same type; they associate with each individual concept a unique individual or no individual at all and with each attribute of degree n a sequence of n individual concepts in every possible circumstance. According to the real denotation assignment of each model, my friend's wife is the woman with whom he is really married according to that model when there is such a person. According to other possible denotation assignments, his wife is another person or even he is not married. In spite of their differences, all possible denotation assignments respect by definition real *meaning postulates* that speakers have internalized in learning their language. According to any, a wife is a married woman. We ignore the real denotation of most concepts and attributes in many circumstances. We can only think of denotations that they could have. When we express concepts and attributes only some possible denotation assignments to them *are then compatible with* the truth of our beliefs. Suppose that according to you my friend's wife is young. In that case, possible denotation assignments according to which she is old are then incompatible with your beliefs. Possible denotation assignments rather than possible circumstances are compatible with the beliefs of agents. So my logic accounts for subjective possibilities.

In my approach, the truth definition is relative to both possible circumstances and denotation assignments. An elementary proposition predicating an extensional property of an individual object under a concept is true in a circumstance according to a denotation assignment in a model when according to that assignment the objec which falls under that concept has that property in that circumstance. Otherwise, it is false in that circumstance according to that assignment. In understanding propositions we in general do not know whether they are true or false. We just know that their truth in a circumstance is compatible with certain possible denotation assignments to their concepts and attributes, and incompatible with all others. Most propositions have then *a lot of possible truth conditions*. Of course, any proposition that is *true in a circumstance according to a model* has to be *true in that circumstance according to the real denotation assignment* of that model. So among all possible truth conditions

of a proposition, its *real Carnapian truth conditions* correspond to the set of possible circumstances where it is true according to the real denotation assignment.

In my view, propositions are *identical* when they make the same predications and they are true in the same circumstances according to the same possible denotation assignments. Such a finer criterion of propositional identity explains why many strictly equivalent propositions have a different cognitive or volitive value. Propositions whose expression requires different predications have a different structure of constituents. So are necessarily equivalent propositions that Rome is a capital and that Rome is a capital and not an erythrocyte. One can express one without expressing the other. My identity criterion also distinguishes propositions that we do not understand to be true in the same circumstances: these are not true according to the same denotation assignments to their constituents. Few necessarily true propositions are *obvious* (or *pure*) *tautologies* that we know *a priori*. In order to be *necessarily true* a proposition has to be true in every possible circumstance according to the real denotation assignment. In order to be *obviously tautological*, a proposition has moreover to be true in every circumstance according to every possible denotation assignment to its constituent senses.[6] Unlike the proposition that Oedipus' mother is a woman, the necessarily true proposition that she is Jocasta is not an obvious tautology. It is false according to possible denotation assignments. We now can explicate subjective and objective possibilities. A proposition is *subjectively possible* when it is true in a possible circumstance according to a possible denotation assignment. In order to be *objectively possible* it has to be true in a circumstance according to the real denotation assignment. Few subjective possibilities are objective.

The logic of action requires a ramified conception of time compatible with indeterminism. Attitudes and actions of human agents are not determined. When they do or think something, they could have done or thought something else. In branching time, a *moment* is a complete possible state of the actual world at a certain instant and the *temporal relation* of *anteriority* between moments is partial rather than linear. There is a single causal route to the past. However, there are multiple future routes. Consequently, the set of moments of time is a *tree-like frame* of the following form:



[6] Obvious tautologies are called pure tautologies in most of my previous papers.

A maximal chain *h* of moments of time is called a *history*. It represents a *possible course of history of our world*. Some histories have a first and a last moment. According to these histories the world has a beginning and an end. As Belnap et al. (2001) pointed out, each *possible circumstance* is a pair of a moment *m* and of a history *h* to which that moment belongs. Thanks to histories temporal logic can analyze important modal notions like settled truth and historic necessity. Certain propositions are true at a moment according to all histories. Their truth is then *settled at that moment* no matter how the world continues. So are past propositions and propositions attributing propositional attitudes to agents. Whoever desires something at a moment desires that thing at that moment no matter what happens later. Contrary to the past, the future is open. The world can continue in various ways after indeterminist moments. Thus the truth of future propositions is not settled at such moments. It depends on which historical continuation of that moment is under consideration. When there are different possible historic continuations of a moment, its actual future continuation is not then determined. However, as Occam[7] pointed out, if the world continues after a moment, it will continue in a unique way. The actual historic continuation of each non final moment will be unique even if it is still undetermined at that very moment. Indeterminism cannot prevent that uniqueness.

Human agents who persist in an indeterminist world, have expectations and make plans. According to phenomenology and philosophy of mind, human agents who are directed by virtue of their intentionality towards things and facts of the world, are intrinsically oriented at each moment of their active life towards the real continuation of the world. We all ignore how the world will continue but we are intrinsically oriented at each moment towards the real continuation of that moment. So we always distinguish conceptually that real continuation from other possible continuations whenever we act or think in the world. Whoever attempts at a moment to achieve a future objective, intends to achieve that objective in the real continuation of that moment. Whoever foresees or wishes to have a future grandchild, foresees or wishes that grandchild in the real future. So in my approach both the moment and the historic continuation of the moment are to be considered in order to evaluate our actions and attitudes oriented towards the future. Consequently[8] our elementary illocutions and propositional attitudes at each moment have or will have a certain satisfaction value even if that satisfaction value is still then undetermined when they have a future propositional content. In order to keep a present promise and execute a present intention to give things later, an agent must give later these things in the real continuation of the world. Other possible historic continuations do not matter.[9]

According to my temporal logic every moment *m* has then a *proper history* h_m in each model. Whenever a moment *m* is the final moment of a history *h*, that history *h* is its proper history h_m. All moments that belong to the proper history of an indeterminist moment have of course the same proper history in each

[7] See Prior (1967).

[8] See my paper "Towards a Formal Pragmatics of Discourse" (Vanderveken 2013).

[9] Belnap N., M. Perloff and Ming Xu who reject the idea that each moment of utterance has a proper history have to strongly complicate the theory of satisfaction. See Belnap et al. (2001, 151).

model. A proposition is *true at a moment m* according to a denotation assignment in a model when it is true at moment m in the history h_m of that moment according to that assignment.

Two moments of time m and m' are *coinstantaneous* when they belong to the same instant. Coinstantaneous moments are on the same horizontal line in each tree-like frame. One can analyze *historic necessity* by quantifying over coinstantaneous moments. The proposition that *P is then necessary* (in symbols $\Box P$) is true at a moment according to a model when P is true at all coinstantaneous moments according to all histories in that model. The notion of historic necessity is stronger than that of settled truth. The represented fact is then not only established but inevitable. According to traditional philosophy there are no inevitable actions and intentions. Moreover the possible causes and effects so to speak of actions of any agent at a moment are limited to those which are possible outcomes of the way the world has been up to that moment. As Belnap and Perloff (1992) pointed out, in order to explicate *historical relevance* we must consider coinstantaneous moments having the same past. Such moments are called *alternative moments*. Thus m_1 and m_2 are alternative moments in the last figure. Logical or *universal necessity* is stronger than historic necessity. The proposition that *P is universally necessary* (in symbols: $\blacksquare P$) is true in a circumstance according to a model when P is true in all possible circumstances in that model. In that case the fact represented is always inevitable. A proposition P is *obviously tautological* according to a model when it is true in every possible circumstance according to any possible denotation assignment. The notion of *obvious tautologyhood* is the strongest modal notion. The represented fact is then analytically inevitable subjectively as well as objectively.

2 My New Approach in the Logic of Propositional Attitudes[10]

As I said earlier, propositional attitudes of human agents are about objects that they represent under concepts. Each agent has consciously or potentially[11] in mind a certain set of attributes and concepts at each moment. That set of propositional constituents is of course empty when the agent does not exist. In my view, no agent can have a propositional attitude without having in mind all attributes and concepts of its content. Otherwise, he or she would be unable to determine under which conditions that attitude is satisfied. In order to desire to be bishop one must understand characteristic features determined by meaning of the property of being bishop.

Secondly, possible denotation assignments to propositional constituents rather than possible circumstances are compatible with the satisfaction of agents' attitudes.

[10] See my three papers "A General Logic of Propositional Attitudes" (Vanderveken 2008b), "Beliefs, Desires and Minimal Rationality" (Vanderveken 2009b), and "On the Imperfect but Minimal Rationality of Human Agents" (Vanderveken 2012).

[11] We have unconsciously in mind at each conscious moment of our existence a lot of concepts and attributes that we could in principle express at that moment given our language.

So there corresponds to each agent *a* and moment *m* in each model a unique set *Belief*(*a,m*) of possible denotation assignments to attributes and concepts that are compatible with the truth of beliefs of that agent at that moment. When the agent *a* has no attribute in mind at the moment *m*, *Belief*(*a,m*) is the entire set *Val* of all possible denotation assignments to senses. In that case, that agent has then no attitudes. Otherwise, *Belief*(*a,m*) is always a *non empty proper* subset of *Val*. For whoever has in mind senses respects meaning postulates governing them in his possible use and interpretation of language. So there always are possible denotation assignments to these senses compatible with what that agent then believes. In my view, an agent *a believes a proposition* at a moment *m* when he or she has then in mind all its concepts and attributes and that proposition is true at that moment according to all possible denotation assignments of *Belief (a,m*) compatible with the truth of his or her beliefs at that moment. Our present beliefs directed at the future (previsions, expectations) will become true if things will be as we now believe in *the actual future continuation* of the present moment.

Similarly, to each agent *a* and moment *m* there corresponds in each model a unique non empty set *Desire*(*a,m*) of possible denotation assignments to attributes and concepts that are compatible with the realization of all desires of that agent at that moment. There is however an important difference between desire and belief. Agents can believe, but they cannot desire, that objects have properties or entertain relations without believing that they could be otherwise. For any desire contains a *preference*. Whoever desires something distinguishes two different ways in which represented objects could be in the actual world. In the preferred ways, objects are in the world as the agent desires, in the other ways, they are not. The agent's desire is realized in the first case, it is unrealized in the second case. Thus in order that an agent *a desires the fact represented by a proposition P* at a moment *m*, it is not enough that he or she has then in mind all attributes and concepts of *P* and that the proposition *P* is true at that moment according to all denotation assignments of *Desire(a,m)* compatible with the realization of his or her desire at that moment. That proposition must moreover be false in at least one circumstance according to that agent. Otherwise that agent would not then prefer the existence of the represented fact.

My explication of belief and desire is compatible with philosophy of mind. It accounts for conscious and unconscious attitudes. Whoever has a conscious belief or desire has consciously in mind all attributes and concepts of its propositional content. Whoever has an unconscious belief or desire has unconsciously in mind some of its attributes and concepts. He or she could then express these senses thanks to his or her language. My approach also accounts for the fact that *human agents are neither logically omniscient nor perfectly rational*. We do not have in mind all expressible concepts and attributes.[12] So we *ignore* tautological as well as necessary truths. Our knowledge is limited: we ignore how objects are in a lot of circumstances espe-

[12] We ignore the meaning of certain words. Moreover the languages that we speak have limited expressive capacities. We regularly enrich our languages in order to express new concepts and attributes that we discover.

cially in future circumstances. Many assignments associating different denotations to attributes in these circumstances are then compatible with our beliefs. We have moreover false beliefs and unsatisfied desires. So the real denotation assignment is often incompatible with the satisfaction of our beliefs and desires. Possible denotation assignments compatible with our beliefs and desires can even violate essential properties of objects. In that case we have necessarily false beliefs and insatisfiable desires. My analysis explains why we are sometimes *inconsistent*.

Predicative logic also explicates why propositions true in the same circumstances can have a different cognitive or volitive value. Some have different structures of constituents. So are logically equivalent propositions that mothers are women and that mothers are not ordinals. Their expression requires different *acts of predication*. Others are not true according to the same possible denotation assignments. So are necessarily true propositions that whales are whales and that whales are mammals. We do not understand them as being true in the same conditions. Thus we can assert or believe necessary truths without asserting or believing others. Among all necessary truths, few are *obvious tautologies* like the proposition that whales are whales. We believed in the past that whales were fishes.

However in my approach, *human agents always remain minimally rational*: they *cannot be totally irrational*. First of all, agents cannot believe or desire everything since in every model some possible denotation assignments are compatible with the satisfaction of their beliefs and desires. Moreover, whoever possesses certain beliefs and desires is *eo ipso* committed to possessing others. Indeed all possible denotation assignments compatible with our beliefs and desires respect meaning postulates. Human agents are therefore *minimally logically omniscient*: we cannot have in mind an obviously tautological proposition without knowing for certain that it is necessarily true. Represented objects could not be otherwise according to us. Similarly, obvious contradictions (negations of obvious tautologies) are false in every possible circumstance according to any agent. We can neither believe nor desire obvious contradictions. Some hope that arithmetic is complete (a necessarily false proposition if Gödel's proof is right). But agents could never believe or desire both the completeness and the incompleteness of arithmetic (an obvious contradiction). Moreover we cannot desire the existence of facts represented by obvious tautologies. In order to desire facts we must believe that these facts could not occur. One can desire to drink; one can also desire not to drink. But no one could desire to drink or not drink.

2.1 Analysis of Psychological Modes and Possession Conditions of Attitudes

Descartes in his treatise on ***Les passions de l'âme*** (Descartes 1953) analyzed a large number of propositional attitudes. Contemporary logic and analytic philosophy only consider a few paradigmatic attitudes such as belief, knowledge, desire

and intention. Could we use Cartesian analysis to develop a larger theory of all propositional attitudes? Searle in Chap. 1 of ***Intentionality*** criticized Descartes who tends to reduce all such attitudes to beliefs and desires. Indeed many different kinds of attitudes e.g. fear, regret and sadness reduce to the same sums of beliefs and desires. Moreover, our intentions are much more than a desire to do an action with a belief that we are able to do it. Clearly all cognitive attitudes (e.g. conviction, faith, confidence, knowledge, certainty, presumption, pride, arrogance, surprise, amazement, stupefaction, prevision, anticipation and expectation) are beliefs and all volitive attitudes (e.g. wish, will, intention, ambition, project, hope, aspiration, satisfaction, pleasure, enjoyment, delight, gladness, joy, elation, amusement, fear, regret, sadness, sorrow, grief, remorse and terror) are desires. But psychological modes divide into other components than the *basic categories of cognition* and *volition*. Let me now present these new components.

Many complex psychological modes have a *proper way* of believing or desiring, proper *conditions on their propositional content* or proper *preparatory conditions*. First of all, we feel our beliefs and desires in a lot of ways. Many modes require a special *cognitive or volitive way* of believing or desiring. Thus, *knowledge is* a belief based on strong evidence that gives confidence and guarantees truth. Whoever has an *intention* feels such a strong desire that he or she is disposed to *act* in order to realize that desire. From a logical point of view, a *cognitive or volitive way* is a function f_{ϖ} which restricts basic psychological categories. Like illocutionary forces, modes also have *propositional content* and *preparatory conditions*. Like *predictions, previsions and anticipations* are directed towards the future. *Intentions* are desires to carry out a present or future action. From a logical point of view, a *condition on the propositional content* is a function f_{θ} that associates which each agent and moment a set of propositions. The propositional content conditions of *predictions* and *previsions* associate with each agent and moment the set of propositions which are future with respect to that moment. Moreover any agent of a propositional attitude or of an elementary illocution *presupposes* certain propositions. Certain of these presuppositions are propositional presuppositions that depend on their propositional content. Whoever refers to the king of Belgium presupposes that there is one and only one king of Belgium. All illocutions and attitudes with the same propositional content have the same propositional presuppositions. Other presuppositions depend on the psychological mode and illocutionary force. They are determined by so called preparatory conditions. Thus *promises* and *intentions* have the preparatory condition that the agent is then able to do the action represented by their propositional content. Whoever promises and intends to do something presupposes that he or she can do it. His or her attitude and illocution would be *defective* if that proposition were then false. In the illocutionary case the speaker who presupposes can lie in order to mislead the hearer. In the psychological case however the agent cannot lie to him or herself. Whoever has an attitude both believes and presupposes that its preparatory conditions are fulfilled. A preparatory condition is a function f_{Σ} associating with each agent, moment and propositional content a set of propositions that the agent would presuppose and believe if he had then an attitude with that preparatory condition and propositional content. The sets of cognitive and volitive ways, of propositional

content and of preparatory conditions are *Boolean algebras*. They contain a *neutral* element and they are closed under the operations of *union* and *intersection*.

On the basis of my analysis, one can formally distinguish different kinds of attitudes like fear, regret and sadness which apparently reduce to the same sums of beliefs and desires. Identical psychological modes have the same components. Possession conditions of propositional attitudes are entirely determined by components of their mode and their propositional content. By definition, an agent *a possesses a cognitive* (*or volitive*) *attitude of the form M(P) at a moment m* when he or she has then a *belief* (or *desire*) with the propositional content *P*, he or she feels then that belief (or *desire*) that *P in the cognitive (or volitive) way* ϖ_{M} proper to psychological mode *M*, the *proposition P then satisfies propositional content conditions* $\theta_{\mathrm{M}}(a,m)$ and finally that *agent then presupposes and believes all* propositions $\Sigma_{\mathrm{M}}(a,m,P)$ determined by *preparatory conditions* of mode *M* with respect to the content *P*. Thus an agent *intends that P* at a moment when proposition *P* then represents a present or future action of that agent, he or she desires so much that action that he or she is committed to carrying it out and moreover that agent then presupposes and believes to be able to carry it out. Whoever has an intention intends to act sooner or later. Sometimes the agent intends to act at the very moment of the intention. He or she has then an *intention to act in the present* (what Searle (1982) calls an *intention in action*). Sometimes the agent has a *prior intention*: he or she intends to act at a posterior moment. Most agents who have an intention at a moment have previously formed that intention or they form it at that very moment. They have committed themselves to doing the intended action. Whoever has the *intention to act in the present* forms his or her intention at the very moment of that intention. So agents of intentions in action perform the very act of forming these intentions.

An *attitude strongly commits an agent to another at a moment* when he or she could not then have that attitude without having the second. Thus whoever believes that it will rain tomorrow then foresees rain tomorrow. Some attitudes strongly commit the agent to another at particular moments. Whoever believes now that it will rain tomorrow foresees rain tomorrow. The day after tomorrow the same belief won't be a prevision. It will be a belief about the past. An attitude *contains another* when it strongly commits any agent to that other attitude at any moment. There are *strong and weak psychological commitments* just as there are strong and weak illocutionary commitments (see Searle and Vanderveken 1985). One must distinguish between the overt possession of an attitude and a simple psychological commitment to that attitude. Whoever believes that every man is mortal is weakly committed to believing that Nebuchadnezzar is mortal, even if he has not Nebuchadnezzar's concept in mind and if he or she does not then possess the second belief. No one could simultaneously believe the first universal proposition and the negation of the second.

Psychological modes are not a simple sequence of a basic psychological category, a cognitive or volitive way, a propositional content condition and a preparatory condition. For their components are not independent. Certain components *determine* others of the same or of another kind. Thus the volitive way of the mode of *intention* determines the propositional content condition that it represents a present or future action of the agent and the preparatory condition that that agent is then able to

carry out that action. The two primitive modes of *belief* and *desire* are the simplest cognitive and volitive modes. They have no special cognitive or volitive way, no special propositional content or preparatory condition. Complex modes are obtained by adding to primitive modes special cognitive or volitive ways, propositional content conditions or preparatory conditions. Thus the mode of *prevision* $\mathrm{M}_{foresee}$ is obtained by adding to the mode of belief the propositional content condition θ_{future} that associates with each agent and moment the set of propositions that are future with respect to that moment. $\mathrm{M}_{foresee} = [\theta_{future}]Belief$. The mode of *hope* is obtained from that of *desire* by adding the special cognitive way that the agent is then uncertain as regards the existence and the inexistence of the represented fact and the preparatory condition that that fact is then possible. The mode of *satisfaction* is obtained from that of *desire* by adding the *preparatory condition* that the desired fact exists. The mode of *pleasure* has, in addition, the *volitive way* that the satisfaction of the desire puts the agent in a state of pleasure and the preparatory condition that it is good for the agent. Because all operations on modes add new components, they generate stronger modes. Attitudes $M(P)$ with a complex mode M *contains* attitudes $M'(P)$ whose mode M' have less components. A lexical analysis of terms for attitudes based on my componential analysis explains which name stronger psychological modes. I have drawn semantic tableaux in order to show comparative strength between modes.[13]

Notice that contrary to truth functions, modal and temporal propositions as well as propositions attributing attitudes to agents contain more elementary propositions than their arguments. They serve indeed to predicate new *modal, temporal, epistemic and volitive attributes* to objects of reference. In thinking that God cannot make mistakes we predicate of Him the modal property of infallibility. In thinking that God created the world we predicate of Him the past property of having created the world. In thinking that the pope believes that God exists, we predicate of God the epistemic property of being existent according to the pope. Whoever wishes that God forgives him predicates of God the property that he would prefer His pardon.

2.2 Analysis of Satisfaction Conditions of Propositional Attitudes

The general notion of *satisfaction condition* in logic is based on that of *correspondence*. Agents of propositional attitudes and elementary illocutionary acts are directed towards facts of the world represented by their propositional content. Most often they establish a correspondence or fit between their ideas and things in the case of attitudes and between their words and things in the case of illocutions. Their attitudes and illocutions have for that reason *satisfaction conditions*. In order that the propositional attitude or elementary illocution of an agent at a moment is *satisfied*, there must first of all be a correspondence between that agent's ideas or words and represented things in the world in the history of that moment. The propositional

[13] See the tableaux at the end of my paper "Formal Semantics for propositional attitudes" (Vanderveken 2011).

content must represent a fact that exist at that moment or will exist in the world in its real historic continuation.

As I already said, agents live in an indeterminist world. Their future is open. At each moment where they think and act, they ignore how the world will continue. However, their attitudes and actions are always directed by virtue of their intentionality toward the real historic continuation. Whenever parents refer to their next child they refer to their next child in the real future. In order that a present attitude or illocution directed at the future be satisfied, it is not enough that things will be at a posterior moment as the agent now represents them. They must be so later in the real future. So the *satisfaction* of propositional attitudes and elementary illocutionary acts of an agent at a moment requires the *truth at that very moment* of their propositional content. The notion of *satisfaction* is a generalization of the notion of *actual truth*[14] that takes into account the direction of fit of attitudes and illocutions. The relation of fit or of correspondence is symmetrical: if a proposition fits the world then the world fits that proposition. However there is more to the notion of satisfaction than to that of actual truth because one must consider the *direction of fit* from which the correspondence must be achieved between the mind and the world in the analysis of satisfaction of attitudes, just as one must consider the direction of fit from which the correspondence must be achieved between language and the world in the analysis of satisfaction of illocutions.

There are four possible directions of fit between ideas and things, just as there are four possible directions of fit between words and things. Just as *assertive illocutions* have the *language-to-world direction of fit*, *cognitive attitudes* have the *mind-to-world direction of fit*. They are *satisfied* when their propositional content fits the world. In that case the agent's ideas[15] must correspond to things as they are then in the world. On the contrary, volitive attitudes have the opposite *world-to-mind direction of fit* just as *commissive* and *directive illocutions* have the opposite *world-to-language* direction of fit. They are satisfied only if the world fits their propositional content. In that case represented things in the world must correspond to the agent's ideas.

Each direction of fit between mind and the world determines which side is at fault in case of dissatisfaction. In the cognitive and assertive cases, the agent is at fault in the case of dissatisfaction. So when the agent realizes that there is no correspondence between his or her ideas and represented, that agent immediately changes his or her beliefs and is ready to revise his or her assertions. This is why the truth and falsehood predicates apply so well to satisfied *cognitive attitudes* and *assertive illocutions*. A *belief* and an *assertion* at a moment are *satisfied* when they are *then true* and *unsatisfied* when they are then false. *Satisfaction* and *dissatisfaction* amount to *actual truth* and *actual falsehood* in the case of *cognitive attitudes* and *assertive illocutions*. However, the truth predicates do not apply at all to *volitive*

[14] We need an actuality connective for a right account of satisfaction conditions. A proposition of the form *Actually P* is *true* in a circumstance *m*/*h* when it is true at the moment *m* according to its history h_m.

[15] In the case of illocutions the agent's ideas are the ideas that he or she expresses by his or her words.

attitudes whose direction of fit goes from things to mind just as they do not apply to *commissive* and *directive illocutions* whose direction of fit goes from things to language. For the world and not the agent is at fault in the case of dissatisfaction of volitive attitudes and commissive and directive illocutions. In that case, the agent can keep his desires and remains dissatisfied. He can repeat his previous commissive and directive illocutions. So we use other predicates of satisfaction. *Satisfied wishes* and *desires* are *realized*; *satisfied hopes* and *aspirations* are *fulfilled*, and *satisfied intentions*, *projects* and *plans* are *executed*. *Satisfied promises* and *vows* are *kept*, *satisfied orders* and *commands* are *obeyed*, *satisfied requests* are granted, etc.

Most often, agents having a volitive attitude desire the existence of the fact represented by the propositional content *no matter how that fact turns to be existent in the world*. So most volitive attitudes that agents have at a moment are *satisfied* when their content is then true, no matter for which reason. Things are then such as the agent desires them to be, no matter what is the cause of their existence. The only exceptions to this rule are *volitive attitudes* like *will*, *intentions*, *projects*, *plans* and *ambitions* whose proper volitive way requires that things fit the agent's ideas because he or she wants them in that way. Like commissive and directive illocutionary acts (orders, commands, pledges and promises), such volitive attitudes have *self-referential satisfaction conditions*. Their satisfaction requires more than the existence of the fact represented by their propositional content. It requires that that fact turns to be existent in order to satisfy the agent's attitude. In order to execute a prior intention and to keep a previous promise, an agent must do more than carry out later the intended and promised action in the real future; he or she must carry out that action because of that previous intention and promise. If the agent does not act for that reason, (that agent has forgotten his or her previous intention and promise or he or she does not act freely), that agent does not then execute the prior intention (or keep then the previous promise). Like illocutionary logic, the logic of attitudes can explain such a self-referential satisfaction by relying on *intentional causation*. The attitude and illocution of the agent are then a *practical reason* why the represented fact turns to be existent.[16]

As Searle pointed out in ***Intentionality***, certain *volitive* modes like *joy*, *gladness*, *pride*, *pleasure*, *regret*, *sadness*, *sorrow*, and *shame* have like expressive illocutions the *empty direction of fit*. Agents who have such attitudes do not want to establish a correspondence between their ideas and represented things in the world. They just take for granted either correspondence or lack of correspondence between their ideas and things. In the case of *joy*, *gladness*, *pride* and *pleasure*, the agent believes that the desired fact exists. In the case of *regret*, *sadness*, *sorrow* and *shame*, he or she believes on the contrary that it does not exist. The first attitudes have the special preparatory condition Σ_{Truth} that their propositional content is then true. The second attitudes which contain a desire of the inexistence of the fact represented by their propositional content have the opposite preparatory condition $\Sigma_{\text{Falsehood}}$ that their content is then false. Volitive attitudes with such special *preparatory condition* have the *empty direction of fit* because their agent could not intend to establish a

[16] My logic of action has a reason connective to express intentional causation.

correspondence. This is why they do not have *satisfaction conditions*. Instead of being satisfied or dissatisfied, they are just *appropriate* or *inappropriate*. They are inappropriate when their preparatory condition of actual truth or falsehood is wrong or when their proper psychological mode does not suit the fact represented by their content. No agent should be ashamed of an action that he has not made or that is exemplary and good for all.

As Candida de Sousa Melo (2002) pointed out, declaratory acts of thought have the *double direction of fit between mind and things*. In making verbal and mental *declarations*, the speaker changes represented things of the world just by way of thinking or saying that he is changing them. Whoever gives by declaration a name to a new thing acts in such a way that that thing has then that name. In such a case, an act of the mind brings about the represented fact. Because *attitudes are states and not mental actions*, they could not have the double direction of fit.

3 Intentionality in the Logic of Action

The aim of this section is to give a formal account of the intentionality of human agents and to explicate the nature of their intentional and basic individual actions. By way of performing individual actions at a moment, agents bring about facts in the world. They make then true propositions representing these facts. Whenever they act intentionally they are moreover directed towards facts that they attempt to bring about in the world. The logical constant of Belnap's logic is the connective *stit* ("sees to it that") which serves to express propositions according to which an agent *a does P* (in symbols [*a stit P*]). Because attempts are constitutive of intentional actions, my logic contains a new logical constant *Tries* of *attempt* in order to express propositions of the form [*a Tries P*] according to which the agent *a tries* to do *P*. Notice that propositions that attribute actions and attempts to agents predicate new *agentive attributes*. In thinking that a police officer is making the hostages free, we attribute to that officer the agentive property of freeing hostages. Prefixes like "en" serve to compose agentive predicates in English. To enable is to make able and to enrich is to make rich. Similarly in thinking that a person is making an attempt to be elected, we attribute to that person the agentive property of being a candidate for an election. We need an analysis of agentive attributes in the logic of action.

I will first make basic considerations about individual actions and attempts. When an agent performs the individual action of bringing about a fact at a moment, he or she performs that action at that moment no matter how the world continues. The truth (or falsehood) of propositions of the form [*a stit P*] is then well established at each moment. Agents can repeat individual actions of the same type at different successive moments in a possible course of the world. They can request and request again. Agents also perform individual actions of the same type at alternative moments. When a player is in a checkmate position at a moment in a chess game, that player is a loser at all alternative moments where he or she makes a move in that game. Moments of time are logically related by virtue of actions of agents at these moments. From

a logical point of view, to each agent a and moment m there always corresponds in each model the set $Action^a_m$ of coinstantaneous moments m' which are *compatible with all the actions* that agent a performs at the moment m. They are all, as Chellas (1992) would say, "under the control of—or responsive to the actions of" of that agent at that moment. When an agent a does not act at all at a moment m, all moments coinstantaneous with m are compatible with his or her actions at that moment. However, when he or she does P at moment m, the proposition P is true at all moments $m' \in Action^a_m$ according to any history. In my view, the relation of compatibility with actions is reflexive, symmetric and transitive. So when a moment is compatible with all actions of an agent at another moment, that agent performs exactly the same actions at these moments. Of course because of indeterminism, the same actions of that agent can have different physical effects (that are not actions) in the world at different moments which are compatible with what he or she does at that moment. Every agent persists in the world. What an agent does at a moment depends on how the world has been up to that moment. This is why the relation of compatibility with actions satisfies the so called *historical relevance condition*. Only alternative moments having the same past as m can belong to $Action^a_m$. Moreover, as Belnap said, *the world goes on*. Agents act in the same world. So at least one moment m' belongs to both sets $Action^a_m$ and $Action^b_m$ for any two agents a and b.

Thanks to the new compatibility function, logic can start to analyze individual actions. The proposition that *P is true given what agent a does* (in symbols ΔaP) is true in a circumstance m/h according to a model when proposition P is true at all moments $m' \in Action^a_m$ compatible with the actions of agent a at m according to all histories h'. Chellas (1992) tends to identify the very notion of action with the normal modal operation corresponding to Δ. However each proposition of the form ΔaP is true whenever proposition P is historically necessary. But no agent could bring about an inevitable fact. Inevitable facts exist no matter what we do. So as Belnap pointed out, proposition [*a stit P*] is stronger than ΔaP; it implies that P could be false.

In their logic of agency, Belnap and Perloff (1992) use the logic of branching time and von Neumann's theory of games. Agents make free choices in time. The notion of acting or choosing at a moment m is thought of as constraining the course of events to lie within some particular subset of the possible histories available at that moment. Belnap and Perloff first studied actions that are guaranteed by a past choice of the agent. They made a theory of the so called *achievement stit*. However often agents succeed in doing things that they had no prior intention to do. They spontaneously attempt to do them. Moreover sometimes they do things that they would not have wanted to do. Belnap et al. (2001) came to study later actions directed at the future that are guaranteed by a present choice of the agent. They made a theory of the *deliberative stit*.

My logic of agency is more general; it deals with individual actions made at the very moment of the agent's choice, no matter whether these actions are oriented towards the present or the future. Attempts require a present choice of their agent. Every intentional action contains a present intention in action, few execute a prior intention. Most successful spontaneous attempts to move parts of one's body cause the movement at the very moment of the attempt. We emit sounds when we try to

emit them in the contexts of oral utterances. Belnap's analysis of action in terms of ramified time has the merits of taking very seriously into consideration the temporal and causative order of the world. I follow his approach under many aspects. But I want to take into account the proper *intentionality* of agents that Belnap ignores. For that reason, agents carry out too many actions in his logic. Suppose that a proposition strictly implies another which is not then necessary. According to his analysis, an agent cannot make the first proposition true without also making the second true, even when the second proposition represents a fact that no agent could bring about or even try to bring about at that moment. Thus whoever repeats an action sees to it that he does that action and has done it in Belnap's logic.

Let me now repeat the principles of my approach.[17] In my logic, intentional actions are primary as in philosophy. Some of our actions are involuntary. But any agent who performs unintentionally an action could in principle have attempted that action, and that unintentional action is generated by his or her intentional actions. The basic actions of agents are their primary attempts that are means to make all their other attempts; they generate all their other actions whether intentional or not. Agents know and intend few effects of their basic actions. A lot of their actions are then involuntary. However, not all unintended effects of intentional actions are involuntary actions, but only those that are historically contingent and that the agent could have attempted. In moving we inevitably agitate subatomic particles. Sometimes we are mistaken and we fail. Such events which happen in our life do not constitute actions. Indeed we could not move without agitating particles and our mistakes and failures could not be intentional.[18]

My logic of action contains a theory of *attempt* and of *action generation*. In my analysis, *attempts* are *actions* that agents *make* (rather than attitudes that they have). Attempts are *actions* of a very special kind: personal, conscious, intentional, free and successful. Only the agent can make his or her individual attempts. No one else can make them. Thus when two agents succeed in doing the same action (e.g. to drink) they do it thanks to different personal attempts (in that case different body movements). Attempts are intrinsically intentional actions. There are no involuntary attempts. When an agent makes an attempt, he or she makes that attempt in order to do something else. Attempts are *means* to achieve *ends*. Whoever attempts to make an attempt succeeds in making that attempt, but he or she can fail to reach his or her objective. An attempt is essentially a mental act. Whoever attempts to raise the arm can fail because of an external force. But he or she has anyway mentally made that attempt in forming consciously his or her present intention to raise the arm. Among intentional actions, attempts have then particular success conditions. It is enough to try to make an attempt in order to make it *eo ipso*. Direct attempts by an agent to move parts of one's body are real basic actions. When an agent *forms* the present intention to make a direct movement, an attempt is caused by the very formation of

[17] See Vanderveken (2005b, 2008a).

[18] Goldman (1970) notices that certain act properties like misspeaking, miscalculating, miscounting seem to preclude intentionality. Such properties are not really act properties. We "suffer" mistakes. We do not make them.

that present intention, no matter whether he or she is in a standard condition or not (Goldman 1970, 65). In case the agent of an attempt fails to reach his or her objective, his or her attempt is then *unsatisfied*. In order to make a satisfied attempt, one must make a good attempt in a right circumstance. Whoever attempts to invite a certain person fails when he uses a wrong name or speaks to the wrong person. When agents attempt to perform illocutions, the satisfaction conditions of their attempts are in that case the so called success conditions of their attempted illocutions. Agents often have an *experience of their attempt* when they fail (Searle 1982). Such an experience presents or represents the satisfaction conditions of that attempt.[19]

My logic of action accounts for the minimal rationality of agents who are neither perfectly rational nor entirely irrational. Minimal rationality in action is related to the ways in which agents determine satisfaction conditions of their attempts. We can intend and attempt to do impossible actions. However there are impossible actions that we can neither intend nor attempt to do. My approach represents adequately satisfaction conditions of intentions and attempts. To each agent a and moment m there correspond two non empty sets: $Intention^a_m$ and $Attempt^a_m$ in every model. $Intention^a_m$ contains all denotation assignments to senses which are compatible with the execution of all intentions of that agent at that moment; $Attempt^a_m$ is the set of all pairs of denotation assignments to senses which are respectively compatible with the realisation and the satisfaction of his or her attempts at that moment. Attempts like intentions have the *world-to-mind direction of fit*. Only realized attempts can be satisfied. Consequently all denotation assignments to senses compatible with the satisfaction of attempts of an agent at a moment are compatible with their realisation at that very moment: $id_2Attempt^a_m \subseteq id_1Attempt^a_m$. (In my symbolism, for any Cartesian product $X \times Y$, $id_1(X \times Y) = X$ and $id_2(X \times Y) = Y$.) In my view, any agent of an attempt forms the present intention to make then his or her attempt. Because that attempt has an objective, he or she also forms the intention to achieve that objective at that moment or later in the real historic continuation. Formally, $id_2Attempt^a_m \subseteq Intention^a_m \subseteq Desire^a_m$. Moreover, because attempts are actions, each agent makes the same attempts at all moments compatible with his or her actions. Thus $id_1Attempt^a_m = id_1Attempt^a_{m'}$ when $m' \in Action^a_m$. And similarly for $id_2Attempt^a_m$. There is no action without attempt. Consequently $Action^a_m$ is the set of all coinstantaneous moments with m when all possible denotation assignments to senses belong to the set $id_1Attempt^a_m$. Different agents can attempt to achieve the same objective (to push the car). However no agent can make the attempt of another agent. Each agent does something irreducibly personal when he or she makes an attempt. That agent forms then his or her present intention of making that attempt. No one else can form that intention. So $id_1Attempt^a_m \neq id_1Attempt^b_m$ when $a \neq b$.

In order that an agent a try to make a proposition true at a moment m according to a model, it is necessary but not sufficient that that proposition is true at that moment m in history h_m according to all denotation assignments of $id_2Attempt^a_m$. Agents never

[19] Direct attempts of moving one's body contain a *presentation* and attempts of making an illocution a *representation* of their satisfaction conditions. Searle (1982) does not really consider the fact that attempts are themselves actions.

intend and attempt to make true propositions that are obviously tautological. They only intend and attempt to carry out present or future actions. Because attempts are intentional actions, they have the same propositional content conditions as intentions. The set of propositions representing the objectives of an agent a at moment m according to a model is included in the set $\theta_{\text{intention}}(a,m)$. In my approach, a proposition of the form [a *Tries* P] according to which an agent a attempts to bring about the fact represented by P is true in a circumstance m/h when firstly, that agent forms the conscious intention that P at that moment (and consequently at all alternative moments $m' \in Action^a_m$ compatible with his or her actions then) and secondly, the proposition P is true at moment m in the history h_m according to all denotation assignments of $id_2\ Attempt^a_m$ compatible with the satisfaction of his or her attempts at that moment.[20] Every attempt at a moment is then well established and his or her agent believes then to be able to reach its objective. Moreover the agent attempts to make all his or her attempts. Because agents are minimally rational they never attempt nor intend to do things that they know to be necessary or impossible. They also never attempt nor intend to do something in the past.

As one can expect, the set $Goal^a_m$ of all propositions representing objectives aimed by an agent a at the moment m is empty in a model when the agent a is unconscious or does not act at all at that moment according to that model. Attempts are indeed conscious actions. Each agent attempts to achieve the same goals at every moment compatible with his or her actions. $Goal^a_m = Goal^a_{m'}$ when $m' \in Action^a_m$. Any non empty set $Goal^a_m$ is moreover finite. Human agents can only make a finite number of predications and consequently they can only form finitely many intentions and make finitely many attempts. The objectives of an agent at a moment are either present or future. Whenever the agent attempts to achieve a present objective, he or she either succeeds or fails at the very moment of his or her attempt. That attempt is then either satisfied or unsatisfied. In case the agent has a future objective (he requests an answer), he or she forms a prior intention and it is not then determined whether the attempt will or not be satisfied. All depends on what will happen in the real future. No agent can succeed in achieving at a moment a future objective. Remember that future propositions are false at each final moment. The most that agents can do is to act in such a way that the future fact that they intend to bring about will sooner or later come into existence, if the world continues, no matter how. Whoever hurts mortally an adversary will provoke his or her death if the world goes on, given actual physical laws. Agents can make more or less good contributions to the achievement of their future objectives. By making a move in a chess game one can put the adversary in an inevitable losing position. In that case it is then settled that one will be the winner if the game is pursued.

At each moment of action, any agent makes a few very basic attempts whose objectives are present and entirely personal to him or her. So are our primary attempts to move directly parts of our body or to make purely mental acts of conceptual thought

[20] Bratman (1987) criticizes the principle that whoever attempts to do something intends to do it. But his counter-examples do not work or they concern attempts that are not momentary but last during an interval of time.

like a judgement in soliloquy. Let $BasicGoal^a_m$ be the set of all propositions of $Goal^a_m$ that represent the very basic attempts of the agent a at a moment m. Two non empty sets $BasicGoal^a_m$ and $BasicGoals^b_m$ are disjoint when their agents a and b are different. All our attempts at a moment are related by the relation of being means to achieve our goals at that very moment. Our few basic attempts at each moment are therefore primary in a double sense. First, they are not effects of other attempts. Second, they all together cause all our other attempts because they are made for that purpose.

As philosophers pointed out, human agents act intentionally and especially they form their attitudes and they make their attempts for certain *practical reasons*, because they have then certain beliefs, desires, intentions and objectives and also because of simultaneous and sometimes anterior actions, illocutions and attitudes. They have cognitive attitudes and believe propositions for *theoretical reasons*. Their actions and attitudes are often motivated by several reasons. They keep a previous promise not only because they have put themselves under the obligation to keep it but also in order to please the hearer and get a favour. They suppose or believe that a proposition is true because of their previous experience and of background or social knowledge. However they would not make their attempts and they would not have their cognitive attitudes if they had no practical and no theoretical reasons at all. Indeed their practical and theoretical reasons are the very *intentional causes* of their attempts and cognitive attitudes. For each agent a and moment m let $Reasons^a_m$ be the set of propositions representing all theoretical and practical reasons of that agent at that moment according to a model. Each agent has of course the same practical reasons to make his or her attempts at each moment compatible with his or her actions. Among the practical causes of any attempt there is the agent's conscious intention to make that attempt. There are also his or her basic attempts that cause at that moment all his or her other attempts. $BasicGoal^a_m \subset Reasons^a_m$.

As I said earlier, agents succeed in performing attempted actions, when they make good attempts in right circumstances. It remains to explicate fully the notion of *success.* As Davidson and Searle pointed out, in order that an agent *succeeds* in bringing about a fact, it is not enough that he or she tries and that the fact occurs. The attempted fact must be *caused* by his or her own attempt. Otherwise, the agent failed. Sometimes the agent's attempt is *the* cause why the attempted fact occurs. Often however there is *causal overdetermination*. This happens when several agents bring about the attempted fact, or when the agent brings about the fact because of several simultaneous attempts. In such cases the agent's attempt under consideration is just *a practical reason* among others why the attempted fact occurred. One cannot then assert counterfactually that if the agent had not made that particular attempt, the fact would not have occurred. Like illocutionary logic, the logic of action must consider agents' *practical reasons* in order to explicate *intentional causation* and satisfaction-conditions of attempts. Attempts like commissive and directive illocutions have the things-to-mind direction of fit. In order that an agent succeeds in achieving an objective, his or her attempt must be a practical reason of his or her success.

However the logic of action has to consider other causes than agents' practical reasons in order to explicate success. As Goldman pointed out, certain attempt tokens

generate other action tokens in various ways. In addition to intentional causes, *natural causes*, *conventions* as well as *particular facts existing in the situation* in which they act, enable agents to achieve their objectives. An agent who moves the hand touches material objects that constitute an obstacle. He or she would not touch such objects in another situation where they would not be present. Whoever flips the switch succeeds in turning on the light when the electric lighting system works. The agent's attempt physically causes light because of other facts (there is electric transmission) existing in the situation where he or she acts. His or her attempt and these other facts all together *physically cause* the intended effect given laws of nature. As I said earlier, the logic of ramified time takes into consideration the causative and temporal order of the world. Thus all pertinent natural laws thanks to which agent acts at a moment hold by hypothesis at all coinstantaneous moments.

Agents also succeed in achieving their objectives because of established conventions. According to *conventions*, certain action tokens *count as* constituting others in certain situations. By raising the hand, one succeeds in voting for a proposition in a meeting where participants follow such a convention. Agents of course know their practical reasons as well as the conventions that they follow in making their attempts. But they do not know all relevant physical laws; they often ignore particular physical causes and facts of the situation and sometimes even established conventions which enabled them to achieve their objectives. Logic has to take into consideration such *other reasons* for success.

Now when an agent succeeds in achieving an objective *because of* certain conventions, natural causes and particular facts, these conventions are established and these causes and particular facts exist by hypothesis in what philosophers call the *situation* where the agent acts. These established conventions, causes and particular facts exist even at all moments compatible with the actions of that agent at that moment, no matter whether or not he or she is aware of them. At each moment m where an agent a acts, propositions that represent all *other reasons* for his or her success at that moment are then true at all alternative moments $m' \in Action^a_m$. Moreover, given preceding considerations, when the achievement of the agent's objective is due to natural causes, conventions and particular facts existing in the situation of his or her attempt, it is then historically necessary that these natural causes, conventions and facts cause the existence of attempted facts at all moments which are coinstantaneous with the moment of action of that agent.

One can express and interpret in my logic propositions according to which the agent's attempt to bring about certain facts is *a practical reason* for the truth of certain propositions.

In my symbolism such propositions are of the form $[\rho[\Delta aQ][aTriesP]])$: they mean that Q is true given what agent a does *because* he or she tries P. In my approach, a proposition of the form $[\rho[\Delta aQ][aTriesP]]$ is true in a circumstance m/h when firstly, both propositions $[\Delta aQ]$ and $[aTriesP]]$ are true in that circumstance, secondly, Q is not then historically necessary (it is false in at least one circumstance whose moment is coinstantaneous with m) and thirdly, for some proposition R true in all circumstances m'/h' compatible with a's actions at m, it is then historically necessary that both $[aTriesP]$ and R implies $[\Delta aQ]$, that is to say when both $[aTriesP]$

and R are true in a circumstance coinstantaneous with m, so is $[\Delta aQ]$. The new proposition R represents all particular facts, natural causes and established conventions existing in the situation of the agent that are other reasons why the agent brings about the fact represented by Q.

On the basis of preceding considerations, I define as follows *success* and *failure*. In my object-language the two formulas [*a succeeds P*] and [*aFailsP*]) which mean respectively that agent *a succeeds* and that agent *a fails* in doing *P* are abbreviations.

$$[a\ succeeds\ P] =_{\text{def}} ([aTriesP]) \wedge [\Delta aP] \wedge (\neg\Box P) \wedge [\rho[\Delta aP][aTriesP]].$$
$$\text{And } [aFailsP] =_{\text{def}} ([aTriesP]) \wedge ([\neg\Delta aP] \vee (\Box P) \vee (\neg\rho[\Delta aP][aTriesP])).$$

No agent can succeed or fail in doing something unless he or she makes an attempt. So we do not succeed in performing our unintentional actions. We just perform them.

How could we now explicate the general notion of an individual action (whether intentional or not)? Given the principles of my approach, I advocate the following definition: an agent *a acts so as to bring about P* at a moment when firstly, the proposition *P* is true given what he or she then does, secondly, *P* is then historically contingent, thirdly, the agent *a* could then try *P* and fourthly, he or she brings about *P because of* a present attempt at that moment. In other words, a proposition of the form [*a stit P*] is true in a circumstance *m/h* according to a model when the proposition *P* is false in at least one coinstantaneous circumstance, but *P* is true at all alternative moments $m' \in Action^a_m$ according to all histories, there is at least one coinstantaneous moment m' where $P \in Goal^a_{m'}$ and the proposition $[\rho[\Delta aP][aTriesQ]]$ is true in circumstance *m/h* for at least one proposition $Q \in Goal^a_m$. In my conception of action, there is no action without a simultaneous attempt of the agent. What agents do at each moment has to be caused by their intentional actions at that very moment. Thus dead agents do not act anymore even if their actions can still have effects after their death. According to philosophers, certain basic intentional actions generate all our other actions. So are in my approach the sums of our basic attempts at each moment. Because they are attempts, we succeed in performing our basic actions whenever we attempt them. Using the counterfactual conditional, one can say that if an agent had not made his or her basic attempts at that moment he or she would not then have done anything. In my logic the conjunction $(P_1 \,\&\, \ldots \,\&\, P_n)$ of all $P_k \in BasicGoal^a_m$ represents the basic action of agent a at moment m. Whenever $P_k \in BasicGoal^a_m$, there is no other $Q \in Goal^a_m$ such that the attempt that Q is a practical reason for P_k.

4 Fundamental Valid Laws

In his paper on "Desire, Deliberation and Action", Searle (2005) expressed skepticism about the logic of practical reason. Of course, because of their proper things-to-mind direction of fit, desire and other volitive modes have properties

like indetachability and unavoidable inconsistency which complexify their formal explication. Agents of deliberations are not forced to commit themselves at the end to given actions. They can revoke their prior intentions and not attempt to execute them. When they make an attempt, they can fail. Such properties do not at all prevent the development of the logic of practical reason. My logic of attitudes and action explicates them entirely. Searle is moreover forced to admit the existence of internalized logical relations of psychological and illocutionary commitment and incompatibility because of the very principles of his philosophy of mind. According to him, any agent of an attitude and of an intentional action has in mind the satisfaction conditions of that attitude and the success conditions of that action. Consequently, agents cannot have certain attitudes without having others and they cannot make certain actions without making others and having certain attitudes.

In my approach, there is a proper logic (a recursive theory of possession and satisfaction) for volitive attitudes just as there is a proper logic (a recursive theory of success and satisfaction) for attempts and commissive and directive illocutions[21] which all have the things-to-mind direction of fit. All kinds of attitudes and actions are logically related by virtue of their felicity conditions. My logic explicates formally specific properties of attitudes and illocutions with the things-to-mind direction of fit. It also revises the current logical conception of rationality in explicating why agents are imperfectly rational, why they are sometimes inconsistent and why they do not make all valid theoretical inferences. It moreover explicates why they are not logically omniscient, why they can ignore obvious tautologies as well as necessary propositions. In my logic, the set of beliefs is neither closed under tautological nor under strict implications. Indeed many propositions tautologically and strictly imply other propositions with new concepts or attributes that agents might not have in mind. Agents also ignore how propositions are related by strict implication. My logic also explicates why agents always remain minimally rational in thinking and in acting and solves psychological and illocutionary paradoxes like the paradoxes of the liar and of the sophist. They are indeed minimally consistent; they cannot believe that an obvious tautology is false. So agents know that certain facts could not occur without others.

My predicative logic explicates a new *strong* propositional *implication* that is much finer than Lewis' strict implication and important for the analysis of strong and weak psychological commitments. Formally, a proposition *strongly implies* another when firstly whoever expresses that proposition can express the other and secondly, it cannot be true in a circumstance according to a possible denotation assignment unless the other proposition is also true in that circumstance according to that assignment. Strong implication is finite, tautological, paraconsistent, decidable and *a priori known*. Whoever believes a proposition also believes any proposition that it strongly implies. He or she knows that it could not be true otherwise. Strong implication is also partially compatible with desires, intentions and attempts. Whenever a proposition strongly implies another, whoever attempts or intends to make it true also attempts

[21] See the two volumes of my book *Meaning and Speech Acts* (Vanderveken 1990, 1991).

or intends to make true the other in case that other proposition represents then a possible goal of that agent.

Of course, the logic of desire and intention is very different from that of belief. Agents can both intend to do something and believe that their intended action will have a certain effect without *eo ipso* desiring and intending to produce that effect. One can reject an offer and believe that one will irritate the agent of that offer without desiring and intending to provoke such an attitude. There is sometimes a conflict between the intentions and beliefs of an agent at a moment. Certain possible denotation assignments to senses compatible with the execution of the agent's intentions at a moment are not compatible with the truth of his or her beliefs at that moment. For unwanted effects of the intended action do not occur according to the first assignments. Agents know that some of their beliefs could be false. This can even occur when they believe that it is settled or even inevitable that their action will have a certain unwanted consequence. Bratman and Searle have given a lot of convincing examples. A prior intention to do something P and a belief that it is then necessary that if P then Q do not commit the agent to a prior intention to do Q. We know that we can wrongly believe that certain facts are inevitable. We would then be happier if such facts would not occur. So Kant's principle: "Whoever intends to achieve an end thereby will the necessary means or effects that he or she knows to be part of the achievement of that end" does not apply to *prior intentions*.

However because agents are rational they have to minimally coordinate their cognitive and volitive states in trying to act in the world. So a restricted form of Kant's principle "Any agent who wills the end is committed to willing the necessary means" applies to attempts which are intentions in action. In case the agent of an attempt knows that in order to succeed he or she has to do something else, that agent will try to do that other thing. In other words, whoever attempts to achieve an end attempts to use means that he or she knows to be necessary. Such a restricted Kantian principle is valid in my logic of action. When P and $Q \in \theta_{\text{Present Intention}}(a, m)$ and the agent a knows at moment m that $\Box(P \Rightarrow Q)$, if $P \in Goal^a_m$ then $Q \in Goal^a_m$. Let me give an example. Any agent knows that in order to invite someone one has to make a request. Thus whoever tries to make an invitation *eo ipso* tries to make a request. His or her attempted request then *constitutes* his or her attempted invitation.

As Goldman pointed out, certain action tokens *generate* others causally, conventionally, simply and by extension. My logic of action explicates the various forms of action generation. It can also characterize how illocutions which are the primary units of meaning and communication in the use of language relate to other speech-acts (acts of utterance, propositional acts, attempts at performing illocutions, and perlocutionary acts). Attempts at performing illocutions are new fundamental speech-acts in my taxonomy. They are constitutive of *meaning*. Speakers attempt to publicly perform illocutions by emitting signs. It remains to explicate *how* and *under what conditions they succeed* and how successful illocutions *generate* others (invitations contain requests) and have perlocutionary effects (the hearer is sometimes influenced). At the basis of communication, agents attempt to move parts of their body and this *generates* in the sense of Goldman in various ways their speech-acts. *Generation* in communication is first physically causal (we orally utter sentences in producing

sounds), next conventional (sentence-meaning serves to determine attempted illocutions). Generation is sometimes *simple* (speakers succeed to perform attempted illocutions when they use appropriate words in the right contexts) or *by extension* (they sometimes indirectly perform non-literal illocutions). In order to explicate different kinds of speech-act generation, I intend to integrate illocutionary logic within the logic of action.

References

Belnap, N., and M. Perloff. 1992. The way of the agent. *Studia Logica* 51(3/4): 463–484.
Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future. Agents and choices in our indeterminist world*. Oxford: Oxford University Press.
Bratman, M. 1987. *Intentions, plans and practical reason*. Cambridge, MA: Harvard University Press.
Brentano, F. 1993. *Psychology from an empirical standpoint*. London: Routledge and Kegan Paul.
Chellas, B.F. 1992. Time and modality in the logic of agency. *Studia Logica* 51(3/4): 485–518.
Cherniak, C. 1986. *Minimal rationality*. Cambridge, MA: M.I.T. Press.
Davidson, D. 1980. *Essays on action and events*. Oxford: Oxford University Press.
Descartes, R. 1953. Les passions de l'âme. In *Œuvres et lettres*. La Pléiade, Paris: Gallimard.
de Sousa Melo, C. 2002. Possible directions of fit between mind, language and the world. In *Essays in speech act theory*, ed. D. Vanderveken, and S. Kubo. Amsterdam: Benjamins.
Frege, G. 1977. "Thoughts", "negation" and "compound thoughts. In *Logical investigations*. New Haven, CT: Yale University Press.
Goldman, A.I. 1970. *A theory of human action*. Princeton, NJ: Princeton University Press.
Hintikka, J. 1971. Semantics for propositional attitudes. In *Reference and modality*, ed. L. Linsky. Oxford: Oxford University Press.
Prior, A.N. 1967. *Past, present and future*. Oxford: Clarendon Press.
Searle, J.R. 1982. *Intentionality*. Cambridge: Cambridge University Press.
Searle, J.R. 2005. Desire, deliberation and action. In D. Vanderveken (2005c), 49–78.
Searle, J.R., and D. Vanderveken. 1985. *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.
Vanderveken, D. 1990. *Meaning and speech acts*, vol. 1. Cambridge: Cambridge University Press.
Vanderveken, D. 1991. *Meaning and speech acts*, vol. 2. Cambridge: Cambridge University Press.
Vanderveken, D. 2005a. Propositional identity, truth according to predication and strong implication. In D. Vanderveken (2005c), 185–216.
Vanderveken, D. 2005b. Attempt, success and action generation: a logical study of intentional action. In D. Vanderveken (2005c), 316–342.
Vanderveken, D. (ed.). 2005c. *Logic, thought and action*. Dordrecht: Springer.
Vanderveken, D. 2008a. Attitudes, tentatives et actions. In *Actions, Rationalité & Décision*, ed. D. Vanderveken, and D. Fisette, 39–73. London: College Publications.
Vanderveken, D. 2008b. A general logic of propositional attitudes. In *Dialogues, logics and other strange things, Volume 7 of series tributes*, ed. C. Dégrémont, et al., 449–483. London: College Publications.
Vanderveken, D. 2009a. Aspects cognitifs en logique intensionnelle et théorie de la vérité. *Dialogue* 48(1): 103–128.

Vanderveken, D. 2009b. Beliefs, desires and minimal rationality. In *Logic, ethics and all that Jazz. Essays in Honour of Jordan Howard Sobel of Uppsala Philosophical Studies*, ed. Johansson, Österberg and Sliwinski, vol. 57, 357–372. Uppsala: Uppsala University.

Vanderveken, D. 2011. Formal semantics for propositional attitudes. Special issue. *Science, Truth and Consistency of Manuscrito* 24(1): 323–364.

Vanderveken, D. 2012. On the imperfect but minimal rationality of Human Agents. In *Rationality and its limits*, ed. A. Guseynov, and V. Lektorsky, 136–159. Russian Institute of Philosophy Print: Proceedings of the Meeting in Moscow of the International Institute of Philosophy, Moscow.

Vanderveken, D. 2013. Towards a formal pragmatics of discourse. *International Review of Pragmatics* 5(1): 34–69.

Group Strategies and Independence

Ming Xu

Abstract We expand Belnap's general theory of strategies for individual agents to a theory of strategies for multiple agents and groups of agents, and propose a way of applying strategies to deal with future outcomes at the border of a strategy field. Based on this theory, we provide a preliminary analysis on distinguishability and independence, as a preparation for a general notion of dominance in the decision-theoretical approach to deontic logic.

Based on branching time and a theory of agents and choices, Belnap has developed a general theory of strategies in Belnap (1996b) and Belnap et al. (2001).[1] A simple form of this theory identifies a strategy for an agent with a partial function from moments to the choices available for the agent at those moments, which is found useful by different authors in conceptual analysis and technical development concerning "strategic acts".[2] Horty develops a simpler but similar theory of strategies in Horty (2001), and applies it to his study of "strategic acts" and "strategic oughts". The work presented here concerns both "strategic acts" and "strategic oughts", perhaps with an emphasis on the latter in the background. This paper is the first step of a project to connect Belnap's theory and the decision-theoretical approach to deontic

[1] I would like to give thanks to Nuel Belnap for his comments and encouragements, and to Yan Zhang for several discussions and for catching errors in early drafts of this paper. For the theory of branching time, see Prior (1967) and Thomason (1970, 1984); and for the theory of agents and choices, see, e.g., Belnap (1991, 1996a) and Belnap et al. (2001)

[2] For example, Belnap shows that whenever a doing takes place, there exists a strategy of refraining from that doing (see chapter 13 of Belnap et al. 2001), Müller applies this theory of strategies to deal with continuous actions in Müller (2005), and Broersen and his colleagues apply this theory in their work to extend alternating-time temporal logic in Broersen et al. (2006).

M. Xu (✉)
Department of Philosophy, Wuhan University, Wuhan 430072, People's Republic of China
e-mail: mingxu01@hotmail.com

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*,
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_16,

logic developed in Horty (2001), in a setting involving multiple agents and groups of agents.[3]

In the decision-theoretical approach to deontic logic, what an agent ought to do is taken to be determined by the result of an evaluation of what she can do against background situations or conditions in the form of a partition. If one action is taken to be better than another under each such background condition, it is then inferred to be better than the other unconditionally, or to "dominate" the other, as is often described.[4] The background conditions, however, are required to be independent of the actions being evaluated. This independence requirement is essential, without which the inference is evidently flawed.[5] In Horty (2001), Horty takes the notion of independence here to be causal independence, and presents the background conditions, when evaluating actions of an agent or a group, as what other agents may do at the same time. In other words, actions at the same time by different agents are taken to be independent of each other.

This approach to deontic logic is continued in Kooi and Tamminga (2008) and later in Tamminga (2013), with a notion of relative dominance and a closer relation to game theory. It has so far been limited, nevertheless, to either single-step group actions, or strategies of a single agent while other agents are assumed absent. The reason for such limitation is, I think, that it is not clear how to deal with the independence requirement in a setting involving actions at different moments by different agents, as Horty seems to suggest in Horty (2001). This paper examines strategies for different groups and some relations between them, based on which we develop a notion of independence of strategies for different groups, by way of an analysis of distinguishability and inactivity. We provide some results concerning independence (in Sect. 9), including a characterization of independence in terms of a set-theoretical relation between groups of agents (Theorems 9.6 and 9.10).

Section 1 briefly presents the background theories of branching time, agents and choices, and Sects. 2 and 3 present our notions of outcomes and fields with outcomes at their "borders". In Sects. 4–6, we discuss group strategies with respect to future outcomes and various related notions. Finally we present a preliminary analysis of the notions of distinguishability, inactivity and independence in Sects. 7–9, as a preparation for a future work on dominance.

[3] Belnap's theory of strategies may also be applied to other approaches to deontic logic. For example, Belnap (1996b) shows the connection between his theory of strategies and Thomason's theory of *ought kinematics* (Thomason 1984).

[4] The kind of inference applied here is sometime called the "sure-thing principle" (see Savage 1954).

[5] See discussions in, e.g., Thomason and Horty (1996) and Horty (2001).

1 Stit Frames

In this section, we briefly present the basic notions in the semantic theory for stit,[6] which constitute a general background for our theory concerning what agents may do relative to future outcomes. Let us start with the branching time theory developed by A. Prior and R. Thomason.

A *tree-like frame* is a pair $\langle T, <\rangle$, in which T is a nonempty set, and $<$ is a strict partial ordering on T (i.e., an irreflexive and transitive relation on T) satisfying the following conditions:

NBB : for all $x, y, z \in T$, if $y < x$ and $z < x$, either $y \leqslant z$ or $z \leqslant y$;
HC : for all $x, y \in T$, $z \leqslant x$ and $z \leqslant y$ for some $z \in T$;

where $x \leqslant y$ iff $x < y$ or $x = y$. The label NBB is for "no backward branching", and HC for "historical connection".[7]

We call members of T *moments* or *points*, for which we use m, u, x, y, z etc., and call each maximal $<$-chain of moments in T a *history* (in $\langle T, <\rangle$). We use h, h' etc. for histories and H, H' etc. for sets of them. In particular, we use H_T for the set of all histories (in $\langle T, <\rangle$). Furthermore, we will apply the following notations and expressions, where $M \subseteq T$, c is a chain (of points), and x a point, in T:

- $H_{\langle M\rangle} = \{h \in H_T : h \cap M \neq \varnothing\}$, histories *passing through* M;
- $H_{[c]} = \{h \in H_T : c \subseteq h\}$, histories *passing completely through* c;
- $H_x = \{h \in H_T : x \in h\}$, histories *passing through* x.

It is plain that $H_x = H_{[\{x\}]} = H_{\langle\{x\}\rangle}$. Sets of histories are *compatible* if their intersection is nonempty. It is easy to see that for all x and y, if neither $x \leqslant y$ nor $y \leqslant x$, then no subset of H_x is compatible with any subset of H_y.

A sequence $\langle T, <, \mathit{Agent}, \mathit{Choice}\rangle$ is a *stit frame* if $\langle T, <\rangle$ is a tree-like frame, *Agent* is a nonempty set of "agents", and *Choice* is a function that assigns to each $\alpha \in \mathit{Agent}$ and each $m \in T$ a partition Choice^m_α of H_m satisfying the following conditions:

NC : for each $K \in \mathit{Choice}^m_\alpha$, each $h \in K$ and each $x \in h$, if $m < x$ then $H_x \subseteq K$;

---

[6] Stit, the acronym of "sees to it that", was taken to name a modal operator used in a rigorous philosophical theory of agency and action developed in a series of articles by Belnap, Perloff and their colleagues, which provides, among other things, formal semantics for sentences involving what agents do. The acronym was soon used to refer to the theory itself, and later to other theories as well that share similar principles, methods and logical tools. Stit theories were developed by a number of people in the late 1980s, and have now become a field to which many people have contributed their works. So I will just mention a few pieces of work, among many others, which basically started this field: Belnap and Perloff (1988), von Kutschera (1986) and Horty (1989). For detailed discussions in stit theories, see, e.g., Belnap et al. (2001).

[7] When $\langle T, <\rangle$ satisfies all conditions above for a tree-like frame except the condition HC, we may call it a *multi-tree-like frame*. For the purpose of this paper, we will focus on structures based on tree-like frames, but our discussions can easily be extended to similar structures based on multi-tree-like frames.

IA : for each function f that assigns to each $\beta \in Agent$ a member $f(\beta)$ of $Choice^m_\beta$, $\bigcap_{\beta \in Agent} f(\beta) \neq \varnothing$.

The label NC is for "no choice between undivided histories" and IA for "independence of agents". A function f is a *selection function at* m if $f(\beta) \in Choice^m_\beta$ for each $\beta \in Agent$. We will use $Select_m$ for the set of all selection functions at m. Thus IA above can be restated as that $\bigcap_{\beta \in Agent} f(\beta) \neq \varnothing$ for each $f \in Select_m$.

Let $\langle T, <, Agent, Choice\rangle$ be any stit frame. We call subsets of *Agent* groups (*of agents*), and use $\mathcal{E}, \mathcal{F}, \mathcal{G}$ etc. to range over them. For each $m \in T$ and each group $\mathcal{G}$, we use $Choice^m_\mathcal{G}$ for $\{\bigcap_{\alpha \in \mathcal{G}} f(\alpha) : f \in Select_m\}$ ($Choice^m_\varnothing = \{H_m\}$),[8] call its members *possible choices for* $\mathcal{G}$ *at* m, and use K, K' etc. to range over them. A *possible choice*, or simply a *choice*, is a possible choice for a group at a point. A group $\mathcal{G}$ (or an agent α) has *vacuous choice at* a point m if $Choice^m_\mathcal{G} = \{H_m\}$ ($Choice^m_\alpha = \{H_m\}$). Provided that $h \in H_m$, we use $Choice^m_\mathcal{G}(h)$ for the unique member of $Choice^m_\mathcal{G}$ to which h belongs. Finally, we let $\overline{\mathcal{G}} = Agent - \mathcal{G}$ for each group $\mathcal{G}$. It is easy to verify the following by applying NC:

Fact 1.1. For each $\mathcal{G}$ and all $x, y \in h$ such that $y < x$, $H_x \subseteq Choice^y_\mathcal{G}(h)$.

By this fact, we introduce the following notation: provided that $y < x$, we use $Choice^y_\mathcal{G}(H_x)$ for the unique $K \in Choice^y_\mathcal{G}$ such that $H_x \subseteq K$.

2 Outcomes

In many cases, it is more convenient to use outcomes rather than histories for conceptual analysis or technical development. In this section we discuss a notion of outcome, derived from Xu (1997). We first present the notion in its original form and then convert it into a notion of history-outcome. Throughout this section and the next, we fix a tree-like frame $\langle T, <\rangle$, with respect to which our discussions are to be understood.

Our notion of outcomes presupposes the following: For each $x \in T$ and each $X \subseteq T$, $x \leqslant X$ ($X \leqslant x$, $x < X$ or $X < x$) iff $x \leqslant y$ ($y \leqslant x$, $x < y$, $y < x$) for every $y \in X$, and in such a case, we say that x is a *lower-bound* (*upper-bound*, *proper lower-bound*, *proper upper-bound*) of X. A subset X of T is *forward* (*backward*) *closed* if for all $x, y \in T$, $x < y$ ($y < x$) and $x \in X$ only if $y \in X$. A *past* in $\langle T, <\rangle$ is a nonempty and properly upper-bounded set p of moments that is backward closed. For each past p, $H_{\lfloor p\rfloor}$ is by definition $\{h \in H_T : p \subseteq h\}$, the set of all histories passing completely through p. For each properly upper-bounded nonempty chain c of moments,we use p_c for the smallest past including c, i.e., $p_c = \{x \in T :$

[8] I do not mean to take the empty set as a group or an agent in the literal sense. The only reason why we call it a group is for technical convenience. One could exclude it from groups and add extra conditions in our technical discussions.

$\exists y \in c(x \leqslant y)\}$. Thus for each $x \in T$ that is not maximal in T, $p_{\{x\}}$ is the past $\{y \in T : y \leqslant x\}$.

We want to add the notion of outcomes to the theories of strategies developed in Belnap (1996b) and Horty (2001) in order to extend their theories to deal with strategy-weighing relative to the values of future outcomes. An "outcome" can be reified either as a set of moments or as a set of histories, with a simple relation between them.

An *outcome* (in $\langle T, <\rangle$) is a nonempty and properly lower-bounded set O of moments that is forward closed and historically connected in O, i.e., for all $x, y \in O$, $z \leqslant x$ and $z \leqslant y$ for some $z \in O$. For each past p, an outcome *at* p is an outcome O such that p is the set of all its proper lower-bounds, i.e., $p = \{x \in T : x < O\}$. For each nonempty chain c of moments in T, an outcome *at* c is an outcome at p_c, and for each $x \in T$, an outcome *at* x is an outcome at $\{x\}$.

In Xu (1997), an outcome O is paired with a past p to form a "transition" $\langle p, O\rangle$, where $p < x$ for every $x \in O$, which is used to characterize a process or change from the state p right before the process to the outcome state O of the process. So an outcome O marks the temporal "location" of the completion of a process in such a way that all histories overlapping O are taken to be just those in which the process completes. For technical simplicity, we do not use the notion of transition explicitly in this paper. We apply its idea extensively, nevertheless. The following fact is easily verifiable.

Fact 2.1. Let p be any past, and let O be any outcome at p. Then for each history h, $h \cap O \neq \varnothing$ only if $h - p \subseteq O$.

Let p be any past, and let $\sim_p$ be a relation between histories in $H_{[p]}$ such that for all $h, h' \in H_{[p]}$, $h \sim_p h'$ iff $x \in h \cap h'$ for some $x > p$. It is easy to verify that $\sim_p$ is an equivalence relation. A *history-outcome at* p is an equivalence class modulo $\sim_p$. A *history-outcome at* a properly upper-bounded nonempty chain c (or at a non-maximal point m) is a history-outcome at the past p_c ($p_{\{m\}}$), and a *history-outcome* is a history-outcome at a past.

Proposition 2.2. For all history-outcomes H and H', either $H \subseteq H'$ or $H' \subseteq H$ or $H \cap H' = \varnothing$.

Proof. Let H and H' be outcomes at p and p' respectively, and suppose that $h_0 \in H \cap H'$ and $h' \in H' - H$. It then suffices to let $h \in H$ and show that $h \sim_{p'} h_0$, which implies that $h \in H'$. Since $h_0, h \in H$ and $h_0, h' \in H'$, there are x and y such that $p < x \in h \cap h_0$ and $p' < y \in h_0 \cap h'$. Because $h_0 \in H$ and $h' \notin H$, $h' \not\sim_p h_0$, i.e.,

$$p \not< z \text{ for each } z \in h_0 \cap h'. \tag{1}$$

Because $x, y \in h_0$, either $x < y$ or $y \leqslant x$. If $x < y$, then $p < y$ since $p < x$, and, since $y \in h_0 \cap h'$, $p \not< y$ by (1), a contradiction. It then follows that $y \leqslant x$, and then $p' < x$, and hence $h \sim_{p'} h_0$. ■

Now we have two kinds of outcomes, whose relation needs to be made clear. To help our discussion, let us refer to the kind of outcomes defined earlier as *moment-outcomes*.

Proposition 2.3. Let p be a past, and let f be a function on the set of all moment-outcomes at p such that for each such outcome O, $f(O) = \{h \in H_T : O \cap h \neq \varnothing\}$. Then f is a one-one correspondence between the set of all moment-outcomes at p and the set of all history-outcomes at p.

Proof. For all $h, h' \in f(O)$, there are $x \in O \cap h$ and $y \in O \cap h'$, and then by the condition of historical connection on O, there is a $z \in O$ such that $z \leqslant x$ and $z \leqslant y$, and hence $z \in h \cap h'$. Since $z \in O$, $p < z$, and hence $h \sim_p h'$. It follows that $f(O)$ is included in an equivalence class modulo $\sim_p$. To see that it is itself an equivalence class modulo $\sim_p$, it suffices to suppose that $h \sim_p h'$ with $h \in f(O)$, and show that $h' \in f(O)$. By definition, $m \in h \cap h'$ for an $m > p$. By Fact 2.1, $h - p \subseteq O$, and then, since $m > p$ and $m \in h$, $m \in O$, and hence $h' \in f(O)$. ■

Belnap and Horty use histories to define various notions in their study of strategies. It is then more convenient to use history-outcomes rather than moment-outcomes in our presentation to show a clear picture of the connection between our theory and theirs. By Proposition 2.3, the two kinds of outcomes are different notions of the same idea.[9] From now on, when we speak simply of outcomes, we mean history-outcomes.

For each past p, we use $Outcm_p$ for the set of all outcomes at p, and for each properly upper-bounded nonempty chain c, we use $Outcm_c$ for $Outcm_{p_c}$, and, finally, for each non-maximal point x, we use $Outcm_x$ for $Outcm_{\{x\}}$. It is easy to see that each history h in $H_{[p]}$ belongs to a unique outcome at p, and thus we use $Outcm_p(h)$ for that outcome. Similarly, for each history h passing completely through a properly upper-bounded nonempty chain c or through a non-maximal point x, we will use $Outcm_c(h)$ or $Outcm_x(h)$ for the outcome at c or x to which h belongs. It is routine to verify the following by applying relevant definitions.

Fact 2.4. Let h be any history, let $\{x\}, c \subseteq h$, both of which are properly upper-bounded, and let c be nonempty. Then $c < x$ only if $H_x \subseteq Outcm_c(h)$, and $c \not< x$ only if $Outcm_c(h) \subseteq Outcm_x(h)$.

Fact 2.5. Let $\langle T, <, Agent, Choice\rangle$ be a stit frame, let $\mathcal{G}$ be any group, and let $x \in h$, where x is not a maximal point. Then $Outcm_x(h) \subseteq Choice^x_{\mathcal{G}}(h)$.

The converse of Fact 2.5 does not in general hold: a possible choice for any group (including *Agent*) at a moment may consist of several outcomes at the moment. How many outcomes can there be at a past p without a maximum? The answer is that there

[9] There is nevertheless a shortcoming in a presentation using history-outcomes. Set-theoretically speaking, moment-outcomes at different moments are always different, while history-outcomes at different moments may turn out to be the same. For example, if c is a nonempty segment of a history in which no histories split at any point, then history-outcomes at points in c remain the same. For more discussions of the notion of moment-outcomes and its applications, see Xu (1997, 2010, 2012) and Brown (2008).

may still be more than a single outcome at p even though for all $h, h' \in H_{[p]}$, $h \in Choice^{x}_{Agent}(h')$ for every $x \in p$, i.e., h and h' are not distinguished by any possible choices at points in p. A stit frame $\langle T, <, Agent, Choice\rangle$ is *agency determinate* if for each past p and each history h passing completely through p, $\bigcap_{x\in p} Choice^{x}_{Agent}(h)$ is a single outcome at p. In certain applications, agency determination or similar conditions are proposed to make the semantic structures ideal in some sense. Since the purpose of the current study is to provide a general theory, we will not include this condition for our general framework.

3 Fields and Outcomes Bordering Fields

An *anti-chain* (*in* $\langle T, <\rangle$) is a nonempty subset i of T such that for all $x, y \in i$, neither $x < y$ nor $y < x$. Let i be any anti-chain in $\langle T, <\rangle$. i *intersects* a history h if $i \cap h \neq \varnothing$. If i intersects h, $i \cap h$ is clearly a singleton, and in such a case, we use $m_{i,h}$ for the unique member of $i \cap h$.

A *field* is a nonempty subset M of T. An anti-chain i *covers* a field M (i is a *cover of* M) if for each $x \in M$, $x \leqslant y$ for a $y \in i$ and i intersects every $h \in H_x$ and $x \leqslant m_{i,h}$.[10] M is *covered* if it is covered by an anti-chain, and is *properly covered* if it is covered by an anti-chain i such that $i \cap M = \varnothing$. A properly covered field has two roles in the current study. The first is to provide a background choice situation for our discussion of strategies, and the second is to constrain the so-called future outcomes that agents or groups may attain.

Covered fields may take various "shapes", and it is their "borders" in the future and the outcomes at the "borders" in which we are interested. A cover of a field guarantees the field to have a "border", and a proper cover even guarantees that there are outcomes everywhere along the "border". They are not accurate, nevertheless, in telling where exactly the "border" is, much less about the outcomes there; for they may contain points in the field as well as points far beyond the "border". We then have to find another way to talk about the outcomes at the "border" of a field.

Let M be any field. M is *inward closed* if for all $x, y, z \in T$ such that $x < y < z$, if $x, z \in M$ then $y \in M$. We use M^+ for the *inward closure* of M, i.e., $M^+ = \{x \in T : \exists y, z \in M(y \leqslant x \leqslant z)\}$. A history h *passes across* M if $\varnothing \neq M \cap h < x \in h$ for some x. It is obvious that h passes across M only if $h \in H_{\langle M\rangle}$, but the converse does not hold in general. An outcome H is an M-*bordering* outcome (or an outcome *bordering* M) if there is a history h passing across M^+ such that $H = Outcm_{M^+\cap h}(h)$.[11] For each field M, we will use $OutcmBdr_M$

[10] When assuming the Axiom of Choice, the clause "$x \leqslant y$ for a $y \in i$" is redundant.

[11] Let h pass across M, i.e., $M \cap h < x \in h$ for an x. If M is not inward closed, there may be a $y \in M$ such that $x < y \notin h$ and $H_y \subset H_x \subseteq Outcm_{M\cap h}(h)$. The outcome $Outcm_{M\cap h}(h)$ should not be taken to be bordering M, and to rule out such outcomes as M-bordering outcomes, we need to use M^+ instead of M in our definition. The definition of M-bordering outcomes in terms of moment-outcomes is simpler: O is an M-*bordering outcome* if $O \cap M = \varnothing$ and O is an outcome at a nonempty chain c in M^+ ($c \subseteq M^+$).

for the set of all M-bordering outcomes. It is easy to see that if a field M is inward closed, then for each outcome H, $H \in \mathit{OutcmBdr}_M$ iff $H = \mathit{Outcm}_{M\cap h}(h)$ for an h passing across M. Furthermore we have the following by definition and NBB:

Fact 3.1. Let h pass across M^+ and $h' \in \mathit{Outcm}_{M^+\cap h}(h)$. Then h' passes across M^+, $M^+ \cap h = M^+ \cap h'$ and $\mathit{Outcm}_{M^+\cap h}(h) = \mathit{Outcm}_{M^+\cap h'}(h')$. Consequently, for each outcome H, $H \in \mathit{OutcmBdr}_M$ iff for each $h \in H$, h passes across M^+ and $H = \mathit{Outcm}_{M^+\cap h}(h)$.

The next fact is a direct consequence of Facts 2.4, 2.5 and 3.1.

Fact 3.2. Let $\langle T, <, \mathit{Agent}, \mathit{Choice}\rangle$ be any stit frame, let $H \in \mathit{OutcmBdr}_M$ with M to be any field, and let $\mathcal{G}$ be any group. For each $x \in M$ and each $K \in \mathit{Choice}^x_{\mathcal{G}}$, either $H \subseteq K$ or $H \cap K = \varnothing$.

The following propositions show some facts concerning fields and outcomes bordering them. The first states that no outcome bordering a field is compatible with another such outcome.

Proposition 3.3. Let M be any field, and let $H, H' \in \mathit{OutcmBdr}_M$. Then $H \neq H'$ only if $H \cap H' = \varnothing$. Consequently, for all $U, U' \subseteq \mathit{OutcmBdr}_M$, $\bigcup U = \bigcup U'$ iff $U = U'$.

Proof. By definition, there are histories h and h' passing across M^+ such that $H = \mathit{Outcm}_c(h)$ and $H' = \mathit{Outcm}_{c'}(h')$ where $c = M^+ \cap h$ and $c' = M^+ \cap h'$. By Proposition 2.2, either $H \subseteq H'$ or $H' \subseteq H$ or $H \cap H' = \varnothing$. If $H \subseteq H'$, $h \in \mathit{Outcm}_{c'}(h')$, and then $H = H'$ by Fact 3.1. Similarly, $H' \subseteq H$ only if $H = H'$. Hence $H \neq H'$ only if $H \cap H' = \varnothing$. ■

Proposition 3.4. Let M be any properly covered field. Then,

(i) for each history h, $h \in H_{\langle M\rangle}$ iff h passes across M^+;
(ii) for each $h \in H_{\langle M\rangle}$, there is an $H \in \mathit{OutcmBdr}_M$ such that $h \in H$;
(iii) for each outcome H, $H \in \mathit{OutcmBdr}_M$ iff $H = \mathit{Outcm}_{M^+\cap h}(h)$ for an $h \in H_{\langle M\rangle}$;
(iv) $H_{\langle M\rangle} = \bigcup \mathit{OutcmBdr}_M$ and $\mathit{OutcmBdr}_M = \mathit{OutcmBdr}_{M^+}$.

Proof. (i) Let $h \in H_{\langle M\rangle}$, i.e., $h \in H_x$ for an $x \in M$. Assume that i properly covers M. Then i intersects h and $x < m_{i,h}$. If $m_{i,h} \leqslant z$ for a $z \in M^+$, $m_{i,h} \leqslant z'$ for a $z' \in M$, and then by definition, $z' < u$ for a $u \in i$, and hence $m_{i,h} < u$, contrary to our assumption that i is an anti-chain. It follows that $M^+ \cap h < m_{i,h}$, and thus h passes across M^+. (ii) For each $h \in H_{\langle M\rangle}$, h passes across M^+ by (i), and then $h \in \mathit{Outcm}_{M^+\cap h}(h) \in \mathit{OutcmBdr}_M$ by definition. (iii) follows from (i) by definition, and (iv) follows from (ii), (iii), and a simple fact that $M^+ = M^{++}$. ■

Proposition 3.4 (ii) and the following establish that for each field M, M is properly covered iff no matter which history we go along through M, we always go into an outcome at the "border" of M. From now on, we use "AC" to mark a proposition or a fact to indicate the dependence of our proof (or a routine proof) on the Axiom of Choice.

Proposition 3.5. (AC). Let M be any field such that for each $h \in H_{\langle M\rangle}$, there is an $H \in OutcmBdr_M$ such that $h \in H$. Then M is properly covered.

Proof. For each $H \in OutcmBdr_M$, we know that there is an anti-chain i_H such that $i_H \cap M = \varnothing$ and i_H intersects all and only $h \in H$. Letting i be the union of all i_H with $H \in OutcmBdr_M$, we know by Proposition 3.3 that i is an anti-chain. It is then routine to verify that i properly covers M. ■

Let M be any properly covered field. By Propositions 3.3–3.4, we know that each history $h \in H_{\langle M\rangle}$ is contained in a unique M-bordering outcome. Thus we will use, for each $h \in H_{\langle M\rangle}$, $OutcmBdr_M(h)$ for the unique M-bordering outcome to which h belongs.

For each point m, $OutcmBdr_{\{m\}}$ is obviously the set of all outcomes at m, i.e., $OutcmBdr_{\{m\}} = Outcm_m$. It is worth noting, however, that for a chain c of points, $OutcmBdr_c$ is not in general the same as $Outcm_c$, and that for a field M, $OutcmBdr_M$ is not in general the same as $\bigcup\{Outcm_c : c \text{ is a maximal chain in } M^+\}$. For example, suppose that $h, h' \in H_x$ and $Outcm_x(h) \neq Outcm_x(h')$ (h and h' share no point after x). Let $M = \{x, y\}$ with $x < y \in h'$. Then we can easily verify that $Outcm_x(h)$ is M-bordering, although not an outcome at the chain $\{x, y\}$.

4 Strategies and Their Admitted Future Outcomes

The semantic account for ought sentences developed in Horty (2001) emphasizes a dominance relation between choices at a single moment, for the same agent or group. Despite its merits, the account has two limitations. On the one hand, what one ought to achieve is often not what she can do in a single choice or action, but in a series of choices or actions. On the other hand, we may take a current choice to dominate another not because the immediate outcomes ensured by the former have higher value than those ensured by the latter. It may be because, when we look further into the future possibilities, the former opens a series of actions leading to future outcomes that have higher values than those to which the latter may lead us. This is what brought Horty to his theory in Horty (2001) of strategic ought with a single agent.

There are nevertheless some problems when Horty approaches his notions of strategic acts and strategic oughts, one of which is related to the notion of independence concerning choices for different agents at different moments. The problems are not really in the theories of strategies developed by Belnap and Horty, but in their applications or relations to other theories. We may then proceed safely to expand their theories of strategies, and discuss the problems in some other place.

This section and the following two expand Belnap's theory of strategies to the extent that we can talk about what different groups may do in the same strategy field. In doing so, we restrict ourselves to "primary strategies", as Belnap calls them, or "irredundant strategies", as Horty calls them. Notions and most terms are inherited

directly from Belnap (1996b) and Belnap et al. (2001). Throughout the rest of this paper, we fix $\langle T, <, Agent, Choice\rangle$ to be a stit frame, relative to which all upcoming discussions are to be understood.

A *strategy for* a group $\mathcal{G}$ *in* a field M is a function s such that $dom(s) \subseteq M$, where $dom(s)$ is the domain of s, and $s(x) \in Choice^x_{\mathcal{G}}$ for each $x \in dom(s)$. A *strategy for* $\mathcal{G}$ is a strategy for $\mathcal{G}$ in a field, and a *strategy* (in a field) is a strategy for a group (in the field). A *strategy for* an individual agent α (*in* a field) is a strategy for $\{\alpha\}$ (in the field).[12] We have assumed that a field is always nonempty, and now we further assume that so is every strategy in every field (with functions to be identified with sets of ordered pairs). Here are some basic notions concerning strategies.

Definition 4.1. Let s be any strategy, h any history, m any moment, H any outcome, and M any field. Then

(i) s *admits* h iff $h \in s(x)$ for each $x \in dom(s) \cap h$,[13]
(ii) $adh(s) = \{h' : s \text{ admits } h'\}$,
(iii) s *admits* m iff $m \in h$ for an $h \in adh(s)$,
(iv) $adm(s) = \{x : s \text{ admits } x\}$,
(v) s *admits* H iff $H \subseteq adh(s)$,
(vi) $ado_M(s) = \{H' \in OutcmBdr_M : s \text{ admits } H'\}$.

Concerning the new notion of admitted outcomes bordering a field M, it is easy to verify by definition that $\bigcup ado_M(s) \subseteq adh(s)$ for each strategy s in M, and hence the following fact holds:

Fact 4.2. Let M be any field in which s is a strategy for a group $\mathcal{G}$. Then for each $H \in OutcmBdr_M$, $H \subseteq \bigcup ado_M(s)$ iff $H \in ado_M(s)$.

Definition 4.3. Let s be any strategy for $\mathcal{G}$ in a field M. Then

(i) s is *primary* iff $dom(s) \subseteq adm(s)$;
(ii) s is *secondary* iff it is not primary;
(iii) s is *backward closed in* M iff for all $x, y \in M$, $x \in dom(s)$ and $y < x$ only if $y \in dom(s)$;
(iv) s is *simple in* M iff it is primary and backward closed in M.

For each group $\mathcal{G}$, we use $P\text{-}Strategy^M_{\mathcal{G}}$ ($S\text{-}Strategy^M_{\mathcal{G}}$) for the set of all primary (simple) strategies for $\mathcal{G}$ in M. The realm of primary strategies is our focus in this paper.[14] Note that for each $s \in P\text{-}Strategy^M_{\mathcal{G}}$ and each $x \in dom(s)$, $s(x) \cap adh(s) \neq \varnothing$ by definition, and hence $adh(s) \cap H_{\langle M\rangle}$ is never empty. Furthermore, the following fact can easily be verified.

---

[12] Belnap calls such a function a *consistent and strict strategy for* α *in* M in Belnap (1996b) and Belnap et al. (2001), while Horty calls it a *strategy for* α *in* M in Horty (2001), though the field M in the latter needs to have a starting point up to which M is backward closed.

[13] In Horty (2001), for s to admit h, it is further required that $h \cap dom(s) \neq \varnothing$.

[14] Secondary strategies are important for a study of conditional ought with respect to future outcomes, though they are not in the scope of our current work. For a brief discussion of secondary strategies, see Belnap (1996b) or Belnap et al. (2001).

Fact 4.4. (AC). For each primary strategy s, each nonempty chain in $dom(s)$ can be extended to a history that s admits.[15]

The following propositions establish some simple connections between admitted histories and admitted outcomes, the first of which states that a strategy in M admits an M-bordering outcome if it admits a member of it.

Proposition 4.5. Let s be a strategy for $\mathcal{G}$ in a field M. Then for each $H \in OutcmBdr_M$, $H \in ado_M(s)$ iff $H \cap adh(s) \neq \varnothing$.

Proof. Letting $H \in OutcmBdr_M$, we show that $h \in H \cap adh(s)$ only if $H \subseteq adh(s)$. Suppose that $h \in H \cap adh(s)$. Let $c = M^+ \cap h$. Then $c \neq \varnothing$ and $H = Outcm_c(h)$ by Fact 3.1. Consider any $h' \in H$ and any $x \in dom(s) \cap h'$. We know that $x \in M \cap h' \subseteq c \subseteq h \cap h'$, and thus by Facts 2.4–2.5, $Outcm_c(h) \subseteq Outcm_x(h) \subseteq Choice^x_{\mathcal{G}}(h)$. Then $Outcm_c(h) \subseteq s(x) = Choice^x_{\mathcal{G}}(h)$ since $h \in adh(s)$, and hence $h' \in s(x)$ since $h' \in Outcm_c(h)$. It follows that $h' \in s(x)$ for every $x \in dom(s) \cap h'$, and hence $h' \in adh(s)$. ■

The following proposition is useful when we extend our results concerning admitted histories passing through a field to similar results concerning admitted outcomes bordering the field.

Proposition 4.6. Let M be any properly covered field, and let s be any strategy in M. Then $\bigcup ado_M(s) = adh(s) \cap H_{\langle M \rangle}$.

Proof. By definition, $\bigcup ado_M(s) \subseteq adh(s)$ and $\bigcup ado_M(s) \subseteq \bigcup OutcmBdr_M$, and hence $\bigcup ado_M(s) \subseteq adh(s) \cap H_{\langle M \rangle}$ by Proposition 3.4 (iv). Consider any $h \in adh(s) \cap H_{\langle M \rangle}$. By Proposition 3.4 (ii), $h \in H$ for an $H \in OutcmBdr_M$, and then, since $h \in adh(s)$, Proposition 4.5 implies that $H \in ado_M(s)$, and hence $h \in \bigcup ado_M(s)$. It follows that $\bigcup ado_M(s) = adh(s) \cap H_{\langle M \rangle}$. ■

5 Pre-Simple Strategies and Complete Strategies

Here we present a brief discussion on pre-simple strategies and complete primary strategies. The proofs of propositions in this section follow closely those in Chap. 13 of Belnap et al. (2001), except that we expand various notions there concerning individual strategies to those concerning group strategies. Readers familiar with the materials in Chap. 13 of Belnap et al. (2001) may skip this section.

Recall that when $y < x$, we use $Choice^y_{\mathcal{G}}(H_x)$ for the unique $K \in Choice^y_{\mathcal{G}}$ such that $H_x \subseteq K$ (see Fact 1.1). A strategy s in a field M is *pre-simple in* M iff s is primary, and for all $x, y \in dom(s)$ and $z \in M$, $z < x$ and $z < y$ only if $Choice^z_{\mathcal{G}}(H_x) = Choice^z_{\mathcal{G}}(H_y)$.

For all strategies s and s' for a group $\mathcal{G}$, s' is an *extension of* s (or s' *extends* s) iff $s \subseteq s'$, where we identify functions as sets of ordered pairs. Note that when speaking

[15] It is also easy to verify that this does not hold for secondary strategies.

of an extension s' of a strategy s, we always presuppose that s and s' are strategies for the same group. Note also that if s' extends s, then by definition, $adh(s') \subseteq adh(s)$. A *simple* (*primary*) *extension* of a strategy s in M is an extension of s that is itself simple (primary) in M. A primary strategy in M may have no simple extension at all in M, but each pre-simple strategy in M does have such an extension.

Proposition 5.1. (AC). s is pre-simple for $\mathcal{G}$ in M iff it can be extended to a simple strategy for $\mathcal{G}$ in M.

Proof. Suppose that s is pre-simple for $\mathcal{G}$ in M. Let $D = \{y \in M - dom(s) : \exists x \in dom(s)(y < x)\}$. Consider any $y \in D$. Because s is pre-simple in M, there is a unique $K_y \in Choice^y_{\mathcal{G}}$ such that $H_x \subseteq K_y$ for each $x \in dom(s)$ with $y < x$. Let $s' = s \cup \{\langle y, K_y\rangle : y \in D\}$. It is easy to verify that s' is a backward closed extension of s in M, and then $adh(s') \subseteq adh(s)$ by definition. To show that s' is primary, consider any $x \in dom(s') = dom(s) \cup D$. Then there is a $u \in dom(s)$ such that $x \leqslant u$, and then, letting c be a maximal chain in $dom(s)$ containing u, we know by Fact 4.4 that $c = h \cap dom(s)$ for an $h \in adh(s)$. For each $y \in h \cap dom(s')$, if $y \in dom(s)$, $h \in s(y) = s'(y)$ since $h \in adh(s)$; and if $y \in D$, $y < z$ for a $z \in c$ by the maximality of c in $dom(s)$, and then $h \in s(z) \subseteq s'(y)$ by definition of s'. It follows that $h \in adh(s')$, and then, since $x \leqslant u \in h$, $x \in adm(s')$. Hence s' is primary.

Suppose that s is not pre-simple in M. If s is secondary, there is an $x \in dom(s)$ such that $H_x \cap adh(s) = \varnothing$, and then for each extension s' of s, $adh(s') \subseteq adh(s)$, and thus $H_x \cap adh(s') = \varnothing$, and hence s' is secondary. Assume that s is primary. Then for some $x, y \in dom(s)$ and $z \in M$, $z < x$ and $z < y$, and $Choice^z_{\mathcal{G}}(H_x) \neq Choice^z_{\mathcal{G}}(H_y)$. Consider any backward closed extension s' of s in M. If $s'(z) \neq Choice^z_{\mathcal{G}}(H_x)$, $H_x \cap adh(s') = \varnothing$, and then $x \notin adm(s')$; and similarly, if $s'(z) \neq Choice^z_{\mathcal{G}}(H_y)$, $y \notin adm(s')$. Since either $s'(z) \neq Choice^z_{\mathcal{G}}(H_x)$ or $s'(z) \neq Choice^z_{\mathcal{G}}(H_y)$, either $x \notin adm(s')$ or $y \notin adm(s')$, which makes s' secondary. ■

A complete strategy in a field is a strategy that is defined everywhere in the field along its admitted histories.

Definition 5.2. Let s be any strategy for $\mathcal{G}$ in a field M. Then

(i) s is *complete along* a history h *in* M iff $M \cap h \subseteq dom(s)$;
(ii) s *completely admits* h *in* M iff s admits h and is complete along h in M;
(iii) s is *complete in* M iff s is complete along every $h \in adh(s)$ in M.

The fact below is a direct consequence of our definitions.

Fact 5.3. Let s and s' be any strategies for $\mathcal{G}$ in M such that s' extends s. Then the following hold:

(i) s is complete along h in M only if s' is;
(ii) s completely admits h in M only if s' does.

We will use $\mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$ for the set of all complete primary strategies for $\mathcal{G}$ in M. It is obvious that $\mathit{CP\text{-}Strategy}^M_{\mathcal{G}} \subseteq \mathit{P\text{-}Strategy}^M_{\mathcal{G}}$. The following facts prove useful:

Fact 5.4. Let M be any field, and let $\mathcal{F}$ and $\mathcal{G}$ be any groups. Then

(i) $\mathit{CP\text{-}Strategy}^M_{\mathcal{G}} \subseteq \mathit{S\text{-}Strategy}^M_{\mathcal{G}}$;
(ii) for each $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$ and each $h \in H_{\langle M\rangle}$, $h \cap \mathit{dom}(s) \neq \varnothing$;
(iii) for each $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{F}}$ and each $s' \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$, $\mathit{dom}(s) \cap \mathit{dom}(s') \neq \varnothing$.

Proof. (i) Let $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$. For each $x \in \mathit{dom}(s)$, since s is primary, $h \in s(x)$ for an $h \in \mathit{adh}(s)$, and then, since s is complete along h, $M \cap \{y : y < x\} \subseteq M \cap h \subseteq \mathit{dom}(s)$. It follows that s is backward closed, and then $s \in \mathit{S\text{-}Strategy}^M_{\mathcal{G}}$. (ii) Let $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$ and $h \in H_{\langle M\rangle}$. If $h \cap \mathit{dom}(s) = \varnothing$, then trivially $h \in \mathit{adh}(s)$, and then $h \cap M \subseteq \mathit{dom}(s)$ since s is complete along h in M, and hence $h \cap M \subseteq h \cap \mathit{dom}(s) = \varnothing$, contrary to that $h \in H_{\langle M\rangle}$. (iii) follows from (i) and (ii). ■

For each strategy s for $\mathcal{G}$ in M, and for each $h \in \mathit{adh}(s)$, let s_h be a function on $\mathit{dom}(s) \cup (h \cap M)$ such that $s_h(x) = s(x)$ for each $x \in \mathit{dom}(s)$, and $s_h(x) = \mathit{Choice}^x_{\mathcal{G}}(h)$ for each $x \in (h \cap M) - \mathit{dom}(s)$. Such s_h is obviously unique, and is a strategy for $\mathcal{G}$ in M which extends s, and we say that s' *extends* s (*completely*) *along* h *in* M iff $s' = s_h$.

Proposition 5.5. Let s be a strategy for $\mathcal{G}$ in M, let $h \in \mathit{adh}(s)$, and let s' extend s along h in M. Then the following hold:

(i) s' completely admits h in M;
(ii) s is simple in M only if s' is.

Proof. (i) It is clear by definition that s' is complete along h in M, and $h \cap \mathit{dom}(s') = D \cup (\mathit{dom}(s) \cap h)$, where $D = (h \cap M) - \mathit{dom}(s)$. Also by definition, $h \in \mathit{Choice}^x_{\mathcal{G}}(h) = s'(x)$ for each $x \in D$, and, since $h \in \mathit{adh}(s)$, $h \in s(y) = s'(y)$ for each $y \in \mathit{dom}(s) \cap h$. It follows that $h \in \mathit{adh}(s')$.

(ii) Let s be simple in M. Then s' is evidently backward closed in M. To show that s' is primary, consider any $x \in \mathit{dom}(s')$. If $x \in h$, $x \in \mathit{adm}(s')$ since $h \in \mathit{adh}(s')$ by (i). Suppose that $x \in \mathit{dom}(s') - h$. Then $x \in \mathit{dom}(s) - h$. Since s is primary, $x \in h'$ for an $h' \in \mathit{adh}(s)$. Now for each $y \in \mathit{dom}(s') \cap h'$, if $y < x$, $y \in \mathit{dom}(s)$ since s is backward closed in M, and if $x \leqslant y$, $y \notin h$ since $x \notin h$, and hence $y \in \mathit{dom}(s)$. It follows that for each $y \in \mathit{dom}(s') \cap h'$, $y \in \mathit{dom}(s)$ and then $s'(y) = s(y)$, which implies that $h' \in s'(y)$ since $h' \in \mathit{adh}(s)$. Hence $h' \in \mathit{adh}(s')$, and then $x \in \mathit{adm}(s')$ ■.

A *complete primary extension* of a strategy s for $\mathcal{G}$ in M is an $s' \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$ such that $s \subseteq s'$. We show below that each pre-simple strategy has a complete primary extension.

Proposition 5.6. (AC). Let S be a nonempty $\subseteq$-chain of simple strategies for $\mathcal{G}$ in M. Then $\bigcup S$ is a simple strategy for $\mathcal{G}$ in M that extends all $s' \in S$.

Proof. Let $s = \bigcup S$. It is easy to see that s is a strategy for $\mathcal{G}$ in M extending all $s' \in S$, and is backward closed in M since each $s' \in S$ is. It then suffices to let $x \in dom(s)$ and show that $x \in h$ for an $h \in adh(s)$. Let c be a maximal chain in $dom(s)$ containing x. *Case* 1, c contains a largest member u. Then $u \in dom(s')$ for an $s' \in S$, and $s'(y) = s(y)$ for each $y \in dom(s)$ with $y \leqslant u$. Since s' is primary, $u \in h$ for an $h \in adh(s')$, and then, because $dom(s) \cap h = dom(s') \cap h$, it follows that $h \in s'(y) = s(y)$ for each $y \in dom(s) \cap h$, i.e., $h \in adh(s)$. *Case* 2, c has no largest member. Let h be any history including c. Consider any $y \in dom(s) \cap h$. Then there is a $z \in dom(s) \cap h$ and an $s'' \in S$ such that $y, z \in dom(s'')$, $y < z$ and $s(y) = s''(y)$. Since s'' is primary, $z \in h'$ for some $h' \in adh(s'')$, and then, since $h, h' \in H_z$ and $y < z$, $h \in s''(y) = s(y)$ by NC (see Sect. 1). It follows that $h \in adh(s)$. ■

Applying Zorn's lemma, Fact 5.3 and Propositions 5.5–5.6 and 5.1, one can routinely establish the following.

Proposition 5.7. (AC). For each $s \in S\text{-}Strategy^M_{\mathcal{G}}$ and each $h \in adh(s)$, $s \subseteq s'$ and $h \in adh(s')$ for an $s' \in CP\text{-}Strategy^M_{\mathcal{G}}$, and hence $adh(s) = \bigcup_{s'' \in S} adh(s'')$ where $S = \{s'' \in CP\text{-}Strategy^M_{\mathcal{G}} : s \subseteq s''\}$. Consequently, each pre-simple strategy for $\mathcal{G}$ in M has a complete primary extension in M.

6 Group-Joining Meets

Bringing in different agents and their strategies provides new perspectives to a study of strategies for different agents in the same fields. As a preparation for our discussions on distinguishability and independence, we deal with some technical notions in this section.

Consider two agents α and β, and their strategies s_α and s_β in a field M with $m \in D = dom(s_\alpha) \cap dom(s_\beta)$. Since $\alpha \neq \beta$, IA requires $s_\alpha(m) \cap s_\beta(m) \neq \varnothing$, and the same can be said about each point in D. Letting s be a function on D such that $s(x) = s_\alpha(x) \cap s_\beta(x)$ for each $x \in D$, we know that each $s(x)$ with $x \in D$ is a member of $Choice^x_{\{\alpha,\beta\}}$, and hence s is a strategy for $\{\alpha, \beta\}$ in M. This strategy is what we call the "group-joining meet" of s_α and s_β.

Definition 6.1. Let s and s' be strategies in M for $\mathcal{G}$ and $\mathcal{F}$ respectively such that $\mathcal{G} \cap \mathcal{F} = \varnothing$ and $dom(s) \cap dom(s') \neq \varnothing$. The *group-joining meet of s and s'*, written $s \sqcap s'$, is the function s^* on $dom(s^*) = dom(s) \cap dom(s')$ such that $s^*(x) = s(x) \cap s'(x)$ for every $x \in dom(s^*)$.

It is easy to see that $s \sqcap s' = s' \sqcap s$ when both are defined. When $\mathcal{G}$ and $\mathcal{F}$ are disjoint while $dom(s)$ and $dom(s')$ are not, we know that for each $x \in dom(s \sqcap s')$, $(s \sqcap s')(x)$ is not only nonempty, but also identical to a member of $Choice^x_{\mathcal{G} \cup \mathcal{F}}$. It then follows that $s \sqcap s'$ is a strategy for $\mathcal{G} \cup \mathcal{F}$ in M.

Before showing some facts concerning strategies and their group-joining meets, we need to show that for all primary strategies s and s' for $\mathcal{F}$ and $\mathcal{G}$ respectively, if

$\mathcal{F}$ and $\mathcal{G}$ are disjoint but $dom(s)$ and $dom(s')$ are not, then $adh(s) \cap adh(s') \neq \varnothing$. To that end, we use the following auxiliary notion. Let s and s' be strategies for $\mathcal{G}$ in M, and let $h \in H_{\langle M \rangle}$ and $c \subseteq h \cap M$. *s' extends s in M along c w.r.t. h* if s' extends s such that $dom(s') = dom(s) \cup c$ and $s'(x) = Choice^{x}_{\mathcal{G}}(h)$ for each $x \in h \cap (c - dom(s))$. Note that if s' is an extension of s in M along c w.r.t. h, then such s' is unique, and $h \in adh(s)$ only if $h \in adh(s')$.

Proposition 6.2. (AC). Let $s_{\mathcal{F}}$ and $s_{\mathcal{G}}$ be primary strategies for $\mathcal{F}$ and $\mathcal{G}$ in M respectively, where $\mathcal{F} \cap \mathcal{G} = \varnothing$, and let $m \in dom(s_{\mathcal{F}}) \cap dom(s_{\mathcal{G}})$.[16] Then $adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}}) \cap H_m \neq \varnothing$, and $ado_M(s_{\mathcal{F}}) \cap ado_M(s_{\mathcal{G}}) \neq \varnothing$ if M is properly covered.

Proof. Let $D_{\mathcal{F}} = \{x \in M : \exists y(x \leqslant y \in dom(s_{\mathcal{F}}))\}$ and $D_{\mathcal{G}} = \{x \in M : \exists y(x \leqslant y \in dom(s_{\mathcal{G}}))\}$. By hypothesis, $m \in D_{\mathcal{F}} \cap D_{\mathcal{G}}$. Let c be a maximal chain in $D_{\mathcal{F}} \cap D_{\mathcal{G}}$ containing m. It is easy to verify that $H_{[c]} \subseteq H_m$ and

$$\text{for each } h \in H_{[c]},\ h \cap dom(s_{\mathcal{F}}) \subseteq c \text{ or } h \cap dom(s_{\mathcal{G}}) \subseteq c. \tag{2}$$

Let $d_{\mathcal{F}}$ and $d_{\mathcal{G}}$ be any maximal chains, extending c, in $D_{\mathcal{F}}$ and $D_{\mathcal{G}}$ respectively. It is easy to see that $d_{\mathcal{F}} \cap dom(s_{\mathcal{F}})$ is a maximal chain in $dom(s_{\mathcal{F}})$ that is co-final with $d_{\mathcal{F}}$, and therefore can by Fact 4.4 be extended to an $h_{\mathcal{F}} \in adh(s_{\mathcal{F}})$ such that $c \subseteq d_{\mathcal{F}} \subseteq h_{\mathcal{F}}$. Similarly, $c \subseteq d_{\mathcal{G}} \subseteq h_{\mathcal{G}}$ for an $h_{\mathcal{G}} \in adh(s_{\mathcal{G}})$. Let $s'_{\mathcal{F}}$ be the extension of $s_{\mathcal{F}}$ in M along c w.r.t. $h_{\mathcal{F}}$, and let $s'_{\mathcal{G}}$ be the extension of $s_{\mathcal{G}}$ in M along c w.r.t. $h_{\mathcal{G}}$. By our note above,

$$h_{\mathcal{F}} \in adh(s'_{\mathcal{F}}) \subseteq adh(s_{\mathcal{F}}) \text{ and } h_{\mathcal{G}} \in adh(s'_{\mathcal{G}}) \subseteq adh(s_{\mathcal{G}}). \tag{3}$$

By Propositions 3.4 (ii) and 4.5, it suffices to show that

$$adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}}) \cap H_{[c]} \neq \varnothing. \tag{4}$$

Case 1, there is no last point in c. By (3), $h_{\mathcal{F}}, h_{\mathcal{G}} \in s'_{\mathcal{F}}(x) \cap s'_{\mathcal{G}}(x)$ for each $x \in c$. It is then easy to see by our case assumption and NC that

$$\text{for each } x \in c,\ H_{[c]} \subseteq s'_{\mathcal{F}}(x) \cap s'_{\mathcal{G}}(x). \tag{5}$$

Assume that $h_{\mathcal{F}} \notin adh(s'_{\mathcal{G}})$, or by (3) there is nothing more to show. Then there is a $y \in h_{\mathcal{F}} \cap dom(s'_{\mathcal{G}})$ such that $h_{\mathcal{F}} \notin s'_{\mathcal{G}}(y)$, and then by (5), $c < y \in h_{\mathcal{F}} \cap dom(s_{\mathcal{G}})$, and hence by (2), $h \cap dom(s_{\mathcal{F}}) \subseteq c$ for each $h \in H_y$. It follows from (5) that $H_y \subseteq adh(s'_{\mathcal{F}})$. Since $s_{\mathcal{G}}$ is primary, there is an $h' \in H_y \cap adh(s_{\mathcal{G}})$, and hence (4) holds.

Case 2, there is a last point z in c. Let $H = s'_{\mathcal{F}}(z) \cap s'_{\mathcal{G}}(z)$ and $X = (\bigcup H) \cap \{y : z < y\}$. Because $\mathcal{F} \cap \mathcal{G} = \varnothing$, $H \neq \varnothing$ by IA. We claim that

[16] The condition that $m \in dom(s_{\mathcal{G}}) \cap dom(s_{\mathcal{F}})$ can be weakened to that $m \in M$ such that $m \leqslant x$ and $m \leqslant y$ for an $x \in dom(s_{\mathcal{G}})$ and a $y \in dom(s_{\mathcal{F}})$.

$$\text{for each } x \in c,\ H \subseteq s'_{\mathcal{F}}(x) \cap s'_{\mathcal{G}}(x). \tag{6}$$

For each $x < z$ and $h \in H$, because $h, h_{\mathcal{F}}, h_{\mathcal{G}} \in H_z$, we know by NC and (3) that $h \in Choice^x_{\mathcal{F}}(h_{\mathcal{F}}) = s'_{\mathcal{F}}(x)$, and similarly, $h \in s'_{\mathcal{G}}(x)$. Hence (6) holds. *Subcase* 2A, there is an $x \in dom(s'_{\mathcal{F}}) \cap X$. Then $c \leqslant z < x \in dom(s_{\mathcal{F}})$, and hence by (2), $h' \cap dom(s_{\mathcal{G}}) \subseteq c$ for each $h' \in H_x$, and then $h' \cap dom(s'_{\mathcal{G}}) \subseteq c$ for each $h' \in H_x$ because $dom(s'_{\mathcal{G}}) - dom(s_{\mathcal{G}}) \subseteq c$. It then follows from $H_x \subseteq H$ and (6) that $H_x \subseteq adh(s'_{\mathcal{G}}) \subseteq adh(s_{\mathcal{G}})$. Since $s_{\mathcal{F}}$ is primary, there is an $h \in H_x \cap adh(s_{\mathcal{F}})$, and then (4) holds. *Subcase* 2B, $dom(s'_{\mathcal{F}}) \cap X = \varnothing$. If there is a $u \in dom(s'_{\mathcal{G}}) \cap X$, an argument similar to that in subcase 2A will show that (4) holds. If $dom(s'_{\mathcal{G}}) \cap X = \varnothing$, then (6) implies (4). ■

The following proposition provides a list of useful facts concerning strategies and their group-joining meets.

Proposition 6.3. Let s and s' be strategies for $\mathcal{G}$ and $\mathcal{F}$ in M respectively, where $\mathcal{G} \cap \mathcal{F} = \varnothing$ and $dom(s) \cap dom(s') \neq \varnothing$, and let $s^* = s \sqcap s'$. Then

(i) $adh(s) \cap adh(s') \subseteq adh(s^*)$;
(ii) $adm(s) \cap adm(s') \subseteq adm(s^*)$ if both s and s' are primary; (AC)
(iii) $adh(s^*) \subseteq adh(s)$ and $adm(s^*) \subseteq adm(s)$ if s and s' are backward closed and s' is complete in M;
(iv) $adh(s^*) = adh(s) \cap adh(s')$ if s and s' are both backward closed and complete in M ;
(v) $ado_M(s^*) = ado_M(s) \cap ado_M(s')$ if M is properly covered in which s and s' are both backward closed and complete;
(vi) $adm(s^*) = adm(s) \cap adm(s')$ if s and s' are both primary and complete in M; (AC)
(vii) s^* is backward closed in M if both s and s' are;
(viii) s^* is primary (simple) in M if both s and s' are; (AC)
(ix) s^* completely admits h in M if both s and s' do;
(x) s^* is backward closed and complete in M if both s and s' are.

Proof. (ii) Assume that s and s' are primary. Consider any $x \in adm(s) \cap adm(s')$. By definition, $h, h' \in H_x$ for some $h \in adh(s)$ and $h' \in adh(s')$. Suppose for reductio that $x \notin adm(s^*)$. Then $adh(s^*) \cap H_x = \varnothing$, and hence by (i) and Proposition 6.2, $y \notin dom(s^*) = dom(s) \cap dom(s')$ for each $y \geqslant x$, and consequently, since $h \notin adh(s^*)$, there is a $z \in h \cap dom(s^*)$ such that $h \notin s^*(z)$ and $z < x$. By definition, $s^*(z) = s(z) \cap s'(z)$. Then $h \in adh(s) \cap H_x$ implies $h \in s(z) - s'(z)$, and $h' \in adh(s') \cap H_x$ implies $h' \in s'(z)$, and then by NC, $H_x \subseteq s'(z)$, and hence $h \in s'(z)$, a contradiction.

(iii) Let s and s' be backward closed, and s' complete, in M. Suppose for reductio that $h \in adh(s^*) - adh(s)$, i.e., $h \notin s(x)$ for an $x \in dom(s) \cap h$, and

$$h \in s^*(y) \subseteq s'(y) \text{ for each } y \in c, \tag{7}$$

where $c = dom(s^*) \cap h$. We first claim that

$$x \notin dom(s'), \tag{8}$$

for otherwise we have $x \in c$ since $x \in dom(s) \cap h$, and then $h \in s^*(x) \subseteq s(x)$, a contradiction. Next, suppose for reductio that $x' \notin dom(s)$ for an $x' \in dom(s') \cap h$. Since $x, x' \in h$, either $x' < x$ or $x \leqslant x'$, and then, since s is backward closed in M, $x' < x$ only if $x' \in dom(s)$, contrary to the supposition of this reductio, and hence $x \leqslant x'$. But s' is also backward closed in M, and hence $x \in dom(s')$, contrary to (8). We conclude from this reductio that $dom(s') \cap h \subseteq dom(s)$, and then $c = dom(s) \cap dom(s') \cap h = dom(s') \cap h$, and hence by (7), $h \in adh(s')$. But s' is complete in M, and hence $h \cap M \subseteq dom(s')$, contrary to (8). The rest of (iii) is straightforward.

(iv) follows from (i) and (iii). (v) For each $H \in OutcmBdr_M$, that $H \in ado_M(s \sqcap s')$ is equivalent to each of the following:

- $H \subseteq \bigcup ado_M(s \sqcap s')$ — Fact 4.2
- $H \subseteq adh(s \sqcap s') \cap H_{\langle M \rangle}$ — Proposition 4.6
- $H \subseteq adh(s) \cap adh(s') \cap H_{\langle M \rangle}$ — (iv)
- $H \subseteq (\bigcup ado_M(s)) \cap (\bigcup ado_M(s'))$ — Proposition 4.6
- $H \in ado_M(s) \cap ado_M(s)$ — Fact 4.2

(i) and (vi)–(x) are easily verifiable by definition and (ii)–(iii). ■

Note that Proposition 6.3 (iii–vi) have little room for a generalization concerning the completeness requirement for strategies. In other words, $adh(s \sqcap s') \subseteq adh(s)$ may fail if s' is not complete in M.[17]

Let $\mathcal{E}$ be any group. For each $s \in CP\text{-}Strategy^M_{\mathcal{E}}$ and each group $\mathcal{G} \subseteq \mathcal{E}$, let $s|_{\mathcal{G}}$ be the strategy for $\mathcal{G}$ in M such that $dom(s|_{\mathcal{G}}) = dom(s)$ and for each $x \in dom(s)$, $s(x) \subseteq s|_{\mathcal{G}}(x)$, i.e., $s|_{\mathcal{G}}(x)$ is the only member of $Choice^x_{\mathcal{G}}$ that includes $s(x)$. We call $s|_{\mathcal{G}}$ *the subordinate strategy for* $\mathcal{G}$ *in* s. Note that because $s \in CP\text{-}Strategy^M_{\mathcal{E}}$, $s|_{\mathcal{G}}$ is obviously backward closed in M, and completely admits every $h \in adh(s)$, and hence is primary. Note also that because $adh(s)$ and $adh(s|_{\mathcal{G}})$ may not in general be the same, $s|_{\mathcal{G}}$ may not in general be a complete strategy for $\mathcal{G}$.

Proposition 6.4. (AC). Let $\mathcal{F}$ and $\mathcal{G}$ be disjoint groups and $\mathcal{E} = \mathcal{F} \cup \mathcal{G}$, let $s \in CP\text{-}Strategy^M_{\mathcal{E}}$, and let $s_{\mathcal{F}}$ and $s_{\mathcal{G}}$ be any complete primary extensions of $s|_{\mathcal{F}}$ and $s|_{\mathcal{G}}$ in M respectively. Then $adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}}) = adh(s)$ and $s = s|_{\mathcal{F}} \sqcap s|_{\mathcal{G}} = s_{\mathcal{F}} \sqcap s_{\mathcal{G}} = s|_{\mathcal{F}} \sqcap s_{\mathcal{G}} = s_{\mathcal{F}} \sqcap s|_{\mathcal{G}}$.

Proof. It follows by definition that $s = s|_{\mathcal{F}} \sqcap s|_{\mathcal{G}}$. Suppose that $h \in adh(s)$. Because s completely admits h in M, $h \in s(x)$ for each $x \in h \cap M$, and then, because

[17] Let $x < y$, let $Choice^y_\beta = \{K, K'\}$ with $K \neq K'$, and let $h \in K$ and $h' \in K'$. Suppose that $h, h' \in K_\alpha \cap K_\beta$ for a $K_\alpha \in Choice^x_\alpha$ and a $K_\beta \in Choice^x_\beta$ (the rest of the choice situation is not essential). Let s_α and s_β be strategies for α and β in $\{x, y\}$ such that $dom(s_\alpha) = \{x\}$ and $dom(s_\beta) = \{x, y\}$, $s_\alpha(x) = K_\alpha$ and $s_\beta(x) = K_\beta$, and $s_\beta(y) = K'$. It is then easy to verify that $h \in adh(s_\alpha \sqcap s_\beta)$ but $h \notin adh(s_\beta)$.

$s(x) = s|_{\mathcal{F}}(x) \cap s|_{\mathcal{G}}(x)$ for each such x, $h \in s_{\mathcal{F}}(x) \cap s_{\mathcal{G}}(x)$ for each $x \in h \cap M$, and hence $h \in adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}})$. Suppose for reductio that $h \in adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}})$ and $h \notin adh(s)$. Then $h \notin s(x)$ for an $x \in h \cap dom(s)$, and $h \in s_{\mathcal{F}}(x) \cap s_{\mathcal{G}}(x)$ because $dom(s) \subseteq dom(s_{\mathcal{F}}) \cap dom(s_{\mathcal{G}})$. By definition, $x \in dom(s)$ implies $s_{\mathcal{F}}(x) = s|_{\mathcal{F}}(x)$ and $s_{\mathcal{G}}(x) = s|_{\mathcal{G}}(x)$, and hence $h \in s|_{\mathcal{F}}(x) \cap s|_{\mathcal{G}}(x) = s(x)$, a contradiction. It follows that $adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}}) = adh(s)$. By Proposition 6.3(iv,viii,ix), $adh(s_{\mathcal{F}} \sqcap s_{\mathcal{G}}) = adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}})$, and both $s_{\mathcal{F}} \sqcap s_{\mathcal{G}}$ and s are primary and complete along their admitted histories in M, and hence $dom(s_{\mathcal{F}} \sqcap s_{\mathcal{G}}) = dom(s)$, from which it follows that $s_{\mathcal{F}} \sqcap s_{\mathcal{G}} = s$. The rest of the proposition is guaranteed by definition. ■

7 Distinguishability

Let $\mathcal{G}$ be any group, and let m be any point. Speaking in an abstract way, what $\mathcal{G}$ can do at m is identified with a set of histories within which $\mathcal{G}$ may intuitively constrain the future course of events to lie, as suggested in Belnap et al. (2001) and Horty (2001). Such a set of histories is, of course, presented as a member of $Choice^{m}_{\mathcal{G}}$. To put the matter in a different way, we may say that what $\mathcal{G}$ can do at best at a moment is identified with a maximal set of histories indistinguishable for $\mathcal{G}$ at the moment, where h and h' are *distinguishable for* $\mathcal{G}$ *at* m if $h, h' \in H_m$ and $Choice^{m}_{\mathcal{G}}(h) \neq Choice^{m}_{\mathcal{G}}(h')$.[18] Concerning what $\mathcal{G}$ can do at a single point, this notion of distinguishability may not seem to provide more than what the notion of choice does, for a maximal set of histories indistinguishable for $\mathcal{G}$ at m is nothing but a member of $Choice^{m}_{\mathcal{G}}$. Concerning what $\mathcal{G}$ can do through a field, nevertheless, the notion of distinguishability does provide more than that of choice. By making choices at various points in a field M, $\mathcal{G}$ may also constrain the future course of events to lie within a set of histories passing through M. Applying the notion of distinguishability, we may differentiate one from another of what $\mathcal{G}$ can do through M, which amounts to identifying each of them with a maximal set of histories indistinguishable for $\mathcal{G}$. Metaphorically speaking, distinguishability displays what $\mathcal{G}$ can do through M at the highest resolution. In this section, we study what groups can do through a field in terms of distinguishability.

Let M be any field, let $\mathcal{G}$ be any group, and let $h, h' \in H_{\langle M \rangle}$. h and h' are *distinguishable for* $\mathcal{G}$ *in* M if they are distinguishable for $\mathcal{G}$ at a point in M, and are *indistinguishable for* $\mathcal{G}$ *in* M otherwise. Intuitively, two histories are distinguishable for $\mathcal{G}$ in M just in case some choices for $\mathcal{G}$ at a point in M can tell them apart. It is easy to see that the relation of distinguishability for $\mathcal{G}$ in M is irreflexive and symmetrical, while the relation of indistinguishability for $\mathcal{G}$ in M is reflexive and symmetrical. Note that for h and h' to be distinguishable for $\mathcal{G}$ at x, they must both pass through x. In other words, distinguishability requires availability. It should then be clear that for h and h' to be distinguishable for $\mathcal{G}$ at m, it is *not* enough to only have that $h' \notin Choice^{x}_{\mathcal{G}}(h)$.

[18] The idea here of distinguishability is from Belnap. See, e.g., Belnap (1991).

For each $H \subseteq H_{\langle M \rangle}$, H is *$\mathcal{G}$-indistinguishable in M* if members of H are pairwise indistinguishable for $\mathcal{G}$ in M. When the field M is clear in the context, we often drop the phrase "in M". The following proposition shows that histories distinguishable for a group are distinguishable for all its super-groups, or by contraposition, histories indistinguishable for a group are indistinguishable for all its sub-groups.

Proposition 7.1. Let M be any field, let $\mathcal{F}$ and $\mathcal{G}$ be any groups such that $\mathcal{F} \subseteq \mathcal{G}$, and let $h, h' \in H_{\langle M \rangle}$ and $H \subseteq H_{\langle M \rangle}$. Then

(i) if h and h' are distinguishable for $\mathcal{F}$, so are they for $\mathcal{G}$; and
(ii) if H is $\mathcal{G}$-indistinguishable, it is $\mathcal{F}$-indistinguishable.

Proof. (i) Suppose that h and h' are distinguishable for $\mathcal{F}$. Then there is an $m \in M$ such that $h, h' \in H_m$ and $Choice^m_{\mathcal{F}}(h) \neq Choice^m_{\mathcal{F}}(h')$. Since $\mathcal{F} \subseteq \mathcal{G}$, $Choice^m_{\mathcal{G}}(h) \subseteq Choice^m_{\mathcal{F}}(h)$ and $Choice^m_{\mathcal{G}}(h') \subseteq Choice^m_{\mathcal{F}}(h')$, and hence, since $Choice^m_{\mathcal{F}}(h) \cap Choice^m_{\mathcal{F}}(h') = \varnothing$, $Choice^m_{\mathcal{G}}(h) \neq Choice^m_{\mathcal{G}}(h')$. It then follows that h and h' are distinguishable for $\mathcal{G}$. (ii) follows directly from (i). ■

Provided that A is a nonempty set, a *classification* of A is a set-theoretical "cover" of A, i.e., a subset $\mathbb{X}$ of $\mathscr{P}(A)$ (the powerset of A) such that $\bigcup \mathbb{X} = A$ and for all $X, X' \in \mathbb{X}$, $X \subseteq X'$ only if $X = X'$. A classification of A is like a partition of A, as they both satisfy the condition of exhaustiveness, i.e., $\bigcup \mathbb{X} = A$. The difference between them is obvious, too. A classification allows its members to partly overlap, whereas a partition needs to satisfy the condition of disjointedness, i.e., its members need to be pairwise disjoint. A classification $\mathbb{X}$ of A is *trivial* if $\mathbb{X} = \{A\}$. Note that if $\mathbb{X}$ is a non-trivial classification of A, then $\bigcap \mathbb{X}$ *cannot* be a member of $\mathbb{X}$ because no member of a classification is a proper subset of another. For the same reason, $\varnothing$ is never a member of any classification of any nonempty set.

Let M be any field, and let $H \subseteq H_{\langle M \rangle}$. For each group $\mathcal{G}$, H is a *maximal $\mathcal{G}$-indistinguishable set* (*of histories*) *through M* (a *$\mathcal{G}$-MIS through M*) if H is $\mathcal{G}$-indistinguishable in M but no proper extension of H in $H_{\langle M \rangle}$ is. We will drop the phrase "through M" when M is clear in the context. Note that a $\mathcal{G}$-MIS through M is a maximal set of histories that $\mathcal{G}$ cannot distinguish in M, and is therefore a minimal set of histories within which $\mathcal{G}$ can constrain the future course of events to lie, i.e., one of what $\mathcal{G}$ can do at best through M.

For each group $\mathcal{G}$, let $\mathbb{A}_{M,\mathcal{G}}$ be the set of all $\mathcal{G}$-MISs through M. Applying an argument similar to that used in Lindenbaum's lemma, one can easily verify that for each $h \in H_{\langle M \rangle}$ and each group $\mathcal{G}$, h is contained in a $\mathcal{G}$-MIS through M. Hence we have the following:

Fact 7.2. (AC). For each field M and each group $\mathcal{G}$, $\mathbb{A}_{M,\mathcal{G}}$ is a classification of $H_{\langle M \rangle}$.

For each group $\mathcal{G}$, let us call the classification $\mathbb{A}_{M,\mathcal{G}}$ of $H_{\langle M \rangle}$ the *classification of $H_{\langle M \rangle}$ determined by $\mathcal{G}$*. The following proposition provides a correspondence between $\mathcal{G}$-MISs and complete primary strategies for $\mathcal{G}$: a $\mathcal{G}$-MIS through M is nothing but the set of histories in $H_{\langle M \rangle}$ admitted by a complete primary strategy s for $\mathcal{G}$ in M. Consequently, members of the classification of $H_{\langle M \rangle}$ determined by $\mathcal{G}$ are sets of histories in $H_{\langle M \rangle}$ admitted by complete primary strategies for $\mathcal{G}$ in M.

Proposition 7.3. Let $\mathcal{G}$ be any group, let M be any field, and let $H \subseteq H_{\langle M\rangle}$. Then H is a $\mathcal{G}$-MIS through M iff $H = adh(s) \cap H_{\langle M\rangle}$ for an $s \in CP\text{-}Strategy^{M}_{\mathcal{G}}$. Hence $\mathbb{A}_{M,\mathcal{G}} = \{adh(s) \cap H_{\langle M\rangle} : s \in CP\text{-}Strategy^{M}_{\mathcal{G}}\}$.

Proof. Suppose that H is a $\mathcal{G}$-MIS. Let $D = M \cap (\bigcup H)$. We claim that

$$\text{for each } x \in D,\ K \cap H \neq \varnothing \text{ for exactly one } K \in Choice^{x}_{\mathcal{G}}. \tag{9}$$

Let $x \in D$. Then $x \in h \cap M$ for an $h \in H$. Letting $K = Choice^{x}_{\mathcal{G}}(h)$, we have that $K \cap H \neq \varnothing$. For each $K' \in Choice^{x}_{\mathcal{G}}$ such that $K' \cap H \neq \varnothing$, $K' = Choice^{x}_{\mathcal{G}}(h')$ for an $h' \in H$, and hence, since H is $\mathcal{G}$-indistinguishable, $Choice^{x}_{\mathcal{G}}(h') = K$. It then follows that (9) holds. Now let s be a function on D such that for each $x \in D$, $s(x) =$ the only $K \in Choice^{x}_{\mathcal{G}}$ such that $K \cap H \neq \varnothing$. Then s is a strategy for $\mathcal{G}$ in M that is backward closed in M. We show below that $H = adh(s) \cap H_{\langle M\rangle}$ and $s \in CP\text{-}Strategy^{M}_{\mathcal{G}}$.

Consider any $h \in H_{\langle M\rangle}$. If $h \in H$, then for each $x \in h \cap dom(s)$, $s(x)$ is the only $K \in Choice^{x}_{\mathcal{G}}$ such that $K \cap H \neq \varnothing$, and hence $h \in s(x)$, from which it follows that $h \in adh(s)$. If $h \in H_{\langle M\rangle} - H$, then, since H is a $\mathcal{G}$-MIS, there is an $h' \in H$ and a $y \in M$ such that $h, h' \in H_y$ and $Choice^{y}_{\mathcal{G}}(h) \neq Choice^{y}_{\mathcal{G}}(h')$, and hence by definition of s, $s(y) = Choice^{y}_{\mathcal{G}}(h')$, and consequently $h \notin adh(s)$. It then follows that $H = adh(s) \cap H_{\langle M\rangle}$, which implies that $dom(s) = M \cap (\bigcup H) \subseteq \bigcup adh(s)$ and that $M \cap h \subseteq dom(s)$ for each $h \in adh(s)$, and hence s is primary and is complete in M.

Next suppose that $s \in CP\text{-}Strategy^{M}_{\mathcal{G}}$. We show that H is an $\mathcal{G}$-MIS with $H = adh(s) \cap H_{\langle M\rangle}$. To show that H is $\mathcal{G}$-indistinguishable, it suffices to let $h \in H$ and $h' \in H_{\langle M\rangle}$ such that $K = Choice^{x}_{\mathcal{G}}(h) \neq Choice^{x}_{\mathcal{G}}(h') = K'$ for an $x \in M$, and show that $h' \notin adh(s)$. Because $h \in adh(s) \cap K$, $s(x) = K$ by the completeness of s, and then, since $h' \in K' \neq K$, $h' \notin adh(s)$. To show further that H is a $\mathcal{G}$-MIS, consider any $h_0 \in H_{\langle M\rangle} - adh(s)$. By definition, $h_0 \notin s(y)$ for a $y \in h_0 \cap dom(s)$, and hence $Choice^{y}_{\mathcal{G}}(h_0) \neq s(y)$. Because s is primary and complete in M, there is an $h' \in adh(s)$ such that $s(y) = Choice^{y}_{\mathcal{G}}(h')$. Since $h' \in H_y \subseteq H_{\langle M\rangle}$, it follows that $H \cup \{h_0\}$ is not $\mathcal{G}$-indistinguishable. Hence H is an $\mathcal{G}$-MIS. ■

As noted earlier, what a group can do at best through a field M is to attain a maximal set of histories indistinguishable for the group in M. Proposition 7.3 provides a "finest" characterization of what a group can do through M: For each group $\mathcal{G}$, to attain a $\mathcal{G}$-MIS H means the same as coordinating its members' efforts in such a way that their joint choices form a complete primary strategy that admits H.

The following proposition shows a relation between the classification of $H_{\langle M\rangle}$ determined by a group and the classifications determined by its sub-groups, which will be useful in our discussion of independence.

Proposition 7.4. (AC). Let M be any field, and let $\mathcal{F}$ and $\mathcal{G}$ be disjoint groups. Then $\mathbb{A}_{M,\mathcal{F}\cup\mathcal{G}} = \{H \cap H' : H \in \mathbb{A}_{M,\mathcal{F}} \wedge H' \in \mathbb{A}_{M,\mathcal{G}}\}$.

Proof. Consider any $H \in \mathbb{A}_{M,\mathcal{F}}$ and $H' \in \mathbb{A}_{M,\mathcal{G}}$. By Proposition 7.3, $H = adh(s) \cap H_{\langle M \rangle}$ and $H' = adh(s') \cap H_{\langle M \rangle}$ for some $s \in \textit{CP-Strategy}^M_{\mathcal{F}}$ and $s' \in \textit{CP-Strategy}^M_{\mathcal{G}}$, and then by Fact 5.4 (iii), $dom(s) \cap dom(s') \neq \varnothing$, and hence $H \cap H' = adh(s \sqcap s') \cap H_{\langle M \rangle}$ by Proposition 6.3(iv). We know by Proposition 6.3 (viii,x) that $s \sqcap s' \in \textit{CP-Strategy}^M_{\mathcal{F}\cup\mathcal{G}}$, and then by Proposition 7.3 again, $H \cap H' \in \mathbb{A}_{M,\mathcal{F}\cup\mathcal{G}}$.

Consider any $H^* \in \mathbb{A}_{M,\mathcal{F}\cup\mathcal{G}}$. By Proposition 7.3, $H^* = adh(s^*) \cap H_{\langle M \rangle}$ for an $s^* \in \textit{CP-Strategy}^M_{\mathcal{F}\cup\mathcal{G}}$. Let $s_{\mathcal{F}}$ and $s_{\mathcal{G}}$ be any complete primary extensions of $s^*|_{\mathcal{F}}$ and $s^*|_{\mathcal{G}}$ in M respectively. By Proposition 6.4, $s^* = s_{\mathcal{F}} \sqcap s_{\mathcal{G}}$, and then by Proposition 6.3(iv), $adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{G}}) \cap H_{\langle M \rangle} = H^*$, while $adh(s_{\mathcal{F}}) \cap H_{\langle M \rangle} \in \mathbb{A}_{M,\mathcal{F}}$ and $adh(s_{\mathcal{G}}) \cap H_{\langle M \rangle} \in \mathbb{A}_{M,\mathcal{G}}$ by Proposition 7.3. ■

When considering what groups can do in a field, we sometime want to identify such doings with sets of outcomes bordering the field rather than sets of histories passing through the field. In such cases, distinguishability may also be applied to outcomes bordering the field. Let $\mathcal{G}$ be any group. For each point x, outcomes H and H' are *distinguishable for $\mathcal{G}$ at x* if there are distinct $K, K' \in \textit{Choice}^x_{\mathcal{G}}$ such that $H \subseteq K$ and $H' \subseteq K'$. Let M be any properly covered field, and let $H, H' \in \textit{OutcmBdr}_M$. H and H' are *distinguishable for $\mathcal{G}$ in M* if they are distinguishable for $\mathcal{G}$ at a point in M, and are *indistinguishable for $\mathcal{G}$ in M* otherwise. Note that for H and H' to be distinguishable for $\mathcal{G}$ at x, they must be both available as possible future outcomes relative to x. Let $U \subseteq \textit{OutcmBdr}_M$. U is *$\mathcal{G}$-indistinguishable in M* if all members of U are pairwise indistinguishable for $\mathcal{G}$ in M. U is a *maximal $\mathcal{G}$-indistinguishable set* (*of outcomes*) *bordering M* (a *$\mathcal{G}$-MIS bordering M*) if U is $\mathcal{G}$-indistinguishable in M but no proper extension of U in $\textit{OutcmBdr}_M$ is. We may drop the phrases "in M" and "bordering M" when M is clear in the context. For each group $\mathcal{G}$, let $\mathbb{C}_{M,\mathcal{G}}$ be the set of all $\mathcal{G}$-MISs bordering M.

The follow proposition proves useful in our upcoming discussions.

Proposition 7.5. Let M be a properly covered field, let $\mathcal{G}$ be a group, and let $U, U' \subseteq \textit{OutcmBdr}_M$ and $H, H' \in \textit{OutcmBdr}_M$ with $h \in H$ and $h' \in H'$. Then the following hold:

(i) H and H' are distinguishable for $\mathcal{G}$ in M iff h and h' are distinguishable for $\mathcal{G}$ in M;
(ii) $U \in \mathbb{C}_{M,\mathcal{G}}$ iff $\bigcup U \in \mathbb{A}_{M,\mathcal{G}}$;
(iii) if $U \in \mathbb{C}_{M,\mathcal{G}}$, then $\bigcup U' \subseteq \bigcup U$ iff $U' \subseteq U$.

Proof. (i) holds by definition and Fact 3.2. (ii) Suppose first that $U \in \mathbb{C}_{M,\mathcal{G}}$. Then $\bigcup U$ is $\mathcal{G}$-indistinguishable by (i). To show that no proper extension of $\bigcup U$ in $H_{\langle M \rangle}$ is $\mathcal{G}$-indistinguishable, consider any $h \in H_{\langle M \rangle} - \bigcup U$. By Propositions 3.3–3.4, we let $H = \textit{OutcmBdr}_M(h)$. Since $h \in H$ and $h \notin \bigcup U$, $H \notin U$, and then, since $U \in \mathbb{C}_{M,\mathcal{G}}$, there is a $H' \in U$ such that H and H' are distinguishable for $\mathcal{G}$, and hence, letting $h' \in H'$, h and h' are by (i) distinguishable for $\mathcal{G}$. Hence $\bigcup U \in \mathbb{A}_{M,\mathcal{G}}$. Next suppose that $\bigcup U \in \mathbb{A}_{M,\mathcal{G}}$. Then U is $\mathcal{G}$-indistinguishable by (i). To show that no proper extension of U in $\textit{OutcmBdr}_M$ is $\mathcal{G}$-indistinguishable, consider any $H \in \textit{OutcmBdr}_M$. If $H \notin U$, Proposition 3.3 implies that $H \cap (\bigcup U) = \varnothing$, and then,

letting $h \in H$ and $h' \in \bigcup U$, h and h' are distinguishable for $\mathcal{G}$ by our supposition, and hence by (i), H and some member of U are distinguishable for $\mathcal{G}$.

(iii) Suppose that $\bigcup U' \subseteq \bigcup U$ with $U \in \mathbb{C}_{M,\mathcal{G}}$. Then $\bigcup U \in \mathbb{A}_{M,\mathcal{G}}$ by (ii). Now suppose for reductio that $H \in U' - U$. Then there is an $H' \in U$ such that H and H' are distinguishable for $\mathcal{G}$ in M. Letting $h \in H \subseteq \bigcup U'$ and $h' \in H' \subseteq \bigcup U$, we know by (i) that h and h' are distinguishable for $\mathcal{G}$ in M, and hence $h \notin \bigcup U$, contrary to that $\bigcup U' \subseteq \bigcup U$. Hence $U' \subseteq U$. ■

Similar to Proposition 7.1 and Fact 7.2, we have the following:

Proposition 7.6. Let M be any field, let $\mathcal{F}$ and $\mathcal{G}$ be any groups such that $\mathcal{F} \subseteq \mathcal{G}$, and let H, $H' \in \mathit{OutcmBdr}_M$ and $U \subseteq \mathit{OutcmBdr}_M$. Then

(i) if H and H' are distinguishable for $\mathcal{F}$, so are they for $\mathcal{G}$; and
(ii) if U is $\mathcal{G}$-indistinguishable, it is $\mathcal{F}$-indistinguishable.

Proof. Apply Propositions 7.1 and 7.5(i). ■

Fact 7.7. (AC). For each properly covered field M and each group $\mathcal{G}$, $\mathbb{C}_{M,\mathcal{G}}$ is a classification of $\mathit{OutcmBdr}_M$.

The following proposition shows that a $\mathcal{G}$-MIS bordering M is nothing but the set $\mathit{ado}_M(s)$ for a complete primary strategy s for $\mathcal{G}$ in M.

Proposition 7.8. Let $\mathcal{G}$ be any group, let M be any properly covered field, and let $U \subseteq \mathit{OutcmBdr}_M$. Then U is a $\mathcal{G}$-MIS bordering M iff $U = \mathit{ado}_M(s)$ for an $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$. Hence $\mathbb{C}_{M,\mathcal{G}} = \{\mathit{ado}_M(s) : s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}\}$.

Proof. Let $U \subseteq \mathit{OutcmBdr}_M$ and $H = \bigcup U$. Then that $U \in \mathbb{C}_{M,\mathcal{G}}$ is equivalent to each of the following:

- $H \in \mathbb{A}_{M,\mathcal{G}}$, Proposition 7.5(ii)
- $H = \mathit{adh}(s) \cap H_{\langle M \rangle}$ for an $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$, Proposition 7.3
- $H = \bigcup \mathit{ado}_M(s)$ for an $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$, Proposition 4.6
- $U = \mathit{ado}_M(s)$ for an $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$. Proposition 3.3

Hence the conclusion holds. ■

The following is a simple consequence of Propositions 7.3 and 7.8.

Proposition 7.9. Let M be a properly covered field, let $\mathcal{G}$ be a group, and let $H \subseteq H_{\langle M \rangle}$. Then $H \in \mathbb{A}_{M,\mathcal{G}}$ iff $H = \bigcup U$ for a $U \in \mathbb{C}_{M,\mathcal{G}}$.

Proof. $H \in \mathbb{A}_{M,\mathcal{G}}$ iff (by Proposition 7.3) $H = \mathit{adh}(s) \cap H_{\langle M \rangle}$ for an $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$ iff (by Proposition 4.6) $H = \bigcup \mathit{ado}_M(s)$ for an $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$ iff (by Proposition 7.8) $H = \bigcup U$ for a $U \in \mathbb{C}_{M,\mathcal{G}}$. ■

The idea in the following proposition is similar to that in Proposition 7.4, but with respect to future outcomes rather than histories.

Proposition 7.10. (AC). Let M be any properly covered field, and let $\mathcal{F}$ and $\mathcal{G}$ be disjoint groups. Then $\mathbb{C}_{M,\mathcal{F}\cup\mathcal{G}} = \{U' \cap U'' : U' \in \mathbb{C}_{M,\mathcal{F}} \wedge U'' \in \mathbb{C}_{M,\mathcal{G}}\}$.

Proof. Let $U \subseteq \mathit{OutcmBdr}_M$ and $H = \bigcup U$. Then $U \in \mathbb{C}_{M,\mathcal{F}\cup\mathcal{G}}$ iff (by Proposition 7.5(ii)) $H \in \mathbb{A}_{M,\mathcal{F}\cup\mathcal{G}}$ iff (by Proposition 7.4) $H = H' \cap H''$ for an $H' \in \mathbb{A}_{M,\mathcal{F}}$ and an $H'' \in \mathbb{A}_{M,\mathcal{G}}$ iff (by Propositions 7.9 and 7.5(ii))

$$H = (\bigcup U') \cap (\bigcup U'') \text{ for a } U' \in \mathbb{C}_{M,\mathcal{F}} \text{ and a } U'' \in \mathbb{C}_{M,\mathcal{G}}. \tag{10}$$

It is easy to verify by Proposition 3.3 that $(\bigcup U') \cap (\bigcup U'') = \bigcup (U' \cap U'')$, and then Proposition 3.3 again, (10) holds iff $U = U' \cap U''$ for a $U' \in \mathbb{C}_{M,\mathcal{F}}$ and a $U'' \in \mathbb{C}_{M,\mathcal{G}}$. ■

8 Inactivity and Busyness

Before we move on to the notion of independence, we present a short discussion of inactivity and "busyness" in a field. This is because, as it turns out, the inactivity of a group plays a special role, often behind the curtain, in our discussion of independence, while the absence of "backward busyness" allows us to have a characterization of independence in terms of a set-theoretical relation between groups.

Let $\mathcal{G}$ be any group, x any point, X any set of points and h any history. $\mathcal{G}$ is *active at* x, or x is a (*real*) *choice point for* $\mathcal{G}$ (α), if $\mathit{Choice}^x_{\mathcal{G}} \neq \{H_x\}$. $\mathcal{G}$ is *inactive at* x if $\mathit{Choice}^x_{\mathcal{G}} = \{H_x\}$, is *inactive in* X if it is inactive at each $x \in X$, and is *inactive along* h *in* X if it is inactive in $h \cap X$. We say that an agent α is *active/inactive at* x, *in* X, or *along* h *in* X, if $\{\alpha\}$ is so. It is easy to see that the empty group is always inactive in every field, and that if $\mathcal{G}$ is inactive at x, in X, or along h in X, then so is every sub-group of $\mathcal{G}$. Furthermore we have the following list of simple facts concerning inactivity.

Fact 8.1. Let M be any field, let $\mathcal{G}$ be any group, and let $s \in \mathit{CP\text{-}Strategy}^M_{\mathcal{G}}$. Then the following hold:

(i) if $\mathcal{G}$ is inactive in M, then $\mathit{CP\text{-}Strategy}^M_{\mathcal{G}} = \{s\}$, $\mathit{dom}(s) = M$ and $\mathit{adh}(s)$ is the set of all histories;
(ii) if M is properly covered and $\mathcal{G}$ is inactive in M, then $\mathit{ado}_M(s) = \mathit{OutcmBdr}_M$;
(iii) if $\mathcal{G}$ is inactive in M and s' is a strategy in M, $\mathit{adh}(s') \subseteq \mathit{adh}(s)$;
(iv) if $\mathcal{G}$ is inactive in $\mathit{dom}(s)$, then $\mathit{dom}(s) = M$ (and hence $\mathcal{G}$ is inactive in M). (AC)

Proof. We show only (iv). Suppose that $\mathcal{G}$ is inactive in $\mathit{dom}(s)$. Consider any $y \in M$. By the Axiom of Choice, $y \in h$ for a history h. Since $s(x) = H_x$ for each $x \in \mathit{dom}(s)$, $h \in s(x)$ for each $x \in h \cap \mathit{dom}(s)$, i.e., $h \in \mathit{adh}(s)$. Because s is complete in M, $h \cap M \subseteq \mathit{dom}(s)$, and hence $y \in \mathit{dom}(s)$. ■

By definition, $\bigcap\mathbb{A}_{M,\mathcal{G}}$ is the set of histories in $H_{\langle M\rangle}$ each of which is indistinguishable for $\mathcal{G}$ from all histories in $H_{\langle M\rangle}$. The inactivity of a group $\mathcal{G}$ along a history h in M, as it turns out, amounts to the indistinguishability for $\mathcal{G}$ in M between h and all other histories in $H_{\langle M\rangle}$. In other words, $\bigcap\mathbb{A}_{M,\mathcal{G}} = \{h \in H_{\langle M\rangle} : \mathcal{G}$ is inactive in $h \cap M\}$, as shown below.

Proposition 8.2. Let $\mathcal{G}$ be any group, let M be any field, and let $h \in H_{\langle M\rangle}$. Then $\mathcal{G}$ is inactive in $h \cap M$ iff $h \in \bigcap\mathbb{A}_{M,\mathcal{G}}$, and consequently $\mathcal{G}$ is inactive in $(\bigcup\bigcap\mathbb{A}_{M,\mathcal{G}}) \cap M$.

Proof. Suppose that $\mathcal{G}$ is inactive in $h \cap M$. Consider any $h' \in H_{\langle M\rangle}$ and any $x \in M$ such that $h, h' \in H_x$. Since $x \in h \cap M$, $Choice^x_{\mathcal{G}} = \{H_x\}$, and hence $Choice^x_{\mathcal{G}}(h) = Choice^x_{\mathcal{G}}(h')$. It then follows that h and h' are indistinguishable for $\mathcal{G}$ in M for each $h' \in H_{\langle M\rangle}$, and hence $h \in \bigcap\mathbb{A}_{M,\mathcal{G}}$.

Suppose next that $h \in \bigcap\mathbb{A}_{M,\mathcal{G}}$. If $Choice^x_{\mathcal{G}} \neq \{H_x\}$ for an $x \in h \cap M$, $Choice^x_{\mathcal{G}}(h) \neq K$ for some $K \in Choice^x_{\mathcal{G}}$, and, since $h' \in K$ for some $h' \in H_{\langle M\rangle}$, h and h' are distinguishable for $\mathcal{G}$ in M, contrary to the supposition that $h \in \bigcap\mathbb{A}_{M,\mathcal{G}}$. Hence $\mathcal{G}$ is inactive in $h \cap M$. ■

We know by Proposition 7.3 that $\bigcap\mathbb{A}_{M,\mathcal{G}} \subseteq adh(s) \cap H_{\langle M\rangle}$ for every $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ (since by definition, $\bigcap\mathbb{A}_{M,\mathcal{G}} \subseteq H$ for every $H \in \mathbb{A}_{M,\mathcal{G}}$). Now we also know by Proposition 8.2 that for each $h \in H_{\langle M\rangle}$, $\mathcal{G}$ is inactive along h in M iff $h \in \bigcap\{adh(s) \cap H_{\langle M\rangle} : s \in CP\text{-}Strategy^M_{\mathcal{G}}\}$. That is to say, all strategies in $CP\text{-}Strategy^M_{\mathcal{G}}$ "overlap" with exactly those histories in $H_{\langle M\rangle}$ that $\mathcal{G}$ is inactive along in M. We can similarly show the following.

Proposition 8.3. Let $\mathcal{G}$ be any group, let M be any properly covered field, and let $h \in H_{\langle M\rangle}$. Then $\mathcal{G}$ is inactive in $h \cap M$ iff $OutcmBdr_M(h) \in \bigcap\mathbb{C}_{M,\mathcal{G}}$, and consequently $\mathcal{G}$ is inactive in $(\bigcup\bigcup\bigcap\mathbb{C}_{M,\mathcal{G}}) \cap M$.

Note that $\bigcap\mathbb{A}_{M,\mathcal{G}}$ cannot be a member of $\mathbb{A}_{M,\mathcal{G}}$ if $\mathbb{A}_{M,\mathcal{G}}$ is a non-trivial classification of $H_{\langle M\rangle}$, as we noted earlier in Sect. 7. Similarly, $\bigcap\mathbb{C}_{M,\mathcal{G}}$ cannot be a member of $\mathbb{C}_{M,\mathcal{G}}$ if $\mathbb{C}_{M,\mathcal{G}}$ is a non-trivial classification of $OutcmBdr_M$.

For each field M, a group $\mathcal{G}$ (or an agent α) is *sooner or later active in M* (*SOL active in M*) if for each $h \in H_{\langle M\rangle}$, there is a choice point $x \in h \cap M$ for $\mathcal{G}$ (α), i.e., $\mathcal{G}$ is not inactive in $h \cap M$. Note that a group can be SOL active in M when some or even all proper sub-groups of it are not. The following is a consequence of Propositions 8.2–8.3.

Corollary 8.4. For each field M and each group $\mathcal{G}$, if $\mathcal{G}$ is SOL active in M, then $\bigcap\mathbb{A}_{M,\mathcal{G}} = \bigcap\mathbb{C}_{M,\mathcal{G}} = \varnothing$.

Now consider a complete primary strategy s' for $\mathcal{F}$ in a field M. It is quite possible that $\mathcal{F}$ is inactive in M, and then by Fact 8.1, $adh(s) \subseteq adh(s')$ for any strategy s in M for any group $\mathcal{G}$. In our discussion of independence, we need to deal with situations where $adh(s) \subseteq adh(s')$ holds somehow for an $s' \in CP\text{-}Strategy^M_{\mathcal{F}}$ and an $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ with $\mathcal{F} \cap \mathcal{G} = \varnothing$. What is a sufficient and necessary condition for $adh(s) \subseteq adh(s')$ to hold? It turns out thatthe inactivity of $\mathcal{F}$ in M is not, but the

inactivity of $\mathcal{F}$ in $dom(s)$ is, such a sufficient and necessary condition, as we show below.

Proposition 8.5. Let M be any field, let $\mathcal{G}$ and $\mathcal{F}$ be disjoint groups, and let $s \in CP\text{-}Strategy^{M}_{\mathcal{G}}$ and $s' \in CP\text{-}Strategy^{M}_{\mathcal{F}}$. Then (i) $\mathcal{F}$ is inactive in $dom(s)$ iff (ii) $adh(s) \subseteq adh(s')$ iff (iii) $adh(s) \cap H_{\langle M \rangle} \subseteq adh(s')$.

Proof. Suppose that (i) holds. If $h \in adh(s) - adh(s')$, $h \notin s'(x)$ for some $x \in h \cap dom(s')$, and then $x \in dom(s)$ because $x \in h \cap M$ and s is complete along h in M, and hence $h \in H_x = s'(x)$ by (i), a contradiction. It then follows that (ii) holds, which clearly implies (iii).

Suppose that (iii) holds. We prove (i) below. We first show that

$$dom(s) \subseteq dom(s'). \tag{11}$$

Consider any $x \in dom(s)$. Because s is primary, $h \in s(x)$ for an $h \in adh(s)$, and then $h \in adh(s) \cap H_{\langle M \rangle}$, and hence $h \in adh(s')$ by (iii). Since s' is complete along h, $x \in dom(s')$. Hence (11) holds. Now suppose for reductio that $\mathcal{F}$ is not inactive in $dom(s)$. Then by (11), $s'(y) \neq K$ for some $y \in dom(s)$ and $K \in Choice^{y}_{\mathcal{F}}$, and hence

$$K \cap adh(s') = \varnothing. \tag{12}$$

Because $\mathcal{F} \cap \mathcal{G} = \varnothing$, there is by IA an $h' \in s(y) \cap K$. We show below that

$$h'' \in s(y) \cap K \text{ for an } h'' \in adh(s). \tag{13}$$

Assume that $h' \notin adh(s)$ (or there is nothing more to show). Then $h' \notin s(z)$ for a $z \in h' \cap dom(s)$. Because $y, z \in h'$, $z \leqslant y$ or $y < z$. We claim that

$$y < z. \tag{14}$$

Since $h' \in s(y)$ and $h' \notin s(z)$, $y \neq z$. Because s is primary, there is an $h^* \in s(y) \cap adh(s)$, and then $y \in h' \cap h^*$, and hence by NC and $h' \notin s(z)$, $z < y$ only if $h^* \notin s(z)$, contrary to that $h^* \in adh(s)$. It follows that (14) holds. Since s is primary, there is an $h'' \in s(z) \cap adh(s)$ and $z \in h' \cap h''$. Because $h' \in s(y) \cap K$, it follows from NC and (14) that $h'' \in s(y) \cap K$, which completes the proof of (13). But $K \subseteq H_y \subseteq H_{\langle M \rangle}$, by which (13) implies that $\varnothing \neq K \cap adh(s) = K \cap adh(s) \cap H_{\langle M \rangle}$, and then by (iii), $K \cap adh(s') \neq \varnothing$, contrary to (12). We then conclude from this reductio that (i) holds. ■

Note that clause (i) in the conclusion of Proposition 8.5 depends on no particular strategies for $\mathcal{F}$ in M. We then have the following as a direct consequence of Proposition 8.5.

Corollary 8.6. Let M be any field, let $\mathcal{G}$ and $\mathcal{F}$ be disjoint groups, and let $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ and $s^* \in CP\text{-}Strategy^M_{\mathcal{F}}$ such that $adh(s) \cap H_{\langle M\rangle} \subseteq adh(s^*)$. Then $adh(s) \subseteq adh(s')$ for every $s' \in CP\text{-}Strategy^M_{\mathcal{F}}$.

Proof. By hypothesis and Proposition 8.5, $\mathcal{F}$ is inactive in $dom(s)$, and then for each $s' \in CP\text{-}Strategy^M_{\mathcal{F}}$, $adh(s) \subseteq adh(s')$ by Proposition 8.5 again. ■

Recall that for each $s \in P\text{-}Strategy^M_{\mathcal{G}}$, $adh(s) \cap H_{\langle M\rangle}$ is by definition never empty.

Corollary 8.7. Let M be any field, let $\mathcal{G}$ and $\mathcal{F}$ be disjoint groups, and let $\mathcal{F}$ be SOL active in M. Then for each $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ and each $H \in \mathbb{A}_{M,\mathcal{F}}$, $adh(s) \cap H_{\langle M\rangle} \nsubseteq H$.

Proof. Let $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ and $H \in \mathbb{A}_{M,\mathcal{F}}$. By Proposition 7.3, $H = adh(s') \cap H_{\langle M\rangle}$ for an $s' \in CP\text{-}Strategy^M_{\mathcal{F}}$. Suppose for reductio that $adh(s) \cap H_{\langle M\rangle} \subseteq H \subseteq adh(s')$. Then by Corollary 8.6, $adh(s) \subseteq adh(s'')$, and then $adh(s) \cap H_{\langle M\rangle} \subseteq adh(s'') \cap H_{\langle M\rangle}$, for each $s'' \in CP\text{-}Strategy^M_{\mathcal{F}}$, and hence $\varnothing \neq adh(s) \cap H_{\langle M\rangle} \subseteq \bigcap \mathbb{A}_{M,\mathcal{F}}$ by Proposition 7.3. This is impossible because $\mathcal{F}$ is SOL active in M, and hence $\bigcap \mathbb{A}_{M,\mathcal{F}} = \varnothing$ by Corollary 8.4. ■

A group $\mathcal{G}$ (or an agent α) is a *busy chooser* if there is an infinite chain of choice points for $\mathcal{G}$ (α) that is both upper- and lower-bounded.[19] The kind of busyness relevant to our current work is "backward busyness". $\mathcal{G}$ (α) is *backward busy in M* if there is a lower-bounded infinite chain c in M satisfying that for each $x \in c$, there is a $y \in c$ such that $y < x$ and $Choice^y_{\mathcal{G}} \neq \{H_y\}$ ($Choice^y_{\alpha} \neq \{H_y\}$). Note that when a group is infinite, the busyness of the group does not imply the busyness of any of its members or sub-groups, but if a group is *not* busy, neither is any of its members or sub-groups.

We know that different strategies in M for the same group $\mathcal{G}$ may "overlap" in the sense of sharing some admitted histories in $H_{\langle M\rangle}$, and we do not know whether each such strategy is "disjoint" with at least one other such strategy. When $\mathcal{G}$ is SOL active but not backward busy in M, nevertheless, the existence of such a "disjoint" strategy is guaranteed for each complete primary strategy for $\mathcal{G}$ in M, as the following proposition shows.

Proposition 8.8. (AC). Let M be any field, in which $\mathcal{G}$ is SOL active but not backward busy, and let $s \in CP\text{-}Strategy^M_{\mathcal{G}}$. Then there is an $s' \in CP\text{-}Strategy^M_{\mathcal{G}}$ such that $adh(s) \cap adh(s') \cap H_{\langle M\rangle} = \varnothing$.[20]

Proof. Let D be the set of all minimal choice points for $\mathcal{G}$ in $dom(s)$, i.e., the set of all choice points $x \in dom(s)$ for $\mathcal{G}$ such that $y < x$ for no choice point $y \in dom(s)$

[19] Busy choosers and busy choice sequences play a special role in various conceptual analyses of agency and technical developments, especially when achievement stit and strategies are involved. See, e.g., Belnap et al. (2001) and Xu (1995).

[20] The hypothesis that $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ and $\mathcal{G}$ is SOL active but not backward busy in M can be weakened to that $\mathcal{G}$ is any group and $s \in CP\text{-}Strategy^M_{\mathcal{G}}$ such that for each $h \in adh(s) \cap H_{\langle M\rangle}$, there is a least choice point in $h \cap M$ for $\mathcal{G}$.

for $\mathcal{G}$. For each $x \in D$, select a $K_x \in Choice^x_{\mathcal{G}}$ such that $K_x \neq s(x)$. Let s^* be a function on $dom(s^*) = \{y \in dom(s) : \exists x \in D(y \leqslant x)\}$ such that $s^*(x) = K_x$ for each $x \in D$, and $s^*(y) = s(y)$ for each $y \in dom(s^*) - D$. Then s^* is a strategy for $\mathcal{G}$ that is backward closed in M. Note that by definition, $\mathcal{G}$ is inactive in $dom(s^*) - D$, from which it follows that

$$s^*(x) \subseteq adh(s^*) \text{ for each } x \in D. \tag{15}$$

For each $y \in dom(s^*)$, if $y \in D$, $y \in adm(s^*)$ by (15); and if $y < x$ for an $x \in D$, $s^*(x) \subseteq H_y = s^*(y)$, and hence $y \in adm(s^*)$ by (15). It follows that s^* is simple in M. Consider any $h \in adh(s) \cap H_{\langle M \rangle}$. Because s is complete along h in M, there is a choice point $z \in h \cap dom(s^*)$ such that $s^*(z) \neq s(z)$, and then $h \notin s^*(z)$, and hence $h \notin adh(s^*)$ by definition. It then follows that $adh(s) \cap adh(s^*) \cap H_{\langle M \rangle} = \varnothing$, and then by Proposition 5.7, we can extend s^* to a complete primary strategy s' for $\mathcal{G}$ in M, and hence $adh(s) \cap adh(s') \cap H_{\langle M \rangle} = \varnothing$. ■

9 Independence

Recall the condition IA: for each moment m, $\bigcap_{\alpha \in Agent} f(\alpha) \neq \varnothing$ for each $f \in Select_m$, where $Select_m$ is the set of all functions each of which assigns each agent α a member of $Choice^m_\alpha$. This is equivalent to the statement that for each moment m, and for all disjoint groups $\mathcal{G}$ and $\mathcal{F}$, $K \cap K' \neq \varnothing$ for all $K \in Choice^m_{\mathcal{G}}$ and $K' \in Choice^m_{\mathcal{F}}$. In our current framework, what each group can do at a moment m are presented as the choices for the group at m, which are taken in a good sense to be causally independent of what others can do at m. Under the condition IA, nothing $\mathcal{G}$ can do at m may "rule out" anything that $\mathcal{F}$ can do at m, where $\mathcal{F}$ and $\mathcal{G}$ are disjoint, nor may anything $\mathcal{G}$ can do at m "force" $\mathcal{F}$ to do one thing at m rather than another. In other words, there is no $K \in Choice^m_{\mathcal{G}}$ such that $K \cap K' = \varnothing$ for any $K' \in Choice^m_{\mathcal{F}}$, nor is there any $K \in Choice^m_{\mathcal{G}}$ such that $K \subseteq K'$ for any $K' \in Choice^m_{\mathcal{F}}$ if $Choice^m_{\mathcal{F}}$ is not a singleton. In general, we may say that for a given partition $\mathbb{A}$ of H_m, $\mathbb{A}$ is independent of what $\mathcal{G}$ can do at m iff

$$\text{for each } K \in Choice^m_{\mathcal{G}},\ K \cap H \neq \varnothing \text{ for each } H \in \mathbb{A}. \tag{16}$$

This notion of independence is a fundamental notion in the decision-theoretical approach to deontic logic, based on which Horty builds his theory of dominance between choices at a point (Horty 2001): A choice K for $\mathcal{G}$ at m dominates another, K', if for a partition $\mathbb{A}$ of H_m, independent of what $\mathcal{G}$ can do at m, K is "better than" K' under each condition presented as a member of $\mathbb{A}$. The partition $\mathbb{A}$ of H_m that Horty uses is $Choice^m_{\mathcal{G}}$, which is, as we said above, independent of what $\mathcal{G}$ can do at m.

As stated earlier, we want to take the decision-theoretical approach to deontic logic to go beyond single-choice-point situations. To that end, we need a notion of

independence more general than (16) above, based on which we can build a more general notion of dominance. In this section we provide a preliminary analysis of independence on our current setting.

In our previous discussion of what groups can do through a field M, we have identified them as sets of histories that are indistinguishable for the groups in M. It would be natural to continue applying such identification in our discussion concerning what one group can do through M being independent of what another group can do through M. In the context of deontic logic, nevertheless, it is more convenient to talk about certain background conditions to be independent of certain strategies, rather than being independent of certain sets of histories. So we will identify what groups can do through M with strategies for the groups in M, and speak of a set of strategies of which a classification of $H_{\langle M\rangle}$ is independent.[21]

A field M is like a big "point", and a set S of strategies for $\mathcal{G}$ in M is like a set of choices for $\mathcal{G}$ at a point. One might then be attempted to define a classification $\mathbb{A}$ of $H_{\langle M\rangle}$ to be independent of S the same way as (16) for a partition of H_m to be independent of $Choice^m_{\mathcal{G}}$: $\mathbb{A}$ is independent of S iff

$$\text{for each } s \in S,\ adh(s) \cap H \neq \varnothing \text{ for each } H \in \mathbb{A}. \tag{17}$$

This won't do, nevertheless. Suppose that there is a history h passing through M and that $\mathcal{G}$ is a group inactive along h in M, which is quite possible. Then, as a consequence of Proposition 8.2, we would have that $\bigcap \mathbb{A}_{M,\mathcal{G}} \neq \varnothing$, and hence for each $s \in \textit{CP-Strategy}^M_{\mathcal{G}}$, $adh(s) \cap H \neq \varnothing$ for each $H \in \mathbb{A}_{M,\mathcal{G}}$. Hence, if we define independence as (17), $\mathbb{A}_{M,\mathcal{G}}$ would be independent of $\textit{CP-Strategy}^M_{\mathcal{G}}$, which is counter-intuitive.[22]

Let s be a strategy in M and let $H \subseteq H_{\langle M\rangle}$. We say that s *guarantees* H if $adh(s) \cap H_{\langle M\rangle} \subseteq H$, and that s *excludes* H if $adh(s) \cap H = \varnothing$. One may also be tempted to define a classification $\mathbb{A}$ of $H_{\langle M\rangle}$ to be independent of a set S of strategies just in case

$$\text{for each } s \in S \text{ and } H \in \mathbb{A},\ s \text{ neither guarantees nor excludes } H. \tag{18}$$

Even though this suggested account is intuitive and simple, a little reflection shows that it would work only in more restricted cases. For example, the trivial classification $\{H_{\langle M\rangle}\}$ of $H_{\langle M\rangle}$ should be taken to be independent of any set of strategies, but it is not so according to (18) because all strategies guarantee $H_{\langle M\rangle}$. The situation becomes

[21] Although identifying what $\mathcal{G}$ can do through M with a strategy s for $\mathcal{G}$ in M is different from identifying it with $ads(s) \cap H_{\langle M\rangle}$ or with $ado_M(s)$, the differences are only technical, not conceptual—they arise from different ways of talking about the same thing.

[22] This becomes clearer if we notice that what $\mathcal{G}$ can do through M can be identified with $\textit{CP-Strategy}^M_{\mathcal{G}}$ as well as $\mathbb{A}_{M,\mathcal{G}}$. Under the circumstance described in the main text above, if we were to define independence as (17), what $\mathcal{G}$ can do through M would be independent of what $\mathcal{G}$ can do through M.

more complicated once we take into consideration that some groups may be inactive in a proper subset of M.

The intuitive idea in our notion of independence is this, which is a slight generalization of (18): Given a classification $\mathbb{A}$ of $H_{\langle M\rangle}$ and a set S of strategies. $\mathbb{A}$ is independent of S if no strategy in S may exclude any member of $\mathbb{A}$, nor may any such strategy guarantee a member of $\mathbb{A}$ without guaranteeing all members of $\mathbb{A}$.

Definition 9.1. Let M be any field, let $\mathbb{A}$ be any classification of $H_{\langle M\rangle}$, and let S be any set of strategies in M. $\mathbb{A}$ is *independent of* S if the following hold:

(i) for each $s \in S$ and each $H \in \mathbb{A}$, $adh(s) \cap H \neq \varnothing$, and
(ii) for each $s \in S$ and each $H \in \mathbb{A}$, $adh(s) \cap H_{\langle M\rangle} \subseteq H$ only if $adh(s) \cap H_{\langle M\rangle} \subseteq \bigcap \mathbb{A}$.

It is easy to verify that in the context above, if $\mathbb{A}$ is independent of S, so is each subset of $\mathbb{A}$ (as long as it is still a classification of $H_{\langle M\rangle}$), which in turn is independent of each subset of S. Definition 9.1(ii) may appear wrong because it allows a strategy s in S to guarantee a member H of $\mathbb{A}$, but actually, it allows s to guarantee H only when s guarantees all members of $\mathbb{A}$. Note that if $\mathbb{A}$ is a partition of $H_{\langle M\rangle}$ (not just a classification of $H_{\langle M\rangle}$), then Definition 9.1 (ii) amounts to that for each $s \in S$, $adh(s) \cap H_{\langle M\rangle} \subseteq H$ only if $H = H_{\langle M\rangle}$ (i.e., only if $\mathbb{A}$ is the trivial partition $\{H_{\langle M\rangle}\}$). Note also that if $\mathbb{A}$ is a non-trivial partition of $H_{\langle M\rangle}$, it is then independent of S iff for each $H \in \mathbb{A}$ and each $s \in S$, neither $adh(s) \cap H = \varnothing$ nor $adh(s) \cap H_{\langle M\rangle} \subseteq H$. That is to say, this account of independence is a generalization of the account (18) suggested above, and the two accounts work exactly the same if we restrict classifications of $H_{\langle M\rangle}$ to non-trivial partitions of $H_{\langle M\rangle}$. A similar remark can be made about the following easy consequence of Corollary 8.4 (compare it to (18)):

Corollary 9.2. Let M be a field, and let $\mathcal{F}$ be SOL active in M. Then for each set S of strategies, $\mathbb{A}_{M,\mathcal{F}}$ is independent of S iff for each $s \in S$ and each $H \in \mathbb{A}_{M,\mathcal{F}}$, $adh(s) \cap H \neq \varnothing$ and $adh(s) \cap H_{\langle M\rangle} \nsubseteq H$.

Proof. We only need to assume that $\mathbb{A}_{M,\mathcal{F}}$ is independent of S, and show that $adh(s) \cap H_{\langle M\rangle} \nsubseteq H$ for all $s \in S$ and $H \in \mathbb{A}_{M,\mathcal{F}}$. Suppose for reductio that $adh(s) \cap H_{\langle M\rangle} \subseteq H$ for an $s \in S$ and an $H \in \mathbb{A}_{M,\mathcal{F}}$. Because $\mathbb{A}_{M,\mathcal{F}}$ is independent of S, $adh(s) \cap H_{\langle M\rangle} \subseteq H'$ for all $H' \in \mathbb{A}_{M,\mathcal{F}}$, and then, since $adh(s) \cap H_{\langle M\rangle} \neq \varnothing$, $\bigcap \mathbb{A}_{M,\mathcal{F}} \neq \varnothing$, contrary to Corollary 8.4. ■

Under the condition of SOL activity, independence is "symmetrical" in the following sense.

Proposition 9.3. Let M be a field, and let $\mathcal{F}$ and $\mathcal{G}$ be disjoint groups that are SOL active in M. Then $\mathbb{A}_{M,\mathcal{F}}$ is independent of $CP\text{-}Strategy_{\mathcal{G}}^{M}$ iff $\mathbb{A}_{M,\mathcal{G}}$ is independent of $CP\text{-}Strategy_{\mathcal{F}}^{M}$.[23]

[23] Had we defined independence as a relation between classifications of $H_{\langle M\rangle}$, we would then have that for all disjoint groups $\mathcal{F}$ and $\mathcal{G}$ that are SOL active in M, $\mathbb{A}_{M,\mathcal{F}}$ is independent of $\mathbb{A}_{M,\mathcal{G}}$ iff $\mathbb{A}_{M,\mathcal{G}}$ is independent of $\mathbb{A}_{M,\mathcal{F}}$.

Proof. Assume that $\mathbb{A}_{M,\mathcal{F}}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$. Consider any $s \in$ *CP-Strategy*$^M_{\mathcal{F}}$ and $H \in \mathbb{A}_{M,\mathcal{G}}$. By Proposition 7.3, there are $s' \in$ *CP-Strategy*$^M_{\mathcal{G}}$ and $H' \in \mathbb{A}_{M,\mathcal{F}}$ such that $adh(s') \cap H_{\langle M\rangle} = H$ and $H' = adh(s) \cap H_{\langle M\rangle}$. By our assumption, $H' \cap adh(s') \neq \varnothing$, i.e., $adh(s) \cap H \neq \varnothing$; and by hypothesis and Corollary 8.7, $adh(s) \cap H_{\langle M\rangle} \nsubseteq H$. It follows from Definition 9.1 that $\mathbb{A}_{M,\mathcal{G}}$ is independent of *CP-Strategy*$^M_{\mathcal{F}}$. ■

Let M be any field, and let m be any point. m is a *starting point of* M if $m \in M$ and $m \leqslant x$ for each $x \in M$. A starting point of a field is obviously unique. It is easy to see that if a field M has a starting point, $dom(s) \cap dom(s') \neq \varnothing$ for all backward closed strategies s and s' in M. We show below that for each $\mathcal{G}$, the classification $\mathbb{A}_{M,\overline{\mathcal{G}}}$ of $H_{\langle M\rangle}$ is independent of what $\mathcal{G}$ can do through M, where we identify what $\mathcal{G}$ can do through M with either *CP-Strategy*$^M_{\mathcal{G}}$ or *S-Strategy*$^M_{\mathcal{G}}$, and in the latter case, M needs to have a starting point.

Theorem 9.4. (AC). Let M be any field, and let $\mathcal{G}$ and $\mathcal{F}$ be disjoint groups. Then the following hold:

(i) $\mathbb{A}_{M,\mathcal{F}}$ is independent of *CP -Strategy*$^M_{\mathcal{G}}$;

(ii) $\mathbb{A}_{M,\mathcal{F}}$ is independent of *S -Strategy*$^M_{\mathcal{G}}$ if M has a starting point;

(iii) $\mathbb{A}_{M,\overline{\mathcal{G}}}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$, and is independent of *S-Strategy*$^M_{\mathcal{G}}$ if M has a starting point.

Proof. (iii) follows directly from (i) and (ii). Letting $\mathbb{A} = \mathbb{A}_{M,\mathcal{F}}$, we only need to show that Definition 9.1(ii) holds with $S =$ *S-Strategy*$^M_{\mathcal{G}}$, and that Definition 9.1(i) holds with $S =$ *CP-Strategy*$^M_{\mathcal{G}}$, and with $S =$ *S-Strategy*$^M_{\mathcal{G}}$ if M has a starting point. Let $H \in \mathbb{A}$. By Proposition 7.3, $H = adh(s^*) \cap H_{\langle M\rangle}$ for an $s^* \in$ *CP-Strategy*$^M_{\mathcal{F}}$.

Consider any $s \in$ *S-Strategy*$^M_{\mathcal{G}}$, and suppose that $adh(s) \cap H_{\langle M\rangle} \subseteq H$. By Proposition 5.7, $adh(s) = \bigcup_{s' \in S'} adh(s')$ where $S' = \{s' \in$ *CP-Strategy*$^M_{\mathcal{G}}$: $s \subseteq s'\}$. Since $(\bigcup_{s' \in S'} adh(s')) \cap H_{\langle M\rangle} \subseteq adh(s^*)$, Corollary 8.6 implies that $(\bigcup_{s' \in S'} adh(s')) \cap H_{\langle M\rangle} \subseteq adh(s'')$ for each $s'' \in$ *CP-Strategy*$^M_{\mathcal{F}}$, and hence $adh(s) \cap H_{\langle M\rangle} \subseteq adh(s'') \cap H_{\langle M\rangle}$ for each $s'' \in$ *CP-Strategy*$^M_{\mathcal{F}}$. It then follows from Proposition 7.3 that $adh(s) \cap H_{\langle M\rangle} \subseteq H'$ for each $H' \in \mathbb{A}_{M,\mathcal{F}}$. Hence Definition 9.1(ii) holds.

Consider any $s \in$ *S-Strategy*$^M_{\mathcal{G}}$. If $s \in$ *CP-Strategy*$^M_{\mathcal{G}}$, $dom(s) \cap dom(s^*) \neq \varnothing$ by Fact 5.4(iii), and if $s \notin$ *CP-Strategy*$^M_{\mathcal{G}}$ and M has a starting point, we also have that $dom(s) \cap dom(s^*) \neq \varnothing$. Then by Proposition 6.2, $adh(s) \cap adh(s^*) \cap H_{\langle M\rangle} \neq \varnothing$, i.e., $adh(s) \cap H \neq \varnothing$. Hence Definition 9.1(i) holds. ■

SOL activity and the absence of backward busyness enable us to establish a characterization of independence (Theorem 9.6) in terms of a set-theoretical relation between groups. We begin with a special case.

Proposition 9.5. (AC). Let M be any field, in which $\mathcal{F}$ is not backward busy and all sub-groups of $\mathcal{F}$ are SOL active, except for the empty group. Then for each group $\mathcal{G}$,

(i) $\mathcal{F} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{A}_{M,\mathcal{F}}$ is independent of $CP\text{-}Strategy_{\mathcal{G}}^{M}$;
(ii) $\mathcal{F} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{A}_{M,\mathcal{F}}$ is independent of $S\text{-}Strategy_{\mathcal{G}}^{M}$, provided that M has a starting point.

Proof. By Theorem 9.4, we only need to assume that $\mathcal{F} \nsubseteq \overline{\mathcal{G}}$, and show that $\mathbb{A}_{M,\mathcal{F}}$ is not independent of $CP\text{-}Strategy_{\mathcal{G}}^{M}$ (and hence, not independent of $S\text{-}Strategy_{\mathcal{G}}^{M}$). By our assumption, $\mathcal{F} \neq \varnothing$. There are two cases.

Case 1, $\mathcal{F} \subseteq \mathcal{G}$. Let $s \in CP\text{-}Strategy_{\mathcal{G}}^{M}$, let $\mathcal{E} = \mathcal{G} - \mathcal{F}$, and let $s_{\mathcal{F}}$ and $s_{\mathcal{E}}$ be any complete primary extensions of $s|_{\mathcal{F}}$ and $s|_{\mathcal{E}}$ respectively (see Sect. 6). Then by Proposition 6.4, $s = s_{\mathcal{F}} \sqcap s_{\mathcal{E}}$ and $adh(s) = adh(s_{\mathcal{F}}) \cap adh(s_{\mathcal{E}})$, and hence $adh(s) \cap H_{\langle M \rangle} \subseteq adh(s_{\mathcal{F}}) \cap H_{\langle M \rangle} \in \mathbb{A}_{M,\mathcal{F}}$ by Proposition 7.3. Because $\mathcal{F}$ is SOL active in M, Corollary 9.2 implies that $\mathbb{A}_{M,\mathcal{F}}$ is not independent of $CP\text{-}Strategy_{\mathcal{G}}^{M}$.[24]

Case 2, $\mathcal{F} \nsubseteq \mathcal{G}$. Let $\mathcal{E} = \mathcal{F} \cap \mathcal{G}$ and $\mathcal{E}^* = \mathcal{F} \cap \overline{\mathcal{G}}$. Then $\mathcal{F} = \mathcal{E} \cup \mathcal{E}^*$, and $\mathcal{E} \neq \varnothing \neq \mathcal{E}^*$ because $\mathcal{F} \nsubseteq \overline{\mathcal{G}}$ and $\mathcal{F} \nsubseteq \mathcal{G}$. Let $s \in CP\text{-}Strategy_{\mathcal{G}}^{M}$. It suffices to show that $adh(s) \cap H = \varnothing$ for an $H \in \mathbb{A}_{M,\mathcal{F}}$. Letting $s_{\mathcal{E}}$ and $s_{\mathcal{G}-\mathcal{E}}$ be any complete primary extensions of $s|_{\mathcal{E}}$ and $s|_{\mathcal{G}-\mathcal{E}}$ respectively, we know by Proposition 6.4 that $s = s_{\mathcal{G}-\mathcal{E}} \sqcap s_{\mathcal{E}}$ and

$$adh(s) = adh(s_{\mathcal{G}-\mathcal{E}}) \cap adh(s_{\mathcal{E}}). \tag{19}$$

Let $s_{\mathcal{E}^*} \in CP\text{-}Strategy_{\mathcal{E}^*}^{M}$, and by Fact 5.4 (iii), let $s' = s_{\mathcal{E}} \sqcap s_{\mathcal{E}^*}$. By Proposition 6.3 (iv, viii, x), $s' \in CP\text{-}Strategy_{\mathcal{F}}^{M}$ and $adh(s') = adh(s_{\mathcal{E}}) \cap adh(s_{\mathcal{E}^*})$. Since $\varnothing \neq \mathcal{E} \subseteq \mathcal{F}$, $\mathcal{E}$ is by hypothesis SOL active but not backward busy in M, and then by Proposition 8.8, there is an $s''_{\mathcal{E}} \in CP\text{-}Strategy_{\mathcal{E}}^{M}$ such that

$$adh(s_{\mathcal{E}}) \cap adh(s''_{\mathcal{E}}) \cap H_{\langle M \rangle} = \varnothing. \tag{20}$$

Applying Fact 5.4 (iii), we let $s'' = s''_{\mathcal{E}} \sqcap s_{\mathcal{E}^*}$. Then by Proposition 6.3(iv, viii, x) again, $s'' \in CP\text{-}Strategy_{\mathcal{F}}^{M}$ and $adh(s'') = adh(s''_{\mathcal{E}}) \cap adh(s_{\mathcal{E}^*})$, and hence

$$adh(s) \cap adh(s'') \cap H_{\langle M \rangle} = \varnothing \tag{21}$$

by (19) and (20). Finally, Proposition 7.3 implies that there is an $H \in \mathbb{A}_{M,\mathcal{F}}$ such that $adh(s'') \cap H_{\langle M \rangle} = H$, and then $adh(s) \cap H = \varnothing$ by (21). ■

Now we are ready to establish a general characterization of independence in terms of a set-theoretical relation between groups.

[24] When $\mathcal{F} \subseteq \mathcal{G}$, a weaker condition, that $\mathbb{A}_{M,\mathcal{F}}$ is non-trivial, suffices for $\mathbb{A}_{M,\mathcal{F}}$ not to be independent of $CP\text{-}Strategy_{\mathcal{G}}^{M}$. In fact, we can show that $\mathbb{A}_{M,\mathcal{F}}$ is trivial if Definition 9.1 (ii) holds with $\mathbb{A} = \mathbb{A}_{M,\mathcal{F}}$ and $S = CP\text{-}Strategy_{\mathcal{F}}^{M}$: Suppose that Definition 9.1(ii) so holds. Consider any $h \in H_{\langle M \rangle}$. By Proposition 7.3, $h \in adh(s) \cap H_{\langle M \rangle} \in \mathbb{A}_{M,\mathcal{G}}$ for an $s \in CP\text{-}Strategy_{\mathcal{G}}^{M}$. The argument in the main text shows that $h \in adh(s) \cap H_{\langle M \rangle} \subseteq H$ for an $H \in \mathbb{A}_{M,\mathcal{F}}$, and then by our supposition, $h \in adh(s) \cap H_{\langle M \rangle} \subseteq \bigcap \mathbb{A}_{M,\mathcal{F}}$. It then follows that $H_{\langle M \rangle} \subseteq \bigcap \mathbb{A}_{M,\mathcal{F}}$, and hence $\mathbb{A}_{M,\mathcal{F}}$ is trivial.

Theorem 9.6. (AC). Let M be any field, in which no group is backward busy, let $\mathcal{E}$ be the group of all agents that are inactive in M, and let all other agents be SOL active in M. Then for all groups $\mathcal{F}$ and $\mathcal{G}$,

(i) $\mathcal{F} - \mathcal{E} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{A}_{M,\mathcal{F}}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$;
(ii) $\mathcal{F} - \mathcal{E} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{A}_{M,\mathcal{F}}$ is independent of *S-Strategy*$^M_{\mathcal{G}}$, provided that M has a starting point.

Proof. Let $\mathcal{F}$ and $\mathcal{G}$ be any groups, and let $\mathcal{F}^* = \mathcal{F} - \mathcal{E}$. Then all sub-groups of $\mathcal{F}^*$ are by hypothesis SOL active in M, except for $\varnothing$, and hence by Proposition 9.5, the conclusions hold with $\mathbb{A}_{M,\mathcal{F}}$ to be replaced by $\mathbb{A}_{M,\mathcal{F}^*}$. It is easy to verify that $\mathbb{A}_{M,\mathcal{F}\cap\mathcal{E}} = \{H_{\langle M\rangle}\}$, and hence $\mathbb{A}_{M,\mathcal{F}^*} = \mathbb{A}_{M,\mathcal{F}^*\cup(\mathcal{F}\cap\mathcal{E})} = \mathbb{A}_{M,\mathcal{F}}$ by Proposition 7.4. ■

When dealing with a classification of outcomes bordering a field, we may similarly define its independence of a set of strategies in the following way.

Definition 9.7. Let M be any properly covered field, let $\mathbb{C}$ be any classification of $OutcmBdr_M$, and let S be any set of strategies in M. $\mathbb{C}$ is *independent of* S if the following hold:

(i) for each $s \in S$ and each $U \in \mathbb{C}$, $ado_M(s) \cap U \neq \varnothing$, and
(ii) for each $s \in S$ and each $U \in \mathbb{C}$, $ado_M(s) \subseteq U$ only if $ado_M(s) \subseteq \bigcap\mathbb{C}$.

The idea in Definition 9.7 is clearly the same as that in Definition 9.1, except that for this new notion of independence to make sense, the strategy field needs to be properly covered. Furthermore, we have the following:

Proposition 9.8. Let S be any set of strategies in a properly covered field M, and let $\mathcal{F}$ be any group. Then (i) $\mathbb{C}_{M,\mathcal{F}}$ is independent of S iff (ii) $\mathbb{A}_{M,\mathcal{F}}$ is independent of S.

Proof. Suppose that (i) holds. Consider any $s \in S$ and $H \in \mathbb{A}_{M,\mathcal{F}}$. Then $H = \bigcup U$ for a $U \in \mathbb{C}_{M,\mathcal{F}}$ by Proposition 7.9, and then $ado_M(s) \cap U \neq \varnothing$ by (i), and hence $(\bigcup ado_M(s)) \cap (\bigcup U) \neq \varnothing$. It follows from Proposition 4.6 that $adh(s) \cap H_{\langle M\rangle} \cap H = adh(s) \cap H \neq \varnothing$. Hence Definition 9.1(i) holds with $\mathbb{A} = \mathbb{A}_{M,\mathcal{F}}$. Suppose that $adh(s) \cap H_{\langle M\rangle} \subseteq H$ $(= \bigcup U)$. Then $ado_M(s) \subseteq U$ by Propositions 4.6 and 7.5(iii), and hence by (i), $ado_M(s) \subseteq U'$ for each $U' \in \mathbb{C}_{M,\mathcal{F}}$. This implies that $\bigcup ado_M(s) \subseteq \bigcup U'$ for each $U' \in \mathbb{C}_{M,\mathcal{F}}$, and then by Propositions 4.6 and 7.9, $adh(s) \cap H_{\langle M\rangle} \subseteq \bigcap\mathbb{A}_{M,\mathcal{F}}$. Hence Definition 9.1(ii) holds with $\mathbb{A} = \mathbb{A}_{M,\mathcal{F}}$, and hence (ii) holds.

Next suppose that (ii) holds. Let $s \in S$ and $U \in \mathbb{C}_{M,\mathcal{F}}$. By Proposition 7.8, $U = ado_M(s'')$ for an $s'' \in$ *CP-Strategy*$^M_{\mathcal{F}}$, and then, letting $H = \bigcup U$, we know that $H = adh(s'') \cap H_{\langle M\rangle} \in \mathbb{A}_{M,\mathcal{F}}$ by Propositions 4.6 and 7.3. By (ii), $adh(s) \cap H \neq \varnothing$, and then $adh(s) \cap adh(s'') \cap H_{\langle M\rangle} \neq \varnothing$, and hence by Propositions 4.6 and 3.3, $\bigcup(ado_M(s) \cap ado_M(s'')) = \bigcup(ado_M(s) \cap U) \neq \varnothing$, and consequently $ado_M(s) \cap U \neq \varnothing$. It follows that Definition 9.7(i) holds with $\mathbb{C} = \mathbb{C}_{M,\mathcal{F}}$. Suppose that $ado_M(s) \subseteq U$. Then by Proposition 4.6, $adh(s) \cap H_{\langle M\rangle} \subseteq \bigcup U = H$, and hence

by (ii), $adh(s) \cap H_{\langle M \rangle} \subseteq H^*$ for each $H^* \in \mathbb{A}_{M,\mathcal{F}}$. This implies by Proposition 7.5(ii) that $adh(s) \cap H_{\langle M \rangle} \subseteq \bigcup U^*$ for each $U^* \in \mathbb{C}_{M,\mathcal{F}}$, and then by Propositions 4.6 and 7.5(iii), $ado_M(s) \subseteq \bigcap \mathbb{C}_{M,\mathcal{F}}$. Hence Definition 9.7(ii) holds with $\mathbb{C} = \mathbb{C}_{M,\mathcal{F}}$, and hence (i) holds. ■

Applying Proposition 9.8, we can easily establish the following "duals" of Theorems 9.4 and 9.6.

Theorem 9.9. (AC). Let M be any properly covered field, and let $\mathcal{G}$ and $\mathcal{F}$ be disjoint groups. Then the following hold:

(i) $\mathbb{C}_{M,\mathcal{F}}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$;
(ii) $\mathbb{C}_{M,\mathcal{F}}$ is independent of *S-Strategy*$^M_{\mathcal{G}}$ if M has a starting point;
(iii) $\mathbb{C}_{M,\overline{\mathcal{G}}}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$, and is independent of *S-Strategy*$^M_{\mathcal{G}}$ if M has a starting point.

Theorem 9.10. (AC). Let M be any field, in which no group is backward busy, let $\mathcal{E}$ be the group of all agents that are inactive in M, and let all other agents be SOL active in M. Then for all groups $\mathcal{F}$ and $\mathcal{G}$,

(i) $\mathcal{F} - \mathcal{E} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{C}_{M,\mathcal{F}}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$;
(ii) $\mathcal{F} - \mathcal{E} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{C}_{M,\mathcal{F}}$ is independent of *S-Strategy*$^M_{\mathcal{G}}$, provided that M has a starting point.

The following is an easy consequence of Theorems 9.6 and 9.10 and Definitions 9.1 and 9.7.

Corollary 9.11. (AC). Let M be any field, in which no group is backward busy, let $\mathcal{E}$ be the group of all agents that are inactive in M, and let all other agents be SOL active in M. Then for all groups $\mathcal{F}$ and $\mathcal{G}$,

(i) $\mathcal{F} - \mathcal{E} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{A}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$ for each classification $\mathbb{A}$ of $H_{\langle M \rangle}$ such that $\mathbb{A} \subseteq \mathbb{A}_{M,\mathcal{F}}$ iff $\mathbb{C}$ is independent of *CP-Strategy*$^M_{\mathcal{G}}$ for each classification $\mathbb{C}$ of $OutcmBdr_M$ such that $\mathbb{C} \subseteq \mathbb{C}_{M,\mathcal{F}}$;
(ii) if M has a starting point, then $\mathcal{F} - \mathcal{E} \subseteq \overline{\mathcal{G}}$ iff $\mathbb{A}$ is independent of *S-Strategy*$^M_{\mathcal{G}}$ for each classification $\mathbb{A}$ of $H_{\langle M \rangle}$ such that $\mathbb{A} \subseteq \mathbb{A}_{M,\mathcal{F}}$ iff $\mathbb{C}$ is independent of *S-Strategy*$^M_{\mathcal{G}}$ for each classification $\mathbb{C}$ of $OutcmBdr_M$ such that $\mathbb{C} \subseteq \mathbb{C}_{M,\mathcal{F}}$.

This completes our preliminary study on independence. In order to achieve a general notion of dominance in the current setting, we need to consider some issues involved in independence and the sure-thing principle. We leave those issues to a future study.

References

Belnap, N. 1991. Before refraining: Concepts for agency. *Erkenntnis* 34: 137–169.

Belnap, N. 1996a. Agents in branching time. In *Logic and reality: Essays in pure and applied logic, in memory of Arthur Prior*, ed. B.J. Copeland, 239–271. Oxford: Oxford University Press.

Belnap, N. 1996b. Deontic kinematics and austere strategies, ed. P. K. T. Childers and V. Svoboda, *Logica '96: Proceedings of the 10th International Symposium*, Filsosofia, 21–40.

Belnap, N., and M. Perloff. 1988. Seeing to it that: A canonical form for agentives. *Theoria* 54: 175–199.

Belnap, N., M. Perloff, and M. Xu. 2001. *Facing the future: Agents and choices in our indeterminist world*. Oxford: Oxford University Press.

Broersen, J., A. Herzig, and N. Troquard. 2006. A stit-extension of ATL, ed. M. Fisher, 69–81. In *Proceedings of Tenth European Conference on Logics in Artificial Intelligence, JELIA 06*. *Lecture Notes in Artificial Intelligence*, vol. 4160. Berlin: Springer.

Brown, M. 2008. Acting, events and actions, ed. R. van der Meyden, and L. van der Torre, 19–33. In *Deontic Logic in Computer Science: 9th International Conference, DEON 2008*, Luxembourg, Luxembourg, July 2008, Proceedings. *Lecture Notes in Artificial Intelligence*, vol. 5076. Berlin: Springer.

Horty, J.F. 1989. An alternative stit operator. Manuscript, Department of Philosophy, University of Maryland.

Horty, J.F. 2001. *Agency and deontic logic*. Oxford: Oxford University Press.

Kooi, B., and A. Tamminga. 2008. Moral conflicts between groups of agents. *Journal of Philosophical Logic* 37: 1–21.

Müller, T. 2005. On the formal structure of continuous action. In *Advances in Modal Logic*, vol. 5, ed. R. Schmidt, I. Pratt-Hartmann, M. Reynolds, and H. Wansing, 191–209. London: King's College Publications.

Prior, A. 1967.*Past, Present and Future*. Oxford: Oxford Univerity Press.

Savage, L. 1954. *The foundations of statistics*. Second revised edition. New York: Dover Publications.

Tamminga, A. 2013. Deontic logic and strategic games. *Erkenntnis* 78(1): 183–200.

Thomason, R.H. 1970. Indeterminist time and truth-value gaps. *Theoria* 36: 264–281.

Thomason, R.H. 1984. Combinations of tense and modality. In *Handbook of philosophical logic*, vol. 2, ed. D. Gabbay, and F. Guenthner, 135–165. Dordrecht: Reidel Publishing Company.

Thomason, R.H., and J.F. Horty. 1996. Nondeterministic action and dominance: Foundations for planning and qualitative decision. In *Proceedings of the Sixth International Conference on Theoretical Aspects of Rationality and Knowledge (TARK-96)*, vol. 2, 229–250. San Francisco: Morgan Kaufmann Publishers.

von Kutschera, F. 1986. Bewirken. *Erkenntnis* 24: 253–281.

Xu, M. 1995. Busy choice sequences, refraining formulas and modalities. *Studia Logica* 54: 267–301.

Xu, M. 1997. Causation in branching time (I): Transitions, events and causes. *Synthese* 112: 137–192.

Xu, M. 2010. Combinations of stit and action. *Journal of Logic, Language and Information* 19(4): 485–503.

Xu, M. 2012. Actions as events. *Journal of Philosophical Logic* 41: 765–809.

Biographical Interview

Nuel Belnap

Abstract Biographical interview with Nuel Belnap, conducted in Utrecht, June 17 and 19, 2012. Interviewer: Thomas Müller. Edited, with the help of NB, in Pittsburgh, March 2013.

1 School Days

TM: Let's start with your school days. How many children were you at home?

NB: Four children. Three older sisters.

TM: Was it tough?

NB: No, amicable. Except of course there was this sister closest to me. We would have lots of arguments… what you do when you're young.

TM: And you lived near Chicago, for the whole time you went to school?

NB: Yes, we lived in Winnetka till I went to college. And then some. My parents still had the house, and they had a room for me.

TM: What was the school system like?

NB: Preschool when you were 5, in most cases. And first grade at 6. And, I was thinking about this yesterday when someone was talking about local customs, the rule was clearly that when you were in the first and second grade, you wore short pants.

TM: Ok!

NB: And then when you got to 3rd and 4th grade, you could wear knickers. In 5th grade you could wear long pants. Firm rules. They weren't even school rules, but everybody did it!

N. Belnap (✉)
1001 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA15260, U.S.A.
e-mail: belnap@pitt.edu

T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action,*
Outstanding Contributions to Logic 2, DOI: 10.1007/978-3-319-01754-9_17,

TM: So no school uniform or anything …

NB: No.

TM: And short pants even in the winter? Or how did that go?

NB: Yes, also in winter, but you had snow suits …

TM: … that you would put over for the walk to school?

NB: Yes. We didn't live very far from the school. Certainly no more than half a mile, but I don't remember how much less it was. Very close to the elementary school.

TM: And then that's 6 years, or how long did that go on for?

NB: Well, some people went on to junior high school in 6th grade – I did. And some in 7th. Junior high school is up to 8th grade, so that was 6th, 7th and 8th.

TM: And that was still close to where you lived?

NB: It was further, but it was surely less than a mile.

TM: That was already during the war, then?

NB: Well, it didn't begin during the war.—We had war bonds. I would go around with some kind of placard on my front and back that made fun of Adolf. The schools were excellent, they were nationally renowned, and progressive. Very much influenced by John Dewey.

TM: And did that mean coed?

NB: Oh yes, all schools were coed. But that went without saying, really, for public schools in that time period.

TM: And then, after you finished junior high?

NB: I went to high school. New Trier Township High School. And all those schools were really top-drawer. It was a wealthy suburb.

TM: So it was like that already then, that the school districts have a big influence on the value of property?

NB: Yes, that's true. I guess. I wasn't very much aware of the value of property.

TM: Did you get to pick special subjects, or was that a one size fits all idea?

NB: Not at all in grade school, and I don't think we had much choice in junior high either, there might have been some elective or something. But we had kind of a standard type of selection in high school, which is to say, … there wasn't much you could do. And I don't really remember my selection principles at all.

TM: How was that with languages? Did you get to pick those yourself?

NB: Yes, you did get to pick your languages yourself. And I made a bad choice, I chose Spanish, because it was easy. And I learned what a mistake that was when I got to college and everyone learned French. I had to start over. But I learned a lot in high school. So much so that it really let me coast through about two years of college.

TM: That would mean you weren't really interested in many things that were happening in college because you'd had most of them, in a way?

NB: I guess. I don't remember my emotional structure. But I enjoyed high school.

TM: You must have been very strong at math at that time.

NB: Yes, I took a lot of math, and I got good grades and everything. But no prizes. A lot of math, and I took the standard sciences, biology, chemistry, physics. I told you what I remember about my physics: An instructor who always said: "probably actually".

TM: And any ancient languages, was that at high school too?

NB: I did take Latin, for … I think I only took it for a year. And that was hard. My languages were always very hard for me. I didn't work that hard at the Latin so I didn't learn much.

TM: Did you get any advice from school on what to pursue at college? I mean, the American college system is so much different from what I'm used to—in Germany the idea is more that you choose a subject from the start and that's what you will do your MA in.

NB: I know, it's not like that at all. The University of Illinois, like many colleges, had a program of a sort of general studies for freshmen and sophomores, and that's what I took. I didn't have to, but that's what I took.

TM: Like great books?

NB: I took great books in high school actually, and I loved it.

TM: That's a very nice idea, I think, to read great books in high school.

NB: Wrestling with Aristotle was the high point of my education, really. When I was a senior. The *Ethics*. So I sort of coasted through college as well—I was overprepared.

TM: That's how things go. So it wasn't really clear that it would be philosophy.

NB: No, I majored in philosophy for lack of anything else to do.

TM: Where there philosophy courses in high school, apart from the great books program?

NB: No.

TM: But that had some philosophy on it.

NB: Some, but … well, we read some Shakespeare, a miscellaneous collection of great books. Later on when I was in Edgewood, a suburb of Pittsburgh, I taught great books at the third grade level. That was fun. Fairy stories, and what does it mean to be "born on a lucky day". We had fun. So then I went into the service, and really had no idea what I wanted to do, so I thought I had to pick out something. So I picked out going to graduate school in philosophy, after my two years of Air Force service.

TM: Where did that take you, the service?

NB: Oh that's when I programmed for the IBM 701 computer, in Washington. Six weeks in Texas getting basic training, crawling under the machine guns. But that wasn't much. And then Washington for two years. And my college girlfriend was in Washington as well, and so we got married. It was about like that! I had no idea what school to pick for grad school.

TM: Did you apply to several?

NB: Yes I did, I don't remember which. Probably half a dozen or so, they all accepted. I picked out Yale because I had a first cousin that was in New Haven. That was what tipped the balance.

TM: It's like that, I think. Let's go back a little further, you said you had very good schools, but what about home—did you pick up an interest in books there?

NB: No, it wasn't a bookish family. I read all the time, of course …

TM: … but that was you.

NB: Yes. And my sisters did too. I don't remember too much about the older sisters, but certainly Dorothy read a lot too. But that came from the school.

TM: And the math interest, was there anything from home that would ring with that?

NB: Nothing at home at all. I joined the math club in high school.

TM: What did you get to do there?

NB: There's a presentation I remember. I went through this proof of $e^{i\pi} + 1 = 0$. And I just read a bunch of general purpose books that had that in them, I was fascinated with it.

TM: I have heard that the American mathematics education is much different from what we get in Europe, in the sense that you get introduced to proofs rather late. How was that with you? Because now you're proving things all the time.

NB: I had only had proofs in geometry, Euclid.

TM: That was done, but …

NB: … but nothing else.

TM: So how did you learn to prove things?

NB: I guess, whatever, I don't know … maybe I never did learn! I never had a course that asked for proofs.

TM: In logic then, in grad school, that's what you would do.

NB: Well, as a freshman, in the beginning logic of grad school we did proofs, of course. I meant to say the logic text I studied for my major was Cohen and Nagel. And I took a final examination of some kind, for honors in philosophy or something, and part of it was on logic. And what did they ask me …? Something about syllogisms. And I hadn't a clue! I said, "I will be happy to answer this question if you explain the terminology", which they did.

TM: And then you could, I guess.

NB: Yes.

TM: Do you remember any teachers that were important for getting you somewhere academically? From the school days, I mean.

NB: I had a lot of good teachers, but is there any that stand out? I don't know. I remember Mr. Skarda did mathematics, and he said the one thing you're never gonna remember is what fraction 83 and a third is. And that was fixed firmly in my mind. And I had a good geometry teacher, Ms. Galley.

TM: So math and geometry were different subjects?

NB: Geometry was a mathematics course, but it was a term devoted to geometry. In that sense a separate course. We never got into anything advanced, never got into even pre-calculus.

TM: Any probability theory, statistics?

NB: I guess elementary, but I don't remember too clearly. Or not at all maybe. We had algebra as freshmen, geometry as sophomore. And I had two more years of mathematics, but I can't remember exactly what they were. College algebra probably.

TM: So where did you learn set theory?

NB: Not in high school. Teaching it at Pitt.

TM: And you taught from Pat Suppes's book?

NB: I did. … External to the school system, in the service, my boss was Thomas Steel, who was head of our section or whatever it was called. And he asked me what philosophy was, and I hadn't a clue. I still don't know! So I said, ok Tom, tell me what mathematics is. And he said, well it begins with the following axiom … and then he had Quine's textbook, *Mathematical logic*, which was axiomatic. And he gave me those axioms and taught me a lot about mathematics. I enjoyed that, I spent a lot of time with Tom, when I was supposed to be working.

TM: You kept that up for a very long time.

NB: My association with Tom? Yes, I haven't seen him for several years now, and before then there was a large number of years. And I may never see him again, he lives out in the Philadelphia area somewhere. I had a trip that took me in that vicinity, so I saw him several years ago. Not likely to recur. He was really a big influence on me, and he's the one that brought me along to these international meetings that had something to do with computer languages.

TM: The trips to Europe in the '60s.

NB: He organized that for me. And I was absolutely useless on those committees. And it turns out Dana Scott was on one of the committees, and I was bowled over that he had some things to say, effortlessly. He influenced me a lot. More when we overlapped at Oxford. He gave me some stuff to read, I read his papers. I was fascinated by his lambda calculus stuff.

2 From BA at Illinois to Grad School at Yale

TM: Let's look at your university education more closely. The BA is from the University of Illinois. What was your subject at that time?

NB: I majored in philosophy, but in a desultory way, as I said—in a relaxed way, I didn't take it very seriously. I did it for lack of anything better to do.

TM: And was that something the family was happy with at the time?

NB: Well … they supported me in whatever I wanted to do, but my father would have rather I got ready for law school. I did take a couple of law courses, but they didn't suit me.

TM: It's funny, because of the meticulousness of the work you're doing. What wasn't good about law, what didn't work well with you?

NB: I don't know what to say, I wasn't really interested in anything at that time. I mean, I did my work and I got strong grades, but I really wasn't grabbed by anything until I got to graduate school, after my service.

TM: So the BA was interrupted for service?

NB: No, I was in the air force for two years, '52 to '54, but that was after college.

TM: And then you went on to graduate school at Yale?

NB: That's right.

TM: And that's when things changed?

NB: During my first year in graduate school I got interested in philosophy.

TM: Do you remember any decisive moment in that development?

NB: No. I was much taken by metaphysics and Paul Weiss, and I also studied with Arthur Pap and Henry Margenau, and that was very interesting. And I took a year long course on Whitehead which I much enjoyed, taught by Nathaniel Lawrence.

TM: That was all at Yale then. You also went on to do your PhD there. Did you have to prepare a piece of work to finish your MA, or was that course work?

NB: Course work. The MA was just after two years.

TM: And they hired you from grad school there?

NB: Well, I had a year of Fulbright in 1957–1958, studying with Canon Robert Feys at Louvain, and that's what really got me interested intellectually. He gave me an article by Ackermann to read, and that's the first time anybody had ever given me something to work on. Before that it was all course work. I worked very hard at that. Ackermann's "Begründung einer strengen Implikation", from the 1956 *Journal of Symbolic Logic*.

TM: So you got to read that in Europe.

NB: Yes. It's just a short article, but it fascinated me and I got interested in logic in that way. I had taken a lot of logic courses when I was at Yale, from Frederic Fitch. I must have taken about eight courses—I don't remember how many—a lot. Basic logic. I don't know if I took them all either, or whether I just sat in. But most terms I was doing some logic.

TM: With Fitch.

NB: Yes. Rulon Wells taught me my first logic course, we used Fitch's book. And that was interesting.

TM: I see when you work on something that you use that system—the method of subproofs—you have such a good way of employing it. You take out a page and then

you write down formulae; you work from both ends and you see where you need to fill in steps.

NB: Yes, I learned that from Fitch's book.

TM: So you interrupted your time at Yale while already working on your PhD there officially, to go to Europe on the Fulbright?

NB: Yes. And I had a PhD topic but I didn't work on it at all.

TM: What was the topic?

NB: Existence—the nature of existence.

TM: We're working on that now!

NB: And then I went on to study with Feys and that's when I got interested and hooked on the academic life.

TM: And you took that back with you to Yale?

NB: Yes, I had proven something but I didn't know what it was. And I asked Feys if he knew anybody that, when I got back to the States, could help me. And he said: "Certainly, Gödel". And I was totally intimidated. I didn't look up Gödel, but I did look up Alan Anderson. He had taught one of Fitch's lectures before I went to Belgium, and he was a wonderful teacher. I was smitten, and so when I got back I asked him whether he knew anybody who knew about Ackermann's *Strenge Implikation*. Alan was a very sweet man and he said (*singing*): "I do". I still have that image clearly. And that got us started working on what turned into the program on relevance logic.

TM: You defended your PhD at Yale in 1960 then.

NB: Sort of. It was early in 1960 that Alan said, "Why don't you write your dissertation on the stuff that we've been working on?" And I was quite surprised to learn that I could do that.

TM: Because you had been given the other topic.

NB: I thought it was cast in stone.

TM: And who was your official supervisor?

NB: Alan. Before that it was Weiss.

TM: You could transfer that, but you still thought you had to work on the old topic.

NB: I wasn't working on it, but I thought it had to be my dissertation topic. But … I've forgotten the dates, but it was in maybe February or March, 1960, that I had this conversation with Alan, and I immediately just gathered up all the work I'd done and made a dissertation out of it. And it was effortless, I did that in six weeks. I felt very lucky, watching other people struggle. I had it all done, and it wasn't threatening—I was just having fun with Alan. We worked together very closely.

TM: And while you were working on your dissertation, you were already employed as an instructor at Yale, so you taught your first courses there?

NB: That's right, I was employed off the boat, so to speak, coming back from Belgium in '58.

TM: And then once you had defended your dissertation they made you an assistant professor there.

NB: Yes, from ’60 to ’63.

3 From Yale to Pittsburgh

TM: So you started academic life as an assistant professor in 1960, and you had lots of very good colleagues at the time. Yale was a very strong department.

NB: It was a strong department, yes. We had a strong chairman, Charles Hendel, but he left in 1959. And things started to fall apart, and stayed that way for decades.

TM: Right … So what courses did you teach? Anderson was there to teach logic, was Fitch still around?

NB: Yes, all the time.

TM: And you were also teaching logic?

NB: Yes. But not only logic. I usually taught a general philosophy course of some kind. Not an advanced course because that was really for the older guys.

TM: And did you have many students then at Yale? How was the student population? I mean, you had grown up with them as it were, in your days as a graduate student. But you were teaching mostly undergrads?

NB: Yes, I was teaching undergraduates all the time. I didn’t teach any graduate students and graduate courses at Yale.

TM: Were they different, as students, compared to the students you were together with in your days when you did your BA at the University of Illinois? Was that different, undergrads at Yale and at Illinois?

NB: I had nothing to do with the undergraduates at Illinois, so I didn’t know. One of my fraternity brothers introduced me to philosophy, in a way—Jack Karns. There were people coming back from the service that were five or six years older than I was. My fraternity house was way south of the campus, half a mile or more. And Jack wanted to take this course, and he wanted somebody to walk with. And it was a philosophy course, from Max Fisch. Most of it I hadn’t any clue as to what was going on, but he did have us read a little Whitehead, and I really liked that a lot. And that was the first interesting thing in philosophy that came my way. I took logic there but the logic was Cohen and Nagel, not much beyond syllogisms.

TM: That was different at Yale, with Fitch.

NB: Yes, Fitch’s book was on propositional logic and quantifiers. His own inimitable system. But I had none of that at Illinois. So Jack got me interested in philosophy, I give him credit.

TM: What was your fraternity?

NB: Alpha Tau Omega. And when I graduated I entirely lost interest in the fraternity.

TM: Over the years I've come across so many people who tell me that they got their first logic from you, that you taught them. I mean many distinguished philosophers. Did that start at Yale already? I think I remember most of the stories were about Pittsburgh. Did you have any memorable undergraduate students at Yale, with whom you kept in touch?

NB: Undergraduates … Yes, well, kept in touch—I don't know about that. Somehow Alan wrangled me a research assistant. We had some National Science Foundation contracts and we hired undergraduates through that. Most of them came and went, but an outstanding example was Jon Barwise. He was my research assistant for one year. He was in his formative stage. We had a good time. And then there was Neil Gallagher, who didn't stay in philosophy. And John Wallace, who later went to Minnesota, was working on these NSF grants with Alan and me.

TM: You also had this grant by the *Office of Naval Research.*

NB: Yes (*laughs*). Omar Khayyam Moore, a social psychologist, had this grant from the program that the Navy had, of some kind, I've forgotten the details. But they were willing to support me for a while. That was after my dissertation was written, it paid for putting the dissertation together in a distributable form. Omar paid for that. My dissertation was on relevance logic, entailment, and was published by the *Office of Naval Research, Group Psychology Branch*. And I had to write some kind of preface, that one's one of my favorite prefaces actually. I wrote, I don't remember the words at all, but I remember the theme was: "It has not been conclusively proven that this material is totally irrelevant to social psychology".

TM: Self-applying the system, as it were, to the preface.

NB: Omar was a very interesting guy, and he later came to Pittsburgh.

TM: He had a position at Yale at the time?

NB: Yes. I persuaded somebody or other to move him to Pittsburgh.

TM: A lot of people were moved. So you were going along with Alan in '63, or had he left a little earlier?

NB: No, it was Wilfrid Sellars who was the person that Pitt wanted, and two of us young assistant professors, Jerry Schneewind and I, hung on his coattails to get to Pitt. And that's how I got to Pitt. I was so thrilled. I had been destined for a career consulting for the System Development Corporation, and I had actually signed up for a job with them. When I was in graduate school I went out there some summers, and I was working with Thomas Steel. He was my boss so to speak. I had met him when we were in the service, as I told you, and he was head of our unit, which was working on ciphers or codes or something like that.

TM: When you went to Pittsburgh it could have happened that you would have gone to work with the System Development Corporation.

NB: At that time it was the only job offer I had.

TM: So your initial appointment as an assistant professor at Yale was only good for two or three year?

NB: Yes, I was in my final year …

TM: … but then you could join Sellars …

NB: There was a policy at Yale, they'd see if they could reappoint you, but they declined to reappoint me. So I always said they fired me. Jerry Schneewind got an extra year, if he wanted it, but he didn't want it. And then Wilfrid brought Jerry and me to Pitt. And I was just thrilled. Adolf Grünbaum and Nick Rescher were here, they were the ones that were already here when we went. The department was a very local street car type department, and gradually we made it to international status.

TM: Absolutely. So Alan stayed at Yale for a while?

NB: Alan did. I think in '64 he went to Manchester, he had a Fulbright, working with Arthur Prior. And it was while he was there that I persuaded the Pitt people, that was the hard part, to bring him. At that time Pitt had a very loose administrative structure, it was just run by the chancellor, and there weren't any committees. There were three deans, and the point was that they had no power at all.

TM: So it was really all through the chancellor that you had to …

NB: … and the vice-chancellor, Charlie Peake. A great academic. And then it was a question of getting Charlie to bring Alan. That was in '65. And we had been working together at Yale and we just continued working at Pitt. He died too young, in '73, that's all.

TM: That's true. So the first volume of *Entailment* you finished together?

NB: Yes, he had his hands on every bit of it.

TM: Your cooperation had been a bit interrupted, I guess, when he went to Europe on a Fulbright, and then because of the distance between Yale and Pittsburgh. But then you joined forces again there.

NB: That's right. Well, we had kept in close touch. We were working on things together; the way we worked when we worked together was cheek by jaw. We just sat down and wrote sentences together.

TM: I don't think it's very common for philosophers to do that.

NB: I don't think so either.

TM: But it's a very nice, very intense way of working.

NB: We had a really good time.

TM: You were hired as an associate professor, with tenure already in '63?

NB: With the promise of tenure, but not tenure. They gave me tenure, I don't know, a year later, and I was made a full professor three years later, in '66, after I'd had my six years an assistant professor.

TM: There's the Sellars Room in the Cathedral of Learning, on 10th floor. Was that Sellars's office?

NB: No. Adolf had an office on the 20th floor or something. Wilfrid's office was downstairs, 2nd or 3rd floor. A wonderful office. He was provided with a secretary and a suite, as was Adolf, and Nick. Alan and I fortunately were able to hire secretaries through National Science Foundation grants.

TM: At that time, it's not as if you sit down and type something in LaTeX. It must have been truly different—lots of typewriter work.

NB: Phyllis, I remember, my first secretary—this was back at Yale—she gave us her first page, and it was just full of mistakes, and very sweetly Alan said to her, "This won't quite do". And she never made another mistake. We were very fortunate to have secretaries.

TM: Yes, you really relied on that kind of support I suppose. Did you also share an office ever, Alan and you?

NB: No. I lived in his office, so to speak, a large part of the time. But no, we never shared an office. I had an office at Yale northerly on the campus, in a big old building, an enormous office that had a dark room.

TM: So you could do your photography.

NB: Well, sort of. I evinced interest in hooking up the plumbing, so to speak. And the next day they came and took the dark room out.

TM: Where was your office at Pitt initially?

NB: We were in the Schenley Hall. Pitt bought the Schenley Hotel and we were on the seventh floor and every office had a bathroom, a hotel bathroom. Which was terrific, because you didn't have to waste time going down the hall. My first person on the other end of the bathroom was Storrs McCall, and we worked together a lot, we had a lot of fun.

TM: So that was the time when the department was really building up to become the strong department that it then became.

NB: I neglected to mention Kurt Baier.

TM: He was there already?

NB: Yes, he was. He came a couple of years after Nick and Adolf, but before I came by a year or two. And he was chairman. A wonderful chairman he was. You hardly knew he was chairing. We had meetings in the hall. "Yeah, that's a good idea, let's do it." Not bureaucratic like it is today. And his wife Annette—those were bad times for females. She was treated very poorly for a long, long time. She got some low class employment at Carnegie Mellon University for a while, and then finally Pitt hired her back on a proper professorship. And then she thrived.

TM: How close were the ties with CMU doing all these years?

NB: They've gradually gotten closer. But they were close even then, there was a lot of back and forth.

TM: It's fortunate to have these two institutions so close together. I can see that with Kohei Kishida, for example, how well it worked out.

NB: Wonderful.

4 Employment History at Pitt

TM: Let's go over the employment history at Pitt. You arrived there in '63 and you stayed there your whole career, but you wore many hats at Pitt, as it were. So the first employment was in philosophy. But a few years later you entered sociology.

NB: No, I did that right at the beginning. I taught a course that was about half philosophers and half sociologists. We had a lot fun. We all started at the beginning because nobody knew the other topic at all. It was a big seminar, about 20 people in it. That was invigorating.

TM: What would you do?

NB: Well, I read stuff, I didn't know any social science when I took the job. But I claimed I would be very pleased to teach the philosophy of the social sciences. That seemed to be a requirement for getting the job. And I had a good time doing it, I enjoyed that. I taught that for quite a few years, don't remember how many. And with these big classes of interested people. Our graduate program at Pitt was very different than it is now and has been for probably decades. We brought in about 20 students a year at the beginning. And very good ones. Bas Van Fraassen was in one of our classes. And Mike Dunn and Peter Woodruff. Bob Meyer was already there, as a graduate student—he was a hang-over so to speak.

TM: At Pitt they made you a professor of sociology in '67, a year after they made you a full professor in philosophy. And you had that job for twenty years, a joint appointment between philosophy and sociology. And then the joint appointment with philosophy of science started in 1971. Was that around the time they had a separate program in History and Philosophy of Science?

NB: It would've been about then. Larry Laudan got appointed as a member of the history department, and that wasn't going to work out. So Pitt built around him a History and Philosophy of Science department, which thrived.

TM: It's really a model, I guess, for many many other departments.

NB: Every once in a while somebody would suggest to the chancellor, Posvar at the time, that the two departments ought to be combined. And Posvar would say, "What a bad idea! Now we've got two world-class departments, and you want to make it one?"

TM: From among the hats at Pitt that you wore the named professorship stands out, named after Alan Ross Anderson, starting in 1984. How did that come about?

NB: Well, Mrs. Anderson gave quite a lot of money … I collected as much money as I could after Alan died in 1973, I got quite a bit, but it was from graduate students and colleagues and it didn't amount to much. But then Mrs. Anderson gave a substantial amount of money, I forgot what fraction of a million. And that's how I came to be the Alan Ross Anderson distinguished professor for philosophy.

TM: And then there is another hat at Pitt: Professor in the Intelligent Systems program, what was that?

NB: That was an Artificial Intelligence-like program. Rich Thomason came to do that. I had taught him at Yale, and he had taught me at Yale. He had been out to California and he brought back all these weird ideas about maximal consistent sets. And he taught me all that stuff, at Yale. We brought him to Pitt and kept him as long as we could.

TM: That was about the time when Pittsburgh underwent a major transformation because the steel mills were closing. And they reinvented Pittsburgh as a place for high-tech. When did the steel mills begin to close?

NB: They were closing already in the ’60’s. There wasn’t a year in which they all closed, it was spread out. My sense of history is very poor, that’s about all I can say. They were still there when I came in ’63, but they were disappearing.

TM: You’ve spent such a long time at Pitt, there’s a full page of departmental positions. Have you ever been head of department?

NB: I’ve once been chairman, acting chairman, 1974. But I didn’t do much. I was trying to recruit Solomon Feferman.

TM: You must have been involved in many of the hires that Pittsburgh did over the years. Any notable stories?

NB: I remember hardly anything about it. The first hires—I guess Bob Brandom was in the early ’70s. Myles Brand and Bob Brandom came in at that time.

5 Visiting Professorships

TM: You spent most of your academic life at Pitt, but you had quite a number of visiting professorships along the way, the first of which brought you to California in ’73. Can you talk a little about that?

NB: That was Irvine. Irvine was recently thriving, but it was thriving. I taught there just one term, in the winter fortunately. I took my family out and we lived in Laguna Beach, they found us this wonderful house, on the beach, or close to the beach. That’s mostly what I remember. I taught Bressan’s *General interpreted modal calculus* out there, I remember that.

TM: Did you have many students?

NB: Yes, I did. I don’t remember how many, but it was a noticeable number. I wasn’t doing graduate teaching at Irvine, I think.

TM: Was that the time you had the motorcycle?

NB: The biggest one, yes. I started with a scooter and moved my way up. I got to California with one of those great, big ones. Somebody was willing to rent it to me for three months or something—or to sell it to me, with the agreement to buy it back.

TM: Then you spent quite some time at Bloomington, Indiana, three times.

NB: Pitt at the time was on the trimester system. They had already been on the trimester system when I came, which turned out not to succeed because people

didn't want to go to classes for a third term. It was one of Litchfield's good ideas that didn't work out. But it did mean that I could have a trimester off. And I used that to go to teach at Indiana University in the falls of '77, '78 and '79, as I recall.

TM: Who was at Indiana at that time? Did you connect with people there?

NB: I did, my former student Mike Dunn was the principal attraction, we collaborated a lot there. I guess he's the only one I collaborated with then. But they had a lively History and Philosophy of Science department, and I spent a lot of time with them. Ron Giere was there. I organized a weekly meeting, a lunch meeting, between the two departments. And that was great. They didn't have all that much opportunity institutionally to talk to each other, so I felt I was doing a service there.

TM: And then a visiting professorship brought you to Leipzig, in 1996, the Leibniz professorship at the *Zentrum für Höhere Studien*.

NB: Yes, that was a wonderful term. I guess it came about because of Heinrich Wansing, who had drifted through Pittsburgh a few years earlier, and had this idea of how to do a Gentzen-style calculus, and I had worked on that quite a bit, on that style of calculus. I said "that's not going to work". And he was really bowled over. I persuaded him that it wouldn't work. So we went out and did it a different way. That was when I was developing Display logic. And then he was at Leipzig. He wasn't a man with power, but he was persuasive, I guess. I think that's how it went. Meggle was there, I did not have so much to do with him, but I saw a lot of Pirmin Stekeler-Weithofer.

TM: That must have been an interesting time there, just a few years after German reunification and the city still very much in transformation.

NB: Oh yes, there was an enormous apartment size crane on every corner. They put me on the 6th floor of a building, which was terrific. I was at first intimidated by the prospect of doing six floors, no lift, but I liked it. My classes were very small. Pirmin used to come regularly. And there were a couple of other students who came. One other from the faculty, but I can't remember who it was. It would be a class of about four.

6 Professional Service

TM: Let's look at your professional service. You served on the board as a program committee chairman for the *Association for Symbolic Logic*. And that was still during your time at Yale.

NB: It was.

TM: You also did a lot of other service to the ASL as well, right?

NB: I guess it was for about a decade.

TM: And then there is the *Society for Exact Philosophy*. You helped found this in the early '70's?

NB: I was one of the founding members, but I wouldn't say I helped found it. I came to the first meeting, and I was the vice-president and president for a while, and program coordinator, etc. So in the early days I took a hand in it. We had one meeting at Pitt. It was joint Canadian/American by design, and we traded off between the two countries as to where the meetings were. It was very nice to have some close association with some Canadians, Mario Bunge for example.

TM: It was picking up the tradition of the Vienna Circle, right?

NB: Yes, the spirit was to be that of the Vienna Circle.

TM: This was the American/Canadian cooperation, but you also played a role in the British *Mind Association* for quite a while, as their American outpost.

NB: Alan had been the U.S. treasurer of it, which meant he kept a few funds in the bank, $100 or $200. Every once in a while they would ask him to pay for something. I inherited that job from Alan, for about twenty years. It was not a position of power, I payed a bill or two every year.

TM: You're also a long-time member of the *American Philosophical Association*. And since fairly recently, you're a member of the *American Academy of Arts and Sciences*.

NB: I was surprised.

TM: You were elected in 2008, together with the Coen brothers, right? I think it was the year they accepted the Coen brothers.

NB: Yes, that's right. I never met them.

TM: Any stories about the APA? Did you go to the meetings regularly?

NB: I did for a decade or two. Sometimes the Western or the Mid-Western, but mostly the Eastern. Pretty regularly. Alan and I would submit a paper, something like that. In the beginning these were smaller meetings. They could be held in a university.

TM: Now it's the job market and so it's this huge event. Did you go there, for Pitt, to hire people?

NB: Yes. But it wasn't as much fun.

7 Journals

TM: Let's look at journals. The earliest involvement with a journal that I find on your list is with the *American Philosophical Quarterly*. You were on their editorial board for over a decade.

NB: Nick Rescher was the editor and he would pass along papers for me to read.

TM: And then I think you entered the editorial board of the *Journal of Philosophical Logic* when it was founded. You're still on their advisory board. Were you involved in setting up the journal?

NB: I was on the executive committee, or what it was called—the board of governors. We had meetings, frequently at Nick's office or at my office, the three of us. Gerald Massey was on the board, it was a pretty close knit group.

TM: And Pitt was running quite a number of important journals.

NB: Wilfrid Sellars was running *Philosophical Studies*, I don't think there was anybody else.

TM: And the *Notre Dame Journal of Formal Logic* was also set up around the time and you were involved with that from the beginning?

NB: I don't know what the beginning was, but it was a much less significant relationship. I had very little to do—Sobociński did everything.

TM: Then you're on the editorial board of *Philosophy of Science*.

NB: Today I don't do anything for these journals anymore, but some of them keep me on the masthead, I guess. I don't know which ones do or which ones don't.

TM: *Studia Logica*, how did that come about? I mean, that was set up in Poland.

NB: I read papers when asked to. I was never involved in the day-to-day activities.

TM: So it was really the *Journal of Philosophical Logic*, where your strongest involvement was, and the *American Philsophical Quarterly* before that. Your list also mentions the *Philosophical Research Archives*.

NB: I was just reading papers for the APQ, for Nick, I didn't participate in any of the administration or anything like that.

TM: You've done a lot of refereeing in your years.

NB: I have, I did a lot of refereeing. But I haven't for the last decade, or five or six years.

TM: You've done your share.

NB: That's what I tell them.

TM: Do you have a particular style that you would recommend? What should a referee do?

NB: No, I don't have any contribution to make about that.

TM: I sometimes get sent a "proof of the squaring of the circle" or something. Have you come across those things as well? Proofs that Cantor's proof is wrong? That's a favorite.

NB: The theme sounds familiar, but I don't remember any hands on activity.

8 Prizes and Fellowships

TM: Maybe the next thing to go over would be the list of prizes and fellowships. It starts with something pre-doctoral, from Yale.

NB: That was a book prize, I split it with somebody, money to buy books with.

TM: And then you held a fellowship at Yale, in the year before you went to Belgium.

NB: I did, and Alan really promoted that. I didn't have a fellowship when I went to Yale. I was on the GI bill, they would pay for your graduate education as well.

TM: And then you had the Fulbright Fellowship. How did Fulbright work in those times?

NB: You wrote an application and a committee looked through the applications, and you'd sign up for which country you'd like to visit. I thought I probably wouldn't do very well in the competition and so I didn't choose any of the English-speaking countries. I knew a little French, so I chose Belgium.

TM: Any ties to Belgium? Had you been there?

NB: No. I looked it up ahead of time, and must have been talking to people, I don't quite remember. I corresponded with Feys, quite a bit. It was great. I took my wife and my two-year old, and we lived on the Chaussee de Vleurgat.

TM: In Louvain?

NB: No, in Brussels. There was no housing to be had in Louvain at that point.

TM: So you commuted by train?

NB: Yes, about twelve miles.

TM: Then, as you said, Alan helped you for the Morse research fellowship which you held at Yale, just before leaving.

NB: Yes, I had the final year off, no teaching.

TM: And then you also had a Guggenheim fellowship in '75–'76, which you preferred over another grant, from the *National Endowment for the Humanities*. And that was to work on *Entailment*?

NB: That's certainly what I was working on. It went together with half a sabbatical. I don't remember going any place, it just paid for the groceries for my family of six, at that time. And I really don't remember my application.

TM: And then you were at Stanford for a while, in 1982–83, as a fellow at the *Center for Advanced Study in the Behavioral Sciences*.

NB: That was my next sabbatical. One term on a sabbatical, one term fellowship, matched.

TM: And how was Stanford then? Did you interact with many people there?

NB: Not many, but Pat Suppes was there …

TM: Was Jon Barwise there?

NB: No, but Solomon Feferman.

TM: So that's how you know him?

NB: We had tried to hire him at Pitt in 1974. He was very much underpaid at Stanford. I think we knew that he wasn't going to take the Pitt job, but we were going to facilitate his living conditions. So we had him out.

TM: And then the next fellowship I see is in 1988, from the AAAS. That was, I guess, the next sabbatical.

NB: It must have been. It's certainly about five years later.

TM: So that's the deal you get, once every five years you get one term off and you try to match it …

NB: Something like that. It was six terms, as I recall, at Pitt. And then one term off.

9 Honors

TM: Let's go over the honors. There is the *Festschrift* for your 60th birthday, *Truth or Consequences*, edited by Mike Dunn and Anil Gupta, that came out in 1990. There was a special issue of the *Journal of Philosophical Logic*, twenty years later. It came out in 2010, put together by Philip Kremer and Heinrich Wansing then. And at Leipzig, ten years in between, so it's evenly spaced, they made you a Doctor phil. *honoris causa*. So you went there for a ceremony—how as that?

NB: Oh, I enjoyed it very much! It was partly seeing old friends, and Krister Segerberg did a biographical spiel, and they played some music.

TM: You also got a *Chancellor's Distinguished Research Award* from the University of Pittsburgh. That went together with your visiting professorship at Leipzig?

NB: No, that was just cash. I spent it on audio-equipment for my office.

TM: How long had you had that office for, the 10th floor office? I think quite a while …

NB: I couldn't tell you, we moved over to Schenley Hall, and then we moved over to the Cathedral, but I couldn't tell you what year.

TM: But that's when they gave you 1028-A?

NB: Yes. It was Kurt Baier's office that I moved into. I can't visualize myself any place else.

TM: And then there is becoming a fellow of the *American Academy of Arts and Sciences*, in 2008. From among these honors, is there any one that you remember especially fondly? Do you connect any of these with a feeling that you were on the right track, doing good work? Or is that more in collaborations that you got that feeling?

NB: I'm not sure what your question is, so the answer is "No", or else it's "Yes".

TM: I think what I'm trying to get at is, sometimes you work on something for a long time and then you get some external recognition for it and that helps you get a feeling of "I'm doing the right thing", of pursuing a fruitful research line.

NB: I don't think my grants ever had that much effect on my self-opinion. I guess I was … it's stupid to say, but I guess that I was confident that I was doing ok. I was gratified of course, by the various awards. I also got some kind of medal from a place in Finland; Pörn invited me there, when I was working on action theory and stuff. But I forget what it was.

TM: You must have been in almost every European country academically, as a visitor.

NB: You have the list!

TM: We'll go through the list of talks, maybe we'll have the map of Places Visited By Nuel Belnap. There is another list of grants, consultantships and research fellowships. And it starts with the *National Science Foundation* funding that you were talking about at Yale, in 1962–1963, for summer undergraduate research that you directed. And then the consultancy for the *Office of Naval Research*, on "problem solving and social interaction".

NB: Alan and Omar.

TM: That's what paid for your PhD thesis as a book. And then the next thing is your involvement with the System Development Corporation in California that you said was your job offer.

NB: I went out there in the summer for some period, as you can see for quite a few years.

TM: Was that the time that you got into computer programming, or did you have earlier experience with that?

NB: Computer programming had been my first job, in the Air Force.

TM: Oh ok, so that's how early it started really. That was in the mid 1950's.

NB: Early '50's. Very early. I graduated in '52, and then I went to Washington, worked for the NSA.

TM: "No Such Agency". What type of computers did they have?

NB: They gave us the first large-scale IBM computer, literally the size of the prototype that IBM kept. It was the first machine they sold. The NSA were about the only people who could afford it. And that was an interesting job, I really enjoyed programming. The IBM 701 had 32 instructions. One of them was a "No OP", and one of them was "Stop".

TM: Quite a lot you can do with 5 bits. How did you enter a program?

NB: Punch cards, that's how they worked.

TM: Assembler programming on punch cards.

NB: No, that was pre-Assembly.

TM: Really writing the codes for the instructions …

NB: Yes, the actual numbers. What I remember there is that for six months we had to program in octal, and finally they gave us a way that we could program in decimal. That was just wonderful.

TM: I can imagine, it really wrecks your brain. Good training for a logician, and certainly something to make you resource-sensitive. Do you see a link with the logical systems you're interested in?

NB: No, I don't.

TM: So you had a lot of experience in computer-related work, when you went to work for the System Development Corporation.

NB: I did, but I didn't use it for them really, although I did do a little programming, because they had a machine that I could program on. I wrote a program to test matrices. That's the only one I wrote there. Then later I wrote a program in FORTRAN to help assemble an index.

TM: That was like a standard piece of software with many people, for a long time.

NB: Well, not with many, just a few friends. That was fun too. But I did that at Pitt. I would tell you what I did for the NSA but they would shoot me.

TM: We don't want to risk that! The list of grants has quite some NSF funding, and that would give you a research assistant, or would it also buy you out of teaching?

NB: It didn't buy me out of teaching, they were summer grants as I recall. But it would get me some help, and I don't remember how that worked any more. Alan and I did all that together. Sometimes I would be the principal investigator, and sometimes I would be associate investigator. We switched roles.

TM: So that was really joint work with Alan. You were also very early in using computers in research in the humanities, in the 1960's, working with a grant from IBM.

NB: I did have a grant from them, yes. I taught a little course, there were never many people on it. We read whatever there was to read.

TM: Which wouldn't be much at that time. But it nicely goes together with involvement in philosophy of the social sciences, in a way.

NB: Sure.

TM: And then the computing story continues with the involvement in the *International Federation of Information Processors*.

NB: Tom Steel was head of that section, whatever the section was, "Formal description of computer languages", and he "acquired" me.

TM: And that brought you over to Europe a couple of times, for meetings.

NB: Yes … Vienna, Sardinia, Copenhagen, … Vienna again.

TM: And then it seems that in a similar context you spent one term at Oxford too?

NB: Yes, in my sabbatical. I got some money from them, and I was there for one term, Hilary term, in 1970.

TM: Which college were you at?

NB: Wolfson. Well, I didn't live in the college.

TM: So we're even Oxford co-collegiates. That was fairly recently set-up then, I think, Wolfson. Did you go punting?

NB: I didn't go punting. A substantial amount of bicycling, some of it in the mud.

TM: And you also spent time at the Australian National University, in Canberra.

NB: Yes, I had a term there. That was on a sabbatical.

TM: Who did you work with there? Was that the relevance logic community?

NB: Yes, Bob Meyer was there and I overlapped actually for a couple of weeks with Mike Dunn. In Oxford I overlapped for a couple of weeks with Dana Scott, and that was very valuable to me.

TM: Did you meet him there or had you met him before?

NB: He came to my first paper, in 1963 I guess, at the University of Chicago, when I was trying to get a job. He was the only one in the audience, I think, who followed whatever I was saying. I would see him every once in a while, off and on. And of course he eventually came to CMU.

TM: You also spent some time in Moscow, at the Academy of Sciences.

NB: I did, that was a fairly short visit, in 1991, it wasn't a term or anything like that. I don't know, two or three weeks.

TM: Who did you work with there, or who got you over?

NB: I didn't work with anybody in particular. This one person, Vojshvillo, he taught at Moscow State University and he invited to one of his seminars. And I talked a little bit about whatever I was talking about at the time, but then … there was this decision problem for the system *R* for relevance implication. And Vojshvillo had provided a decision procedure for this, in *Studia Logica* in 1983. But recently my former student Alasdair Urquhart had published a proof that there was no decision procedure. Vojshvillo spoke no English but he asked me in Russian what I thought of that. And I said … it was potentially embarrassing, but I got off without embarrassment, I said, "But look, he's my friend". That was interesting. My daughter Mary Jo came with me on the trip that time, we had a good time.

TM: To Moscow? '91 was exciting, there was lots happening … it was really Wild West in some respects. That was in March? It must have been really cold.

NB: It was, there was ice in the streets.

TM: And in the same year then, to make up for it maybe, you stayed in Europe, or you went on to Italy, to visit Padua? In '91 it says you were a visiting professor there.

NB: Aldo Bressan invited me.

TM: So you worked with Bressan there?

NB: We talked. We didn't really do any joint projects, but we talked.

TM: And Alberto Zanardo was around at the time, was that how you got to know him?

NB: It is how I got to know him. There were two of them, Bressan's students, and I forget the other one's name, he was a physicist, who did a little work on relativity theory.

TM: There's the odd one on the list: work as a consultant for Westinghouse, that's the elevator company, right?

NB: Yes, that was a one-shot deal.

TM: So what did you do, design a new elevator brake?

NB: I gave a lecture, and I gave it on … they didn't know what to do with me … what they were interested in was building robotic submarines or something like that, that

was the general topic. But I didn't talk about that, I just talked about relevance logic and how this contradiction tolerant system could be of some use. This was through a neighbor of ours in Pittsburgh.

10 Doctoral Students

TM: Let's go over the list of your doctor students. Those are all Pittsburgh PhD's, right? Or is there any involvement in dissertations running somewhere else?

NB: There is one down there from the University of Indiana, Daniel Cohen. The rest is Pittsburgh.

TM: It's a list with very distinguished people on it, and it nicely reflects your interests over the years. Many of those people you have collaborated with. Michael Dunn was already at Pitt when you came?

NB: No, he was, I think, in my first class.

TM: So he was quick to finish, in '66, and you only came in '63.

NB: He was. Those were the days.

TM: So you could finish a PhD in three years at that time.

NB: Well, you could. Michael took three. "The algebra of intensional logics".

TM: And then there is Carlo Giannoni, "Conventionalism in logic".

NB: He was a hold-over from the old Pitt department, as was Bob Meyer. And I took over Carlo as a kind of a charity case, so to speak, he didn't have anywhere else to go. I was never interested in conventionalism in logic. Bob of course I worked with a lot. "Topics in modal and many-valued logic".

TM: And there is Jim Carson in '69, "Logics of space and time", so that really prefigures some later day interests. Kent Wilson? "Are modal statements really metalinguistic?"

NB: Again he was a hold-over from the old department. Just being helpful.

TM: Peter Woodruff, "Foundations of three-valued logic".

NB: He dropped out of philosophy. That's philosophy's loss. A very smart guy, but he could never write anything.

TM: And then there's Dorothy Grover, "Topics in propositional quantification", and Ruth Manner, "Conditional forms: assertion, necessity, obligation and commands". Garrel Pottinger, "A theory of implications"; Alasdair Urquhart, "The semantics of entailment". So this is really a lot of work on the relevance logic project here. Jonathan Broido, "Generalization of model theoretical notions and the eliminability of quantification into modal contexts"; Arnold Vandernat, "First-order indefinite and generalized semantics for weak systems of strict implication".

NB: What he did was invent S9. I think it would have been part of his dissertation. He just published a book, recently, that came across my desk.

TM: And there is Robert Birmingham, "Law as cases".

NB: We had a lot of fun together. Again, he did it in three years. And that was his third post-graduate degree.

TM: So he had a law background when he came?

NB: He had law background when he came and a PhD in economics.

TM: And then there is Anil Gupta, "The logic of common nouns: an investigation in quantified modal logic". Did you work together with him on the Bressan manuscript?

NB: No, that was earlier—the book came out in 1972, and Anil did his PhD in 1977.

TM: But he took your Bressan classes.

NB: And he went beyond those, yes.

TM: Glen Helman "Restricted lambda abstraction and the interpretation of some non-classical logics"; Zane Parks, "Studies in philosophical logic and its history". You told me the story of his defence, which is probably not for the record?

NB: Not for the record. Good story.

TM: There is Daniel Cohen, "The logic of conditional assertion", whom you supervised with Mike Dunn, who had already moved to Indiana. And there is Jay Garfield, "Cognitive science and the ontology of mind", that's interesting.

NB: It was interesting, he wasn't professionally interested in logic at all. We had a good time together.

TM: And then there is Jeff Horty, "Some aspects of meaning in non-contingent language", and Michael Kremer, "Logic and truth". Aldo Antonelli, "Revision Rules: An investigation into non-monotonic inductive definitions", Mitch Green, "Illocutions and attitudes", and Philip Kremer, "Real Properties, Relevance Logic and Identity". It was running in the family, the Nuel thing.

NB: I had the three of them, the father and two sons, carry my desk upstairs, that big heavy thing.

TM: Then there is Ming Xu, who came from China to do his PhD at Pitt.

NB: Yes, and he stayed for much longer than planned.

TM: But now he's back, and he's big in China, right?

NB: I don't really know how big he is, it's a little hard to judge.

TM: Well, China is so big.

NB: Yes.

TM: There is Stephen Glaister, "Belief revision". And John MacFarlane, "What does it mean to say that logic is formal"—you were on the committee but he wasn't officially your PhD student? And then you have to add Kohei Kishida, "Generalized Topological Semantics for First-Order Modal Logic".

NB: That's correct.

TM: That's quite a list. A few of those finished in three years, but some hung on for a lot longer …

NB: As you get later, they get longer.

TM: So who is the longest?

NB: Zane, I guess. He wrote several papers, among them a nice one in *Journal of Philosophical Logic*, in '72.

11 Publications

TM: One more list to go: Let's go over your publications. I'm of course very happy about that publication list because it got me the Erdős number 3, through your 1967 paper with Spencer, who later wrote a book with Erdős.—The first paper, from 1955, is influenced by Weiss?

NB: I guess so, I always debated whether I should put that on my list of publications or not. I hope nobody looks it up. I don't think anyone can find it.

TM: We'll try.

NB: I don't think I want you to try.

TM: And then the first real paper already made into the *Journal of Symbolic Logic* immediately, in 1959. It's an abstract, right? A two-page abstract on Ackermann's *Strenge Implikation*.

NB: There's the paper a year later, it's called "Modalities in …" instead of "A modification of Ackermann's 'rigorous implication' ".

TM: That's then the real paper, but you reported the result before. And then there is a technical report that appears in *Zeitschrift für Mathematische Logik*. And then in 1960, lots of things in *JSL* and technical reports for your grant, with the *Office of Naval Research*.

NB: And reviews of some kind or another. A book note.

TM: On Pat Suppes's *Axiomatic set theory*. And most of your formal-logical publications are co-authored with Alan Anderson.

NB: Indeed.

TM: Most of that material made it into the book, *Entailment*?

NB: Yes, probably all of it. We didn't want to throw anything out.

TM: Tell me about the "simple proof of Gödel's completeness theorem" in *JSL* 1959?

NB: It was just … instead of Gentzen consecutions you just add disjunctive formulas and did the obvious thing. I mean unending disjunctions. We were anticipated with that format though by Schütte.

TM: We see you doing your job writing book reviews for the *Review of Metaphysics*, for the *JSL* … then there is "Entailment and relevance". And then there is the metalogical paper, "Tonk, plonk and plink", in reply to Prior's "Runabout inference ticket". That's a nice piece.

NB: An afternoon's piece.

TM: It started something, right?

NB: It has a lot of credit that it doesn't deserve.

TM: Why do you think so? There are some very original ideas in there.

NB: It's sloppy.

TM: It was good enough for *Analysis* … Then there is a paper on intuitionism with Hugues Leblanc. How did that come about?

NB: I forget. It was a supervaluational paper of some kind. We had some kind of result about what you could do in which notation or something. It was mostly Hugues's paper.

TM: And there is more work building up to *Entailment*, and then there is the first paper on your work on questions, for the System Development Corporation. Is that the nucleus for the later book, *The logic of questions and answers*?

NB: Yes indeed.

TM: How did that crop up out of your consultancy work?

NB: Well, when I went out there I was afraid they were going to assign me some project. So I decided to bring a topic with me. I had read this little paper of David Harrah's, "A logic of questions and answers", and I said I'd like to work on that, and they said "fine".

TM: And then you had to write a report.

NB: A few years later, yes.

TM: There's more book reviews, and another paper with Hugues Leblanc and with Rich Thomason on intuitionism.

NB: We met in a hotel room in New York and worked that out.

TM: And then there is more reviews and more work on *Entailment*, and also a *Journal of Philosophy* paper on "Questions, answers and presuppositions" in 1966. That's early, I guess, for formal work on presuppositions. How did you get to work with the notion of presuppositions?

NB: It was already in the book. The book hadn't come out yet, but it was already in my earliest research, the simple logical ideas.

TM: Ask a stupid questions and get a stupid answer, the main theorems, I remember that.—And then in '67 we have the item that makes the link to Joel Spencer, later a coauthor of Paul Erdős's.

NB: Yes. This was when he was an undergraduate, in '67. We recruited him on one of these National Science Foundation support schemes. In the summer, or some part of the summer.

TM: And you wrote this up and he went on into mathematics from there.

NB: Yes, on to the Courant Institute.

TM: So we're in 1967 … A lot of the foundational work for *Entailment* is still going on. There is a reprint of your piece on "Tonk". And then there is work on distributive

lattices by you together with Michael Dunn. And a longer outline of *Entailment* with Alan. There is also work on the substitution interpretation of the quantifier by you and Mike Dunn. I'm just going over these thing, and whenever you want to comment …

NB: For a while Lennart Åqvist and I had a kind of dog-and-pony show. He would read a paper at one conference and I would read one on the next.

TM: Was he in the U.S.?

NB: He visited, yes. I don't remember, I think he was there for a term or what. They didn't treat him well.

TM: At Pitt?

NB: No, in Scandinavia, in Sweden.

TM: There's another piece on questions, in this nice volume *The logical way of doing things* edited by Karel Lambert. There is a lot of Pittsburgh in that volume. And there is a result that you published with Storrs McCall, in 1970, "Every functionally complete m-valued logic has a Post-complete axiomatization". In 1970 there is another piece which I think builds up to *The logic of questions and answers*? The piece on "Conditional assertion and restricted quantification"?

NB: No. Conditional assertion was a separate project. And restricted quantifiers was a quantified version of that.

TM: And then in '71 all you do is write a review, but you prepare Bressan's manuscript, right?

NB: I don't know what I was doing.

TM: But around that time you must have been working on getting that manuscript published.

NB: One might hope so.

TM: Because then in '72 the book, Bressan's *General interpreted modal calculus* comes out, with your preface, and I'm sure it would never be there, but for your work on this.

NB: I suppose that's true. My contribution was really minor even given that. No logical contribution at all.

TM: I guess you would have clarified a few things in the manuscript.

NB: I don't think so. I don't remember, but I don't think so. I was just translating from Italian into English.

TM: Well.—There is joint work with Dorothy Grover, in 1973, and this is connected with the project of the prosentential theory of truth already?

NB: Yes.

TM: And there is another piece on conditional assertion …

NB: … and restricted quantification. I forget what the difference is between those two papers. I hope there's a difference.

TM: Well this is in a book, and the other is in a journal. And then there is a *Journal of Philosophical Logic* paper on interrogatives. And then, was it in '73 or '74 that Alan died?

NB: Would have been ’73.

TM: Then you have your long piece on the prosentential theory of truth in *Philosophical Studies*, in which you tell the nice story of how it got in there, despite its length. I think that’s also one of the papers that many philosophers have read, because it’s not very technical.

NB: I think, I don’t know if it’s read anymore, but it was read for a long time.

TM: I think it’s being reprinted. And then you have the “Useful four-valued logic”.

NB: That was really with malice aforethought. I worked out that title, and the reason is that it sounded so pretentious that I thought people would think there was a topic there that they should deal with. It didn’t look just like a result.

TM: But that’s what it *is*, useful, right, the title is descriptive.

NB: Yes, that’s fair. I tricked them into reading the paper and all.

TM: And Ryle was there?

NB: He was there! Did I ever tell you about that? He came and congratulated me, he said that was the best paper he ever heard …

TM: There you go! That’s a real best paper award I think.—In 1975 you finished *Entailment*, Volume 1. It was already announced as Volume 1, because there was such a large body of material.

NB: Yes there was. Princeton University Press was quite reluctant to let us have that title because they were afraid we would Church them.

TM: Oh, that’s how you can use “to Church”. That has happened, indeed, with the *Introduction to mathematical logic*. But you didn’t.

NB: We didn’t. It took fifteen years but we didn’t.

TM: Then here is a result on the piece of software you wrote for the System Development Corporation.

NB: Yes. I don’t know what I wrote it for.

TM: It’s about testing matrix-claims. So this is really about logical matrices?

NB: It’s a trivial paper, I wouldn’t say it’s about logic at all. I didn’t remember what the spiel was of that article.

TM: And then you have the next book appearing, with Thomas Steel, *The logic of questions and answers*. There’s really lots of books appearing. That was translated into Russian, later on?

NB: Yes.

TM: So that’s ’76. Can you say something about “The two property”, in *The relevance logic newsletter*? Or maybe about the newsletter?

NB: Oh, the newsletter was just a very small publication that existed for a few years. The two property was simply that all the theorems had to have an even number of variables. It had to do with a little fragment; I had conjectured that its only theorem was $A \to A$, including fat formulas, but always $A \to A$.

TM: And then there is "How a computer should think", in 1977, which is the first paper on the useful four-valued logic?

NB: I don't know, first or second, there were two of them.

TM: Yes, because then there's the paper with the title "A useful four-valued logic".

NB: I put those together for the book.

TM: And there is an abstract with Michael McRobbie.

NB: Currently the president of Indiana University.

TM: There you go. Succeeding Mike Dunn, or how is this?

NB: No, Michael was never president.

TM: But he had some higher office there as well?

NB: He was the Dean of Informatics. He attracted McRobbie.

TM: So we're in 1978 and your book indexing software BINDEX gets published. And then in '79 your piece with McRobbie that you had reported on, on tableaux for relevance logic, was published. And then more work that builds up towards the second volume of *Entailment*, I guess? "A consecution calculus for positive relevant implication with necessity".

NB: That was for Volume two.

TM: And then there is a piece on the development of modal and relevance logics in Agassi's *Modern logic*, from 1980: "Modal and relevance logic: 1977".

NB: They didn't like it that I put the year on it, but I insisted.

TM: What is it about the year?

NB: That's how far my little history went.

TM: And then, even though they dismantled your darkroom, you still got a photograph published in Haugeland's book, *Mind design*.

NB: I did.

TM: What's on the photo actually? I never looked it up.

NB: It's the head and shoulders of John Haugeland.

TM: Then there is an application of your logic of questions and answers in the Montague grammar project, and a piece on teaching logic and relevance logic.

NB: I don't remember how that came about either.

TM: "Logika voprosov i otvetov".

NB: That was at a conference or meeting.

TM: So there are some Russian translations of your work, by Smirnov. And then there is display logic. So was Heinrich Wansing already around, had you met him back then in '82 when you published your first work on display logic?

NB: I don't remember, I'd certainly done the work before I met with Heinrich.

TM: And then the paper on "Gupta's rule of revision theory of truth", that became the nucleus for your joint book with Anil Gupta, on the revision theory of truth?

NB: Although I am listed as a co-author of *The revision theory of truth*, it was really Anil's book.

TM: Then there are papers on the semantics of questions, this is in the linguistics community, right? With von Stechow as one of the editors. And then there is an abstract on display logic in *JSL*, and there is more work for the second volume of *Entailment*. And more work on the semantics of questions. And then you have a joint paper with Anil in 1987, in the *Journal of Philosophy*: "A note on extension, intension, and truth". It's quite noteworthy for a paper with a logic focus to appear there?

NB: I guess, I don't remember. It was largely Anil's paper.

TM: And here's another book you brought to life, Charles Hamblin's *Imperatives*, in 1987. You did the editing on that one as well?

NB: No, I just wrote the foreward.

TM: That's an important book.

NB: I think it is, yes. Though it's not much read today.

TM: You put it on the list of required reading for *Facing the future*.

NB: That's right!

TM: And then there is a German translation of the prosentential theory of truth paper, in *Der Wahrheitsbegriff*, edited by Lorenz Puntel. And then 1988 sees the first paper in the "seeing to it that" program, *stit*, with Mickey Perloff. And then there are many more papers in that line to come. But also the relevance logic program is going on: You write for the *Directions in relevance logic* volume edited by Norman and Sylvan. And then in 1990 you have a paper with Gerry Massey, on semantic holism. How did that come about?

NB: He brought me a logic problem of some kind and I solved it. But he wrote up the stuff about the history—he put in a kind of medieval historical context, with lots of scholars doing something or other. It was nice what he did.

TM: And there's a paper, also in 1990, "Declaratives are not enough". I think I've never read that, a shame.

NB: Not really. It's just a kind of a rehash of taking questions and imperatives seriously.

TM: There is further development in the display logic program. And a second paper on *stit* by you and Mickey. And in the year after, 1991, a paper that you authored for *Erkenntnis*, on the same topic, "Before refraining: concepts for agency". And also in the first DEON workshop, *Deontic logic in computer science*, you had a paper on *stit*. You seeded it into the computer science community very early. I think of DEON as a computer science conference mostly. So Jean-Jules Meyer and Roel Wieringa edited that. In 1992 there is more on developing *stit*-theory.

NB: And twenty years later people start reading that paper!

TM: Good for them!—If you have an *annus mirabilis*, it's 1992? There is "Backwards and forwards in the modal logic of agency", in *Philosophy and phenomenological*

research. Also the second volume of *Entailment* comes out, and your "Branching space-time" paper appears in *Synthese*. How long had you been working on that?

NB: I had been working on it for several years, I'm sure. But memory dims.

TM: And it's really an outgrowth of the *stit*-project, in a way.

NB: It is.

TM: I learned this very late, because I didn't connect initially, but it makes perfect sense, of course. It's all about causal independence and how you model that.

NB: Exactly.

TM: Moving on, in '93 there is the book with Anil Gupta, on *The revision theory of truth*, and also you're promoting *stit* with Mickey Perloff. And I guess in connection with the work on circular definitions for the revision theory book, there is your paper "On rigorous definitions"?

NB: No, that was entirely separate. I mean, this paper mentions circular definitions in about a paragraph at the end. But mostly it's just … the literature on rigorous definitions is so poor, I thought I wasn't going to interfere with anyone by adding. I shouldn't say poor, but thin.

TM: I think our Utrecht PhD student Sebastian Lutz was able to make good use of that. And then there is your paper with Mitch Green, "Indeterminism and the thin red line", which is a classic, I guess.

NB: In certain circles.

TM: The circles are growing. And the next paper is on substructural logics.

NB: "Life in the undistributed middle". I tried to find that paper, I guess yesterday or the day before, but it's not accessible from where I am.

TM: And there is a paper on analytic tableaux, for linear logic is that?

NB: Yes.

TM: And more work building up toward *Facing the future*. There is also the other approach to *stit*, deliberative *stit* rather than the achievement *stit*, with Jeff Horty, in a paper for the *Journal of philosophical logic*, 1995. And a piece in the *Festschrift* for Ruth Barcan Marcus. Did you know her well?

NB: Academically, but pretty well. Over a lot of years. I have always felt affection for her.

TM: Unfortunately I never met her, I think she had some very good influences. So there is a paper on "The display problem" in a volume on proof theory that Heinrich Wansing organized. And then the second paper on BST, "Branching space-time analysis of the GHZ theorem". This was when the Hungarians came to invade Pittsburgh, László Szabó and Miklós Rédei.

NB: That was not a good paper.

TM: Well, I think it was really an important step. And then you went to the Prior memorial conference, there is a paper on "Agents in branching time" in the proceedings volume, edited by Jack Copeland. It may actually be that this is the first time that

I saw your name, something around that time, because I had looked at that volume. And there is a paper on "The very idea of an outcome"; the notes connected to that paper, I think, brought Tomasz Placek into BST.

NB: That's right.

TM: Then here is work in '97 building up to the deontic part of *Facing the future*. Did you have the idea of wrapping that all up in a book by then? It must have been at around that time.

NB: It must have been. Mickey and I had the idea fairly early, that we were writing chapters of a book.

TM: And you were getting them out as articles. There is another *dstit* paper with Jeff Horty; he had his book out in the same year as *Facing the future*. And there is this very nice paper on "Concrete transitions" in '98, digging out this notion of a transition in von Wright, which is not easy to find in his work, I think. Did he react to the paper?

NB: I don't think so. I read it at his conference, yes. But I don't recall that he had anything to say about it.

TM: In his work, transitions are really buried in the idea of "and next", which is much less subtle then what you make of it. "Truth by ascent"', from 1999, I think I haven't read that. What's in there?

NB: Intimations of the revision theory.

TM: And then this is a piece that I guess will be very hard to find, "Modest notions of free will and indeterminism", in the *Proceedings of the Creighton Club*, 1999—but it's one of the few pieces in which you say something about the role indeterminism can play for us.

NB: It comes out in some other papers.

TM: Yes, it does. Then 2000 has the invited paper to the *Advances in Modal Logic* conference in Leipzig. This is your piece on "Double time references: Understanding speech-act modalities in an indeterministic setting". So this is, as it were, maybe the next step in the (anti-)Thin red line project. And then in 2001 we have the book, *Facing the future*, together with Mickey Perloff and Ming Xu. You did the typesetting for that all by yourself?

NB: Yes.

TM: I think most of the following items I know pretty well. With an intermission of a couple of years, there is the next paper on applying the branching space-times ideas and working out one of the central notions in the BST-framework, namely the idea of modal correlations, or as you call it, "funny business". And that's when we met, at the Workshop on non-locality and modality in Kraków in 2001 that Tomasz Placek organized with Jeremy Butterfield. Can you say something about the period in between, from 1992 when BST appeared, until that paper? What made you pick up this idea again? It seems like in the late '90s most of your work is really on what becomes *Facing the future*, and BST is not part of that. But it's of course clear in a

sense that BST is going to be the next step. It's interesting that it wasn't on the map for a number of years, and then it becomes really big after ten years.

NB: I don't remember how that went.

TM: The next paper continues the analysis of funny business, "No-common-cause EPR-like funny business in branching space-times", in *Philosophical Studies*, 2003. And then there is your paper from the trip to Guangzhou, China, "Agents in branching space-times". And a paper on non-classical logics from the same trip, right?

NB: Just so.

TM: And then a similar paper, "Agents and agency in branching space-times", appears in Daniel Vanderveken's volume, *Logic, thought and action*, in the book series *Logic, Epistemology, and the Unity of Science*. And then, for me this is maybe the most important paper, the "Causae Causantes" paper in the *British journal for the philosophy of science* from 2005.

NB: I think it's one of my most important papers. It will take another twenty years before anybody reads it …

TM: But then it will have been there, sitting there for people to discover. I think it's great, it's really a masterpiece, a very good paper. It did a lot with me. And then there is your paper with a biographical title, "Under Carnap's lamp", written under Carnap's lamp. The lamp is now in Konstanz, you gave it to their archive, right?

NB: Yes.

TM: And there was some ongoing work on BST; you worked with Matt Weiner in the '90s, Matt had some results and he added a postulate about the relative ordering of suprema of a chain, and there are ideas about building up a probability calculus in BST. You wrote this up and put it together into a nice form, "How causal probabilities might fit into our objectively indeterminsitic world", *Synthese* 2006.

NB: Yes.

TM: And in the *Festschrift* for Hugues Leblanc in 2005 there is a paper on a "Branching histories approach to indeterminism and free will", one of the other places where your approach to free will comes to the fore.

NB: Yes, and I think this has some of the one we passed over on free will …

TM: … the Creighton Club … What I find fascinating now going over this list, which of course I've looked at before, is to see how long some of the lines are that you draw, within your work. Because in 2006 we really have the first paper on Bressan. That's almost 35 years after having brought out the book. I think, we take this list, and we go to the University administrators and we tell them, "You see, *this* is how research goes".

NB: Give it time!

TM: You don't throw everything overboard every three years! The 2006 paper is a nice piece—extremely useful, I think, for making Bressan accessible.

NB: I'm disinclined to think that, I think that what you and I are working on now is making Bressan accessible, but I don't think the 2006 paper does much …

TM: Well, we will see. Then there is "Prosentence, revision, truth and paradox", in *Philosophy and Phenomenological Research* in 2006—again, picking up an earlier topic.

NB: That was, I think, something having to do with Tim Maudlin's book on truth. I'm not quite sure I remember. It was an occasional piece anyway.

TM: Ok, so the next one, 2007 … we're almost done. This is the piece on "Propensities and probabilities", in *Studies in History and Philosophy of Modern Physics*, that you wrote for the proceedings of the 2005 Kraków workshop on branching space-times.

NB: Tomasz Placek found so many mistakes in it.

TM: But there is a new version of it. And there is this very nice piece motivating BST, that you so kindly wrote for the Stuhlmann-Laeisz *Festschrift* that I edited, which is now also out in an updated form in *Synthese* 2012. And also a very nice piece on parameters of truth for my little book on time, *Philosophie der Zeit*. And I remember very fondly, of course, the paper that we published with Kohei Kishida, on "Funny business in branching space-times: infinite modal correlations". And then the written list stops, but we know that there is a lot more, of course, and a lot more to come. There is all your work on topological issues in BST, from the collaboration with Tomasz Placek, and recently, our joint work on "Case-intensional first order logic"; your contribution to this book, on internal cases, is part of that enterprise, which is continuing.—Thanks, we've done the full list of publications!

NB: Have we!

TM: What's the most important paper? I guess, by your reactions to it, the *Causae Causantes* paper from 2005 is one candidate …

NB: I think so.

TM: The "Useful four-valued logic" paper from 1978 is another one?

NB: That's been a very popular paper. Probably the most read paper, but I don't know important it was.

TM: Thanks, Nuel.