GESTURE STUDIES 4

Integrating Gestures

The interdisciplinary nature of gesture

edited by
Gale Stam and Mika Ishino

John Benjamins Publishing Company

Integrating Gestures

Gesture Studies (GS)

Gesture Studies aims to publish book-length publications on all aspects of gesture. These include, for instance, the relationship between gesture and speech; the role gesture may play in social interaction; gesture and cognition; the development of gesture in children; the processes by which spontaneously created gestures may become transformed into codified forms; the relationship between gesture and sign; biological studies of gesture, including the place of gesture in language evolution; and gesture in human-machine interaction. Volumes in this peer-reviewed series may be collected volumes, monographs, or reference books, in the English language.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/gs

Editor

Volume 4

Integrating Gestures. The interdisciplinary nature of gesture
Edited by Gale Stam and Mika Ishino

Integrating Gestures

The interdisciplinary nature of gesture

Edited by

Gale Stam
National Louis University

Mika Ishino
Kansai Gaidai University, Kobe University and University of Hyogo

John Benjamins Publishing Company

Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI: 10.1075/gs.4

Library of Congress Cataloging-in-Publication Data

Integrating gestures : the interdisciplinary nature of gesture / edited by Gale Stam, Mika Ishino.

p. cm. (Gesture Studies, ISSN 1874-6829 ; v. 4)

Includes bibliographical references and index.

1. Language and languages--Study and teaching. 2. Gesture. 3. Second language acquisition. I. Stam, Gale. II. Ishino, Mika. III. Title. IV. Series.

P53.4117I58 2011

808.5--dc22 2010051882

ISBN 978 90 272 2845 1 (Hb ; alk. paper)

ISBN 978 90 272 8720 5 (Eb)

An electronic version of this book is freely available, thanks to the support of libraries working with Knowledge Unlatched. KU is a collaborative initiative designed to make high quality books Open Access for the public good. The Open Access ISBN for this book is 978 90 272 8720 5.

John Benjamins Publishing Co. • P.O. Box 36224 • 1033 ME Amsterdam • The Netherlands
https://benjamins.com

Table of contents

PART I

Nature and functions of gestures

CHAPTER 1

Introduction

Mika Ishino[1] and Gale Stam[2]
Kansai Gaidai University, Kobe University, University of Hyogo[1]
and National Louis University[2]

Interest in gesture has existed since ancient times. However up to the twentieth century, it was primarily studied in two ways – as it related to rhetoric (from Roman times to 1700), i.e., how gestures could enhance a speaker's presentation and as a precursor of oral language (from 1700 to 1900) for the information it could give about language evolution (for an extensive discussion of the history of the field of gesture studies, see Kendon 1982, 2004). It was not until 1941 that gesture began to be studied in a systematic manner in human interaction with the ground-breaking work of David Efron (1941/1972), and it was not until the 1970s with the work of David McNeill (1979, 1981) and Adam Kendon (1972, 1980) that speech and gesture were viewed as aspects of the same process (see Kendon 2004, Stam 2006, Stam & McCafferty 2008), and the field of modern gesture studies was born.

Gestures are ubiquitous and natural in our everyday life, and they convey information about culture, discourse, thought, intentionality, emotion, intersubjectivity, cognition, and first and second language acquisition. Additionally, they are used by non-human primates to communicate with their peers and with humans. Consequently, the field has attracted researchers from a number of different disciplines such as anthropology, cognitive science, communication, neuroscience, psycholinguistics, primatology, psychology, robotics, sociology and semiotics, and the number of modern gesture studies has grown. The purpose of this volume is to present an overview of the depth and breadth of current research in gesture. Its focus is on the interdisciplinary nature of gesture, and the twenty-six chapters included in it represent research in the following areas: the nature and functions of gestures, language development, use in the classroom and in problem-solving, discourse and interaction, and music and dance. Before we present the areas of research, we will present an overview of what gestures are.

What are gestures?

The term 'gestures' has many different meanings, and the gestures that each researcher examines are not always the same. This, of course, can make cross-researcher comparisons difficult at times. Nevertheless, the gestures that each author in this volume deals with are all visible bodily actions employed intentionally and meaningfully. This is a broad definition that covers the many different aspects of gestures.

Kendon (1982) has classified gestures into four types: gesticulation, pantomime, emblem, and sign language. According to the presence or the absence of a language-like property, McNeill (1992: 37) lined up these four types on a continuum and termed it 'Kendon's continuum.' This continuum was later elaborated into four continua by McNeill (2000, 2005). According to this continuum, gesticulations are "idiosyncratic spontaneous movements of the hands and arms accompanying speech" and obligatorily accompany speech (McNeill 1992: 37). Spontaneous gestures are distinct from emblems and sign languages in that they are not regulated by convention and are global, "the meanings of the parts are determined by the whole" and synthetic, "different meaning segments are synthesized into a single gesture" (McNeill 1992: 41). Spontaneous gestures are synchronous with speech and often occur with elements of high communicative dynamism, i.e., contrastive, focused or new information (McNeill 1992, 2002). In addition, their strokes tend to co-occur with prosodic peaks (Nobe 1996, 1998). They perform the same pragmatic functions as speech (Kendon 1980, McNeill 1992). These gestures and their co-occurring speech can represent the same entities, or they can complement each other, where the gestures indicate an aspect present in the speaker's thought, but not expressed through speech.

Spontaneous gestures serve many functions (Stam 2006, in press; Stam & McCafferty 2008) and may serve several functions simultaneously (Heath 1992). They may add information that is not present in individuals' speech or emphasize information that is there (Goldin-Meadow 1999, McNeill 1992). They may serve to lighten speakers' cognitive load (Goldin-Meadow et al. 2001) and improve their performance in other areas. They may help speakers organize spatial information for speaking and aid in the conceptual planning of speech (Alibali et al. 2001). They may also indicate transition in cognitive and language development (Goldin-Meadow & Alibali 1995, Goldin-Meadow & Butcher 2003, Iverson & Goldin-Meadow 2005). In addition, they may be used to retain turns during conversation (Duncan 1972), and listeners may gesture to indicate their active involvement in the conversation (de Fornel 1992). Finally, gestures may indicate speech production difficulties (Feyereisen 1987) and facilitate lexical retrieval (Butterworth & Hadar 1989, Hadar & Butterworth 1997, Krauss & Hadar 1999, Krauss et al. 1995, Morrel-Samuels & Krauss 1992, Stam 2001, in press).

Emblems are culturally codified gestures and include such gestures as the 'OK sign' and the 'two-thumbs-up sign' in the United States or the Dutch gesture for *lekker* 'tasty, yummy' (flat hand moving back and forth roughly parallel to the head at a small distance, 1–2 inches from the ear). The semantic contents of emblems are understandable without speech, though they can co-occur with speech (Morris, Collett, March, & O'Shaughnessy 1979). Emblems are signs, and they have "standards of well-formedness" and "the OK sign must be made by placing the thumb and index finger in contact" (McNeill 1992: 38). Furthermore, they are not part of language in that they do not have syntax as sign languages do. Many emblems go back to Roman times (Morris et al. 1979), and the same form may have various meanings as well as different meanings in different cultures. Emblems are learned gestures and are, therefore, teachable (for reviews and studies on emblems, see Brookes 2001, Calbris 1990, Ekman & Friesen 1969, Kendon 1981, Morris et al. 1979, Ricci Bitti & Poggi 1991).

With pantomime, we find meaningful gestures that are by definition never accompanied by speech. Pantomimes can depict objects, actions or an entire story. These are the types of gestures people make when they are playing a game like charades or when they are asked to explain an action without speech.

Sign languages, such as American Sign Language (ASL), are full-fledged languages. They are composed of signs which are codified gestures that have linguistic properties and are equivalent to lexical words (McNeill 2005). While it is possible to speak while signing, sign language can be fully understood without speech.

Some authors in this volume deal with gestures which spontaneously co-occur with speech, while others deal with gestures which do not accompany speech. The contrast between those gestures that occur with speech and those that occur without have important implications for the essence of what gestures are.

Typology and coding

Spontaneous gestures can be analyzed in terms of their semiotic properties, and several different classification systems have been developed for categorizing them (Bavelas 1992, Cosnier 1982, Cosnier & Brossard 1984, Cosnier & Vaysse 1997, Efron 1941/1972, Ekman & Friesen 1969, Freedman 1972, McNeill 1992, McNeill & Levy 1982). The majority of these are variations of Efron's (1941/1972) original system of batons, ideographs, deictics, physiographs, and emblems (for a detailed discussion of the various classification systems, see McNeill 1992; Kendon, 2004; Rimé & Schiaratura 1991). The system adopted by many authors in this volume is in line with that of Kendon or McNeill.

In relation to their form and meaning, McNeill (1992, 2005) has classified coverbal spontaneous gestures into four major categories: (1) iconics (2) metaphorics (3) beats and (4) deixis. Gestures that provide "a *representation* of the content of an utterance" are termed representational gestures (Kendon 2004: 160) and include iconic and metaphoric gestures. Iconic gestures express images of actual objects and/or actions. Metaphoric gestures, on the other hand, express images of the abstract. Beats stress important words with baton-like movements that are timed to occur with thematic content in discourse and do not depict any imagery. Beats can, however, be superimposed upon iconic or metaphoric gestures. Importantly, beats often manifest pragmatic significance despite their simplicity in form and/or movement. They occur at the meta-level of discourse and highlight information: they may introduce new characters and new themes, summarize action, and accompany repairs. Deictic gestures are not representational; they are pointing movements. Depending on the existence or the presence of their referents, pointing (or deictic) gestures are classified into two types: concrete and abstract deixis (McNeill, Cassell, & Levy 1993). Concrete deixis makes a reference to physically present entities while abstract deixis are points directed towards a seemingly empty space. McNeill, Cassell, and Levy (1993) found that abstract deixis provides new references in space. Contrastively, concrete deixis conveys a reference in its generation. Claiming that "none of these categories is truly categorical," McNeill (2005: 41) has advocated that gestures be analyzed in terms of dimensions, i.e., iconicity, metaphoricity, temporal highlighting, deixis, and social interactivity rather than types because a single gesture often shows multiple dimensions. While emphasizing that it is not easy to determine which categories are dominant or subordinate and that in some gestures, each dimension is not equally displayed, McNeill (2005) introduces the notion of saliency. McNeill mentions that saliency is of theoretical interest and has an impact on the occurrence of "the kind of imagery that occurs" through gesture (McNeill 2005: 43). This claim by McNeill is confirmed in some of the chapters in this volume which employ his typology of gestures.

Areas of research

The research in this volume is divided into six sections or themes: the nature and functions of gesture, first language development and gesture, second language effects on gesture, gesture in the classroom and in problem-solving, gesture aspects of discourse and interaction, and gestural analysis of music and dance.

Nature and functions of gestures

As previously mentioned, gestures are multifunctional: some communicate (Kendon 1994), while others serve cognitive functions. What can be said about the nature of gestures is very much dependent on the paradigm in which they are studied. The chapters in the first section provide us with more insight into the nature and various functions of gesture and give us several models for future gesture research. The studies themselves include gestures that accompany speech as well as those that do not.

Erica A. Cartmill and Richard W. Byrne (Chapter 2) analyze gestures of twenty-eight captive orangutans and show that there are some tight relationships between gesture forms and meanings and that non-human primates can communicate their intentions with one another through gestures. In Chapter 3, David McNeill and Claudia Sowa present evidence from a study in which speech was prevented. Their study sheds light on the ontogenesis of morphemes of gestures as well as the functions of gestures. They demonstrate that in the absence of speech, participants' gestures become more like a language (segmented and analytic) with morphemes (i.e., parings of forms and meaning), syntagmatic values, and standards of form emerging unlike the gestures that co-occur with speech.

Janet Bavelas, Jennifer Gerwing, Meredith Allison, and Chantelle Sutton (Chapter 4) report on a micro-analysis they conducted of grounding steps in dyadic dialogues. Their study shows that participants in discourse make use of abstract pointing gestures to accumulate common ground and indicate understanding. They suggest that their method of analysis could be useful for future research in the understanding of gestures in different situations. In Chapter 5, Autumn Hostetter, Martha Alibali, and Sheree Schrager examine whether speakers' motivation to communicate has an impact on the rate or size of the gestures speakers produce. They find that there is no effect on the frequency of gestures; however, there is an effect on the size of the gestures. Speakers produced a higher proportion of larger gestures when they want their interlocutors to cooperate with them. Their findings suggest that speakers vary the size of their gestures based on whether they want to communicate information clearly or not.

Katharina Hogrefe, Wolfram Ziegler, and Georg Goldenberg (Chapter 6) present a method, the Hamming Distance, for the analysis and transcription of the physiological and kinetic aspects of hand gestures that does not rely on the analysis of the concurrent speech. This method provides gesture researchers a way to measure in how many formal features two gestures differ from each other. Furthermore, they argue that application of this method opens up the potential to conduct quantitative analyses of gestures and is useful when analyzing the data of individuals with severe language disorders.

Many gesture researchers assume that speech and gesture of one person is an integral unit of thinking. Maria Graziano, Adam Kendon, and Carla Cristilli (Chapter 7) argue that speech and gesture among interlocutors is a unified unit of thinking, and they call gestures repeated completely or partially by an interlocutor 'parallel gesturing.' Based on the claim that such 'parallel gesturing' is a gesture-speech ensemble (Kendon 2004), a single-unit of production, they describe parallel gesturing in adult-child conversations and show that parallel gesturing in adult-child conversations serves as a way for interlocutors to show their understanding of the speaker's utterance and alignment to the other's expressive style. Furthermore, they suggest that just as children must acquire adult pronunciation, they must also acquire adult gestures to fit within the gesturing style of their community.

First language development and gesture

The section on first language development and gesture includes research on children from infancy through school age. Researchers in this area work from the assumption that the gestures children produce serve as a window onto their cognitive and/or first language development. Claire Vallotton (Chapter 8) shows that preverbal infants as early as 9 months can create gestural sentences and as early as 10 months can reply to a caregiver's gesture and converse in the gestural mode. Maria Fusaro and Claire Vallotton (Chapter 9) examine infant signs and their environment and find that infants begin to produce gestures modeled by their caregivers when they are about ten months of age. Maria Zammit and Graham Schafer (Chapter 10) suggest that child-directed communication is systematically modified both linguistically and gesturally because it scaffolds language learning. Mats Andrén (Chapter 11) shows that parents give significantly more elaborated responses when children performed *sustained* index finger pointing gestures, and in so doing, he also raises a question of timing of gesture phases. Şeyda Özçalışkan and Susan Goldin-Meadow (Chapter 12) observe the spontaneous gestures of children interacting with their parents from 14 to 34 months of age and find that the number and types of iconic gestures that children produce significantly increase around 26 months.

Kazuki Sekine (Chapter 13) investigates the development of spatial perspectives in preschool age children by looking at how children use gestures in route descriptions, i.e. whether they used *a survey map perspective* which views the environment from a fixed, single viewpoint or *a route map perspective* which takes the form of an imaginary journey. His findings suggest that an understanding of the environment from a bird's-eye viewpoint and the use of a survey map perspective is available as early as 5 years of age, an age much younger than was originally thought such a perspective was acquired, around 8 to 9 years of age. Focusing on the use of representational gestures in narratives, Olga Capirci, Carla Cristilli, Valerio De Angelis, and Maria Graziano (Chapter 14) analyze how children develop their competence in the formal and semantic aspects of gesture. They show that there are formal and semantic properties of gesture children have to acquire in order to develop their communicative competence. In addition, they argue that gesticulation and sign languages, previously identified as the two extremes of "Kendon's Continuum," share some characteristics in common. Hannah Sowden, Mick Perkins, and Judy Clegg (Chapter 15) present a case study of a child with Autistic Spectrum Disorder (ASD), age 2:6 years, interacting with his teacher. As mentioned earlier, speech and gesture is assumed to be an integral unit. However, in children with autism, the development of both language and gesture is impaired. Sowden, Perkins, and Clegg investigate gesture forms, discourse functions of the gestures and the dynamic nature of gesture form and function in the interaction between the child with ASD and the teacher and find that in the beginning the teacher makes use of deictic gestures in order to draw the child's attention and the child immediately imitates the teacher's gestures. Additionally, Sowden, Perkins, and Clegg find that the teacher produces iconic and emblematic gestures in the later phase in the interaction and the child with ASD imitates them as well. They argue that the child's gestures serve a back-channeling function to display his engagement in the interaction.

Second language effects on gesture

The two chapters in the section second language effects on gesture investigate how speaking more than one language affects gesture use. Meghan Zvaigzne, Yuriko Oshima-Takane, Fred Genesee, and Makiko Hirakawa (Chapter 16) investigate whether the presence of mimetics (sound-symbolic words) in language influences children's verbal and gestural descriptions by conducting a cross-linguistic comparison of cartoon narrations by Japanese and French monolingual and bilingual children. While Japanese is rich in mimetics, French is not. The results of their study suggest that the presence of mimetics in Japanese has an impact on co-speech gesture use in the course of the description of motion events; however, this was more evident in the monolingual children than the bilingual ones. Kendra Newbury (Chapter 17) examines the emblematic gesture use of border bilinguals in northern Uruguay, where Portuguese, the traditional language, is being supplanted by Spanish, the national language. She finds that as the speakers shift languages, they also shift emblematic gestures, but that the gesture shift lags behind the linguistic shift.

Gesture in the classroom and in problem-solving

The role that gestures play in communication and cognitive processes both in the classroom and during problem-solving is explored in this section. Susan Gerofsky (Chapter 18) offers an observational analysis of students' elicited gestures of graphs of mathematical functions. Her results show that the students who internalize the graphs and make large gestures are more able to notice mathematically salient features than those whose gestural motions are more restricted. She claims that these findings have implications for the teaching of mathematics in secondary schools. Mitchell Nathan and Martha Alibali (Chapter 19) demonstrate that teachers facilitate intersubjectivity or common ground by their use of gestures in the classroom during conversational repairs and the presentation of a novel (target) representation. They point out that this is done through both linking gestures and gestural catchments. They stress both the personal and social roles that gestures play in establishing intersubjectivity.

Mingyuan Chu and Sotaro Kita (Chapter 20) investigate how gestures reveal the process of problem solving in mental rotation tasks and what role gestures play in the development process. Their results show that when adults solve new problems with regard to the physical world, they experience deagentivization and internalization processes which are similar to the processes that young children experience. In the problem-solving task, adults first simulate the manual manipulation of the stimulus through gestures and then are eventually able to solve the problem without gestures.

Gesture aspects of discourse and interaction

The chapters in this section present evidence of how gestures vary in discourse and interaction. Stephani Foraker (Chapter 21) examines how information structure in discourse is reflected in gestures and whether speakers use different gestures in their presentation of new and given information in discourse. Her study shows that the function of gestures produced reflect differences between new and given information. Katie Wilkin and Judith Holler (Chapter 22) also investigate how gestures reflect information structure in discourse and common ground. Their findings suggest that common ground, i.e., definite articles in their study, is associated mainly with iconic gestures and action information, and no common ground, i.e., indefinite articles, mainly with abstract deictic gestures and entity information.

Claire Maury-Rouan (Chapter 23) examines nonverbal parameters of reported speech and perspective shifts and finds that prosodic cues, head movements, posture shifts, and facial expressions mark reported speech. Furthermore, her findings suggest that a shift in posture, typically a shift in head position mark perspective shifts. Adopting the framework of conversational analysis, Lorenza Mondada and Florence Oloff (Chapter 24) study overlaps in turn-taking. They show how speakers use gestures to display their treatment of different kinds of overlap as being more or less problematic, and whether a speaker continues to gesture is dependent on whether the overlap is viewed as collaborative or competitive. They argue that overlaps need to be looked at from a multimodal perspective as it provides a better understanding of how participants use all resources to manage their talk-in-interaction.

Gestural analysis of music and dance

The two chapters in the section gestural analysis of music and dance provide examples of the type of research that is being done on gesture and the arts. Isabella Poggi (Chapter 25) observes and analyzes a choir conductor's multimodal behavior and his social interaction in music performance. She points out that a conductor as the leader of the choir must pursue common goals shared by the singers and himself to perform beautiful music. Using an annotation scheme, Poggi shows that bodily behavior and facial expressions such as gaze, eye and mouth movements of the conductor play a significant role in his pursuing these goals while conducting. Ellen Campana et al. (Chapter 26) describe an interactive art installation, *Handjabber*, which uses a Laban framework of movement to analyze how people use their bodies to communicate and collaborate. They discuss technical aspects of the installation as well as their experience using the installation to explore participants' metaphoric gestures, body orientation, and interpersonal space.

Conclusion

A wide range of research from various disciplines is represented in this volume. Although it does not cover all fields of current gesture research such as sign languages, neurolinguistics, and artificial intelligence/robotics, it provides a flavor of the type of research that is currently being done on gesture and its interdisciplinary nature. We hope that you enjoy reading the research and are inspired to do some yourself.

References

Alibali, M. W., Kita, S., Bigelow, Lisa J., Wolfman, Chai M. and Klein, S. M. 2001. "Gesture plays a role in thinking for speaking." In *Oralité et Gestualité: Interactions et Comportements Multimodaux dans la Communication*, C. Cavé, I. Guaïtella, and S. Santi (eds), 407–410. Paris: L'Harmattan.

Bavelas, J. B. 1992. "Interactive gestures." *Discourse Processes* 15: 469–489.

Brookes, H. J. 2001. "O clever 'He's streetwise'. When gestures become quotable: The case of the clever gesture." *Gesture* 1 (2): 167–184.

Butterworth, B. and Hadar, U. 1989. "Gesture, speech and computational stages: A reply to McNeill." *Psychological Review* 96 (1) : 168–174.

Calbris, G. 1990. *Semiotics of French Gesture*. Bloomington: Indiana University Press.

Cosnier, J. 1982. "Communications et langages gestuels." In *Les Voies du Langage: Communications Verbales Gestuelles et Animales*, J. Cosnier, A. Berrendonner, J. Coulon and C. Orecchioni (eds), 255–304 Paris: Bordas/Dunod.

Cosnier, J. and Brossard, A. 1984. «Communication non verbale: co-texte ou contexte ?» In *La Communication Non Verbale*, J. Cosnier and A. Brossard (eds), 1–29. Paris: Delacaux et Niestlé.

Cosnier, J. and Vaysse, J. 1997. "Sémiotique des gestes communicatifs." In *Geste, Cognition et Communication, Nouveaux Actes Sémiotiques, 52–54,* J. Cosnier, J. Vaysse, P. Feyereisen, and G. Barrier (eds), 7–28. Limoges: PULIM.

de Fornel, M. 1992. "The return gesture: Some remarks on context, inference, and iconic gesture." In *The contextualization of language*, P. Auer and A. Di Luzio (eds), 159–176. Amsterdam: John Benjamins.

Duncan, S. J. 1972. "Some signals and rules for taking speaking turns in conversations." *Journal of Personality and Social Psychology* 23 (2): 283–292.

Efron, D. 1941/1972. *Gesture and Environment*. Morningside Heights, NY: Kings Crown Press.

Ekman, P. and Friesen, W. V. 1969. "The repertoire of nonverbal behavior: Categories, origins, usage, and coding." *Semiotica* 1: 49–98.

Feyereisen, P. 1987. "Gestures and speech, interactions and separations: A reply to McNeill 1985." *Psychological Review* 94 (4): 493–498.

Freedman, N. 1972. "The analysis of movement behavior during the clinical interview." In *Studies in Dyadic Communication*, A. W. Siegman, and B. Pope (eds), 153–175. New York: Pergamon Press.

Goldin-Meadow, S. 1999. "The role of gesture in communication and thinking." *Trends in Cognitive Science* 3: 419–429.

Goldin-Meadow, S. and Alibali, M. W. 1995. "Mechanisms of transition: Learning with a helping hand." In *The Psychology of Learning and Motivation, Vol. 33*, D. Medin (ed), 117–157. New York: Academic Press.

Goldin-Meadow, S. and Butcher, C. 2003. "Pointing toward two-word speech in young children." In *Pointing: Where Language, Culture, and Cognition Meet*, S. Kita (ed), 85–107. Mahwah, NJ: Lawerence Erlbaum Associates.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D. and Wagner, S. 2001. "Explaining math: Gesture lightens the load."*Psychological Science* 12: 516–522.

Hadar, U. and Butterworth, B. 1997. "Iconic gestures, imagery, and word retrieval in speech." *Semiotica* 115 (1/2): 147–172.

Heath, C. 1992. "Gesture's discreet tasks : multiple relevancies in visual conduct and in the contextualisation of language." In *The contextualization of language*, P. Auer and A. Di Luzio (eds), 101–127. Amsterdam: John Benjamins.

Iverson, J. M. and Goldin-Meadow, S. 2005. "Gesture paves the way for language development." *Psychological Science* 16 (5): 367–371.

Kendon, A. 1972. "Some relationships between body motion and speech: An analysis of an example." In *Studies in Dyadic Communication*, A. W. Siegman, and B. Pope (eds), 177–210. New York: Pergamon Press.

Kendon, A. 1980. "Gesticulation and speech: Two aspects of the process of utterance. In *The Relation Between Verbal and Nonverbal Communication*, M. R. Key (ed), 207–227. The Hague: Mouton.

Kendon, A. 1981. "Geography of gesture." *Semiotica* 37:129–163.

Kendon, A. 1982. "The study of gesture: Some remarks on its history." *Recherches Semiotique/ Semiotic Inquiry* 2: 45–62.

Kendon, A. 1994. "Do gestures communicate? A review." *Research on Language and Social Interaction* 27 (3): 175–200.

Kendon, A. 2004. *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Krauss, R. M. and Hadar, U. 1999. "The role of speech-related arm/hand gestures in word retrieval." In *Gesture, Sign and Speech,* L. Messing and R. Campbell (eds), 93–116. New York: Oxford University Press.

Krauss, R. M., Dushay, R. A., Chen, Y. and Rauscher, F. 1995. "The communicative value of conversational hand gestures." *Journal of Experimental Social Psychology* 31: 533–552.

McNeill, D. 1979. *The Conceptual Basis of Language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McNeill, D. 1981. "Action, thought, and language." *Cognition* 10: 201–208.

McNeill, D. 1992. *Hand and Mind*. Chicago: University of Chicago Press.

McNeill, D. 2000. *Language and Gesture*. Cambridge: Cambridge University Press.

McNeill, D. 2002. "Gesture and language dialectic." *Acta Linguistica Hafniensia* 34: 7–37.

McNeill, D. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.

McNeill, D. and Levy, E. 1982. "Conceptual representations in language activity and gesture." In *Speech, Place, and Action: Studies in Deixis and Related Topics*, R. J. Jarvella and W. Klein (eds), 271–295. New York: John Wiley and Sons.

McNeill, D., Cassell, J. and Levy, E. T. 1993. "Abstract deixis." *Semiotica* 95: 5–19.

Morrel-Samuels, P. and Krauss, R. M. 1992. "Word familiarity predicts temporal asynchongy of hand gestures and speech." *Journal of Experimental Psychology: Learning, Memory and Cognition* 18: 615–623.

Morris, D., Collett, P., Marsh, P. and O'Shaughnessy, M. 1979. *Gestures: Their Origins and Distribution.* New York: Stein and Day.

Nobe, S. 1996. *Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/threshold model of gesture production.* Unpublished doctoral dissertation, University of Chicago.

Nobe, S. 1998. "Synchrony between gestures and acoustic peaks of speech: A cross-linguistic study." In *Oralité et Gestualité: Communication Multimodale, Interaction*, S. Santi, I. Guaïtella, C. Cavé and G. Konopczynski (eds), 543–548. Paris: L'Harmattan.

Ricci Bitti, P. E. and Poggi, I.. 1991. Symbolic nonverbal behavior: Talking through gestures. In *Fundamentals of Nonverbal Behavior*, R. S. Feldman and B. Rimé (eds), 433–457. Cambridge: Cambridge University Press.

Rimé, B. and Schiaratura, L. 1991. Gesture and speech. In *Fundamentals of Nonverbal Behavior*, R. S. Feldman and B. Rimé (eds), 239–281. Cambridge: Cambridge University Press.

Stam, G. 2001. Lexical failure and gesture in second language development. In *Oralité et Gestualité: Interactions et Comportements Multimodaux dans la Communication*, C. Cavé, I. Guaïtella and S. Santi (eds), 271–275. Paris: L'Harmattan.

Stam, G. 2006. *Changes in Patterns of Thinking with Second Language Acquisition*. Unpublished doctoral dissertation, University of Chicago.

Stam, G. and McCafferty, S. G. 2008. "Gesture studies and second language acquisition: A review." In *Gesture: Second Language Acquisition and Classroom Research*, S. G. McCafferty and G. Stam (eds), 3–24. New York: Routledge.

Stam, G. in press. Gestes et recherche de mots en langue seconde. In R. Vion, A. Giacomi and C. Vargas (eds). *La corporalité du langage : Multimodalité, discours et écriture*, Hommage à Claire Maury-Rouan. Aix en Provence : Presses Universitaires de Provence.

CHAPTER 2

Addressing the problems of intentionality and granularity in non-human primate gesture

Erica A. Cartmill[1] and Richard W. Byrne[2]
University of Chicago[1] and University of St Andrews[2]

Any study of communicative gesture must identify which movements are purposeful (*intentionality*) and which examples of movements should be grouped into a single gesture (*granularity*). Where researchers studying human gesture are aided by linguistic context, researchers studying non-human primates must rely on their subjects' movements alone to address these questions. We propose an approach to intentionality and granularity in non-human primate gesture based first on the possibility that only some, but not all individuals that use particular movements do so as intentional gestures, and second on the premise that gestures found to have specific meanings reflect real-world distinctions made by the animals. We apply this approach to the behavior of 28 captive orangutans and identify 64 distinct gestures, 29 of which have specific, predictable meanings.

Introduction

The study of gesture in non-human primates (hereon "primates") presents challenges beyond those encountered in the study of human gesture. Accompanying speech or conversational context can be used to interpret the meanings of human gesture (Iverson & Goldin-Meadow 1998), and it may actually be impossible to understand the meanings of human gestures if they are removed from their spoken context (McNeill 2000). Primate gestures, however, are not produced within a known linguistic framework; it is thus difficult to determine their meanings. Here, we discuss some of the special challenges facing students of primate gesture and propose a systematic approach to studying meanings of gestures. We advocate locating each example of gesture within its communicative and social context, taking into account the behavior of both the gesturer and recipient in communicative exchanges of varying length. We begin by describing two of the most difficult questions facing gesture researchers – (1) how does one know whether a movement is communicative (*intentionality*), and (2) how does one know whether a set of examples constitutes a single gesture (*granularity*). We explain how these problems are approached in human gesture research and suggest how they might be addressed in primate gesture research. To answer the first of the two questions, we describe an analysis of intentionality based on the behavior of each individual; this allows for the possibility that some but not all individuals that use a particular movement do so as a communicative gesture. To answer the second question, we argue that potential gestures exist as meaningful signals for the individuals who use them if they show predictable meanings across multiple examples.

We use findings from our 3-year study of orangutans to illustrate the effectiveness of an individual, context-based approach to studying primate gesture. Our general methodology centers around a study of meaning, based on both the goal of the gesturer and the outcome of the exchange, and includes gestures produced on their own as well as during extended social interactions. Our focus on identifying specific meanings in primate gestures may come as a surprise to those familiar with other work on ape gesture. Most recent studies of ape gesture have focused on the relative flexibility of gestures compared to vocalizations, and have used this contextual flexibility to support gestural origin theories of language evolution (see Arbib et al. 2008, Call & Tomasello 2007, Pollick & de Waal 2007). The ability to employ gestures flexibly in different ways rather than automatically in response to stimuli demonstrates that apes use gestures *intentionally*. However, if gestures are used so flexibly that there is no predictable relationship between form and meaning, then they are not used intentionally *to communicate something*. Our approach to gesture meaning measures the probability that a particular form is successful at achieving a particular social goal: gestures that very frequently achieve a particular goal are deemed to have that meaning. Redirecting the discussion of ape gesture from flexibility to meaning will open up new comparisons to human language and will allow researchers to test the way in which they define ape gestures.

Identifying intentional gestures

Researchers studying human gesture determine that movements are gestures by requiring that they be part of a communicative act (Iverson & Goldin-Meadow 1998, Kendon 2004). When produced concurrently with speech, the communicative nature of the act is clear. When produced in isolation, clues such as eye contact are used to determine that the gesture itself is communicative (Goldin-Meadow 2004), though discourse-level analysis renders this a fairly straightforward task since solitary gestures are most often contextualized within a larger spoken exchange. Primate researchers, on the other hand, must identify which movements are gestures without the help of an overt communicative context.

Since non-effective movements in primates are typically produced without accompanying vocalizations, researchers must determine whether potential gestures themselves constitute a communicative act, relying on social clues and evidence within the movements to identify communicative *intentions*. Eye contact, body orientation, response waiting, and persistence are all used as evidence for intentionally communicative gesturing (Call & Tomasello 2007, Genty et al. 2009, Pika et al. 2005).

But complicating the question of intentionality is the possibility that a movement used by one individual as an intentional gesture might also be used by another, but in a non-intentional way. Our approach to intentionality builds on previous work that attempted to identify the intentionality of primate gestures according to strict criteria (see Call & Tomasello 2007); we make the important addition of requiring that intentionality be identified in each individual's use of a particular gesture. Previously, (see Liebal et al. 2004, Liebal et al. 2006, Pika et al. 2003, Pika et al. 2005) it has been assumed that if a gesture were used intentionally by *one* or a *few* individuals, then it was an intentional gesture for *all* individuals. Like Genty et al. (2009), we exclude all examples of a gesture made by individuals who did not show at least one clearly intentional use of that gesture, thereby allowing for the possibility that some individuals in a population might use a movement as an intentional gesture and some might not.

Addressing the granularity of analysis

To identify meaningful gestures, researchers studying both human and primate gesture must address the question of how to categorize individual examples into definable, meaningful gestures. The way in which a movement sequence is segmented into analyzable units and how those units are categorized into definable gestures (i.e. the "granularity" of analysis) will affect what types of analyses are possible and may significantly impact the conclusions of the study. On the one hand, finely dividing complex movements allows for a more detailed analysis of timing and subtlety of meaning. This analysis is effective in revealing the tight association between speech and movement in human discourse (e.g. McNeill 1992), but risks overlooking broad commonalities in form by focusing too closely on the specific gestural elements and is too laborious to apply to large datasets. On the other hand, considering complex movements as whole units (on a level somewhat analogous to noun or verb phrases in speech) is simpler and is successful in identifying commonalities across many examples (e.g. Goldin-Meadow 2003), but risks defining gesture types too generally to reveal much specificity in meaning.

Imagine, for example, if we were to group all oscillating movements of the head into a single gesture type. In this case, nodding and shaking the head would be considered to be the same gesture, and we would conclude that it had a very ambiguous meaning. The possibility of making this type of error affects both human and primate gesture researchers who must therefore keep the problem of granularity in mind when attempting to determine which movements constitute definable gestures and have particular meanings.

Researchers studying primate gesture must tackle the problem of granularity without accompanying speech providing any clues as to how to segment and categorize movements. If researchers apply too fine a granularity to their definitions of gestures, this would lead to an overestimation of the number of gesture types (Figure 1a). This overestimation could lead researchers to conclude that some gesture types were idiosyncratic or limited to highly-specific situations, when a broader analysis would have ignored these small variations and revealed that all individuals use the same gesture type. Underestimation of gesture types by using too coarse a granularity (Figure 1b) could similarly overlook important variations in meaning by erring in the other direction: lumping many different movements into a single type, when the primates themselves perceive differences between them.

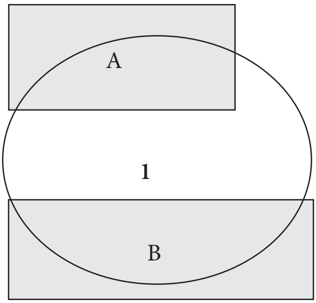


Figure 1a. Gestures defined by too fine a granularity. (The white circles represent gestures 1 and 2 as perceived and used by a group of primates. The grey boxes represent the gestures (A, B, C) as defined by a human observer.)

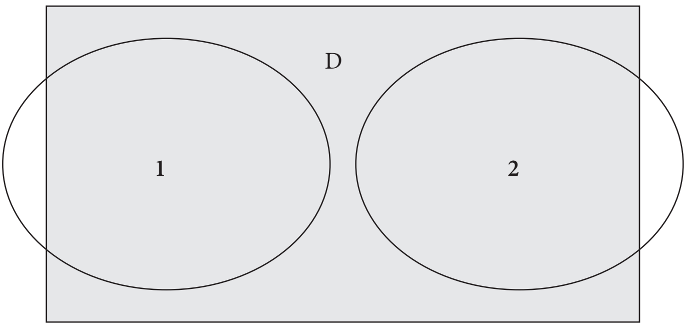


Figure 1b. Gestures defined by too coarse a granularity. (The white circles represent gestures 1 and 2 as perceived and used by a group of primates. The grey box represents the gesture (D) as defined by a human observer.)

The granularity of gesture definitions is of great importance in assessing whether gestures vary between individuals and whether any gestures carry specific meanings. This is a problem common to gesture studies of both humans and primates. Accurately determining the level of analysis is made more complicated by the fact that a structural variable might make a difference to the definitions of some gestures but not to others. For example, whether a movement is performed while holding an object has a large effect in distinguishing *reaching* from *showing* an object, but makes no difference to *pointing* (which could be done with or without an object in hand).

Although intentionality and granularity must both be separately addressed in any study of the meaning of primate gestures, they also interact: movements must be determined to be intentional in order to be defined as gestures, and the granularity of gesture definitions will affect analyses of repertoire size and gesture meanings. Thus, one way to test the adequacy of the gesture definitions at a particular granularity is to determine whether any of the observer-defined gestures have distinct meanings. If they do, it is likely that the granularity of their definitions is not too large. However, attempts to maximize the specificity of gesture meanings by dividing broadly-defined gestures into more narrow ones must be balanced by the desire to avoid defining all gestures as idiosyncratic. If all gestures were defined as idiosyncratic, no further analysis would be possible as each individual's gestures (or even each instance of an individual's gestures) would be considered unique, and thus distinct from all others.

Granularity and gesture meaning

We propose to address granularity through an assessment of gesture meaning: gestures with consistent meanings used by several individuals are deemed to have an appropriate level of granularity, and those without consistent meanings are investigated further to determine whether redefinition of the gesture could increase consistency of meaning. Our attribution of meaning to gestures is systematic and takes into account both the gesture's goal and the recipient's response, a significant departure from analyses of meaning typical in animal communication studies primarily based on the recipient's response (see Hauser 2000). Additionally, we suggest that analysis of meaning should be based on all types of exchanges involving gesture (single gesture events, longer sequences and turn-taking events), whereas some previous studies restricted analyses of meaning to single gesture-reaction events to simplify identification of recipient responses (e.g. Genty et al. 2009). Including all types of gestural exchanges in analyses of meaning is a more naturalistic and more comprehensive approach that should lead to a more representative account of how gesture is used within non-human populations.

Since our approach to evaluating the granularity of the analysis involves identifying consistency in gesture meanings, it is necessary to identify the meanings of the gestures as intended and perceived by the study subjects. We did not expect that each gesture would have a one-to-one correspondence with a particular meaning. However, if primates are using gesture as a primary means of communication, then it should be expected that at least some of their gestures communicate specific meanings. Our study of orangutan gestures led us to conclude that this is, indeed, the case.

Assessing meaning in orangutan gestures

We began our study of orangutan gesture by opportunistically filming social interactions that occurred amongst 28 orangutans at several European zoos. We first selected all movements performed in the presence of other orangutans that did not appear to have a direct function (e.g. reaching towards an object would be included, but picking it up would not). We then grouped all of these movements into "potential gestures" according to their similarities along certain structural variables: modality, body part, movement, force, speed, and use of an object. We then determined which of these potential gestures were used as intentional communicative signals by applying a strict set of intentionality criteria to all examples and retaining only those gestures performed by individuals who had used those particular gestures at least once in an intentional manner. We deemed an example of a gesture to be intentional if it was (1) directed towards another, with (2) the objective of obtaining a particular goal, and (3) employed flexibly rather than as an automatic response to a stimulus (Bruner 1981, Pika et al. 2005, Tomasello & Call 2007). We used the gaze direction of the signaler prior to gesturing to determine whether visual and auditory gestures had a specific recipient. (Tactile gestures were directed at a recipient, by definition.) In order to establish whether the signaler had an intended goal in gesturing, we looked for evidence that the signaler "expected" a response from the recipient; measures of expected response included response waiting, gaze alternation, persistence, and using modalities appropriate to the attentional state of the recipient (e.g. visual gestures when the recipient is looking).

To address the issue of whether or not our definitions of gestures accurately accorded with the perceptions of the species (i.e. whether the granularity was right) we tested our judgments of gesture granularity by comparing gesture form to meaning. Take the earlier example of grouping nodding and shaking of the head as a single gesture. In this case, one could differentiate nodding from shaking by comparing each example's structure to its contextual meaning. Through that juxtaposition, direction of movement would emerge as a dividing variable, splitting an ambiguous gesture into two meaningful ones. By attributing meanings to a set of apparently successful orangutan gestures and determining whether a particular gesture was consistent in its meaning across examples, we were able to identify ambiguous gestures and reassess our definitions of those gestures in an attempt to better match the way in which orangutans used them.

A systematic approach to assessing meaning

We propose that the process of working out the meaning of a primate gesture should combine a measure of gesturer intent with one of recipient response (for more details on this approach, see Cartmill & Byrne 2010). For each act of gesture, we may be able to identify both an apparent goal of the gesturing individual and a subsequent reaction of the recipient. The reaction of the recipient may either fulfill the gesturer's goal or not – and may be a lack of response altogether. If a reaction does not fulfill the gesturer's goal, he or she might continue to gesture until getting the desired reaction or giving up entirely (see Genty et al. 2009). We define a recipient reaction that causes the gesturer to stop gesturing as an *interaction outcome* (Figure 2a). In interactions consisting of a single gesture and reaction, the reaction immediately following the gesture is the interaction outcome. In longer interactions, the final reaction of the recipient is the interaction outcome for all gestures.

In order to determine whether the interaction outcome satisfied the gesturer's goal, the gesturer must be ascribed a goal every time he or she gestures (Figure 2b). In our study, we ascribed a gesturer goal to each example of gesture based only on (1) the general context of the exchange (e.g. whether either one was feeding), (2) our knowledge of the identity of the individuals involved (e.g. whether an infant was gesturing to her mother), and (3) whether the form of the gesture seemed designed to effect a particular response (e.g. a pushing gesture would be more likely to indicate a goal of moving another than a hitting gesture would). Our attribution of goals to gesturers was thus *not* based on the observed responses in that exchange. This meant that we could ascribe a goal to a signaler and then be surprised when a non-expected reaction caused the gesturer to cease gesturing. We did not assume that every gesture in a sequence shared the same goal, though all shared the same *interaction outcome*. We also assumed that a gesturer always intended to elicit an active behavior from a recipient; thus, the goal could never be "no reaction." The goals we attributed to gesturers were: Affiliate/Play, Stop action, Sexual contact, Look towards, Look at/Take object, Share food/object, Co-locomote, or Move away. Once goals had been attributed to each example of gesture, we defined any examples in which the presumed goal matched the interaction outcome as having *goal-outcome matches* (Figure 2c).

In the example gesture sequence shown in Figure 2c, gestures 1 and 3 have goal outcome matches. This means that the gesturer appeared successful in fulfilling her goal of eliciting a particular reaction from the recipient. If gesture 1 and 3 frequently had the same goal-outcome match when they were produced by other individuals or by the same individual at other points, then we would define them as having *meaning*.

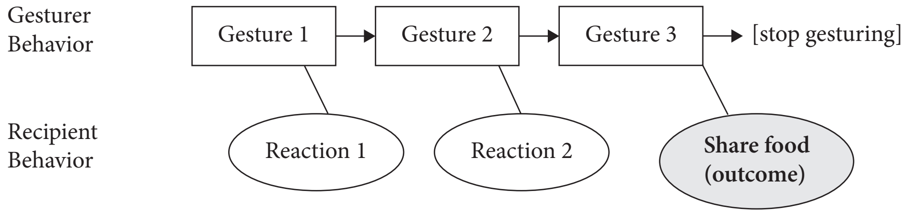


Figure 2a. Directly observable gestures and reactions in a sequence of gestures.

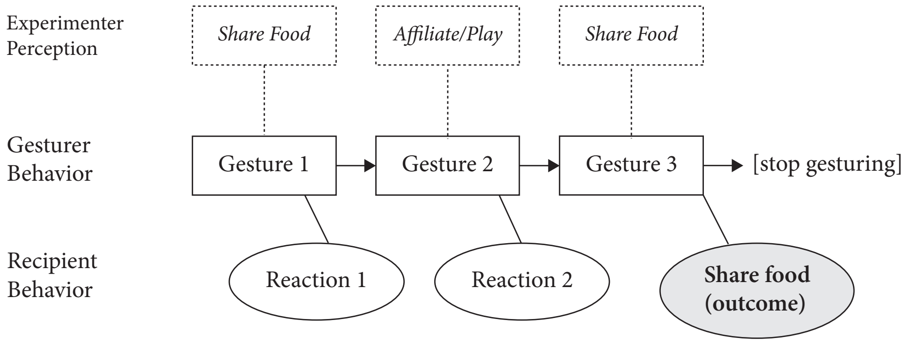


Figure 2b. Gestures, reactions, and experimenter-ascribed goals of the gesturer in a sequence of gestures.

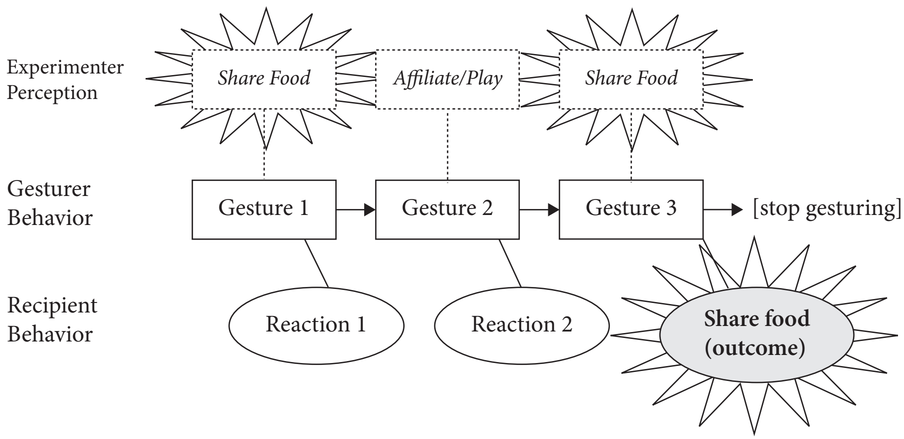


Figure 2c. Goal-outcome matches in a sequence of gestures. Note that both Gesture 1 and Gesture 3 have goal-outcome matches.

Using meaning to evaluate granularity

Determining that a gesture has meaning provides support for the analysis of granularity: if a gesture is found to have the same goal-outcome match in many examples, then it is likely that the gesture exists as a meaningful signal for the primates and is not an artifact of the human observer's interpretation. A lack of meaning for a gesture does not necessarily mean that that gesture doesn't exist. But, if such ambiguous gestures can be combined or subdivided into non-idiosyncratic, meaningful gestures then it is likely that the redefined gestures would provide a more accurate reflection of the real-world gestures. By removing or adding structural variables from the definition of an ambiguous gesture (thereby increasing or decreasing the granularity of the definition), it should be possible to achieve a more accurate definition and determine which variables are important in distinguishing a particular gesture from others.

In our study of orangutan gestures, we used goal-outcome matches as a means of investigating gesture meaning as well as testing the granularity of our definitions. Once we had applied intentionality criteria to all examples of gestures and reduced our dataset to only intentionally-communicative movements, we found that more than half of all observed gestures had goal-outcome matches. Importantly, only 15% had outcomes that conflicted with the presumed goal of the gesturer, the other non-matching cases occurred when the recipient did not respond to the gesturer or looked away.

We defined three degrees of observable meaning for gestures – *tight*, *loose*, and *ambiguous* – based on how frequently they were used with a single goal-outcome match (Cartmill 2008, Cartmill & Byrne 2010). All gestures with tight and loose meanings had one of six meanings: *Affiliate/Play*, *Stop action*, *Look at/Take object*, *Share food/object*, *Co-locomote*, and *Move away*. Where gestures had either loose meanings or were ambiguous, we investigated further in the hope that we could redefine the gestures so as to identify gestures with tight meanings from among the range of ambiguity. We considered including new variables in the definitions, prioritizing different variables, or combining existing gesture types. We found that almost all of the loose or ambiguous meaning gestures in our sample could be redefined by taking into account one of these variables so that a subset of the examples could be defined as a new gesture with a tight meaning. The possibility of new definitions indicated that our original definitions did not always reflect orangutans' perceptual distinctions between gestures. This demonstrates that human observers are liable to make unreliable judgments about what is and is not a gesture in another species and that corrective processes to observers' first attempts can be very valuable.

Though it would have been possible for us to redefine most ambiguous gestures by adding additional structural or social variables, doing so would have resulted in many gestures that were idiosyncratic or were restricted to certain age pairings. We reasoned that social variables in particular – such as the gesturer's identity, age, and relationship to the recipient – should not be used to redefine gestures, since they affected the *use* of gestures (particularly their effectiveness), but not their *forms*. We decided to create

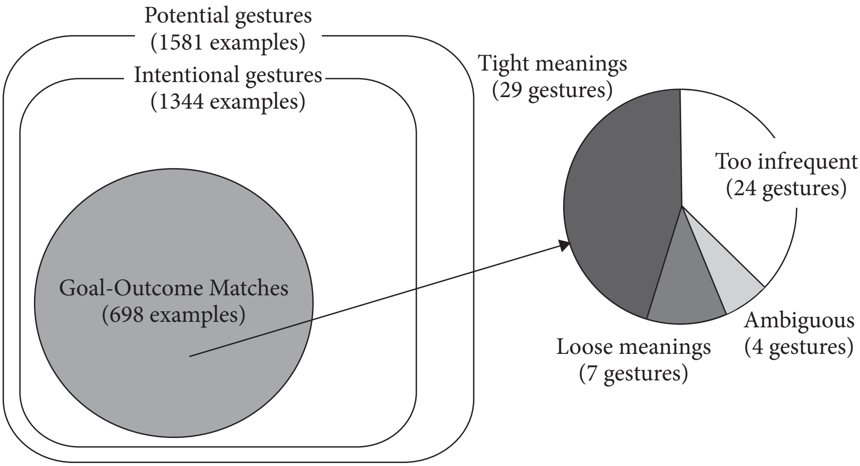


Figure 3. Frequencies of examples of intentional gestures and goal-outcome matches. Examples of goal-outcome matches consist of 64 gestures, categorized into those with tight meanings (29), loose meanings (7), ambiguous meanings (4), and those too infrequent to analyze further (24).

only two new gestures by including the variable "target location" (the place towards which a gesture is directed). When target location was included in the set of defining variables, two new gestures could be defined as having tight meanings. After redefining these gestures, our final set of orangutan gestures consisted of 64 intentional gestures, 29 of which had tight meanings, 7 of which had loose meanings, and 4 of which were ambiguous (for examples of specific gestures and their meanings, see Cartmill & Byrne 2010). The remaining 24 gestures were observed fewer than four times during the study and were deemed to be too infrequent to be included in the analysis of meaning. Figure 3 illustrates our process of narrowing down the observed movements to identify meaningful gestures.

Conclusion

Our approach to studying non-human gesture helps address the problems of intentionality (how do you know whether a movement is communicative?) and granularity (how do you know whether a set of examples constitutes a single gesture?). In our study of orangutans, we deemed movements to be communicative if they met criteria for intentional signals and required that each individual use a potential gesture intentionally before adding that gesture to his or her observed repertoire. We tested the granularity of our definitions of gesture by determining whether any gestures had consistent goal-outcome matches across examples. We concluded that non-idiosyncratic gestures showing this consistency exist as perceptible, meaningful gestures for the orangutans themselves; the successful assignment of tight meaning to 29 (out of 64) gestures supports the granularity of our gesture definitions. It is essential that researchers studying gestures in animals not shy away from discussing intentionality and granularity as it is precisely these variables that allow us to challenge our assumptions and definitions and to more accurately identify how other species perceive and use gesture.

Acknowledgements

The authors thank the staff of Apenheul Primate Park, Durrell Wildlife Conservation Trust, and Twycross Zoo for their support, and J. G. Foster, K. Brown, and M. Cartmill for their reviews of the manuscript and contributions to figure design. Funding for the project came from the University of St Andrews and the Russell Trust.

References

Arbib, M., Liebal, K. and Pika, S. 2008. "Primate vocalization, gesture, and the evolution of the human language." *Current Anthropology* 49 (6): 1053–1076.

Bruner, J. 1981. "Intention in the structure of action and interaction." In *Advances in Infancy Research*, L. Lipsitt (ed), 41–56. Norwood: Ablex.

Call, J. and Tomasello, M. 2007. *The Gestural Communication of Apes and Monkeys*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cartmill, E. A. 2008. *Gestural Communication in Orangutans (pongo pygmaeus and pongo abelii): A Cognitive Approach*. Unpublished doctoral dissertation, University of St Andrews. St Andrews.

Cartmill, E. A. and Byrne, W. 2010. "Semantics of primate gestures: Intentional meanings of orangutan gestures." *Animal Cognition* 13 (6): 793–804.

Genty, E. Breuer, T., Hobaiter, C. and Byrne, R. W. 2009. "Gestural communication of the gorilla (*gorilla gorilla*): Repertoire, intentionality and possible origins." *Animal Cognition* 12 (3): 527–546.

Goldin-Meadow, S. 2003. *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Harvard University Press.

Hauser, M. D. 2000. "A primate dictionary? Decoding the function and meaning of another species' vocalizations." *Cognitive Science* 24 (3): 445–475.

Iverson, J. M. and Goldin-Meadow, S. 1998. *The Nature and Functions of Gesture in Children's Communications*. San Francisco: Jossey-Bass.

Kendon, A. 2004. *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Liebal, K., Pika, S. and Tomasello, M. 2004. "Social communication in siamangs (*symphalangus syndactylus*): Use of gestures and facial expressions." *Primates* 45 (1): 41–57.

Liebal, K., Pika, S. and Tomasello, M. 2006. "Gestural communication of orangutans (*pongo pygmaeus*)." *Gesture* 6 (1): 1–38.

McNeill, D. 1992. *Hand and Mind*. Chicago: The University of Chicago Press.

McNeill, D. 2000. *Language and Gesture*. Cambridge: Cambridge University Press.

Pika, S., Liebal, K., Call, J. and Tomasello, M. 2005. "The gestural communication of apes." *Gesture* 5 (1/2): 41–56.

Pika, S., Liebal, K. and Tomasello, M. 2003. "Gestural communication in young gorillas (*gorilla gorilla*): Gestural repertoire, learning and use." *American Journal of Primatology* 60 (3): 95–111.

Pika, S., Liebal, K. and Tomasello, M. 2005. "Gestural communication in subadult bonobos (*pan paniscus*): Repertoire and use." *American Journal of Primatology* 65 (1): 39–61.

Pollick, A.S. and de Waal, Frans B. M. 2007. "Ape gestures and language evolution." *Proceedings of the National Academy of Sciences of the United States of America* 104 (19): 8184–8189.

CHAPTER 3

Birth of a Morph*

David McNeill[1] and Claudia Sowa[2]
University of Chicago[1] and University of Erlangen[2]

When speech is prevented, gesture morphs emerge de novo. The morphs include standards of good form and syntagmatic values. However, when speech is present, gestures do not attain morph status, do not have standards of form or syntagmatic values.

What is a morph?

Morphemes are the atoms of language, the undecomposable units of form and meaning, fixed, repeatable, listable, and maintained according to convention. We see all these factors except conventions in the two studies to be reviewed, the wordless Snow White narration from Ralph Bloom's (1979) thesis and the gestured motion event descriptions from Gershkoff-Stowe & Goldin-Meadow (2003).

To identify a morph, certain hallmarks can be sought. A morph is a Saussurian sign: a pairing of signifier and signified, the unsplittable two sides of a coin in his metaphor. This holds for all signs, including non-morph gestures. To be a morph, in addition, the sign must be patterned on two levels – Hockett's duality of patterning: patterned both as a meaning and as a *form* (cf. Hockett & Altmann 1968). The signifier may or may not be iconic, but a standard of form or 'pattern' is crucial. It is form patterning that differentiates morphs from metaphors, recurrence, priming, reference, and catchments – all of which also produce gesture recurrence, as we describe. The question is, does the *form* of the gesture meet standards? Does form, *qua* form, reflect something more than iconicity, a standard to which the form is being held?

* Based on a presentation at the 3d conference of the International Society for Gesture Studies, Evanston, IL, June 18–21, 2007. Supported by the Spencer Foundation, the U. S. National Science Foundation STIMULATE program, Grant No. IRI-9618887, "Gesture, Speech, and Gaze in Discourse Segmentation", and the National Science Foundation KDI program, Grant No. BCS-9980054, "Cross-Modal Analysis of Signal and Sense: Multimedia Corpora and Tools for Gesture, Speech, and Gaze Research."

It is not recurrence

Morphs recur but recurrence, while necessary, is not sufficient for a gesture to be a morph – at least 6 causes of gesture recurrence can be identified, only one of which is actually being a morph and meeting standards of good form:

- **Expected metaphoric imagery**, in which a culturally given or 'expected' metaphor appears in gesture form. These are gestures such as the concept of something progressing as a rotation in space or a conduit gesture of presenting a concept or meaning as an object in the hand (the conduit was originally identified with verbal material in Reddy 1979 and Lakoff & Johnson 1980; see Figure 10b for an illustration). Such gestures embody expected metaphors but owe nothing to standards of form. They recur, not because of standards but because the metaphor recurs, and the gestures are iconically depicting the vehicles of these metaphors.
- **Unexpected metaphoric imagery**, in which a metaphor in gesture is created on the fly and then recurs downstream for a period in the discourse. Such a recurring gesture is strictly ephemeral. An example is a metaphor for an 'antagonistic force' described in McNeill (2008), a gesture depicting the rounded shape of a bowling ball as the implement of an 'antagonistic force' in the cartoon story being recounted. The construal of the bowling ball as a metaphor was an individual product, not shared with anyone else. Such recurrence is not a morph itself but is a kind of premonition of one and may be a first step toward a morph standard.
- **Referential iconicity**, in which imagery recurs for the same reference object. The various 'up inside the pipe' gestures later in Figure 9 illustrate the phenomenon. Different speakers hit on similar imagery in which 'Sylvester' is an extended index finger, a gesture triggered initially as an iconic image of his ascent and compression inside a drainpipe that then appeared in other contexts. This also could become stabilized into a kind of proto-morph.
- **Morphology**, in which a gesture is required to meet a standard of form. This is the target case and the only one in which it is appropriate to speak of form standards and syntagmatic value formation. We see such morphs *de novo* in the two experiments.
- **Priming**, in which a prior action makes a similar later action more likely but the evidence does not favor it as a factor in morph birth. While it can produce recurrences, we see in the Snow White experiment that form standards came first, and did so immediately, and thus could not have derived from priming.
- **Catchments**, in which recurring gesture features (not always whole gestures) carry a discourse theme. Again, no form standard is present, just the continuing thematic content recurring in gesture (see the "it down" case study in McNeill 2005).

To summarize, with 6 causes, only one of which is morphemic, recurring gestures alone are not sufficient to confer morph status.

Morph hallmarks

If standards of form are the *sine qua non* of morphs, how can we identify them? To answer the form question, several probes can be used; namely,

- Do people *recognize violations of gesture form*? "OK" with the middle finger on the thumb, instead of the canonical forefinger, may convey precision but it is not the "OK" morph-emblem. See Figure 1 for an example of a canonical "OK" gesture, adhering to the form standard. While the Snow White narrator, not using speech, was sensitive to form violations by his listener, in cartoon narrations with speech, gestures may be more or less transparent but there is no sense in which they can be termed 'not well-formed' in accord with some standard of form for the gesture itself.
- If two gestures have different meanings but similar forms, is there some form difference, however minor, added to at least one of them solely to *maintain distinctiveness*? The addition has no function of its own, as with the crooked little finger of the Warlpiri Sign Language for "truck", added to distinguish it from the otherwise identical sign for "child" (Figure 2, image from Kendon 1988). The finger crook has no other function. We see something similar in Snow White gestures (see 'Ritualization' and Figure 7).
- Do people *have intuitions of good form*? If a gesture appears to be made the 'right way', or if one makes it not in that way and it seems 'wrong', or if it changes meaning, we can attribute it to intuitions of good form. The "OK" sign made with the middle finger rather than forefinger violates one's intuitions of how it should be formed. Intuitions are the individual speaker's experience of the systematics of a code. We infer that the SW listener developed intuitions of how the King and Queen gestures should be formed from his own uses of them. These differed in certain respects from the narrator's versions (see Figure 6); since the differences were consistent, there is this hint that intuitions had arisen rather than slavish imitation (and the narrator, in further confirmation, rejected them as 'violations', which can be called diverging intuitions).

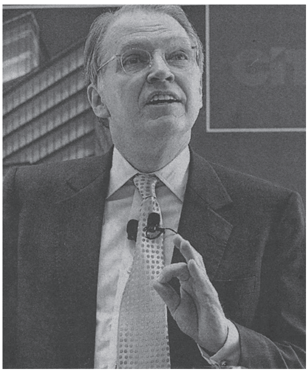


Figure 1. The "OK" emblem adhering to the form standard: forefinger on thumb tip, other fingers extended. Image of former CIT CEO Jeffrey Peek in 2006, from the WSJ, July 22, 2009.

Figure 2. Warlpiri sign for "truck", showing elevated small finger to distinguish it from an otherwise similar form for the unrelated meaning of 'child'. From Kendon (1988).

– Finally, are there *geo-cultural zones* in which different standards have evolved? An example is pointing, which shows cultural specificity, taking different forms across cultures. With Westerners and many others the extended forefinger prototypically performs pointing. While alternatives may be understood, they are not the norm. Elsewhere the norm is a flat hand, and in Laos one norm is lip protrusion, as shown in Figure 3 (Enfield 2001). A future Ralph Bloom experiment, in which the narrator and the listener go their separate ways and use the new morphs with others, could evolve different form standards, and this in fact seems to have begun with the listener's King and Queen gestures.

Figure 3. Jahai (Laos) lip point. From Enfield (2001).

Whence a morph?

It is almost impossible to answer this question with speech alone. Even a novel morpheme like "to Google" fits the established patterns of English. A different approach is to study the emerging home signs of deaf children born to hearing parents (Goldin-Meadow & Mylander 1984) or the successive cohorts of Nicaraguan Sign Language (Senghas & Coppola 2001). A third approach, followed here, is to describe the gestures created by hearing adults when speech is denied. We describe examples of new morphs in the wordless Snow White (SW) narration from Ralph Bloom's 1979 thesis and in the gestured video vignette descriptions from Gershkoff-Snow & Goldin-Meadow (2002). The morphs created in these experiments may reveal aspects of the general process of morph formation, the same in its essentials as those with spoken morphs and signs.

Merely having a gesture symbol does not a morph create. A crucial condition is that the gestures should be the sole vehicles of communicative exchanges, as in SW and the vignettes. When gestures accompany speech, as in Canary Row (CR) cartoon narrations in which speech co-occurred, they recur, as with the extended finger 'Sylvester' gestures in Figure 9, but are cut loose from consistent meanings and are not maintained. Communication creates a social unit in which form standards, analysis, repeatability and combination emerge naturally. To spin the metaphor, communicative exchange is the midwife to the birth of morphs.

Standards of form and their emergence in gesture-only communication

As we have seen, to find gesture morphs we need to distinguish them from repeating gestures – metaphors, iconic gestures and catchments – which may look morphemic but are not. Morphs are more than iconic gestures – they are also shaped by standards of form. A gesture morph implies, among other qualities, that a gesture meets, consciously or not, standards of form and is open to violations, such that changes of form may cancel the morph.

When do standards emerge? Several things must take place. First, the gesture becomes analytic, as opposed to global. Second, it becomes stable and repeatable and thus extractable from context. Third, it is distinguished from other morphs, and fourth, it combines with other gestures to create syntagmatic values.

In SW, in the gesture for the Queen especially, gestures exhibit a number of these morph hallmarks: stable and repeatable, analytic rather than holophrastic, and extractable from context. These qualities are shown in Figs. 4, 5 and 6.

The morphs underwent streamlining with time, form changes that enhance speed and execution, but did not lose their mutual distinctiveness. This is seen in the comparison of Figs. 4 (the first occurrences of Queen and King) and 5 (later occurrences) in which streamlining was accompanied by loss of iconicity but the iron-clad preservation of distinctiveness can be seen.

It is also in this situation that we may see added form differences to distinguish gestures with different meanings whose forms would otherwise converge. In Figure 7, the 'crown' component of 'Queen' reduces to a single brow-sweep but adds an upright index finger, which probably stems from the original up-down 'crown' but now distinguishes 'Queen' from other sweeping motions, and is a microcosm of the addition of the uniconic crooked finger shown for the Warlpiri sign in Figure 2.

In a post-experiment interview, the SW narrator could provide descriptions of the gestures and their distinguishing features, so the contrasts had solidified into conscious form standards. Moreover, the narrator criticized the listener's variations of these forms as 'violations', so the standards were, for him at least, normative.

Finally, there was also dialogic use of 'King' and 'Queen' by the listener (Figure 6). The crucial 'has-breasts' distinction was preserved, as was the two-morpheme structure of each gesture ('has-crown' + 'has-muscles' or + 'has-breasts') but a 'dialect' difference appeared in how the 'has-crown' and 'has-muscles' features were formed – the first without revolution at the head, the second with a downward slice with the hands in front of the chest, suggesting 'flat chested' rather than upraised iconic clenched arms for 'has-muscles'. The primary speaker had used the 'flat chested' form himself earlier but did not continue with it. The meaning of Morph 2 may thus have shifted along with the form shift to something like 'flat-chested' or 'has-*no*-breasts' for King, losing touch with the original 'has-muscles' meaning and making explicit a distinction (has vs. has-no breasts) not encoded by the narrator. So linguistic drift got in motion almost immediately – another microcosm, of the divergence of languages in this case. Had the listener been required to use this morph set with fresh listeners, a kind of experimentally engineered migration, a new branch of the original language and something like the 'geo-cultural' variation of pointing in Figure 4 could have been set in motion.

Ritualization

The 'Queen' offers the best window on how an initially iconic morph can, over time, lose iconicity. 'Queen' never loses the distinctive feature of 'having-breasts' but the other feature, 'crown', which is non-contrastive, steadily turns less iconic, although it never totally disappears. Figure 5 showed some of this process; Figure 7 shows a more complete history, starting with the two hands circling up-and-down for 'crown' but changing to a single hand sweeping across the brow with an upright index finger; this finger possibly adds distinctiveness to an otherwise commonplace movement. The order of gestures also changed from 'crown'→'breasts' to 'breasts'→'crown', and this, together with the possibly linked change to a brow-sweep, made unbroken transitions to succeeding gestures possible – something like fluent signing. Figure 7, at the bottom, shows a smooth transition being effected when the now single 'crown' hand moves down to the front of the body while simultaneously the left hand moved up into the same space and effected a smooth transition into the next (a two-handed hour-glass shape) gesture for 'Snow White'. This smooth transition was made possible by the ritualization of 'crown'.

Clips from Ralph Bloom's "Snow White" wordless narration

Initial "King" and "Queen"

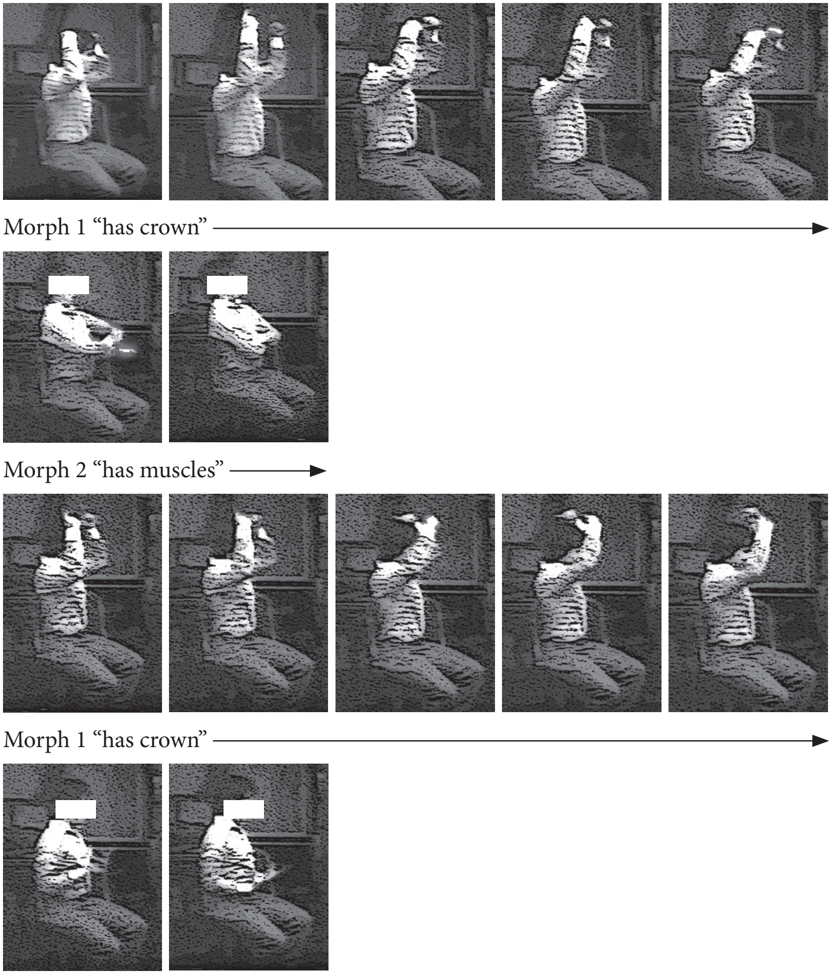
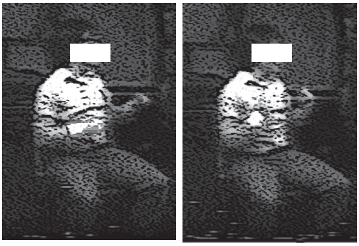


Figure 4. First occurrences of "King" and "Queen" morphs. The gestures are two-morph combinations. Note the immediate contrast of Morph 2: 'has-muscles' vs. 'has-breasts'. Morph 1, 'has-crown', is the same. The two hands rotate around the head, forefingers pointing down, moving up and down as they rotate. The spatial head vs. torso distinction and pointing vs. cups for Morph 1 and Morph 2 are maintained despite later streamlining (see next example). The duration of "King", the first gesture of the pair, was 4.3 seconds. "Queen", the second, was down to 2 seconds, and this acceleration continued. (SW gestures from Ralph Bloom.)

Clips from Ralph Bloom's "Snow White" wordless narration
Later 'streamlined' Queen

Morph 1 "has crown" ⟶



Morph 2 "has breasts" ⟶

Figure 5. The later abbreviated "Queen". The 'has-crown' morph made with a single hand and a partial revolution; the 'has-breasts' morph is still two cupped hands but now inward and not held upward. The changes improve speed but also reduce iconicity, so some movement toward arbitrariness. Duration is down to slightly more than 1 second for the entire two-morph combination, about the span of a spoken word. The head-torso distinction is still present and was never lost during the entire narration.

Birth of a syntagmatic value

We can take this analysis a step further. Not only morphemes themselves but the syntagmatic values of morpheme combinations can be seen emerging in gestures when speech is denied. University students, not deaf but not allowed to speak, devise multi-gesture descriptions. This is not surprising in itself, but it is important that these gesture descriptions appear to involve *de novo* syntagmatic values, not necessarily ones fashioned out of any languages they speak.

When symbols combine within some hierarchically dominant frame, they acquire values that exist because of that combination and exist only there, in a kind of construction (cf. Goldberg 1995). The value of being a direct object in speech is a case; "ball" is not a direct object in itself – it becomes one only in combination with a verb ("toss the ball" and the like). In the Gershkoff-Stowe & Goldin-Meadow (2002) experiment, non-signing hearing participants described video vignettes showing, in the example to be analyzed, a doll seeming to somersault through the air and land in an ashtray comparatively the size of a sandbox (a Ted Supalla 1982 ASL verb of motion

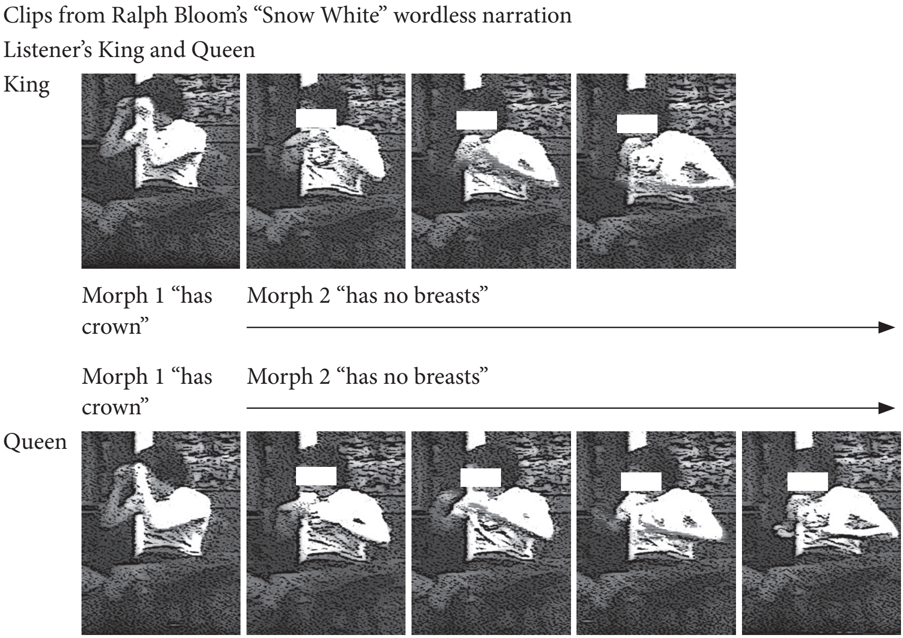


Figure 6. The "Queen" and "King" morphs in a dialogue by the listener. He was attempting to clarify which character, King or the Queen, the narrator had just gestured. The Morph 1-Morph 2 distinction is preserved but a 'dialect' difference has appeared in the 'has-crown' and 'has-muscles' features – the first without revolution, the second a downward slice with the hands in front of the chest. Morph 2 may be an instance of 'language drift', shifting to something like 'flat-chested' or 'has-no-breasts', away from its original 'has-muscles'. The speaker had just before used 'flat-chested' in combination with 'his usual 'has-muscles' and crown'. The listener did not arrive at these features himself, and this seems to be the essence of linguistic drift triggered by contact in microcosm. If so, it suggests an even more robust role for language contact in the diversification of languages. Gaze was directed at the (official) speaker, not at the gestures, showing that the gestures had attained unconscious status as elements in the communicative system. The question speech act was also conveyed non-verbally with a forward head lean that was maintained throughout.

vignette). The key requirement was that participants not use speech; everything was to be conveyed by gestures which the participants themselves created. With an intransitive action like somersaulting, three sequences were found with some frequency: S-M-A, M-S-A, and S-A-M (S = 'stationary object', here the ashtray; M = moving object, here the doll; A = action, here the arc with somersault). These sequences correspond to different 'constructions' (Table 1).

Clips from Ralph Bloom's "Snow White" wordless narration

Position in narration	Frame Number	Gesture	Gesture Features, Sequence
2	1:31;21–1:33;28		Queen: – crown (4 peaks)
			– breasts
3	1:47;03–1:49;03		Queen: different order, reduced crown – breasts
			– crown (2 peaks, 1st with RH only, then LH comes in with G-hand for 2nd peak with BH moving around head)
30	6:35;12–6:38;14		Queen: again different order, further reduced crown – breasts

– crown with RH G-hand, circling around head, index up

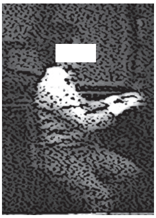
39 9:09;01–9:10;27

Queen: new order continues

breasts

– crown with RH G-hand circling around head with 2 peaks, index again up

Segue to following gesture – left hand starting to rise to meet downward moving right



Start of two-handed 'hour-glass' shape "Snow White"

Figure 7. 'Ritualization' of a gesture morph through 4 stages during SW (termed 'positions' 2, 3, 30 and 39, covering about 9 minutes of the narration). In words, the distinctive 'has-breasts' feature never disappears; the order changes from 'crown-beasts' to 'breasts-crown', probably because the 'crown' undergoes significant reduction (2 hands to 1) and streamlining, which in turn promotes unbroken motion into the following gestures – a syntagmatic effect. The added upright index finger at position 30 distinguishes the brow-sweep that is 'crown' from ordinary brow sweeps.

Table 1. Spontaneous 'moving-object', 'at-a-location' and 'end-state' syntagmatic values

'Construction'	Sequence	Example
MOTION (increasing activity)	S-M-A	ring-doll-somersault
LOCATION (where action occurred)	M-S-A	doll-ring-somersault
RESULT (end-state of action)	S-A-M	ring-somersault-doll

Syntagmatic values are seen in that the same 'M' (doll) gesture, for example, has different values in different combinations:[1]

– *doll* is 'moving-object' in the Motion package; the 'phrase' is M alone.
– *doll* is 'at-a-location' in the Location package; the 'phrase' is (M-S).

and
– *doll* reaches 'end-state' in Result; the 'phrase' is (A-M).

(To see these syntagmatic values, we recommend mimicry – try performing the sequence of gestures in each row while thinking in terms of the meanings, MOTION, LOCATION, or RESULT – and note what the M gesture seems to mean in this construction, its value within this overall pattern.) These new syntagmatic values show regularities beyond any iconicities. Only the S-M-A order is iconic (the sequence corresponds to increasing activity). So it may indeed be possible to have new syntagmatic values in combined gestures without speech. They come forth seemingly automatically.[2] Each syntagmatic value comes with a paired significance (moving-object, at-a-location, end-state). Thus a basic property of a morph combination emerges.

Do syntagmatic values also emerge with the speech-synchronized gestures in Canary Row? No. Instead, CR narrators, when they combine gestures, enrich the imagery but do not create new values that exist only in the combination. In Figure 8, the speaker produced two images of Tweety dropping the bowling ball into the pipe,[3] the second a more elaborate version with two hands that occurred after a question by listener. The imagery is elaborated and shows increasing iconicity – a value intrinsic to the imagery itself. It is not a new value (as direct object is a new value of a noun in a verb phrase) but a more elaborate version of the already-existing iconic picture, and thus is the very opposite of a syntagmatic value that exists only in combination.

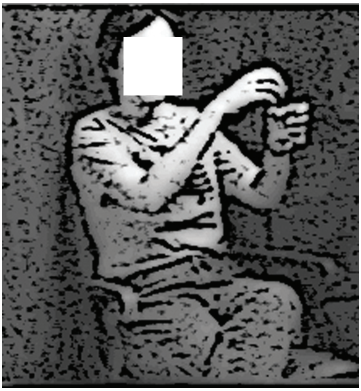
1. An insight due to Amy Franklin.

2. All the more striking, then, that gestures *with* speech are global and (especially) synthetic – resisting, in other words, construction-like tendencies when combined with speech. Cf. Goldin-Meadow et al. (1996).

3. In the recounted episode, the character 'Sylvester' is attempting to reach another character, 'Tweety', by climbing a drainpipe on the inside. He is thwarted when Tweety drops a bowling ball into the pipe and Sylvester and bowling ball meet explosively mid-pipe. He swallows the bowling ball and rolls back out the bottom and onto the street, now a living bowling ball, and eventually into a bowling alley, where he gets a strike.

and *TW* **thr**ows **a bow**ling ball/down in **the*** the th**ing**

Listener: where does he throw the bowling ball?



it's **one of those gu**tter **pi**pes an' he throws **the ball into the** top

Figure 8. Illustrating the non-syntagmatic combination of gestures accompanying speech. Left hand joins right hand in second panel for elaboration of entering-the-pipe imagery triggered by listener's query. The 'value' of the left hand derives from the image as a whole, not from the combination.

Gesture families, preliminary and ephemeral morphs

In CR, gestures show a tendency to stabilize on certain forms, for example, 'Sylvester' becomes a single-finger (pointing) hand for several narrators; 'Granny' is a loose open hand approximating the form called the 'B-hand' in American Sign Language notation, and 'Tweety' is a character viewpoint gesture with various handshapes. Figure 9 shows the single-finger 'Sylvester' gestures by three speakers, in order of their occurrence. These are recurring forms, but the forms are inconsistent. Non-single-finger handshapes are also used for Sylvester, and the single-finger handshape appears for other references. In short, there is gravitation to a certain form, often with an iconic start (the first of the Sylvester single-finger gestures was both deictic and iconic for squeezing into the pipe), but the form does not become fixed, nor is it reserved for one

Jan. | **V. J.** | **Viv.**

and he tries to <um> **climb up**

the* <uh> **cli**mbs up the drainpipe

he tries going **up** the ins**ide** of t**he** **drai**npipe

this time he tries to go up **in**side the rain gutter (from a later scene)

and he comes o**ut the** bottom of the drainpipe (later part of above scene and could be primed)

and then he* then you **see him on some electrical** wires

and he **rolls on down** into **a bow**ling alley (also part of the two-similar-hands 'bowling ball' catchment (see Figure 4), simultaneously showing in one gesture both Sylvester as a character and the drama in which he is taken over by Tweety's b-ball)

and that catapults him up (could be simple deixis)

he comes **swinging throu**gh on a rope (could be iconic for the rope)

little hat (could be simple deixis)

and he's **walk**ing on it (could be simple deixis)

Figure 9. Recurring 'Sylvester' gestures in Canary Row narrations by 3 speakers. In some cases the gesture could be deictic, but this does not conflict with a concurrent 'Sylvester' meaning.

Figure 10a

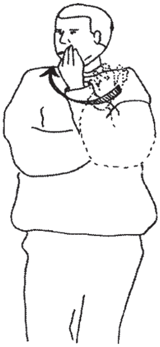


Figure 10b

Figure 10 a. Neapolitan *grappolo* cultural gesture. From Kendon (2004: 230) **b.** Conduit palm up, open hand metaphor with "to **get Tweety**, by an English speaker.

meaning. If we call this polysemy, it is far beyond what one expects in a functioning communicative system. In these respects it is a proto-morph not yet over the threshold of becoming a word or sign.

The gestures also form what Kendon (2004) terms 'gesture families' – gestures sharing one or more form features that cluster around some core meaning. Kendon's examples came from the Neapolitan gesture culture and, it appears to us at least, were centered on one or another kind of metaphor. The *grappolo* (Figure 10a) appears to be a metaphor akin to the conduit gesture. Unlike the English speaker's conduit (Figure 10b), the *grappolo* is structured by standards. It must take the finger-bunch shape (which, conduit-like, encloses a meaning). But in both the Neapolitan *grappolo* and English speaker's conduit a discursive object appears to be held in the hand. Also the pragmatic function of the *grappolo* is subset of the poly-functionality of the un-morphemic conduit – in Kendon's words, "the speaker is trying to clarify or make more specific what is to be considered" (Kendon 2004: 230) – a meaning more specific than the general meaning of the conduit as "a container holding discursive content".

We have form-stabilizations in CR that are perhaps another aspect of morph birth, the co-opting of an initially iconic or metaphoric form by some initially incidental meaning, which then becomes the final meaning: so the rising single-finger hand for Sylvester that initially meant compressing and ascending came to mean, in later occurrences, just Sylvester, unsqueezed and unascending; and just as (we suppose) an initial conduit type metaphoric gesture image was co-opted in the formation of the *grappolo* by a narrower speech-act to gain clarity in an interactive situation. Another view of Figure 9 therefore is that it shows a CR gesture family in its order of emergence. We see the initial iconicity of the gesture and its later focus on what at first was an incidental meaning but which became the sole meaning. While the speaker was seemingly unaware of the recurring gesture forms, there is a kind of form-agglutination taking place that is explained by the concept of a material carrier, a concept from Vygotsky (in Rieber

& Carton 1987: 46). The extended index finger, at first iconically depicting Sylvester's compression and ascent in the pipe, became a material carrier for the total synthesized ensemble of Sylvester, pipe, ascent and compression. Subsequent Sylvester references were still embodied in this gesture form and continued on this basis, though no longer with upness or inness, which 'wore off' as it were, while the material sign (the single finger gesture) remained.

Summary and conclusions: Birth of the static dimension

We've tested the conditions under which a morph/syntagmatic-value threshold is reached, and observe that it is unattainable when there is speech. On the other hand, when speech is absent, morph properties arise automatically. There are new stable recurring forms held to standards, and *de novo* syntagmatic values.

Why is an absence of speech important?

The generalization that fits the cases where morphs and syntagmatic values do emerge is *absence of speech*, and here form comes into its own. How does an absence of speech have these effects? We suggest four factors (in possible causal order):

1. Release of gesture from the imagery-language dialectic of the growth point (see McNeill & Duncan 2000, McNeill 2005). This seems essential, since otherwise gestures are strongly constrained to maintain a semiotic opposition to language, away from any kind of language-like morph status with combinatoric potential.
2. Increased awareness of gesture as a symbolic medium. Without speech, attention naturally falls to gesture as the sole channel, and this in itself can foster morph and combinatoric status enhanced with consciousness of form standards.
3. Swerving to pantomime and other points on Gesture Continuum[4] (see McNeill 2000). As part of the same focus on gesture as the sole channel of communication,

7 A continuum of how speech and gesture relate (formerly called 'Kendon's Continuum', renamed at Kendon's request): As one moves along the continuum, two kinds of reciprocal changes occur. First, the degree to which speech is an obligatory accompaniment of gesture *decreases* from gesticulation to signs. Second, the degree to which gesture shows the properties of a language *increases*. Gesticulations are obligatorily accompanied by speech but have properties unlike language. Speech-linked gestures are also obligatorily performed with speech, but relate to speech in a different manner – sequentially rather than concurrently and in a specific linguistic role (standing in for a complement of the verb, for example). Signs are obligatorily *not* accompanied by speech and have the essential properties of a language. Clearly, therefore, gesticulations (but not the other points along the Continuum) combine properties that are unalike, and this combination occupies the same psychological instant. A combination of unalikes at the same time is a framework for an imagery-language dialectic.

the speaker resorts to mime, and this has properties of combination and recurrence of its own.

4. Ritualization or streamlining to bring gestures in line with the temporal parameters of communication.

The result of this chain of causation can be morph segmentation and syntagmatic combination, and the beginning of new elements of language.

In addition, what has been termed 'shareability' seems crucial (Freyd 1983) – constraints on information that arise because it must be shared. Constraints because:

> It is easier for a individual to agree with another individual about the meaning of a new 'term' (or other shared concept) *if* that term can be described by: (a) some small set of the much larger set of dimensions upon which things vary; and (b) some small set of dimensional values (or binary values as on a specific feature dimension). Thus, terms are likely to be defined by the presence of certain features. (p. 197, italics in original).

In three words, shareability produces discreteness, repeatability, and portability – the semiotic qualities of morphs. In her concluding footnote, Freyd speculates that shareability may be relevant to the intrapsychic workings of individual minds, the dynamic creation of utterances in context, as well as to the interpsychic relations between individuals. We also posit shareability at the moment the SW narrator or a vignette subject creates a novel gesture with which to communicate events to his/her listener. A linguistic dimension of gesture emerges.

With speech, however, the role of gesture changes

When there is speech, however, the gesture must take on a different role. It then is needed to form 'growth points' or units of an imagery-language dialectic that propels thought and speech, and without which everything slows down (as happens when, for example, narrators temporarily lose the thread of the story: gestures lose content as they run out of material and speech gains vacuity). In a growth point, an idea is simultaneously embodied in contrasting semiotic modes. One mode is segmented-analytic (linguistic) and the other is global-synthetic (gestural/imagistic). Both modes must be active for a dialectic to form.

'Gesticulation' is motion that embodies a meaning relatable to the accompanying speech..

'Speech-linked gestures' are parts of sentences themselves. Such gestures occupy a grammatical slot in a sentence.

'Emblems' are conventionalized signs, such as thumbs-up or the ring (first finger and thumb tips touching, other fingers extended) for "OK".

'Pantomime' is a gesture or sequence of gestures conveying a narrative line, with a story to tell, produced without speech.

'Signs' are lexical words in a sign language (typically for the deaf) such as ASL.

Unlike a generative model which says that performance is carried out by 'applying' or 'using' competence, and unlike the Saussurian model, which defines *parole* at the residue after subtracting *langue* (the systemic aspects) from *langage* (the totality of human communicative potential), an imagery-language dialectic (unified in a growth point) defines the dynamic as powered by the opposition of unlike semiotic modes for the same idea, a dynamic in which the static dimension is an essential ingredient. The dynamic is impossible without the static, and vice versa. For this reason, gestures in combinations with speech do not take on morph qualities, including standards of form and syntagmatic values, as each endangers the possibility of an imagery-language dialectic (for extensive discussion of the GP and imagery-language dialectic, see McNeill 2005).

The bioprogram

The creation immediately of paradigmatic oppositions between the K and Q in the SW wordless narration and the equally fast emergence of syntagmatic values in the vignettes experiment suggest an ability at this level specifically geared to language, aspects of a 'bioprogram' for language (the term is from Bickerton 1990). Given the above three-way distinction between 'performance', 'parole' and the place of the static dimension in an imagery-language GP dialectic, we conceive of this bioprogram in different ways. In a dialectic, the morph properties are jointly conceptualized with imagery: both are essential for a dialectic. So whatever explains the origin of one must consider the other. In my own ruminations on this topic (e.g., McNeill et al. 2008), I have concluded that language (here, morphs) and gesture had to evolve jointly; it is not possible that one came before the other, neither gesture-first nor speech-first, and then to explain our current situation of an imagery-language dialectic.

Rethinking the morph

Another implication of the SW and vignettes experiments concerns the conception of the morph itself. It is possible to see standards of form, which we have adopted as the *sine qua non* of the morph together with the other criteria of morph status earlier mentioned, as standards of *actions* rather than of entities in some kind of unchanging semiotic space. This makes the morph into a template for behaving. Behavior is not ever (we believe) meaningless, so this template would naturally include the two sides of the sign, the signifier and signified, but they are no longer 'sides' and are now regarded as sedimented meaningful actions (for the concept of 'sedimentation' see Merleau-Ponty 1942/2006). If we adopt this perspective, the synchronic method, the mainstay of linguistic analysis, comes under scrutiny too. It devolves to uncovering the intuitions of 'good', that is, socially-constituted, conventional (behavioral) forms. Intuitions play a role by signaling that speech (or manual sign) actions are "the way we do things around here". Intuitions can be taken to be the individual's mode of access to these standards, and may correspond to highly entrenched action patterns in the motor orchestrating parts of the brain. The classic *langue* ('competence')-*parole* ('performance') distinction is replaced by the idea of actions meeting standards, and the traditional psycholinguistic position that 'performance' is the limited rendition of 'competence' becomes meaningless: an action cannot be derived (with or without limits) from this or any other standard; this mistakes the relation, *the action is compared to and guided by a standard*. Rather than limit, it enables. The GP theory is the systematization of this multimodal action-based perspective, in which cognitive and linguistic movement is fueled by a dialectic of imagery and linguistic form, that is, by modes of organizing actions of the vocal and manual articulators as they work in concert to co-express a shared idea unit.

References

Bickerton, D. 1990. *Language and Species*. Chicago: University of Chicago Press.

Bloom, R. 1979. *Language Creation in the Manual Modality: A Preliminary Investigation*. A. B. thesis, Department of Behavioral Sciences, University of Chicago.

Enfield, N. J. 2001. "'Lip-pointing': A discussion of form and function with reference to data from Laos." *Gesture* 1: 185–211.

Freyd, J. J. 1983. "Shareability: The social psychology of epistemology." *Cognitive Science* 7: 191–210.

Gershkoff-Stowe, L. and Goldin-Meadow, S. 2002. "Is there a natural order for expressing semantic relations?" *Cognitive Psychology* 45: 375–412.

Goldberg, A. 1995. *Constructions: A Construction Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldin-Meadow, S., McNeill, D. and Singleton, J. 1996. "Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality." *The Psychological Review* 103: 34–55.

Goldin-Meadow, S. and Mylander, C. 1984. *Gestural Communication in Deaf Children: The Effects and Non-Effects of Parental Input on Early Language Development*. Chicago: University of Chicago Press.

Hockett, C. F. and Altmann, S. A. 1968. "A note on design features." In *Animal Communication: Techniques of Study and Results of Research*, T. A. Sebeok (ed), 61–72. Bloomington, IN: Indiana University Press.

Kendon, A. 1988. *Sign Languages of Aboriginal Australia: Cultural, Semiotic and Communicative Perspectives*. Cambridge: Cambridge University Press.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

McNeill, D. and Duncan, S. D. 2000. "Growth points in thinking for speaking." In *Language and Gesture*, D. McNeill (ed), 141–161. Cambridge: Cambridge University Press.

McNeill, D. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.

McNeill, David. 2008. "Unexpected metaphors." In *Metaphor and Gesture*, A. Cienki and C. Müller (eds), 155–170. Amsterdam/Philadelphia: Benjamins.

McNeill, D., Duncan, S. D., Cole, J., Gallagher, S. and Bertenthal, B. 2008. "Growth points from the very beginning." *Interaction Studies* [special issue on proto-language, D. Bickerton and M. Arbib (eds)] 9 (1): 117–132.

McNeill, D. 2009. "Imagery for speaking." In *Crosslinguistic Approaches to the Psychology of Language. Research in the Tradition of Dan Isaac Slobin*, J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura and Ş. Özçalişkan (eds), 517–530. New York: Psychology Press.

Merleau-Ponty, M. 1962. *Phenomenology of Perception* (C. Smith, trans.). London: Routledge.

Merleau-Ponty, M. 2006. *The Structure of Behavior* (A. L. Fisher, trans.). Pittsburgh: Duquesne University Press [original pub. in 1942].

Reddy, M. J. 1979. "The conduit metaphor: a case of frame conflict in our language about language." In *Metaphor and Thought*, A. Ortony (ed), 284–297. Cambridge: Cambridge University Press.

Rieber, R. W. and Carton, A. S. (eds.). 1987. *The Collected Works of L. S. Vygotsky. Vol. 1: Problems of General Psychology*. *Including the Volume* THINKING AND SPEECH (intro. and trans. by Norris Minick). New York: Plenum.

Senghas, A. and Coppola, M. 2001. "Children creating language: How Nicaraguan Sign Language Acquired a Spatial Grammar." *Psychological Science* 12: 323–328.

Supalla, T. 1982. *Structure and Acquisition of Verbs of Motion and Location in American Sign Language.* Unpublished Ph.D. dissertation, University of California, San Diego.

Vygotsky, L. S. 1987. *Thought and Language*. Edited and translated by E. Hanfmann and G. Vakar (revised and edited by A. Kozulin). Cambridge: MIT Press.

Whorf, B. L. 1956. *Language, Thought, and Reality. Selected Writings of Benjamin Lee Whorf*. J. B. Carroll (ed.). Cambridge, MA: The MIT Press.

CHAPTER 4

Dyadic evidence for grounding with abstract deictic gestures

Janet Bavelas, Jennifer Gerwing, Meredith Allison, and Chantelle Sutton
University of Victoria

Speakers use gestures to communicate within a dialogue, not as isolated individuals. We therefore analyzed gestural communication within dyadic dialogues. Specifically, we microanalyzed *grounding* (the sequence of steps by which speaker and addressee ensure their mutual understanding) in a task that elicited abstract deictic gestures. Twenty-two dyads designing a hypothetical floor plan together without writing implements often used gestures to describe these non-existent spaces. We examined the 552 gestures (97% of the database) that conveyed information that was not presented in the accompanying words. A highly reliable series of analyses tracked the immediate responses to these nonredundant speech/gesture combinations. In the vast majority of cases, the addressee's response indicated understanding, and the speaker/gesturer's actions confirmed that this understanding was correct.

1. Studying gestural communication by individuals versus dyads

Laboratory studies of gestural communication usually focus on the speaker and the addressee separately, as encoder or decoder. In encoding studies, the focus is on gesture production in differing conditions (e.g., how visibility influences the speaker's gestures; see review in Bavelas, Gerwing, Sutton, & Prevost 2008, Table 1). Because only the speaker's actions are of interest, the task and the interaction are highly asymmetrical. In these dialogues, the addressee, who may be the experimenter, a confederate, or another participant, often has instructions to respond minimally. Unfortunately, research has shown that constraining the addressee's behaviors may have an unintentional, deleterious effect on the speaker's communicative behaviors (Bavelas, Coates, & Johnson 2000, Beattie & Aboudan 1994).

Decoder studies focus primarily on gesture comprehension (see review in Holler, Shovelton, & Beattie 2009). These designs can be even more removed from dyadic conversation. For example, the participants might watch gestures in brief video excerpts, often without conversational context. Again, evidence from other fields suggests that such a design would affect the addressee's ability to understand the gestures. For example, Schober and Clark (1989) found significantly better comprehension by an addressee who was interacting with the speaker than by someone who heard the same information but did not participate in the dialogue. Thus, the encoder and decoder research designs we have been using are not well suited to investigating conversational gestures, which by definition occur within real dialogues.

Recent research has begun to include experiments with two freely interacting participants (e.g., Bangerter 2004; Bavelas, Chovil, Coates, & Roe 1995; Bavelas, Chovil, Lawrie, & Wade 1992; Bavelas et al. 2008; Clark & Krych 2004; Furuyama 2000; Gerwing & Bavelas 2004; Holler & Stevens 2007; Özyürek 2000, 2002). However, the unit of analysis in many of these experiments has remained individual in the sense that the dependent variable was usually a summary of one participant's gestures (e.g., average rate of speaker's gestures). Such measures of aggregated individual actions are useful or even essential for answering certain experimental questions, but they necessarily remove communicative acts from their sequential context, separating one participant's actions from the immediately preceding and succeeding actions of the other person.

In three of the above studies, the dependent variable did reflect the immediate dyadic sequence in which the gestures occurred. Bavelas et al. (1995, Study 2) demonstrated that addressees responded as predicted to the speaker's spontaneous interactive gestures. Furuyama (2000) illustrated how addressees sometimes incorporated the speaker's previous gesture into their own. Clark and Krych (2004) demonstrated how addressees used gestural actions to indicate their state of understanding of the speaker's directions. In each of these three studies, the primary focus was on a gesture in relation to the immediate dyadic context in which it occurred, and the summary data preserved this unit of analysis.

We propose that the participants in a conversation shape their gestures, like their words, to fit a specific, immediate context. Therefore, the ideal design for revealing whether and how conversational gestures communicate would focus on dyadic sequences and would include (a) two or more participants who can interact spontaneously and as themselves; (b) a symmetrical task to which both can contribute; (c) the gestures of both participants; and (d) an analysis of each gesture in the context and interactive sequence within which it occurred. In pursuit of this ideal, the present study obtained moment-by-moment dyadic evidence of gestural communication using a design that included two real participants, without constraints on their interaction, designing a floor plan together. The gestures could be from either participant, and our analysis of grounding sequences included the responses of both of them.

2. Grounding

Fundamental to Clark's (1996) collaborative model of language use is *grounding* (Clark & Schaefer 1989), a moment-by-moment process by which the participants establish that they understand each other well enough for current purposes. Grounding is an intrinsically mutual responsibility, not an individual process: "Speakers and their addressees go beyond autonomous actions and collaborate with each other, moment by moment, to try to ensure that what is said is also understood" (Schober & Clark 1989: 211). "Moment by moment" means that grounding is a micro-process that is constantly occurring, usually in the background of the dialogue and not just in conclusion.

Our preferred description of a *grounding sequence* involves a rapid three-step interchange between the participants: The person who is speaking at the moment *presents* some information, the addressee *responds* with an indication or display of understanding (or not), and then the speaker *acknowledges* this response by indicating that the addressee's understanding was correct (or not). These steps can involve words, gestures, nodding, gaze, or other actions, singly or in combination.

In the following examples from our floor-plan data, underlined words indicate the location of a gesture. Also, throughout this chapter, we will distinguish between the participants by arbitrarily treating the speaker/gesturer of the moment as female and the addressee at that moment as male.

(1) The speaker was describing their plan, starting at the entrance to the apartment:
Speaker: So we could have, like, you come in.
Addressee: Yeah.
Speaker: There's a kitchen ...

While saying "you come in," the speaker gestured the location of the entrance by placing her two index fingers together on the table. The addressee indicated explicitly that he understood the location by saying "Yeah." Then the speaker/gesturer located "a kitchen" by placing her left hand slightly to the left of where she had placed the entrance. Notice that, instead of explicitly acknowledging the addressee's understanding, the speaker/gesturer presupposed it by continuing her tour of the floor plan.

Addressees also use continuation as a way of indicating understanding:

(2) The participants were reviewing their plan, and the speaker had just used gestures to place the two bedrooms on either side of a hallway.
Speaker: ... and then a bathroom
Addressee: bathroom at the end
Speaker: [nods]

As the speaker said "and then a bathroom," she pointed to a spot at the end of where she had previously placed the hallway. The addressee immediately displayed his understanding by saying "bathroom" simultaneously and then finishing her sentence by naming the location that the speaker had only gestured ("at the end"). The speaker's nod explicitly acknowledged that the addressee had understood correctly.

Recall that the standard for grounding is "well enough for current purposes" (Clark 1996: 221), so the participants may also rely on implicit indications of understanding. Indeed, conversation would sink under its own weight if every step of every grounding sequence were explicit. Instead, participants often minimize their joint effort by more economical implicit responses, as shown in the next example. Note that there were two presentations in this example, and the grounding was entirely implicit in the first one:

(3) Speaker: In my mind, $_1$the bedrooms... $_2$are on this side.
Addressee: [nodding] Ohhh-kay!
Speaker: Yeah.

The speaker/gesturer began the first sequence with the words "In my mind, the bedrooms" as she placed her hand to show the location of one of the bedrooms. She then paused briefly, and the addressee continued to watch her gestures (implicitly indicating understanding). The speaker then said "are on this side" while moving her hand to a location further beyond, where the other bedroom would be. This second presentation of new information served two functions: It presupposed the addressee's understanding of her first gesture, thereby implicitly acknowledging it and ending that grounding sequence, and it presented further new information, initiating a new sequence. This time, the addressee indicated his understanding explicitly (with "Ohhh-kay!" and a big nod), and the speaker/gesturer's acknowledgment was also explicit ("Yeah").

A grounding analysis can also identify points at which mutual understanding does not occur. At each step, either participant can initiate a clarification or repair. That is, the addressee can ask for clarification from the speaker/gesturer. Or the speaker/gesturer can detect that the addressee's understanding is wrong and correct it.

In sum, grounding sequences are an observable, intrinsically dyadic process, focused precisely on the establishment of mutual understanding. They are thus well-suited to examining the communicative value of gestures for interlocutors. Our analysis focused on the grounding process initiated by presentations of nonredundant speech/gesture combinations (i.e., ones where the gesture conveyed information that was otherwise missing from the words), then examined the addressee's immediate response, and then the speaker/gesturer's acknowledgment. We propose that a successful grounding sequence after a nonredundant speech/gesture combination provides observable, local evidence that the participants used these gestures to communicate and mutually considered the gestural information to be part of their accumulating common ground.

3. Abstract deictics

The task used here evoked a different kind of gesture than in many previous experiments, namely, gestures depicting something that does not exist. The participants sat across a bare table and designed a floor plan for a student apartment. As they talked, all of them spontaneously "drew" their plans on the table with their gestures, creating and pointing to hypothetical spatial relationships that had no concrete referent. These gestures were *abstract deictics* (e.g., McNeill, Cassell, & Levy 1993), which are a special kind of pointing. As Kendon (2004) explained, most pointing gestures indicate a space or location that is currently visible or a direction toward a real location that is not yet visible. In contrast, abstract deictics actually create spaces and refer to locations that do not physically exist. Our participants' gestures did not represent any existing space; they depended entirely on the participants' shared understanding of their words and gestures. We expected that, even in these cases, the participants would readily show that they understood each other.

4. Research design and procedures

4.1 Task and hypotheses

Each dyad designed its own layout for a two-bedroom student apartment on the table between them.[1] There were no assigned roles; both participants could contribute to the design of the plan as they wished. We emphasized the goal of mutual understanding by advising them that when they were finished, they would each have to draw the agreed-upon plan independently.

4.2 Method and procedure

A total of 44 University of Victoria students formed 22 dyads (12 female/female, 1 male/male, and 9 female/male). All participants spoke English fluently, were unacquainted, and knew they would be videotaped. In return for participating, they received course bonus credits.

Recording equipment in our Human Interaction Lab included a remotely controlled Panasonic WD-D5000 color camera with a wide-angle lens and a Soundgrabber II omni-directional microphone. We digitized the analog video into AVI format using Broadway (www.b-way.com) and analyzed it with Broadway on an 18-inch ViewSonic GS790 color monitor.

After the participants read and signed a consent form, they had a few minutes to get acquainted with one another. They then did two or three unrelated tasks, including the primary one: The experimenter asked them to design a floor plan for a two-bedroom apartment appropriate for University Student Housing. The floor plan should include (but not necessarily be limited to) the bedrooms, a bathroom, a living room, and a kitchen. The experimenter emphasized that the layout of the apartment was

1. We varied the width of the table the dyad worked on. As predicted, the wider space led participants to move their gestures forward, toward their partner. Because there were no other significant differences, we will not include this variable in the rest of the chapter.

most important, not the dimensions of the rooms or where the furniture went. She also informed them that, later, they would each have to draw the floor plan separately. After answering questions, the experimenter left the participants to design their plan. When they were done, she returned and re-seated them on either side of a partition to make their individual drawings of the plan.

Afterward, the experimenter explained the purpose of the study, answered questions, and gave them a written summary. Finally, they watched their videotape, and each indicated on a permission form whether and how we could use their videotape (e.g., to be viewed only for research, shown to professional audiences, etc.).

5. Analysis and results

5.1 The data set

The purpose of the analysis was to examine grounding sequences that began whenever either of the participants used a nonredundant combination of speech and gesture to present new information about the location of a room or rooms in their floor plan. We limited the potential data to about two minutes of each dyad's discussion of their final floor plan, excluding initial discussions of possible criteria and preliminary layouts. When a discussion of the final plan was substantially longer than two minutes, we analyzed only the first and last minute. During these two minutes, the mean proportion of time spent gesturing was .82 ($SD = .11$).

Within this data set, independent analysts located all gestures that depicted an identifiable room. They excluded gestures that were not about the floor plan; gestures that did not locate an identifiable room within it; gestures that were not analyzable after repeated viewing; and adaptors. They included gestures by the addressee of the moment only if the response added new verbal or gestural information, initiating a new, overlapping grounding sequence. Note that our focus was not on individual gestures but on the presentation of information about rooms in the plan, which could include one or more rapidly contiguous gestures. The inter-analyst reliability for the above decisions ranged from 80% to 97%.

The final data set was 571speech/gesture combinations that depicted identifiable rooms in the floor plan.

5.2 Identifying nonredundant speech/gesture combinations

We focused on *nonredundant* gestures, which contributed information that was missing from the words. Nonredundant gestures required that the addressee apprehend and integrate information from both speech and gesture. A typical nonredundant gesture was

(4) Speaker: Let's say we have the <u>door here.</u>

As she said "door here," the speaker/gesturer traced a line about an inch wide on the table. It was only her gesture that showed precisely where "here" was. Therefore, the gesture was nonredundant with the words. Notice that all of the gestures in Examples 1, 2, and 3 above were also nonredundant.

In contrast, *redundant* gestures conveyed no additional information beyond the words; for example,

(5) Speaker: so we put the bedrooms on the right side and the bathrooms on the left, is that right?

The speaker/gesturer first used her right hand to make a vague pointing gesture to her right; then, she used her left hand to make a similar gesture to her left. Both her words and her gestures depicted "right" then "left," with no additional or more specific information in the gestures, which were therefore redundant with her words. See Gerwing and Allison (2009) for a more detailed explanation of this and other redundancy analyses.

Reliability for redundancy versus nonredundancy across all groups and all gestures was 96.5%.

5.2.1 *Redundancy results*

Redundancy between gestures and words was rare; 552 of the 571 speech/gesture combinations analyzed included gestures that were *not* redundant with the words (mean proportion = .97; *SD* = .05). As illustrated in examples 1 to 4, the gestural information was usually essential to their task (e.g., the location of the rooms).

5.3 Grounding sequences

A grounding sequence consisted of the *presentation* of one of the above 552 nonredundant speech/gesture combinations, the *addressee's response,* and any *acknowledgement* by the speaker/gesturer. Figure 1 depicts the overall analysis.

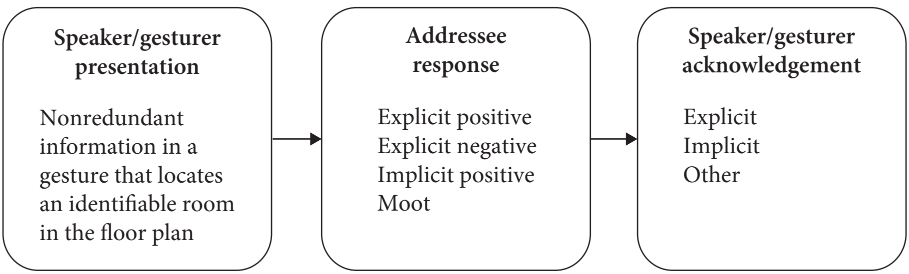


Figure 1. Schematic figure of the three stages of analysis.

5.3.1 *Addressee responses to nonredundant gestures*

Immediately following each of the speaker/gesturer presentations, addressees could respond by indicating whether they understood (or not). As described below, their response could be explicit or implicit. It could also be positive (indicating understanding), negative (indicating lack of understanding), or moot (indeterminate).

5.3.1.1 *Explicit versus implicit addressee responses.* An *explicit addressee response* was one that provided decisive feedback to the speaker/gesturer about whether the addressee had or had not understood the nonredundant speech/gesture combination. Examples included saying "yeah," finishing the speaker/gesturer's sentence, nodding, gesturing the same room, or alternatively, asking a question. However, in the second-by-second tempo of a spontaneous dialogue, it would be inefficient for addressees to respond explicitly to every presentation that they had understood; the participants sometimes rely on implicit indications. *Implicit addressee responses* did not provide overt evidence about the addressee's state of understanding. The addressee simply continued to pay attention and allowed the speaker/gesturer to go on, or the addressee took an action that implicitly built on the speaker/gesturer's presentation without any overt expression of understanding. Two analysts examined what the addressee did immediately following the speaker/gesturer's presentation and decided whether the addressee contributed an explicit or implicit response. Their reliability on a randomly selected set of 19 groups was 89%.

5.3.1.2 *Explicit positive versus explicit negative responses.* Explicit addressee responses could be positive, indicating understanding, or negative, indicating not understanding or requesting clarification. Typical *explicit positive responses* were "yeah," nodding, or gesturing the same room. In an *explicit negative response*, the addressee typically asked for clarification about the relative location of rooms. Based on a randomly selected 20 groups, inter-analyst reliability for distinguishing whether an explicit addressee response was positive or negative was 96%.

5.3.1.3 *Implicit positive versus moot.* Recall that the standard for grounding is "well enough for current purposes" (Clark 1996: 221), so it is efficient for the participants to use some implicit indications of understanding. However, it is more difficult for analysts, who are outside the dialogue, to judge when an implicit response is clearly negative. Therefore, in our analysis, implicit addressee responses could be either positive or moot. *Implicit positive responses* occurred when the addressee did not overtly indicate a lack of understanding. He either continued to watch the speaker/gesturer or said something that built on a presupposed understanding. The remaining cases were *moot*; the addressee was either looking away from the speaker/gesturer or said something unrelated to the previous presentation of information, possibly overlooking or ignoring the speaker's contribution. We deemed these responses to be moot because they were not even implicitly positive. First, two analysts examined all implicit responses and, together, differentiated between positive and moot addressee responses. Then a third analyst did the same analysis independently for six of the 22 dyads. Reliability between the first pair and the third analyst was 91%.

5.2.1.4 *Results for addressee responses.* The results provided strong, moment-by-moment dyadic evidence that addressees understood presentations with nonredundant gestures. The vast proportion of their responses were positive (M = .955, SD = .052) rather than negative or moot (M = .045, SD = .051). A paired-sample t-test indicated that these mean proportions were significantly different ($t_{(21)}$ = 41.7, p < .001). Moreover, even though the dialogues often had rapid or even overlapping exchanges, the addressees were more likely to provide explicit feedback (M = .617, SD = .140) than only implicit feedback (M = .382, SD = .139); again, this difference was significant ($t_{(21)}$ = 3.95, p = .001). As suggested by these mean proportions, the addressees' positive responses were more often explicit (M = .589, SD = .152) than implicit (M = .366, SD = .143); $t_{(21)}$ = 3.60, p < .01. That is, addressees were significantly more likely to provide overt evidence that that they had understood the speaker/gesturer's presentation than inferential evidence. It is noteworthy that explicit negative responses (indicating that the addressee had *not* understood) were extremely rare (M = .029, SD = .036). All of these 17 instances were questions. Our impression was that, in about half of these cases, the addressee was seeking to clarify a genuine misunderstanding. In the remaining cases, the addressee may have understood and was asking a question as a polite way of disagreeing (e.g., "Oh so you walk through the kitchen into the living room?"). Finally, the mean proportion of implicit addressee responses that were moot was also very small (M = .016, SD = .035).

5.3.2 *Acknowledgment by the speaker/gesturer*

In a fully explicit grounding sequence, the speaker/gesturer would acknowledge (or correct) the addressee's indication of understanding. However, constantly stopping the flow of content to acknowledge the correctness of the addressee's understanding would be quite inefficient, violating the principle of *least joint effort* (e.g., Clark 1996, Clark & Krych 2004, Clark & Schaeffer 1989). Indeed, this third step in the grounding sequence does not appear in many versions of the theory (e.g., Clark, 1996). Most versions treat the speaker's confirmation of understanding as the default response, which would therefore be implicit. We tested this assumption by examining what the speakers in our task actually did to close each grounding sequence.

5.3.2.1 *Speaker/gesturer's acknowledgment.* We analyzed three possible responses. An *explicit acknowledgment* was analogous to an explicit addressee response; the speaker/gesturer responded overtly, e.g., saying "right" or "OK," nodding, finishing the addressee's sentence, or repeating the addressee's exact word(s). An *implicit acknowledgment* occurred when the speaker/gesturer's response presupposed that the addressee had understood so far. For example, when the speaker/gesturer simply went on to finish what she had been saying before the addressee had responded, or she clarified the information in the addressee's response without overt acknowledgment (e.g., did not say "yeah"), or she continued by presenting new information. There were also some *other* responses, such as when the speaker/gesturer said or did something unrelated to the addressee's response or the addressee took up the turn before the speaker/gesturer could continue. Two analysts made these decisions independently for19 of the 22 groups, with 87% agreement for all gesture sequences within those groups.

5.3.2.2 *Results for acknowledgments.* The speaker/gesturers' acknowledgment of addressees' understanding was seldom explicit ($M = .151$, $SD = .081$). Instead, they usually acknowledged implicitly, such as by moving on to new information ($M = .767$, $SD = .078$). There were few "other" responses ($M = .081$, $SD = .073$), which strongly suggests that both participants were completing each grounding sequence (albeit implicitly) rather than interrupting it with other actions.

Recall that the 17 instances of explicit negative addressee responses were questions. The speaker/gesturer's response in 15 of these instances was to answer the question or otherwise clarify what she had presented, usually within a few seconds. That is, the speaker acknowledged the state of the addressee's understanding by providing the required information.

5.3.3 *Results for the grounding sequences*

A grounding sequence is a sequence of contingent actions, and Table 1 shows the proportional relationships between the addressees' and the speaker/gesturers' responses. In the most frequent pattern (42% of the sequences), the addressee indicated his understanding explicitly (e.g., saying "yeah" or repeating the words), then the speaker/gesturer followed up implicitly (e.g., continuing on to new information).

In the next most common pattern (36% of the sequences), the addressee responded implicitly (e.g., simply continued to pay attention), and the speaker/gesturer also carried on implicitly. It is noteworthy that in these cases, the speaker/gesturer did not explicitly check on her addressee's level of understanding. The speaker/gesturer seemed to have acted on the default assumption that the information in her speech/gesture

Table 1. Sequential proportions of addressee responses and speaker follow-up responses

Addressee response	Speaker follow-up response	M (SD)
Explicit	Explicit	.14 (.08)
	Implicit	.42 (.13)
	Other	.05 (.06)
Implicit	Explicit	.01 (.02)
	Implicit	.36 (.13)
	Other	.02 (.03)

combination was successfully grounded *unless* the addressee explicitly revealed that it was not. Much less frequently (14% of sequences), both participants grounded explicitly. All of the remaining combinations were rare.

6. Summary

Ultimately, it is the participants themselves who determine the communicative value of their gestures. We tested their mutual understanding using a microanalysis of grounding sequences after each nonredundant speech/gesture combination that located a room or rooms in their proposed floor plan. Gestures are essential in such spatial tasks, and virtually all of the gestures they used carried information that was not in their words. The addressee had to understand both the words and gestures together.

Mutual understanding was potentially even more difficult in this task because the gestures lacked any external anchor or referent. There were no real objects or spaces to point at or manipulate. The dyad had to co-construct and sustain the invisible floor plan with their words and abstract deictic gestures. In spite of the difficulty of their task and the speed of spontaneous dialogue, only 4.5% of the addressees' 552 responses indicated that they had not understood the information that the speaker/gesturer had presented.

The results suggest that this method would be useful both for looking even more closely at how dyads understand each other's gestures and for examining the process in other situations. Grounding is an "opportunistic" process (Schober & Clark 1989) in which the participants seize on whatever works, and solutions to grounding in other contexts could not fail to be interesting.

References

Bangerter, A. 2004. "Using pointing and describing to achieve joint focus of attention in dialogue." *Psychological Science* 15 : 415–419.

Bavelas, J., Chovil, N., Coates, L. and Roe, L. 1995. "Gestures specialized for dialogue." *Personality and Social Psychology Bulletin* 21 (4): 394–405.

Bavelas, J., Chovil, N., Lawrie, D. A. and Wade, A. 1992. "Interactive gestures." *Discourse Processes* 15: 469–489.

Bavelas, J., Coates, L. and Johnson, T. 2000. "Listeners as co-narrators." *Journal of Personality and Social Psychology* 79 (6): 941–952.

Bavelas, J., Gerwing, J., Sutton, C. and Prevost, D. 2008. "Gesturing on the telephone: Independent effects of dialogue and visibility." *Journal of Memory and Language* 58: 495–520.

Beattie, G. and Aboudan, R. 1994. "Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships." *Semiotica* 99: 239–272.

Clark, H. H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. H. and Krych, M. A. 2004. "Speaking while monitoring addressees for understanding." *Journal of Memory and Language* 50: 62–81.

Clark, H. H. and Schaefer, E. F. 1989. "Contributing to discourse." *Cognitive Science* 13: 259–294.

Furuyama, N. 2000. "Gestural interaction between the instructor and the learner in origami instruction." In *Language and Gesture*, D. McNeill (ed), 99–117. Cambridge: Cambridge University Press.

Gerwing, J. and Allison, M. 2009. "The relationship between verbal and gestural contributions in conversation: A comparison of three methods." *Gesture* 9 (3): 312–336.

Gerwing, J. and Bavelas, J. 2004. "Linguistic influences on gesture's form." *Gesture* 4 (2): 157–195.

Holler, J., Shovelton, H. and Beattie, G. 2009. "Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context?" *Journal of Nonverbal Behavior* 33: 73–88.

Holler, J. and Stevens, R. 2007. "The effect of common ground on how speakers use gesture and speech to represent size information." *Journal of Language and Social Psychology* 26 (1): 4–27.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

McNeill, D., Cassell, J. and Levy, E. T. 1993. "Abstract deixis." *Semiotica* 95: 5–19.

Özyürek, A. 2000. "The influence of addressee location on spatial language and representational gestures of direction." In *Language and Gesture*, D. McNeill (ed), 64–83. Cambridge: Cambridge University Press.

Özyürek, A. 2002. "Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures." *Journal of Memory and Language*, 46: 688–704.

Schober, M. F. and Clark, H. H. 1989. "Understanding by addressees and overhearers." *Cognitive Psychology* 21: 211–232.

CHAPTER 5

If you don't already know, I'm certainly not going to show you!

Motivation to communicate affects gesture production

Autumn B. Hostetter,[1] Martha W. Alibali[2] and Sheree M. Schrager[3]
Kalamazoo College,[1] University of Wisconsin-Madison[2] and The Saban Research Institute, Childrens Hospital Los Angeles[3]

The present study aimed to determine if variations in a speaker's motivation to communicate influence the frequency or size of the gestures the speaker produces. We observed the gestures produced by speakers as they gave route directions to a listener who they believed would use the information either to cooperate with them in a later game, compete with them, or merely play simultaneously. Gesture rates were not affected. However, speakers produced a higher proportion of gestures that were large in size when they expected their listener to cooperate with them than when they expected their listener to compete with them. These findings suggest that gestures are shaped in part by speakers' desire to communicate information clearly to their listeners.

Introduction

Speakers frequently produce representational gestures that depict an image of the spatial objects, properties, or relationships that they are describing (Alibali 2005; Alibali, Heath, & Myers 2001; Krauss 1998). There is some controversy about whether such gestures actually contribute significantly to listeners' comprehension of spoken messages. Some evidence suggests that listeners glean very little from speakers' gestures (Krauss, Dushay, Chen, & Rauscher 1995; Krauss, Morrel-Samuels, & Colasante 1991) while other evidence suggests that listeners comprehend better when speakers use gestures (Kelly, Barr, Church, & Lynch 1999; Kendon 1994; Riseborough 1981 Rogers, 1978). The communicative effectiveness of representational gestures is likely mediated by several factors, including the redundancy of the gestures with speech (Kelly & Church 1999), the clarity of the speech signal (Graham & Argyle 1975; McNeil, Alibali, & Evans 2000; Rogers 1978) and the size of the gestures (Beattie & Shovelton 2005).

However, regardless of whether listeners actually benefit from gestures, speakers sometimes produce their gestures as though they want listeners to attend to them. Melinger and Levelt (2004) asked speakers to convey information about both the size and shape of stimuli. They found that speakers occasionally depicted information about one of the dimensions in their gestures without also articulating the information in speech. This suggests that the speakers were intentionally using their gestures to communicate necessary information. Further, many studies have demonstrated that speakers alter the form and quantity of their gestures depending on the position and knowledge of their audience, suggesting that speakers take their listeners' perspectives into consideration when planning and producing representational gestures (Gerwing & Bavelas 2004, Holler & Stevens 2007, Jacobs & Garnham 2007, Özyürek 2002). For example, Gerwing and Bavelas (2004) found that speakers produced larger, clearer gestures when they were describing information that was new than when they were describing information that had been mentioned before. Similarly, Holler and Stevens (2007) found that speakers were more likely to produce gestures when describing information that was unknown to their listeners than when describing information that was known. It seems, then, that speakers produce gestures that are more frequent and larger when they believe that their listeners may have difficulty comprehending. The present study investigates whether this consideration for the listener is always present and manifested in gesture, or whether it depends on the speaker's motivation to communicate clearly.

The factors that influence the quantity and form of speakers' gestures are a matter of theoretical debate, with some theories describing gestures as being shaped primarily by cognitive factors (see, for example, de Ruiter 2000; Kita 2000; Krauss, Chen, & Gottesman 2000) and others describing gestures as being shaped by more social factors (see, for example, Bavelas, Chovil, Coates, & Roe 1995; Kendon 2004). Recently, a framework has been proposed that considers gesture as being influenced by both cognitive and social factors. According to the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali 2008), gestures are overt manifestations of the perceptual and motor simulations that underlie thinking and speaking. Whenever speakers think about spatial information, their neural and cognitive systems activate the perceptual and motor states that are involved in actually perceiving and interacting with spatial information (e.g., Barsalou 1999; Glenberg & Kaschak 2002; Wexler, Kosslyn, & Berthoz 1998).

Although such simulations always underlie spatial thinking and speaking, the GSA framework proposes that speakers can change the likelihood that a particular simulation will be expressed as an overt gesture by changing their *gesture threshold*. The gesture threshold is conceptualized as the minimum amount of simulated action that is needed for the motor system to produce an overt gesture. Speakers can maintain a high threshold, and thereby prevent the majority of their simulations from being produced as overt gestures, if they do not wish to gesture in a particular situation. This may be particularly likely when speakers are in situations where they feel that gestures are rude or inappropriate or when they are being intentionally vague. Similarly, speakers may also maintain a low threshold, and thereby increase the number of simulations that come to be expressed as gestures. This may be particularly likely when speakers are in situations where they believe a gesture would be strongly helpful in conveying their meaning or in situations where they are particularly motivated to communicate clearly.

The purpose of the present study is to examine whether variations in speakers' motivation to communicate information lead to differences in gesture production. Toward this aim, we asked speakers to describe route information that they believed would be relevant to their success in a subsequent game. In one condition, speakers were told that the person they were communicating with would be cooperating with them in the game, thus increasing their motivation to communicate the route information clearly. In a second condition, they were told that their addressee would be competing with them in the game, thus decreasing their motivation to communicate the route information clearly. In a control condition, they were told that their addressee would be playing the game simultaneously, but that their success in the game in no way depended on the other person's performance.

Two dependent variables are of interest. First, speakers may change the *frequency* of their gestures when they are motivated to communicate clearly. According to the GSA framework, speakers can inhibit their action simulations from being realized as overt gestures, and they should be less likely to do this when they are more motivated to communicate clearly about the spatial information they describe. Thus, speakers who believe that communicating successfully will improve their own success in a future game should inhibit fewer simulations and ultimately produce more representational gestures than speakers who believe that communicating successfully will actually be detrimental to their own future success in the game. This is in line with previous studies in which speakers changed their gesture frequency depending on the knowledge of their audience (Alibali & Nathan 2007, Holler & Stevens 2007, Jacobs & Garnham 2007). Second, speakers may also change the *size* of their gestures depending on their motivation to communicate clearly. The action simulations involved in describing spatial route information may be so strong that they are difficult to suppress entirely, even when a speaker sees expressing such information as potentially detrimental to his or her own future success in the game. Simulations may still be expressed as gestures, but on a smaller scale than they otherwise would be. Indeed, previous research has shown that speakers produce larger gestures when their audience is more likely to benefit from them (Gerwing & Bavelas 2004) and that larger gestures are more communicatively effective than smaller gestures (Beattie & Shovelton 2005). It is expected that speakers will produce larger gestures when communicating clearly is important to their own success in a game than when it is irrelevant or detrimental.

Method

Participants

Sixty-eight native English speakers volunteered to participate in exchange for extra course credit. The sample was largely Caucasian, with 9% of participants claiming an ethnicity other than white (Asian or Hispanic). Data from 19 participants were not included in the final analyses either because their data were not properly recorded ($n = 1$), because they reported being suspicious of some aspect of the experimental setup (the camera, the confederate, the cover story, or the interest in gesture; $n = 13$), because they did not correctly follow the instructions for describing the routes ($n = 3$), or because they failed the manipulation check that tested their understanding of the game's rules ($n = 2$). There was no difference in the number of participants excluded from each experimental condition. The final sample included 49 participants (39 female, 10 male).

Materials

A map of a fictitious town was created in Appleworks 6.0 (see Figure 1). The map depicted 10 buildings and locations (e.g., factory, library, park) as well as several landmarks (e.g., river, fountain). The map was printed in color on an 8.5 x 11 in. sheet of paper and laminated. A list of five routes accompanied the map (e.g., Factory → Library; Shopping Mall → Home, etc.). A questionnaire was also created to test participants' knowledge of the game's rules.

Procedure

Two experimenters alternated between the experimenter and confederate roles. The confederate for each session posed as a participant and arrived in the waiting room five minutes prior to the start of the experiment. The experimenter led the participant and the confederate to the testing room together.

Participants were told that the study was about people's ability to navigate new spatial layouts and that there were two conditions in the study: a map condition and a verbal condition. The participant in the map condition would study a map of a fictitious town. The participant in the verbal condition would hear a verbal description of the town's layout given by the participant in the map condition. Both participants would then play a video game that took place in the fictitious town. In this video game, the players would each control a taxi cab and earn points by successfully delivering passengers to their requested locations. Following this brief overview and the participant's signed consent to participate, the experimenter pretended to randomly assign the participant to the map condition and the confederate to the verbal condition.

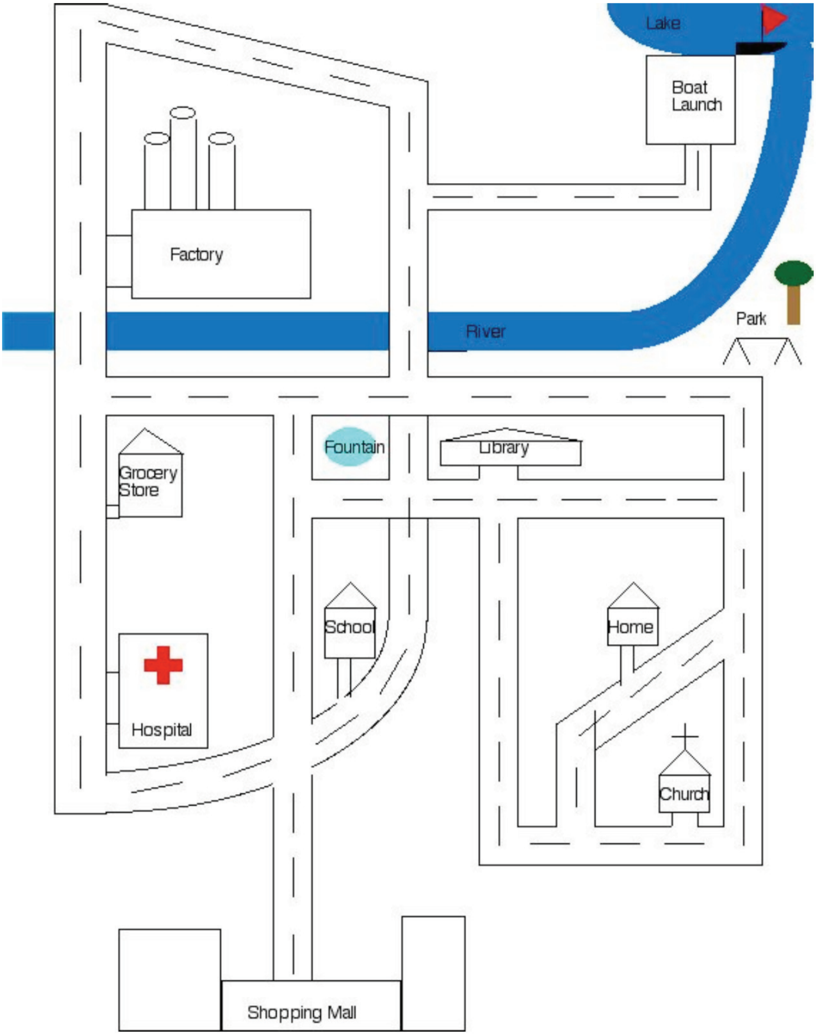


Figure 1. The map of the fictitious town that participants were asked to study. Specifically, they were asked to learn and describe five routes: Factory → Library; School → Grocery Store; Park → Boat Launch; Hospital → Church; Shopping Mall → Home.

The participant then received the map along with the list of five routes that would supposedly occur frequently in the video game. The experimenter stressed the importance of paying particular attention to the routes on the list, including landmarks that are passed along the way so that the routes would seem familiar during game play. The experimenter then left the room for five minutes while the participant studied the map and list. The confederate remained in the testing room and worked quietly on a word-find puzzle.

When the experimenter returned, she took away the map and list of routes and explained more about the taxi driver game. She stressed the necessity of staying on the designated roads while delivering passengers. She also explained that the two would be playing the game simultaneously, and each would be able to see the other player's taxi cab on the screen as well as his or her own cab. The experimenter then discreetly consulted a random assignment schedule and administered one of the three experimental manipulations. In the *neutral condition*, the experimenter stated that the two drivers should try not to be too distracted by one another as their scores would be calculated completely independently. Each player's score would be based solely on the number of fares he or she delivered successfully, regardless of how well the other person had done. In the *cooperative condition,* the experimenter stated that the two drivers should watch out for one another and try not to get in each other's way because both drivers were part of the same team. Every time the other driver delivered a fare successfully, the score of both drivers would increase. In the *competitive condition*, the experimenter stated that the two drivers should think of themselves as drivers for rival cab companies who were competing for fares. They should try to beat one another to passengers, as every time one driver successfully delivered a fare, the other driver's score would decrease.

Following this manipulation, the participant and the confederate filled out a questionnaire to test their understanding of the game's rules that had just been described. This questionnaire included several filler questions, as well as two questions of interest. Each question was followed by three options. The first question was *How will the other person's score affect your score during the game?* with the options (*a) not at all (their performance does not affect my score)*, (*b) negatively (if they deliver passengers successfully, my score will decrease),* and (*c) positively (if they deliver passengers successfully, my score will increase)*. The second question was *During the game, how should you treat the other person's cab?* with the choices (*a) stay out of the other person's way*, (*b) try to get in the other person's way and beat him/her to passengers*, and (*c) ignore what the other person is doing*. For each question, the correct answer depended on the experimental manipulation each participant had received. For example, participants in the competitive condition should select b for the first question and b for the second question while participants in the cooperative condition should select c for the first question and a for the second question.

The experimenter next explained that the participant in the verbal condition needed to receive a verbal description of the town's layout. The experimenter stated that she would go through the list of common routes one at a time and ask the participant who had just studied the map to give as detailed a description as possible of how to best navigate each route. The participant should include landmarks where possible, try to be specific about things like whether the route required a left or right turn, and take as much time as needed for each description. The experimenter also explained that the descriptions would be audio taped so that they could be checked later for accuracy. The experimenter maintained the cover story by briefly instructing the confederate to pay close attention to the routes and to visualize what each one might look like in the game.

The experimenter then pressed record on the audio tape recorder and prompted the participant to describe the first route. When the participant finished describing the first route, the experimenter prompted the participant with the second route, and so on until all five routes were described. During the descriptions, the confederate remained oriented toward the participant at all times and gave occasional small nods to indicate understanding. The hidden video camera was positioned to record a head-on view of the participants during their descriptions.

Following the five descriptions, the experimenter explained the true purpose of the study and gave the participants an opportunity to withdraw their video data. All declined. Finally, participants completed a debriefing questionnaire where they reported whether they were suspicious of the camera, the confederate, the interest in gesture, or the video game.

Data coding

Participants were screened for inclusion based on their answers to the debriefing questionnaire, their answers to the manipulation questions, and their adherence to the instructions to describe routes that did not deviate from the town's designated roads. Each route from the remaining participants was then assigned an accuracy score ranging from 0 to 4, according to the following rubric. *Incomplete* (0) was assigned to descriptions that were not complete, such as when participants stopped midway through their descriptions and said that they did not remember anymore. *Inaccurate* (1) was assigned to descriptions that were not an accurate reflection of how to travel between the two locations. For example, participants described a different route than the one asked or misremembered the location of one of the two buildings. *Fairly Accurate* (2) was assigned to descriptions that described the correct locations of the buildings but did not provide an accurate account of how to get from one to the other. For example, some participants misremembered the correct sequence of turns involved in a route. *Accurate* (3) was assigned to routes that described an accurate route between the named locations. *Accurate with Details* (4) was assigned to routes that described a correct route and included one or more details, such as landmarks.

Each route description was also coded for accompanying gestures. Each gesture that occurred was described and categorized with respect to *type* and *size*. Gestures could be one of two *types*: representational or beat. Representational gestures were those that conveyed semantic information about the accompanying speech. For example, a movement to the left with the phrase "you turn left on the next street" was coded as a representational gesture. Beat gestures were those that did not convey semantic information about the accompanying speech. For example, a bimanual up and down movement on the word *end* in "you go to the end of the street" was coded as a

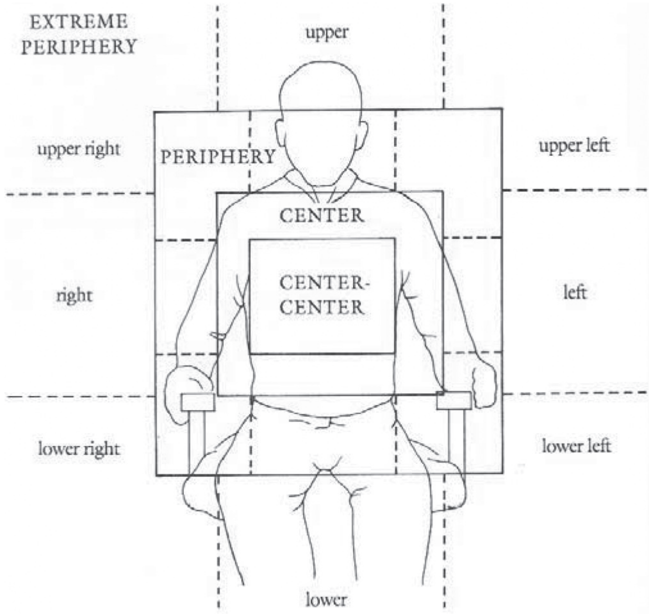


Figure 2. The typical gesture space of an adult speaker used to code the size of representational gestures. Copyright 1992 by the University of Chicago. Reprinted with permission.

beat gesture. Both representational and beat gestures were converted to rates per 100 words for each route.

The *size* of each representational gesture was also coded. Following McNeill (1992) and Beattie and Shovelton (2005), we consulted the diagram depicted in Figure 2 to classify size by determining the number of spatial boundaries each gesture crossed. We then calculated the proportion of representational gestures produced by each participant that crossed one or more boundaries.

Reliability

Three coders worked independently to code the data of the participants. Once all participants' data had been coded by one of the three coders, one of the coders reviewed the codes assigned to 18 participants (approximately 37% of the data) by the other two coders in order to establish reliability. Agreement for coding the accuracy of each route was 87%. Agreement for segmenting individual gestures from the stream of manual activity was 94% ($N = 695$). Agreement for classifying each gesture as representational or beat was 96%, and agreement for classifying each representational gesture ($N = 559$) as crossing a boundary or not was 82%. The codes assigned by the original coders were used in all cases.

Results

We begin by comparing the accuracy and amount of speech produced by speakers when they believed they would be competing, cooperating, or playing simultaneously with the confederate. We then compare the frequency and size of the gestures produced by speakers in the three conditions.

Analysis of speech

Accuracy of Speech. The accuracy ratings assigned to each route were analyzed with a one-way ANOVA, which revealed a significant effect of condition, $F(2, 46) = 5.56$, $p = .007$. Participants who believed their addressees would be cooperating with them described routes more accurately than participants who believed their addressees would be competing against them or merely playing at the same time (see Figure 3). This suggests that our manipulation did influence speakers' motivation to communicate as we intended.

However, this difference in speech accuracy could confound the gesture analyses; that is, speakers in the cooperative condition may gesture differently from those in the competitive and neutral conditions because their speech is richer and more accurate. Previous research suggests that gestures help speakers produce speech that is more informative (Hostetter, Alibali, & Kita 2007) and more image-evoking (Rimé, Schiaratura, Hupert, & Ghysselinckx 1984) than speech produced without gestures. Thus,

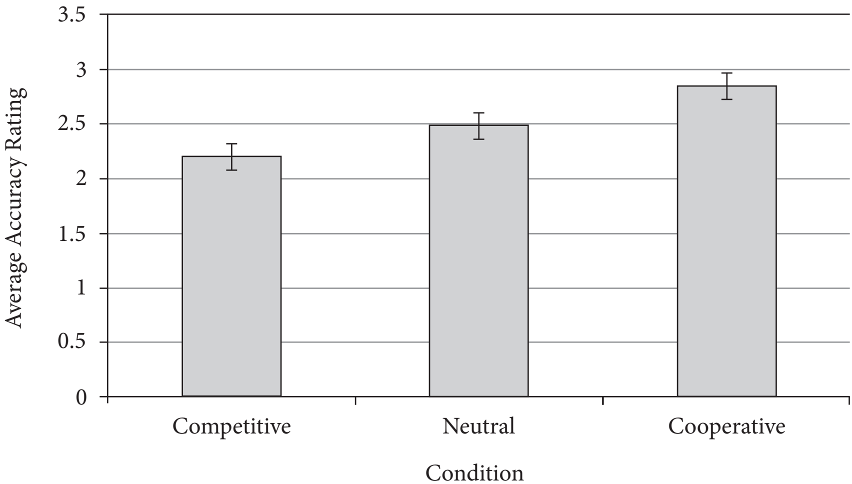


Figure 3. The average accuracy ratings assigned to all routes in each of the three experimental conditions. Error bars represent standard errors of the means. See text for complete description of coding rubric.

in the present experiment, speakers who believed they would be cooperating may have gestured differently for self-oriented reasons (i.e., to help themselves produce more accurate and informative speech) rather than for listener-oriented reasons (i.e., to create an image that can be referenced by the listener). While the self-oriented functions of gesture are certainly interesting, our focus in the present experiment is on listener-oriented changes in gesture. It is thus important not to confound motivation to communicate with the accuracy of the speech produced. We therefore limited all further analyses to only those routes from each participant that were rated as either accurate or accurate with details ($N = 159$).

Amount of Speech. A one-way ANOVA compared the average number of words produced by participants as they described accurate routes in each of the three conditions and revealed no significant differences, $F(2, 48) = 1.041$, $p = .36$. Participants who believed they would be cooperating with their listeners did not produce more words ($M = 71.51$, $SD = 24.30$) than participants who thought they would be competing ($M = 58.94$, $SD = 25.01$) or playing simultaneously ($M = 68.31$, $SD = 27.45$) with their listeners.

Analysis of gesture

Frequency of Gesture. A 3 (condition: competitive, cooperative, neutral) x 2 (gesture type: representational v. beat) repeated measures ANOVA revealed no significant effects involving experimental condition on gesture rates during accurate route descriptions. Contrary to our hypothesis, speakers who believed they were cooperating with their listeners did not produce more representational gestures per 100 words ($M = 9.49$, $SD = 3.45$) than did participants who believed they would be competing against ($M = 9.72$, $SD = 5.93$) or playing simultaneously with ($M = 12.19$, $SD = 4.59$) their listeners, $F(2, 46) = 2.49$, $p = .09$. Not surprisingly given the highly spatial nature of this route description task, speakers did produce representational gestures at a higher rate ($M = 10.50$ per 100 words, $SD = 4.82$) than beat gestures, ($M = 2.10$, $SD = 2.49$), $F(1, 46) = 162.70$, $p < .001$. There was no interaction between condition and gesture type, $F(2, 46) = 3.70$, $p = .71$.

Gesture Size. We next compared the size of the representational gestures produced in each of the three conditions when speakers were describing accurate routes. The one-way ANOVA revealed a marginal effect of condition, $F(2, 46) = 2.54$, $p = .09$. A planned Fisher's LSD comparison revealed that participants who believed they would be cooperating with their listeners produced a higher proportion of gestures that crossed a boundary ($M = 0.24$, $SD = 0.19$) than did participants who believed they would be competing against their listeners ($M = 0.11$, $SD = 0.11$), $p = .03$. There were no differences involving the size of the gestures produced by participants who believed they would play simultaneously with their listeners ($M = 0.18$, $SD = 0.15$). See Figure 4.

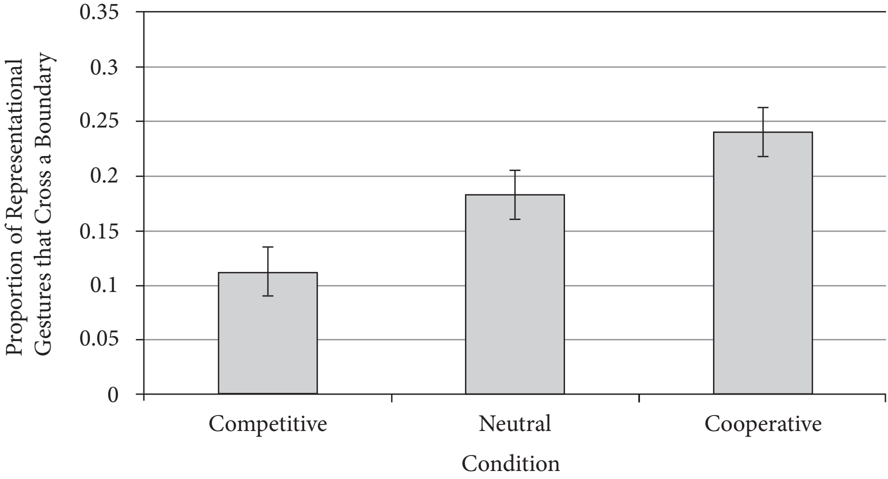


Figure 4. The average proportion of gestures that crossed a spatial boundary in each of the three experimental conditions. Error bars represent standard errors of the means.

Discussion

Do speakers alter the quantity or form of their gestures when they are motivated to communicate clearly? We found no evidence to suggest that speakers alter the quantity of their gestures depending on their communicative motivation; speakers produced similar rates of gestures regardless of whether they expected to cooperate, compete, or play simultaneously with their listeners. However, speakers produce larger gestures when their listeners' understanding will benefit their performance in a later game than when their listeners' understanding is detrimental to their own performance. Although previous research has suggested that large gestures are more communicatively effective than small gestures (Beattie & Shovelton 2005) and that speakers are more likely to produce large gestures when the information they describe is unknown to their listener (Gerwing & Bavelas 2004), this is the first evidence that speakers produce larger gestures when improving their listeners' comprehension is in their own best interest.

The present study supports the theoretical stance taken by many (e.g., Bavelas, et al. 1995, Kendon 2004) that representational gestures are shaped by aspects of the social situation. The social situation is an important determinant of whether gestures are produced in the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali 2008). The GSA framework claims that representational gestures are the by-product of cognitive simulations that recreate perceptual and motor states, but that the overt expression of such simulations is influenced by social factors. According to the GSA framework, speakers can increase the number of simulations they express as gestures by lowering their gesture threshold, i.e., by lowering the minimum amount of action simulation needed for the motor system to produce an overt gesture. In the present study, we found no evidence that speakers changed the frequency of their gestures depending on how motivated they were to communicate; instead, we found that motivation to communicate led speakers to change the size of their gestures. This suggests that the conceptualization of the gesture threshold as outlined by Hostetter and Alibali (2008) may need to be expanded. Rather than blocking simulated action from being produced as a gesture at all, a low threshold may instead attenuate the size of the gesture that is produced.

Although the manipulation used in this study was somewhat artificial, there are certainly many situations in the real world in which speakers have a genuine interest in communicating successfully to their listeners. For example, to be successful in their professions, teachers, doctors, and salespeople must communicate information clearly to their students, patients, or clients. The ways in which professionals with strong communicative motivations use gestures to accomplish their communicative goals await further study, but the present data suggest that such individuals may be particularly likely to produce gestures that are large in size. Further data is needed to determine whether these larger gestures actually aid the comprehension of listeners in professional settings.

In conclusion, the present study suggests that speakers alter their gestures depending on their motivation to communicate information clearly. When it is in speakers' best interest to explain information clearly, they produce gestures that are larger in size and, consequently, more likely to benefit their listeners. When it is not in speakers' best interest to explain information clearly, speakers do not reduce their gesture rates, but they produce gestures that are smaller in size. Thus, if someone doesn't want to tell you something, their small gestures probably won't show you either.

References

Alibali, M. W. 2005. "Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information." *Spatial Cognition and Computation* 5: 307–331.

Alibali, M. W., Heath, D. C. and Myers, H. J. 2001. "Effects of visibility between speaker and listener on gesture production." *Journal of Memory and Language* 44: 169–188.

Alibali, M. W. and Nathan, M. 2007. "Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson." In *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron and S. J. Derry (eds), 349–365. Mahwah, NJ: Erlbaum.

Barsalou, L. W. 1999. "Perceptual symbol systems." *Behavioral and Brain Sciences* 22: 577–660.

Bavelas, J. B., Chovil, N. Coates, L. and Roe, L. 1995. "-Gestures specialized for dialogue." *Personality and Social Psychology Bulletin* 21: 394–405.

Beattie, G. and Shovelton, H. 2005. "Why the spontaneous images created by the hands during talk can help make TV advertisements more effective." *British Journal of Psychology* 96: 21–37.

de Ruiter, J.-P. 2000. "The production of gesture and speech." In *Language and Gesture*, D. McNeill (ed), 284–311. Cambridge, UK: Cambridge University Press.

Gerwing, J. and Bavelas, J. B. 2004. "Linguistic influences on gesture's form." *Gesture* 4: 157–195.

Glenberg, A. M. and Kaschak, M. P. 2002. "Grounding language in action." *Psychonomic Bulletin and Review* 9: 558–565.

Graham, J. A. and Argyle, M. 1975. "A cross-cultural study of the communication of extra-verbal meaning by gestures." *International Journal of Psychology* 10: 57–67.

Holler, J. and Stevens, R. 2007. "The effect of common ground on how speakers use gesture and speech to represent size information." *Journal of Language and Social Psychology* 26: 4–27.

Hostetter, A. B. and Alibali, M. W. 2008. "Visible embodiment: Gestures as simulated action." *Psychonomic Bulletin and Review* 15: 495–514.

Hostetter, A. B., Alibali, M. W. and Kita, S. 2007. "Does sitting on your hands make you bite your tongue? The effects of gesture inhibition on speech during motor descriptions." In D. S. McNamara and J. Gregory Trafton (eds), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 1097–1102. Mahwah, NJ: Erlbaum.

Jacobs, N. and Garnham, A. 2007. "The role of conversational hand gestures in a narrative task." *Journal of Memory and Language* 56: 291–303.

Kelly, S. D., Barr, D. J., Church, R. B. and Lynch, K. 1999. "Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory." *Journal of Memory and Language* 40: 577–592.

Kelly, S. D. and Church, R. B. 1999. "A comparison between children's and adult's ability to detect conceptual information conveyed through representational gestures." *Child Development* 69: 85–93.

Kendon, A. 1994. "Do gestures communicate? A review." *Research on Language and Social Interaction* 27: 175–200.

Kendon, A. 2004. *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Kita, S. 2000. "How representational gestures help speaking." In *Language and Gesture*, D. McNeill (ed), 162–185. Cambridge, UK: Cambridge University Press.

Krauss, R. M. 1998. "Why do we gesture when we speak?" *Current Directions in Psychological Science* 7: 54–60.

Krauss, R. M., Chen, Y. and Gottesman, R. F. 2000. "Lexical gestures and lexical access: A process model." In *Language and Gesture*, D. McNeill (ed), 261–283. Cambridge, UK: Cambridge University Press.

Krauss, R. M., Dushay, R. A., Chen, Y. and Rauscher, F. 1995. "The communicative value of communicative hand gestures." *Journal of Experimental Social Psychology* 31: 533–552.

Krauss, R. M., Morrel-Samuels, P. and Colasante, C. 1991. "Do conversational hand gestures communicate?" *Journal of Personality and Social Psychology* 61: 743–754.

McNeil, N. M., Alibali, M. W. and Evans, J. L. 2000. "The role of gesture in children's comprehension of spoken language: Now they need it, now they don't." *Journal of Nonverbal Behavior* 24: 131–150.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

Melinger, A. and Levelt, W. J. M. 2004. "Gesture and the communicative intention of the speaker." *Gesture* 4: 119–141.

Özyürek, A. 2002. "Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures." *Journal of Memory and Language* 46: 688–704.

Rimé, B., Schiaratura, L., Hupert, M. and Ghysselinckx, A. 1984. "Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level." *Motivation and Emotion* 8: 311–325.
Riseborough, M. G. 1981. "Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication." *Journal of Nonverbal Behavior* 5: 172–183.
Rogers, W. T. 1978. "The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances." *Human Communication Research* 5: 54–62.
Wexler, M., Kosslyn, S. M. and Berthoz, A. 1998. "Motor processes in mental rotation." *Cognition* 68: 77–94.

Author note

Autumn B. Hostetter, Department of Psychology, Kalamazoo College, 1200 Academy Street, Kalamazoo, MI 49006, Autumn.Hostetter@kzoo.edu. Martha W. Alibali, Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson Street, Madison, WI, 53706, mwalibali@wisc.edu. Sheree M. Schrager, Community, Health Outcomes, and Intervention Research Program, The Saban Research Institute, Childrens Hospital Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA 90027, sschrager@chla.usc.edu.

This work was supported in part by a collaborative research award to Autumn Hostetter and Sheree Schrager by the Psychology Department at the University of Wisconsin-Madison. We thank Laura Dunn and Joshua McAuliffe for their help with data collection and coding.

CHAPTER 6

Measuring the formal diversity of hand gestures by their hamming distance

Katharina Hogrefe, Wolfram Ziegler and Georg Goldenberg
Clinical Neuropsychology Research Group (EKN), Hospital Bogenhausen

Based on the assumption that the formal diversity of gestures indicates their potential information content, we developed a method that focuses on the analysis of physiological and kinetic aspects of hand gestures. A form-based transcription with the Hamburg Notation System for Sign Languages (HamNoSys, Prillwitz et al. 1989) constitutes the basis for the calculation of a measure of the formal diversity of hand gestures. We validated our method in a study with healthy persons, who retold the same short video clips first verbally and then without speaking. The silent condition was expected to elicit higher formal diversity of hand gestures since they have to transmit information without support from language (Goldin-Meadow et al. 1996). Results were in line with our expectations. We conclude that the determination of the formal diversity of hand gestures is an adequate method for gesture analysis which is especially suitable for analysing the gestures of persons with language disorders.

Introduction

Over the past decades neuropsychological research on spontaneous gesturing in language impaired patients has led to contradictory outcomes. These discrepancies might be partly due to the variety of methods applied for analysing gestures. Most studies evaluate the number or communicative functions of gestures (McNeill 1992). The mere number of gestures does not allow researchers to draw conclusions about the potential information content of the produced gestures. Assigning the communicative function does so, but this method often depends on the analysis of the accompanying verbal utterances which may be insufficient or misleading in patients with language disturbances.

In this paper, we describe a form-based approach for the evaluation of gestures which enables a quantitative comparison between subjects. Hand gestures are transcribed with a modified version of a notation system which was originally developed for sign languages – the Hamburg Notation System for Sign Languages (HamNoSys, Prillwitz et al. 1989). For statistical analysis, we used a measure from *Information and Coding Theory* (Jones & Jones 2000): The *hamming distance* which determines in how many formal features (e.g. handshape, location of the hand with respect to the body) two gestures differ from each other.

In the following section, gesture transcription and the calculation of the hamming distance are described. Further, the interrater-agreement is determined, and the method is validated on the basis of a study with twelve healthy subjects.

Gesture transcription

This method focuses on the analysis of hand gestures. A movement between two rest positions was defined as a gesture or as a sequence of gestures. Body-focused movements, which involve some kind of self-stimulation (Freedman 1972) and usually display a non-phasic structure were excluded from the analysis as we were interested in gestures with communicative content.

Handedness

As our method was developed for the analysis of the data of neuropsychological patients, the issue of the handedness of a gesture is of particular importance. Many of these patients suffer from hemiparesis and can use only one hand for gesturing. To make the method equally suitable for patients with and without hemipareses, all gestures were transcribed as if they were performed unilaterally with the right hand. A code at the beginning of the transcription of each gesture indicated which hand was actually used:

1. unilateral right hand gesture
2. unilateral left hand gesture
3. both hands parallel (acting synchronously or alternating)
4. both hands, right hand dominant
5. both hands, left hand dominant

In cases of left hand gestures and both hand gestures with left hand dominance, the movement of the left hand was mirrored for the notation. This “normalization” of handedness was important for the calculation of the diversity of hand gestures with the hamming distance as persons who are able to use both hands are likely to obtain higher hamming distances (see below).

HamNoSys

The Hamburg Notation System for Sign Languages (Prillwitz et al. 1989) was developed in the tradition of an earlier notation system for sign languages (Stokoe 1960) but is more detailed and tries to maintain an iconic relationship between the symbols and their referents. "Like the phonetic alphabet for spoken languages, HamNoSys should be capable of describing all signs in all sign languages" (Prillwitz et al. 1989: 6). As such the notation system is capable of describing all physiological possible characteristics of signs, and it should also be capable of describing speech-accompanying hand gestures. Version 2.0 of HamNoSys contains a set of approximately 160 symbols. There are two newer versions (3.0, 4.0) which provide extensions to the version 2.0. These extensions mainly tap aspects that were not important for our purposes such as nonmanual components, (for example, see http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html). Hence, the method presented in this paper refers to the version 2.0.

The reduced HamNoSys symbol set

We slightly modified HamNoSys and reduced the symbol set for the transcription of spontaneous gestures. Symbols used for the detailed notation of very specific handshapes like for instance the symbols for distinct parts of the finger (e.g. joint, nail) were left out of the set as such fine-graded distinctions were not expected to play a role in spontaneous gesturing. Two transcibers were involved in the choice of the symbols. Finally, a selection of 105 symbols remained (an overview of the complete set is presented in Hogrefe 2009).

The notation of a gesture with HamNoSys includes the depiction of the configuration of the hand at the beginning of the *stroke* (for a detailed description of the structural organization of gestures compare Kita et al. 1998 and Seyfeddinipur 2006: 82 f.). This *starting point configuration* (as termed in HamNoSys) results from the *handshape* and the *orientation of the hand* (captured by the parameters *extended finger orientation* and *palm orientation*) as well as the *location* of the hand with respect to the body. Further, possible actions of the hand are notated with the parameters *movement* and *repetition*.

Handshape. The notation of the handshape consists of symbols specifying the basic types of handshapes as well as diacritic symbols for the position of the thumb and the bending of the fingers. Figure 1 shows the basic types *fist, flat hand*, and variations of *separated fingers*. These basic types can be modified by the symbol for the extended thumb or the symbol for thumb crossing.

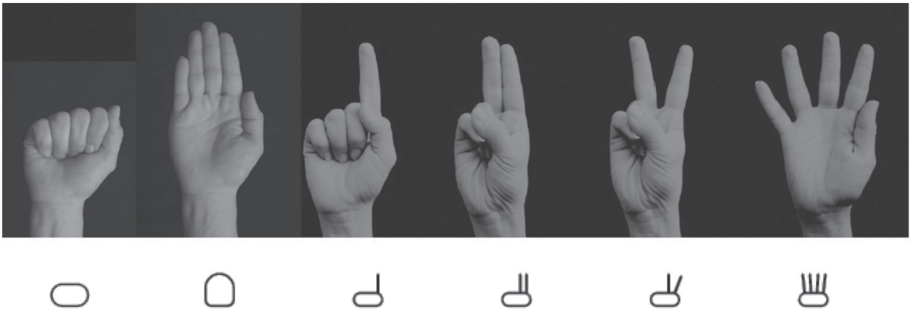


Figure 1. Basic handshapes: fist, flat hand, separated fingers.

Other types of handshape are *thumb combination* handshapes in which the position of the thumb and its relation to the other fingers determines the structural configuration of the entire hand. One example for this is the ring-gesture, where the tip of the thumb and tip of the index finger touch each other to build the form of a ring. HamNoSys distinguishes between closed types, where the thumb is in contact with one or more other fingers and open types, where the thumb does not get in touch with other fingers (Prillwitz et al. 1989: 9f). Further, the bending of the digits can be indicated by adding diacritic symbols for *flat, round* and *sharply bent*.

Hand Orientation: Extended finger orientation and palm orientation. The description of the orientation of the hand results from the notation of the two parameters *extended finger orientation* and *palm orientation*. This leads to a three-dimensional depiction of the hand. Two degrees of freedom are determined with the *extended finger orientation* which corresponds to the direction pointed to by the fingers when fully extended (compare Figure 2).

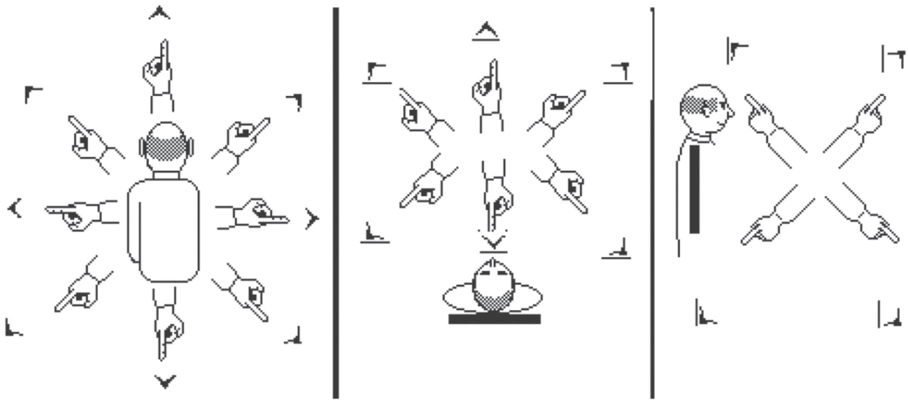


Figure 2. Orientation of the extended fingers (taken from http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html).

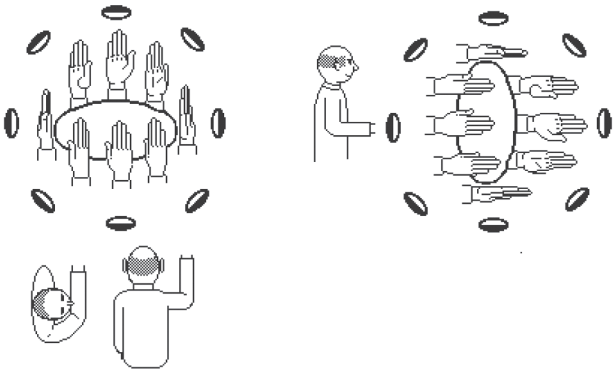


Figure 3. Orientation of the palm (taken from http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html).

The vertical and horizontal lines (body referent lines) refer to the orientation of the fingers with respect to the body of the speaker. Symbols can be combined for the description of double diagonal orientation, e.g. *away from the body to the left* and *downward and away from the body.* For determining the third degree of freedom of the orientation of the hand HamNoSys offers eight symbols for the *palm orientation*. The symbols are ovals; the darkened side indicates the direction of the palm.

Location. Location describes the position of the hand with respect to the body. Most of the symbols in this category refer to specific locations of the body or the head.

In HamNoSys, the torso is divided into three larger layers whereas there is a more differenciated segmentation with eleven different signs for the more specific positions at the head (e.g. eyes, nose, mouth, forehead etc.). Apart from the symbols which refer to the parts of the body, there are symbols which specify the position of the hand with regard to the respective body part in more detail, e.g. *on the left/right side of*, *in contact with*, or *with outstretched arm*. Those additional symbols for the detailed specification of the position of the hand only apply when the position is outside of the neutral gesture space in front of the upper part of the body.

	Nose		Head		Neck
	Ear		Top of head		Shoulder
	Cheek		Forehead		Chest
	Mouth		Eyebrows		Stomach
	Chin		Eyes		Below stomach

Figure 4. Examples of HamNoSys symbols for the category location.

↑	Straight movement
↻	Circling movement
∩	Curved movement
∼	Wavy-lined movement
ʌʌ	Zigzag movement

Figure 5. Examples of HamNoSys symbols for the category movement.

Action. Actions are coded with the categories *movement* and *repetition*. They are used for describing changes of the hand position after the beginning of the stroke. They denote different types of movement like straight, curved, waved, zigzag, or circular movements. See Figure 5 for a selection of these symbols.

Arrows can represent straight movements, or they can be combined with other symbols to indicate the direction of the movement. Further, there are symbols which describe the size of movement (*large, small*). Finally, a single repetition or multiple repetitions with stable or changing starting point can be transcribed.

Whereas translational movements of the hand change the position of the hand relative to body and external space, changes of the hand orientation can be produced without a translational movement, for instance, by rotation of the lower arm at the elbow joint. In this case, the change of the hand orientation is transcribed by means of a substitute symbol in the categories *extended finger orientation* or *palm orientation*.

Note that for batons, which have a biphasic structure and do not comprise a stroke, only the most accentuated point is denoted. Hence, for these gestures the categories *movement* and *repetition* remain empty (indicated by zero).

Figure 6 illustrates the transcription of a gesture. After the notation of the number of the gesture and the onset time of the stroke, it was indicated that this gesture was transcibed as if it had been performed with the right hand ("1"). In the next column we find the notation of the original hand choice. In this example, both hands act in parallel ("3"). Further we find the six categories of the HamNoSys transcription:

– *Handshape*: flat hand
– *Extended finger orientation*: upwards and to the right
– *Palm orientation*: upwards
– *Location*: in front of the right shoulder
– *Movement*: straight to the right
– *Repetition*: no repetition.

Figure 6. Example gesture transcription.

The input programm HamNoChart

In the project "Spontaneous Gesturing in Patients Suffering from Brain Damage" (German Research Foundation, DFG GO 968) we developed the input program HamNoChart (Zierdt et al. 2006) for a computer based gesture transcription with HamNoSys. In this system, notation symbols are displayed on the screen and can be entered into the transcription window by mouse click.

HamNoChart offers the possibility to select which symbols are needed for a particular purpose. Only the selected symbols are shown on the screen, which makes transcription less error-prone. The program possesses two saving functions: first, the transcript can be saved in a txt-format. The txt-file displays the symbols in the unicode format, and it can be imported into a word document. Second, HamNoChart can create a data file which transforms the HamNoSys symbols into numerical codes. This file allows the statistical analysis of the data with the programs MATLAB and SPSS.

A measure of diversity: The hamming distance

We aimed to develop a quantitative measure for the description of the information content of spontaneously produced gestures and prove its usefulness. On the basis of a HamNoSys transcript of a given number of gestures, the formal diversity of the gestures is determined. For this purpose, we applied a measure from the Information and Coding Theory (Jones & Jones 2000), namely the hamming distance. The hamming distance measures in how many features two gestures differ from each other. Figure 8 displays an example transcription of three gestures. In this example, gesture 1 differs from gesture 2 in four features, resulting in a hamming distance of 4. Gesture 1 differs from gesture 3 in one feature, resulting in a damming distance of 1. For each gesture the mean hamming distance is calculated. For the given example the mean hamming distance for gesture 1 is 2.5.

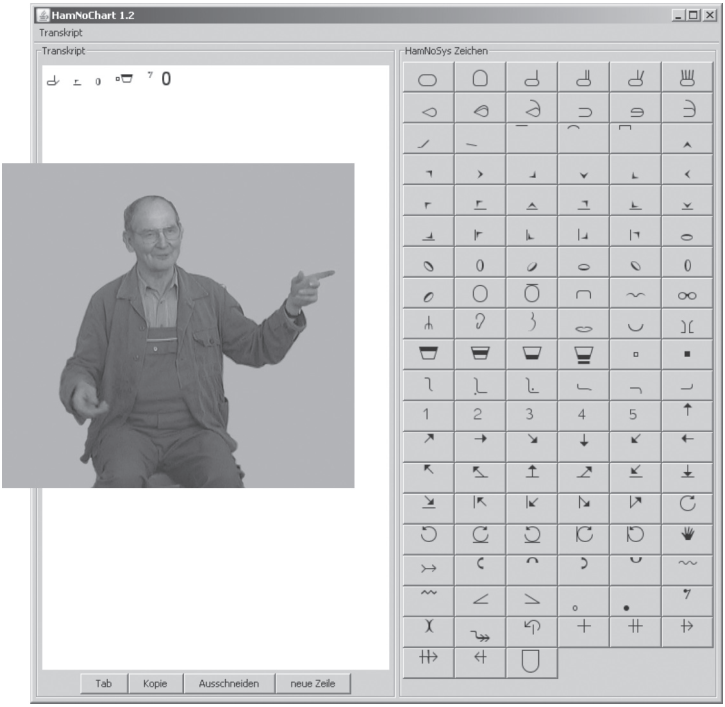


Figure 7. Gesture transcription with the input program HamNoChart (Zierdt et al. 2006).

This procedure is conducted for all gestures in a sample. Then the grand average of all gestures is determined for each subject. A low value indicates that many similar gestures were produced whereas a high value reflects a high formal diversity of gestures.

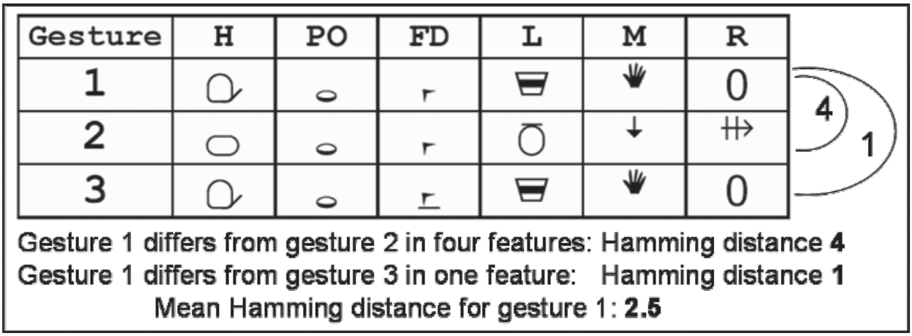


Figure 8. Calculation of the hamming distance for one gesture in a short example transcript.

Interrater-reliability of the method

In a master's thesis, Kögl (2006) evaluated the method as described so far. She collected data of five persons and determined together with the first author of this paper the interrater-reliability. Gestures were elicited in a narration paradigm. Video recordings of the participants served as basis for the gesture transcription (for a detailed description of the method see below). Participants were two patients with left hemisphere lesions, one with mild (LBD1) and one with severe aphasia (LBD2), one patient suffering from right hemisphere brain damage (RBD), and two healthy persons. One of the healthy persons (KON1) retold the stories verbally and the other healthy subject (KON2) retold the stories without speaking only by gesturing. Twenty-five gestures of each person were transcribed. Hence, interrater-reliability was established on the basis of a total of 125 gestures which were coded by two independent raters.

In a first step the onsets of the strokes as identified by the two raters were compared. Then the HamNoSys transcription as described above was conducted on the basis of the onset coding of rater 1. In the following sections, the obtained results for the interrater-reliability will be described.

Onset of the stroke

As the configuration of the hand at the beginning of the stroke is the basis for further transcription, in a first analysis the raters identified the onset time of the stroke. In six cases of the 125 coded gestures the raters differed with respect to the question of whether a movement had to be considered as a gesture or not. For the movements which were identified as gestures by both raters, the coded onset times were compared. In 90.4% of the judgements, the raters differed in no more than four frames, and in 33% they selected exactly the same frame. Statistical analysis showed a significant correlation of the coded onsets in frames between the two raters (Pearson, $r = 0.795$, $p < .001$).

Handedness

For 87.2% (109) of the gestures both raters agreed on the handedness.

Transcription of single gestures

There was a total agreement of the HamNoSys transcription of the single gestures in all six feature categories in 35.2% (44) of the gestures. In a further 33.6% (42) the raters agreed on five of the six features categories.

Different feature categories

We analyzed the agreement for each of the feature categories. In Table 1 the total number and the percentage of equally transcribed symbols and symbol combinations are listed. The highest disagreement was found in the notation of the *extended finger orientation*, whereas the highest agreement was reached in the category *repetition*.

Hamming distances

We calculated the hamming distances (grand averages) for the five subjects on the basis of 25 gestures per person. The range of hamming distances obtained from the transcription of both raters was very similar (4.22 to 4.79 versus 4.18 to 4.97; see Figure 9), and the rank correlation between them was perfect (Spearman, r = 1, $p < .01$).

Table 1. Interrater-agreement in the six analyzed feature categories over a total of 125 gestures (total number and percentage of equally transcribed gestures)

	Hand shape	Extended finger orientation	Palm orientation	Location	Movement	Repetition
Number	102	85	106	102	89	116
Percentage	81.6%	68%	84.8%	81.6%	71.2%	92.8%

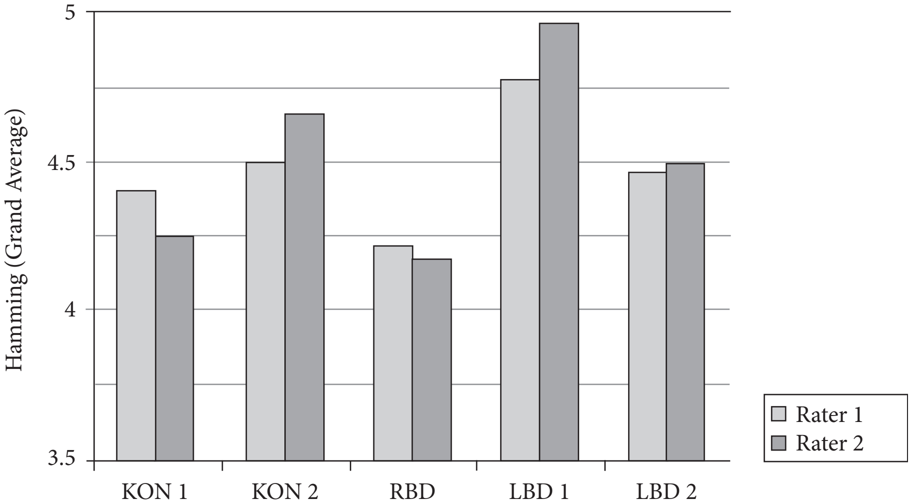


Figure 9. Hamming distances for five subjects calculated on the basis of 25 gestures for rater 1 and rater 2.

Application of the method in a pilot study

We tested the validity of our method in a pilot study where gestures of healthy individuals were recorded and transcribed in two conditions which were expected to cause significant differences with respect to the hamming distances.

Subjects were asked to retell short video clips verbally and without speaking. In the verbal condition gestures were not necessary for conveying the content of the stories. Hence, large inter-individual differences of the diversity of hand gestures were expected. We expected less inter-individual differences in the nonverbal condition where all subjects were forced to use gestures for conveying the content of the stories. Furthermore, we expected overall higher hamming distances in the nonverbal condition because gestures take over the sole communication of the message (Goldin-Meadow et al. 1996).

Subjects

Twelve healthy subjects, eight women and four men, participated in this study. All participants were native speakers of German. The age range was between 23 and 58 with a mean of 41 years.

Material

Stimulus material consisted of ten short video clips. Four clips were part of a *Mr Bean* story, and six clips belonged to two cartoon stories of the *Sylvester and Tweety* series. The duration of the clips varied between 30 and 90 seconds.

Procedure

The video clips were presented on a laptop computer. Immediately after each clip the subject was asked to recount the story from memory. In the verbal condition subjects were asked to retell the story in a vivid and descriptive manner. In the nonverbal condition subjects were required to depict the content of the stories without speaking, only by using their hands. All narrations were videotaped from a frontal position. The first clip served as a warm-up film, and the experimenter gave feedback and asked questions for animating the subjects to retell the story in a more vivid way if necessary. Throughout the narrations of the other nine clips, the experimenter solely made a confirmative utterance like “yes” and “okay”. The experimenter sat opposite to the subject and avoided producing hand gestures.

Data analysis

Sixty-three gestures were transcribed per subject and condition. We calculated the grand average of the hamming distance for each subject in the verbal and the nonverbal condition and compared them.

Results

The grand average of the hamming distances displayed individual differences within the groups as well as a clear difference between the two conditions. Figure 10 shows the obtained values in both conditions. In the verbal condition the grand averages of the hamming distances varied from 1.62 to 4.55 with a mean value of 3.74 and a standard deviation of 0.8. In the nonverbal condition the range was much smaller. It ranged only from 4.22 to 4.95 with a mean value of 4.63 and a standard deviation of 0.25. In all subjects the diversity of gestures was higher in the nonverbal than in the verbal condition. The increase of diversity in the nonverbal condition was statistically significant (Paired Samples T-Test: $t = -3.8$; $p < .005$).

Discussion

The verbal condition yielded the broadest range as well as the highest variance of average hamming distances. In this condition, we also found the lowest value of 1.62. Healthy subjects do not have to rely on nonverbal means of communication to convey the content of a story. Hence, in this condition, inter-individual differences appear with some persons producing a lot of different gestures whereas other speakers produce

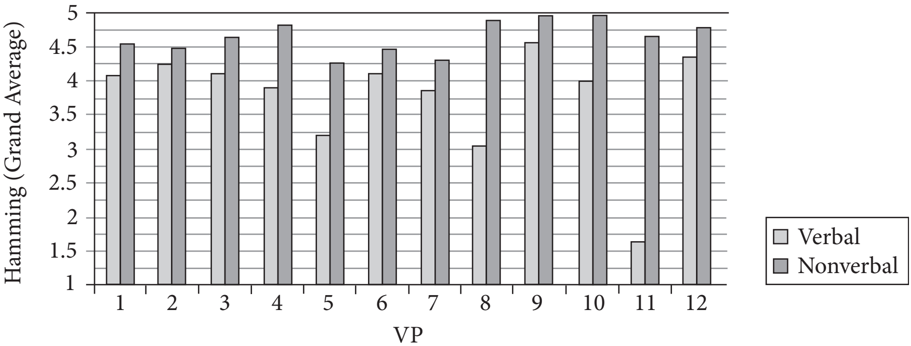


Figure 10. Hamming distances for twelve healthy subjects in a verbal and a nonverbal condition.

a lot of similar gestures. We assume that speakers who produce more lexical or meaning-laden gestures along with speech reach a higher formal diversity than speakers who produce mainly beat gestures which are always quite similar in form. For instance, the participant with the lowest value (participant 11) produced nearly exclusively uniform baton gestures.

All participants obtained higher average hamming distances in the nonverbal condition. The above mentioned person who obtained the lowest value of 1.62 in the verbal condition achieved in the nonverbal condition a hamming distance (grand average) of 4.64, which was slightly above the group mean for this condition. This result is in line with our expectations: Gestures become more diverse when the transmission of the content relies completely on them. Furthermore, individual differences in the nonverbal condition were less pronounced than in the verbal condition. This result is consistent with the expectations, too. The inter-individual differences in gesture production that cropped up in the verbal condition were masked in the nonverbal condition. Here, the participants were explicitly asked to retell the stories in the manual modality. Individual differences decreased because all participants were likewise dependent on the use of hand gestures for conveying the relevant aspects of the stories.

We presume that the hamming distance reflects the degree to which a person encodes meaning aspects in gesture. The results of this pilot study lead to the assumption that the hamming distance is an adequate measure of formal diversity which can be seen as an indicator for the potential informational content of hand gestures.

Conclusion

The aim of this project was to develop a method for the transcription of gestures that does not rely on the analysis of the concomitant verbal utterance and offers the possibility to conduct quantitative analyses of gestures. The described method can be used for the analysis of gestures that are produced along with speech as well as for the analysis of speech-replacing gestures. The reduced symbol set of HamNoSys offers a transcription analogous to the phonetic alphabet capturing the physical features of the gestures. Finally, the diversity and hence the potential information content of gestures can be evaluated by the calculation of the hamming distance. The method is especially suitable for the data of patients with severe language disorders. However, it can be used also for the data of healthy speakers. More recent approaches in gesture research make attempts to characterise gestural forms. In some studies, the determination of physiological and kinetic aspects of the appearance of gestures constitutes the basis for further analyses that address function and meaning of gestures (e.g. Müller 2004, Lausberg & Sloetjes 2009). We claim that a form-based evaluation of gestures should precede analyses that address function and meaning of gestures. These different levels of gesture analysis can reveal different aspects of the mechanisms that underlie gesture production (for a study with aphasic speakers see Hogrefe under review).

References

Freedman, N. 1972. "The Analysis of Movement Behavior during the Clinical Interview". In *Studies in dyadic communication,* A. E. Siegman and B. Pope (eds), 153–175, New York: Pergamon Press.

Goldin-Meadow, S. McNeill, D. and Singleton, J. 1996. "Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality." *Psychological Review*, 103: 34–55.

Hogrefe K. 2009. *Aphasie, Apraxie und Gestik: Zur Produktion von Handgesten bei Patienten mit linkshemisphärischer Hirnschädigung*. Unpublished doctoral thesis. Universität Potsdam.

Hogrefe, K. Ziegler, W., Weidinger, N. and Goldenberg, G. Under review. Non-verbal communication in severe aphasia: Influence of aphasia, apraxia, or semantic processing.[2] Cortex.

Jones, G. A. and Jones, J. M. 2000. *Information and Coding Theory*. London: Springer.

Kita, S., van Gijn, I. and van der Hulst, H. 1998. Movement phases in signs and co-speech gestures, and their transcription by human coders. In *Gesture and Sign Language in Human-Computer Interaction*, Ipke Wachsmuth and Martin Fröhlich (eds), 23–35, Berlin: Springer.

Kögl, O. 2006. *Zum Vergleich sprachbegleitender und sprachersetzender Gesten – Eine explorative Erhebung*. Unpublished master's thesis. Ludwig-Maximilians-Universität München.

Lausberg, H. and Sloetjes, H. 2009. "Coding gestural behavior with the NEUROGES-ELAN system." *Behavior Research Methods*, 41 (3): 841–849.

McNeill, D. 1992. *Hand and Mind. What Gestures Reveal about Thought*. Chicago, London: The University of Chicago Press.

Müller, C. 2004. "Forms and uses of the Palm Up Open Hand: A case of gesture family?" In *The Semantics and Pragmatics of Everyday Gestures*, C. Müller and R. Posner (eds), 233–356, Berlin: Weidler Verlag.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. 1989. *Hamburger Notationssystem für Gebärdensprachen. Eine Einführung*. Hamburg: Signum Verlag.

Seyfeddinipur, M. 2006. *Disfluency: Interrupting Speech and Gesture*. Nijmegen, NL: MPI Series in Psycholinguistics.

Stokoe, W. C. 1960. "Sign Language Structure: An Outline of Visual Communication System of the American Deaf." In *Studies in Linguistics Occasional Papers No.8*. Buffalo, NY: University of Buffalo.

Zierdt, A. Hogrefe, K., Ziegler and W. Goldenberg, G. 2006. *HamNoChart*. Unpublished input program for the computer based gesture transcription with HamNoSys. München: Clinical Neuropsychology Research Group (EKN), Municipal Clinic München GmbH.

This study was founded by the German Research Foundation (DFG, project GO 968).

CHAPTER 7

'Parallel gesturing' in adult-child conversations

Maria Graziano[1], Adam Kendon[2] and Carla Cristilli[3]
SESA-Università degli Studi Suor Orsola Benincasa, Université Stendhal-Grenoble 3,[1] University of Pennsylvania,[2] and Università degli Studi di Napoli "L'Orientale",[3]

Sometimes a speaker repeats an interlocutor's gesture, at least partially. Such 'parallel gesturing' illustrates how gestures can enter into the conversational exchange along with speech. Here we describe examples observed in adult-child conversations (children between 3 and 9 years). Four contexts are noted: (1) adult or child repeats speech and gesture of the other's utterance in displaying understanding; (2) the adult repeats the child's gesture, often with modification, when offering the child a more complete or correct expression of what he or she just said; (3) the adult repeats the child's gesture when matching the child's expressive style; (4) either adult or child parallels the other's gesture when expressions of similar discourse type are reciprocated. Children, like adults, can pay attention to each other's gestures, as well as to words. Differences between adult and child in how a 'paralleled' gesture is performed shows that gestural performance, like speech, involves maturation.

Keywords: gesture, imitation, children, conversation

Introduction

Sometimes a next speaker repeats, completely or partially, a gesture made by the immediately preceding speaker. This phenomenon, here termed 'parallel gesturing', described by de Fornel (1992) as 'return gesture' and by Kimbara (2006) as 'gesture mimicry', is interesting, as these authors show, because it demonstrates how gesture can be relevant for the interaction process. It shows that the preceding speaker's gesture contributed to the next speaker's understanding of what was said, and the act of paralleling the previous speaker's gesture can be a way in which the current speaker displays both cognitive and affective commonality with the other. Tabensky (2001) studied next speaker re-phrasings of previous speaker's utterances and observed gesture re-phrasing as a part of this. As she argued, this shows that speakers respond to each other's gesture-speech ensembles as integrated units.

The studies mentioned deal with adults, but parallel gesturing in conversations between adults and young children is also reported. One of us (Cristilli & D'Agostino 2005, Cristilli 2007) analyzed interactions between teachers and children, aged between two and a half and six years. The aim was to examine how the teacher or the child used gesture in relation to different kinds of didactic interventions made by the teacher. For example, a child was asked to re-tell an episode in a story or was asked to name or explain something, as when looking at pictures in a storybook. The teacher intervened from time to time to help the child, and the use of gesture in these cases was examined. The teacher, in re-stating something the child had said as a way of confirming its correctness, also often repeated the child's gesture. Further, the teacher, in repeating with some modification what the child had said, as a way of expanding it into a more adequate expression, repeated also with some modification, the gesture the child had used. In these cases, the child then repeated the teacher's gesture, modifying their own previous one to be more like that of the teacher. Cases were also observed in which a child, following a teacher's telling of a story, when repeating part of it, revised their gestures so that these became more similar to those of the teacher. In such cases the child appears to use the teacher's gesture as a model for their own performance.

Such examples show that both teacher and child are paying close attention to each other's gestures as well as to each other's words. This means that it is the gesture-speech ensemble (Kendon 2004) that is the unit of expression that the child and teacher deal with. Gesture, in these cases, is not treated as an ignorable 'add on', but as integral to the expressive forms being developed and used.

Here we describe examples of 'parallel gesturing' in conversations between an adult and a young child. Although these conversations were not explicitly didactic, further instances of parallel gesturing following usages described by Cristilli were found. Here we emphasize the role of parallel gesturing in the interaction process, suggesting that it can serve as a way for the participants to display to one another that they share in common an expressive style, that they are 'on the same wavelength' together. Paralleling a gesture of one's conversational partner is part of the process of 'frame attunement' by which the participants come to sustain a common cognitive alignment to the current conversational focus, thus participating in the conversational 'working consensus' (Goffman 1961, 1974; Kendon 1985).

Parallel gesturing in child-adult conversations also allows us to compare child and adult gesture performance (see also Cristilli 2007). In our examples, the manner in which the adult performs a gesture when paralleling that of the child, tends to conform to the conventional form of the gesture, as used in the local culture (in this case, Neapolitan), while the child's version of the same gesture is more like an attempt to produce something closer to the object or action that forms the 'model' from which aspects of what is represented in the gesture is derived. That is, it is more pantomimic or more concrete.

In the following we describe six examples which are drawn from twenty nine instances noted in our recordings and are deemed to be representative of the general features of the parallel gesturing we have observed. These examples have been drawn from fourteen video recordings of conversations between an adult and child which were made as part of a study of narrative skills in young children between 3 and 9 years of age.

The child was asked to tell the story of an animated cartoon that the child and adult had previously viewed together. The conversations all took place in an environment highly familiar to the child, either at the child’s home or at school. The adult, in all cases, was someone the child knew well, either the teacher or someone who was a good friend of the child’s family. The recordings were made in or near Naples, Italy. All participants are native speakers from this area.

The animated cartoon used is from a television series known as “Pingu”, which is about a family of penguins. In the episode used here the family is getting ready for Christmas. They are making Christmas biscuits, decorating the Christmas tree, and wrapping and exchanging presents. In the conversations the adult often asks questions or makes suggestions, helping the child to recall the details of the story.

In presenting the examples, we give the original with an English translation in the line immediately above with the following transcription conventions: (.) indicates a short pause; _ indicates vocal prolongation; ë is schwa; apostrophe indicates truncation. Below the original a notation showing the phase structure of relevant gesture phrases is provided, showing how it aligns with speech. This notation is based on that used in Kendon (2004), which should be consulted for a full explanation. The *preparation* of the gesture phrase is marked as ^^^; the *stroke* is marked as ***; *post-stroke hold* is marked as <u>****</u>; *recovery (return to rest position)* is marked as ###.

We first present four examples (Examples 1 to 4) in which the next speaker (hereafter ‘Interlocutor’) repeats, partially or completely, the ensemble of gestural and spoken action of the previous speaker (hereafter ‘Speaker’). Here we see how parallel gesturing may contribute to the display of shared understanding. Differences between child and adult gesture performance can also be examined. Examples 5 and 6 are then described in which the Interlocutor’s gesture shares features with that of the Speaker, without being a complete or partial repetition of it. In these cases we see how the Speaker’s gesture may contribute to or shape the development of what the Interlocutor says next.

Example 1

In this example the child repeats the adult’s gesture and does so, it seems, both as a display of understanding and as a way of showing that other’s expressive style is shared.

D (5:6 years)

02.15

and how did they make it so you could not look?

M: e come fanno a non guardare?

they closed_ that thing (.) that thing that's next to the key?
D: hanno chiuso_ quella cosetta (.) quella cosetta che sta vicino alla chiave?

eh! the the the eyelet eh?
M: eh! la il l'occhiello eh?
|^^^***************##|

the eyelet
D: l'occhiello
|^*********************##|

the eyelet
M: l'occhiello

D describes how the mother penguin shuts her children in the house and locks the door, while she and father go outside to decorate the Christmas tree. Because she wants to surprise the children, the mother covers the keyhole of the door of the house with a snowball so that the children cannot peep out. In the present extract, M asks how the mother has prevented the children from looking outside. D replies "*hanno chiuso _ quella cosetta (.) quella cosetta che sta vicino alla chiave* – they closed _ that little thing (.) that little thing that's next to the key". The child does not know the expression "*buco della serratura* – keyhole" and uses instead "*cosetta* – little thing". He speaks with a rising intonation, as if asking M to give him the proper term. M responds to this and provides the term "*occhiello* – eyelet" (which, in fact, is not the correct term!).

As M says "*occhiello*" she lifts her left hand, with index finger and thumb extended so they are parallel to one another, as if to define a small space, bringing it to about eye level, thus presenting a small space to look through. D then repeats "*occhiello*" but, at the same time, does a gesture very similar to M's: he lifts his hand to his eye (his right hand), with index finger and thumb forming a small circle. He thus repeats M's entire gesture-speech ensemble. In doing this, he certainly displays his *understanding* of M's utterance, but in responding with gesture and word together, he also enters into the style of it: he shows he shares M's "expressive level".

We noted that D's gesture is "very similar" to M's gesture. How it differs, however, is instructive. M lifts her hand to the level of her eye, her paralleled thumb and index finger suggesting a small space. D, on the other hand, makes a circle with his thumb and index finger and brings it close to his eye, acting out more fully the idea of a keyhole actually being looked through. This sort of difference seems characteristic. That is, the difference between the child's gesture and that of the adult he parallels is that the child's gesture often seems more like an attempt to imitate the actual shape of something or an actual pattern of action, whereas the adult's gesture is more schematic. We shall see this difference again in our other examples.

Example 2

Mo (3:4 years)

00.14

when she had rolled out all the pastry (.) what does she take? what does she do?

M: quando ha steso tutta la pasta (.) cosa prende? cosa fa?

she puts them in an oven

Mo: li met dentro un forno

|^^^^***********************************##|

she puts them all into the oven (.) and then?

M: li mette tutti dentro al forno (.) e poi?

|^^^^^*********************###|

In Example 2, it is the adult who follows the child. The adult both modifies what the child says, slightly re-phrasing his spoken component so that it conforms to a more adult form of expression, but she also does the same for the child's gesture. With the help of M, Mo is explaining how the mother penguin prepared biscuits. M asks: "*Quando ha steso tutto la pasta, cosa prende, cosa fa?* – When she has rolled out all the pastry, what does she take, what does she do?" Mo replies "*li met' dentro un forno* – she puts them in an oven." As he says this, he extends both his arms forward horizontally, his hands spread open, the palm of his left hand partially resting on the palm of his right hand. This looks very much like a representation of putting something forward into something. It seems semantically coherent with his verbal expression. M then repeats Mo's words, modifying them somewhat. At the same time she performs a gesture similar to that of Mo, but in her version the hands, with palms facing downwards and fingers spread, are held in parallel, not in contact as they are moved forward (see Figure 1).

Here, both the verbal and gestural expressions of the child are paralleled in the adult's next turn, but in a manner which is closer to a "standard" form. In speech, M corrects and expands slightly what the child said, pronouncing the verb "*mette* – puts" correctly (Mo said "*met*"), she adds the pronoun "*tutti* – all" and changes to the definite form the article that Mo had used before the word "*forno* – oven", here combining it with a preposition: "*al forno*", literally "to the oven". Mo had said "*un forno* – an oven". By saying "*al forno*" M refers to the specific oven which is in the penguins' house and which can be seen in the cartoon, rather to any oven, as the child's expression might suggest. M's re-formulation of Mo's verbal expression is thus a re-formulation in the direction of a more correct form.

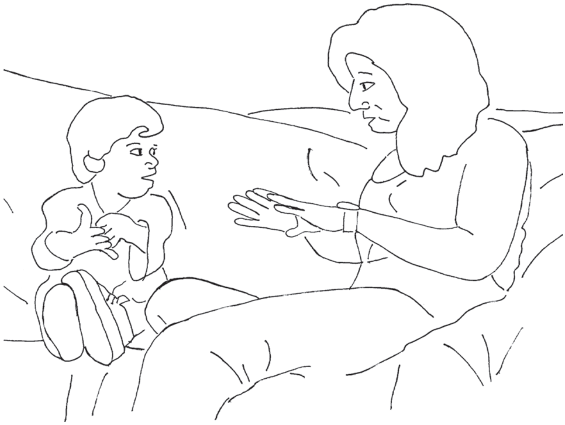


Figure 1. Example 2. Child and adult gesture as they refer to putting biscuits into the oven.

As for the gesture, both that of Mo and that of M represent the idea of putting something into something. As described, M's gesture is realized with two open hands with palms facing down, held side by side, a standardized "putting in" gesture (in another recording, M, using the same words as used here, performs the same gesture). Thus, just as in her words, M re-does Mo's verbal expression to be closer to an adult expression, she does the same for his gesture.

In these two examples the parallel gesturing happens when the Interlocutor confirms what the Speaker has said, repeating both verbal and gestural components of the utterance. By doing this, understanding of the other is displayed, but there is also a display of understanding of the other's way of expressing what is said. Also, in both cases note how it is the *gesture-speech ensemble* that is reproduced as a unit, not just one or other component separately.

Example 3
In Example 3 the adult repeats a gesture produced by the child, also repeating exactly the child's words, in a context in which this serves both to confirm what the child has said and also to show that the adult is entering *with* the child in the same "expressive level", perhaps in this way encouraging the child to go on with her story telling.

F (5 years)
00.00

the mother the mother of_ ë__ the mother was preparing the biscuits
F: la mamma la mamma__di_ ë__ la mamma preparava i biscotti

M: mh!

she cooked them she put them in the oven and she burned herself
F: le cucinava le metteva nel forno e si è scottata
|^^^^******##|

she burned herself when when she opened the oven 'cos the biscuits were ready
M: si è scottata quando quando ha aperto il forno ché i biscotti erano pronti
|^*********##|

This example comes at the beginning of F's account of the 'Pingu' movie. When she says that the mother penguin has burned herself (because she tried to take the biscuits out of the oven without gloves), she makes a gesture in which the hand, posed with fingers spread, is raised towards the side of her face and is moved up and down rapidly. This reproduces the movement of the mother penguin in the film after she tried to take the biscuits out of the oven without putting on her oven gloves. She lifts her flipper near her beak and blows on it as she shakes it up and down. The gesture performed by F, together with the word she uttered while doing so, is immediately repeated by M. In this case we have an example of true parallel gesturing: the form of the hand, the place of execution and the pattern of movement are substantially the same in both participants, more so than is the case in Examples 1 and 2.

The adult follows her repetition of F's gesture-speech ensemble by elaborating the circumstance in which this burning of the flipper occurred, filling in a detail for the child and in this way, perhaps, leading her discourse forward. Here she seems to be following a strategy common in didactic situations, in which the teacher expands a child's utterance (see Cristilli 2007). This is also a common technique in conversations among adults when collaborating in topic development. However, we may note that the adult's repetition of the child's gesture here was not done, it seems, as part of a strategy to display understanding or agreement. Rather, it seems to be an example of the adult entering the child's level of expression as a way of creating solidarity, or *rapport*.

In Examples 1–3 we have examples in which the gestural repetition, whether by child or adult, is combined with a repetition of the concurrent words. That is, the Interlocutor repeats the whole utterance ensemble, treated as a unit, rather than just picking up on one or other component of it. Tabensky (2001: 232–233), referring to her observations, remarked that there is no repetition of gesture when there is exact repetition of the associated words (or a repetition with slight modifications). She suggests that this is because gesture is usually involved in the production of one's own meaning. If one merely repeats another's words, gesture is unlikely to be used. However, in our examples there are large differences between the participants in expressive skill and in ability to maintain sustained attention. In conversations like this, extra efforts must be made, especially by the adult, to establish and maintain with the child a shared perspective on the conversational focus. Parallel gesturings, in the examples described, appear to be done not just to display *cognitive* understanding, but also to show a sharing of expressive style.

Symmetry between interactional partners in gestural and other kinds of bodily actions, as well as matching verbal expressions, has often been noted at conversation beginnings (for example in greetings), when participants must find ways to bring into alignment each other's attention so that a 'working consensus' can be established (Goffman 1961, 1963), a process also termed 'frame attunement' by Kendon (1985). We suggest that the 'full' parallel gesturing (indeed, complete utterance paralleling) seen in these conversations is the result of the more explicit kinds of 'frame attunement work' that is needed for a successful adult-child conversation of the type examined here.

Example 4

Example 4 is another example in which the child parallels the adult's gesture, but here the adult's gesture is produced with a phrase that is verbally incomplete, although complete in meaning when the gesture is included. Here the adult does what teachers often do with small children when they speak a sentence but leave the last word of the sentence unpronounced, so that the child might say the right word and fill in the slot (see Cristilli 2007). In this way the child's understanding is confirmed, and also it helps the child to use a word in an appropriate place in a sentence.

C (3:3 years)

01.00

and because they want?

M: e perché li vogliono?

|^^^^^^^*********************#|

they want to eat them

C: li vogliono mangiare

|^^^^^******************************|

they want to eat them

M: li vogliono mangiare

As M says "*e perchè li vogliono?* – and because they want to ...?" she lifts her hand with the fingers extended but drawn together so that their tips are in contact, and orients with the 'bunched' finger tips toward her mouth. This gesture is well known in Naples and environs and is glossed with "*mangiare* – eat". It has been in use for a very long time (see p. 261and Plate VII in de Jorio 2000 [1832]). The form of M's speech, especially her intonation, shows that a word is missing at the end of the sentence which the child should supply. M's gesture gives what is missing. C responds accordingly, saying: "*li vogliono mangiare* – they want to eat them", repeating, thus, the last part of M's phrase, but now completing it with the expression "*mangiare* – to eat". The child has 'read' correctly the adult's gesture. Even at the age of three this child has thus completely understood this widely used gesture and has understood it to have a lexical equivalent. The child has also understood that M's gesture is used here to supply the verbal item that M, in her speech, had shown was missing.

However, as C says "*li vogliono mangiare*" she lifts her hand toward her mouth, fingers drawn together, thus performing a gesture that, though similar to that of M, differs from it, and does so in just the way we expect, as already suggested, according to observations given above. C's performance is closer to a pantomime of putting something small into the mouth. This is in contrast to the form of the gesture performed by M where the hand is shaped according to a conventional form (the form of the hand M uses is the same as that shown in Plate VII of de Jorio from 1832). As in the gesture described from the nineteenth century, the hand is directed toward the mouth, but does not touch it. Although the *base* of this gesture is surely the act of putting a morsel of food in the mouth, in its *realization* it is highly schematized. Its *referent* is not "putting a morsel of food in the mouth" but the much more abstract notion of "eating". C, it would seem, has not acquired the standardized adult form of this gesture, using instead something closer to a pantomime of putting something in the mouth.

We now present two final examples that illustrate partial paralleling. In Example 5 we see gesture paralleling insofar as the Speaker's gesture serves to establish for the Interlocutor a certain mode of gesture use within the current discourse context, here using the hands as counting devices. In Example 6, in a sort of 'gesture re-phrasing' (cf. Tabnesky 2001), the Speaker's gesture appears to suggest to the Interlocutor features of the topic under consideration not heretofore referred to. The Interlocutor takes these up, thus shifting the focus of the conversation slightly.

Example 5

A (4:1 years)

00.00.35

and and and_ a fam' how many people how many penguins were there?
I: e e e__ una fami' quante perso' quanti pinguini c'erano?
|**********************######|

two
A: due
|***********|

only two?
I: solo due?
|*************|

three (.) wait
A: tre (.) appetta
|**| |^**********************#^|

so how many of them were there?
I: quindi quanti ne erano?

one two three and four

A: uno due tre e quattro
|^**|***|****|******##|

four
I: quattro

In this Example, A says she does not remember the story well. To help her, the teacher asks how many penguins there were, lifting her hand to display four fingers in a well-known gestural expression for "four" – the number of characters in the story. The child replies with "*due* – two" and holds up her hand displaying just two fingers. She has adopted the expressive method of the teacher, but has not paralleled the teacher's actual gesture. The teacher then says "*solo due?* – only two?" but again displays "four" (although this time the hand is not raised so high). A now responds with "*tre (.) appetta* – three (.) wait" as she says "*tre* – three" she holds up her hand, this time with three fingers extended, then, after she has said "*appetta* – wait" she again holds up her hand, now with four fingers extended. At this point she now *shows* four – and this, apparently, is her answer. The teacher, still wanting a spoken reply, continues with "*quindi quanti ne erano?* – so how many of them were there?" A responds by grasping each of her four extended fingers in turn by the other hand, folding each digit down, saying, as she does so: "*uno, due, tre, quattro* – one, two, three, four". The teacher, confirming this, says "*quattro* – four".

Here the teacher, by showing the gesture "four" with her first question, offers the use of fingers as a way to display numbers in this context. The child adopts this use, but does not just imitate the teacher's gesture, since she derives her answer in her own manner, although using 'number display' gestures. Parallel gesturing is manifested through the taking up of a certain way of using gesture that the Speaker had initiated. This mode of using gesture is maintained throughout the sequence in which the issue of the number of penguins is being discussed.

Example 6

E (4:1 years)
00.00.26

and _ what was the mother preparing?
I: e_ che cosa sta preparando la mamma?

the stars the biscuits
E: le 'telline i biscotti

these biscuits are? what are they like?
I: questi biscotti sono? come sono?

with half moons and with stars
E: con mezze lune e con le 'telline
|^^^^^^^**/*/**/**##|

what has she used to make these_ (.) these biscuits?
I: cos'ha usato per fare questi_ (.) questi biscotti?
|^^^^^^^^**/******/*************|

the dough
E: l'impasto

In Example 6, E had begun with a much summarized version of the penguin story, so the teacher asks questions to help him develop a more detailed account. As the extract begins she asks the child what the mother was making. He says she was making biscuits. The teacher asks what these were like – "*questi biscotti sono? come sono?* – these biscuits are? How are [they]? [i.e. what are they like?]". He says they are like half moons and stars: "*con mezze lune e con le 'telline* – with half moons and with stars". As he says "*con le 'telline* – with stars", he lifts both hands up to be level with the top of his head, holding them so that the thumb and index finger of each hand touch those of the other, both hands thus making the shape of a circle. With the hands held like this, palms facing downwards, he moves both towards his right, making a succession of lowering movements as he does so. The gesture thus suggests several round shapes distributed in space. These movements, however, may also refer to using a pastry cutter to produce the biscuits. Now the teacher asks "*cos'ha usato per fare questi – (.) questi biscotti?* – What has she used to make these – (.) these biscuits?" As she asks this, the teacher lifts her right hand, posed with the fingers spread but flexed (a "claw" shape) and makes a succession of lowering movements moving her hand rapidly from one position to another as she does so. The teacher's gesture thus shares features with E's gesture, although it is performed in 'normal' space, not at head level, and with one hand, not two, and the handshape is, as noted, like a 'claw'. This handshape could be interpreted to refer to the biscuit mould (*formina)* the mother used in cutting the biscuits.

Note that the teacher, in her second question, displaces the focus of interest from the form of the biscuits, which E referred to, to the instrument used to make them. This displacement of focus may have come about because the teacher picked up on an aspect of E's gesture (the successive downward movements that he makes) that can be related to the actions done with the biscuit cutter when making biscuits. In E's gesture the form of the biscuits and the multiplicity of them seem prominent. However, these movements may also relate to the actions of cutting biscuits (which E does not refer to verbally) and they may have prompted the teacher to pick up on this theme which, accordingly, becomes the motif of her gesture and to which she refers as she asks "*What has she used to make these biscuits?*"

As Kimbara (2006) has said, a gesture is selective in the features it refers to and, as a result, it highlights those features it selects and not others. Interlocutors who parallel a Speaker's gesture, thus, select the features highlighted in the paralleled gesture in their own gesturing, and this can influence their conception of what is being referred to. In the present case it is possible that, in picking up on the movement feature of E's gesture, the teacher's attention was directed toward the action of cutting biscuits, leading her to adopt the modifications she shows in her expression.

Conclusions

Parallel gesturing shows that the Interlocutor takes into consideration both components of the Speaker's utterance. This means that, in such cases, the gesture-speech ensemble is treated as a single unit of production.

The repetition of the gesture-speech ensemble can be a way in which the Interlocutor shows understanding of the Speaker's utterance (Examples 1and 2.) However, paralleling the Speaker's gesture as well as speech, may be a way for the Interlocutor to display alignment to the other's expressive style as well (Example 3). By paralleling the Speaker's gesture, the Interlocutor enters into the same expressive style as the Speaker. Paralleling of this type may be done in interactions where extra work is needed to build or maintain rapport. It serves to facilitate the process by which participants sustain a common alignment to the conversational focus (Goffman 1961, 1974; Kendon 1985). It may often occur in adult-child conversations, where the child is not yet fully ready to sustain a cooperative focus. This prompts the adult to help the child to do so.

Besides the "full gesture paralleling" of Examples 1, 2, 3 and 4, there were two examples of 're-phrasing' (following Tabensky 2001). In Example 5, the adult introduced a particular mode of using gesture (as an enumeration device), which was then adopted by the child, although used in her own way in developing her response. In Example 6, the adult, in her gesture, took up a feature of the child's gesture but performed her gesture so that together with the associated speech she shifted the conversational focus. It was as if something in the child's gesture suggested a new direction for the development of the topic which the adult brought out in her own gesture. Gestural expression can thus enter into the process by which the topical focus of a conversation evolves.

In "full" gesture paralleling (Examples 1, 2, 3 and 4) we have noted some differences in how the child performs the gesture when compared to the adult. In these cases, the adult's performance is closer to a socially shared style than that of the child. In Examples 2 and 4, for instance, the adult's gesture is conventional. The child's gesture, in contrast, has more of the character of a pantomime of the action referred to. There are other differences, too. In Example 6 the child's gesture is performed at the level of his head, a spatial zone used much less often in adult gesturing, at least in conversations with only a few participants.

These kinds of differences are interesting for they suggest some of the features of gesture performance that children must acquire if they are to fit their gesturing within the style of the adult community into which they will grow. Just as children must learn to pronounce words so that they are no longer deemed 'childish', this is also true of gesture performance. Some of the examples described here (and also in Cristilli 2007) suggest that children attend closely to the gestures of adults and that they model their gestural performance after what they see among adults.

References

Cristilli, C. and D’Agostino, A. 2005. “A proposito di rapporti fra codici e di strategie universali della comunicazione: linguaggio verbale e linguaggio gestuale nell’interazione adulto-bambino.” *Atti del IV Congresso della Associazione Italiana di Linguistica Applicata (AltLA), Modena 19–20 febbraio 2004* Perugia: Guerra Edizioni, 114–43.

Cristilli, C. 2007. “La rilevanza dell’input gestuale in relazione a quello verbale nella comunicazione insegnanti-bambini.” In *Interazioni comunicative nello sviluppo tipico e atipico: implicazioni per la ricerca e il trattamento.* Special Issue of *Psicologia clinica dello sviluppo,* M. C. Caselli and E. Longobardi (eds), 429–455.

de Fornel, M. 1992. “The return gesture: Some remarks on context, inference and iconic gesture.” In *Contextualization of Language*, P. Auer and A. di Luzio, (eds), 159–176. Amsterdam and Philadelphia: John Benjamins.

De Jorio, A. 2000. *Gesture in Naples and Gesture in Classical Antiquity. A translation of* La mimica degli antichi investigata nel gestire napoletano (1832)*, and with an Introduction and Notes, by Adam Kendon*. Bloomington and Indianapolis: Indiana University Press.

Goffman, E. 1961. *Encounters*. Indiannapolis: Bobbs-Merrill.

Goffman, E. 1963. *Behavior in Public Places*. New York: The Free Press of Glencoe.

Goffman, E. 1974. *Frame Analysis*. Cambridge, MA: Harvard University Press.

Kendon, A. 1985. “Behavioral foundations for the process of frame attunement in face-to-face interaction.” In *Discovery Strategies in the Psychology of Action*, G. P. Ginsburg, M. Brenner and M. von Cranach (eds), 229–253s. London: Academic Press.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kimbara, I. 2006. “On gestural mimicry.” *Gesture* 6: 19–61.

Tabensky, A. 2001. “Gesture and speech re-phrasings in conversation.” *Gesture* 1: 213–235.

PART II

First language development and gesture

CHAPTER 8

Sentences and conversations before speech?

Gestures of preverbal children reveal cognitive and social skills that do not wait for words

Claire D. Vallotton
Michigan State University

The author wishes to acknowledge the children, parents, student caregivers, and staff of the Center for Child and Family Studies at the University of California, Davis for their time and patience with this study. Thanks also to the dedicated and talented research assistants who collected and coded these data. Preparation of this manuscript was supported in part by the National Institute of Child Health and Human Development, grant number 1 F32 HD050040-01.

Before first words, children use gestures to communicate and represent concepts. This study investigated two questions: Can infants pair gestures together to create two-gesture *sentences*? Further, can preverbal children engage in conceptually focused gesturing *conversations*? I observed 10 infants for 8 months during interactions with caregivers and coded all gesturing behavior. I used longitudinal growth modeling to analyze the developmental trajectories of gesturing sentence and conversation length. Infants formed 2-gesture sentences as early as 9 months and 3-gesture sentences at 1 year. Infants engaged in 4-turn conversations as early as 11 months; maximum gesture conversation length was 16 turns. Infants' early gesturing frequency and variety predicted later sentence length; however, caregivers' gesturing sentence length suppressed child's sentence length.

Keywords/phrases: child development, gesture, symbolic gesture, communication, representation, infant sign

Gesture as a window into preverbal cognitive and social skills

Sabrina (11.67 months) and her caregiver sat in the infant classroom of the UC Davis child development laboratory, where the university students who care for the children are taught to use a variety of gestures with the children. Sabrina crawled to the family picture board and pointed to the picture of her family. She pulled it off the board. Her caregiver *said, "You found a picture of your family." As Sabrina pointed to each of her family members, her caregiver said, "I see you are thinking about .... (named family member)." Then Sabrina pointed to the picture of another child's family. Her caregiver pulled that photo down, and again the caregiver talked about all family members, following Sabrina's pointing. Sabrina continued to pull down and point at every family picture; the caregiver talked about each until all the pictures were on the floor, then said, "There are no more pictures." Sabrina picked up the picture of her own family, and smiled.*[1]

In this observation, the preverbal child successfully engaged her caregiver in a kind of dialogue about her family and those of her peers. Sabrina demonstrates skilled and intentional communication using the flexible *point* gesture to engage and draw language from her caregiver. How will this type of interaction change as Sabrina gains a diversity of more referent-specific gestures? Will she string gestures together to make gesturing *sentences* representing more complex ideas? Will she engage in conceptually focused gestural turn-taking, or *conversations,* with her caregiver?

Before speaking their first words, children develop many communication and representation skills seen in their use of gestures. From a child development perspective, gestures reveal cognitive and social capacities in preverbal children that scientists and caregivers would miss if they waited for children to speak. This study investigates two such capacities: the cognitive capacity to string symbols together to represent more complex concepts and the social capacity to engage in meaningful and mutual dialogue. Both capacities are apparent shortly after children begin to use words. However, I contend that they are present earlier in development and revealed through children's gesturing behavior in gesture-rich environments.

Development of combining symbolic representations

The ability to represent a concept using a symbol is critical not only for language but also for cognition in general. In early childhood, representations can be seen in symbolic play, gestures, and eventually words. These representations become more complex as they are combined and elaborated into symbolic play scenarios and increasingly longer sentences. Symbolic gestures are those used to represent a referent in its absence. They are built out of actions that are either performed on the referent (e.g. throwing motion represents *ball*), by the referent (e.g. flapping arms represents *bird*), or in routines related to the referent (e.g. hands creating circle overhead represents *sun,* learned in song routine). Gestures learned in particular contexts are slowly de-contextualized to represent a concept in its absence (Bates et al. 1980, Werner & Kaplan 1963).

1. Observed by a student caregiver in the UC Davis laboratory school and recorded as an "anecdotal note," systematic participant observations used in training.

Typically developing children begin combining two words around 18 months when they have a vocabulary of 20 to 40 words; deaf children exposed to a signed language begin combining two signs around this same age and with the same vocabulary (20–40 signs) (Caselli 1983). Caselli never observed hearing children combining two gestures nor young deaf children combining two vocal words and concluded that the ability to combine symbols in the same modality depends on the modality of input (1983). However, just prior to producing two-word sentences, typical children will combine a single gesture, usually *point*, with a single word creating a two-concept cross-modal sentence (Goldin-Meadow & Butcher 2003).

Given that hearing children can combine a gesture with a word and that they are capable of learning many symbolic gestures prior to speech, will a child who regularly uses symbolic gestures combine them to form gestural sentences? Will they do this at an age before we expect them to combine words (18 months)?

Development of turn-taking in communication

The ability to engage in turn-taking with another person is a critical skill for successful communication. As early as two months old infants respond contingently to caregivers in face-to-face interactions (Murray & Trevarthen 1986). By 6 months infants intentionally communicate with adults, drawing adults' attention to themselves, and will persist in their attempts until they know they have the adult's attention (Wagner 2006). Infants' communicative skills grow as they incorporate more behaviors into their repertoire of communication tools, including a variety of gestures (Crais, Douglas, & Campbell 2004). By nine months infants interpret adults' gestures as intentional acts indicating the adult's focus of attention and use gaze-following, pointing, and imitation to join in the adults' attentional focus (Tomasello 1999). Around one year infants not only follow another's gaze and pointing but use pointing gestures to share both attention (Liszkowski et al. 2004) and information (Liszkowski, Carpenter, Striano, & Tomasello 2006). In the daily life of a one year old these pointing gestures, often accompanied by vocalizations, are clear attempts to communicate and usually set off an interactional sequence with the adult that may include sharing attentional foci, information, and meaning (Jones & Zimmerman 2003).

Rutter and Durkin (1987) documented the development of vocal turn-taking; they found that the number of turns babies took during interactions with mothers more than doubled between 12 and 24 months. However, they did not assess the number of turns focused on a given topic or during a distinct interchange one might call a *conversation*. Examining infants' use of eye contact to cue a change in turn and their interruptions of mothers' turns, the authors concluded that between 12 and 18 months the coordination of turn-taking relies upon the mother; after 18 months infants began to interrupt less and use gaze more regularly indicating that it is mother's turn (Rutter

& Durkin 1987). These findings seem to indicate that infants under 18 months may not be able to engage in reciprocal turn-taking in any modality, gestural or verbal.

Symbols expand the scope of conversational topics because they enable dialogue about things beyond the here and now. By two years old, children engage in coordinated *verbal* turn-taking with mothers, though mothers still produce more responses. Importantly, when mothers produce a greater number of responses, children produce fewer, perhaps as if they can't get a word in edgewise (Kaye & Charney 1981). The point gesture is an integral part of the development of communicative turn-taking *about* something, an object that is the focus of attention for child and adult. However, pointing is typically limited to communication about proximal objects. If preverbal children had a variety of symbolic gestures to initiate and sustain interactions about a variety of concepts, could they engage in conceptually focused symbolic turn-taking? That is, can preverbal children have *conversations* in the gesture modality in which the child and interaction partner take repeated conversational turns using gestures? Further, would a greater number of adult gestures result in fewer initiations or responses by the child, shortening the number of turns in a conversation?

Current study: Development of gestural sentences and conversations in preverbal children

Given that preverbal children are capable of using a variety of symbolic gestures prior to speech (Acredolo & Goodwyn 1988), I examined whether they could use these gestures in the cognitively and socially complex ways that they would use words in early language development. Specifically I asked:

1. Can infants combine gestures to create gestural *sentences*?
   a. At what age do infants use 2-gesture and 3-gesture combinations?
   b. Does adult modeling of gesture sentences promote infants' gesture sentences?
2. Can preverbal children engage in conceptually focused gestural *conversations*?
   a. At what age do children reply to adult gestures with their own gestures?
   b. When do infants engage in 4-turn gestural turn-taking?
   c. Does adult gesturing behavior support or suppress infants' gestural turn-taking?

Methods

Gesture-rich environment

I documented the development of gestures in 10 hearing infants who were in the infant classroom at the UC Davis Center for Child and Family Studies. In this classroom, adult caregivers modeled the use of specific gestures to represent salient concepts from the children's environment, for example, tapping fingers against mouth to represent *eat* or tracing index finger from eye down cheek to represent *sad*. Adults were explicitly taught to use a set of symbolic gestures and were instructed to pay attention and respond to infants' gestures. Infants were not explicitly taught to use gestures, but learned them from the adults. This gesture-rich environment provided a unique opportunity to investigate complex uses of gestures by preverbal children.

Participants

The 10 infants were between 4 and 12 months old at the start of the study and 12 to 20 months by the end of the 8 months of data collection. Adult participants were 24 university students studying child development and serving as the infants' caregivers as part of a required internship experience. Student caregivers spent two days each week in the class; there were typically 5 student caregivers and one head teacher in the classroom.

Data collection

Infants and caregivers were videotaped in spontaneous interactions during normal program routines. Each interaction was filmed for 5 minutes; infants were filmed an average of 40 times each over the 8 months. On average, infants were filmed a total of 200 minutes, or approximately 1% of their 360 hours in the classroom.

Coding and transcription

I used microanalytic coding – coding every relevant change in behavior through every second recorded – to capture all gestures by children and caregivers. For the purpose of coding, gestures were defined as intentional, communicative motor behaviors performed in the context of an interaction; markers of interaction context included body orientation or eye gaze towards an interaction partner. For each gesture recorded, coding captured which gesture was performed, who performed it, and when it occurred within the episode. Thus, it was possible to derive a sequence of gestures for one person or a sequence of gestures between two people. Gestures were subsequently coded as serving one of four conversational purposes: (1) Initiation: gesture not preceded by another gesture within 5 seconds[2] (e.g. Infant gestures *bird*); (2) Continuation: gesture

2. Five seconds was used as a conservative yet somewhat arbitrary marker of conversational timing. Through a review of gesturing episodes it was determined that if a child or caregiver were to respond to another's gesture, it would happen within 5 seconds, and those behaviors occurring after 5 seconds were not responses as indicated by changes in attention and gesture content.

preceded by a gesture by same individual within 5 seconds (e.g. Infant gestures *bird* then *points*); (3) Imitation: gesture preceded by same gesture by different person within 5 seconds (e.g. Infant *points* then caregiver *points*); (4) Reply: gesture preceded by different gesture by different person within 5 seconds (e.g. Caregiver *points* then infant gestures *bird*). I used these conversational context codes to determine whether infants or caregivers performed a gestural sentence and how long each sentence was (initiation followed by one or more continuations); and whether there was a gestural conversation (at least one reply or imitation after an initiation) and how many turns were in each conversation.

Coder training and reliability. Coders were naïve to the hypotheses of the current study. They were taught to recognize the gestures through written descriptions and visual demonstrations. Inter-coder reliability was assessed using Cohen's (1960) Kappa. Coders obtained a Kappa of .75 or above on five consecutive episodes before coding independently. Agreement was reassessed on 15% of episodes; Kappa was greater than .82 across all observations.

Variables

Time-invariant. For each infant, there is a variable describing the following:

- Age of entry into the classroom
- Early gesture frequency and variety (average per episode between 10 and 12 months)

Time-varying. For each interaction observed, there is a variable describing the following:

- Infant age
- Infants' and adults' average gestural sentence length
- Infants' and adults' longest sentence length
- Average number of turns per conversation (each turn within 5 seconds of previous)
- Longest conversation

Because so many of the observations included no gesturing by the children, the numerical data are erratic. To smooth the data for statistical modeling, I created running averages for each of the time-varying gesturing variables by averaging three observations together; for example, values in episodes A, B, and C were averaged to create observation 1; values in episodes B,C, and D were averaged to create observation 2; and so on. Further, I created *lagged* running averages for caregiver gesturing variables to capture infants' prior exposure to gestures from adult models. For example, the average caregiver gesturing frequency in observation 1 (average from episodes A, B, and C) was used to predict the level of gesturing in observation 4 (average of episodes D, E, and F).

Analysis

I used multi-level growth models (Singer & Willett 2003) with observations nested within children over time to describe the average developmental trajectories of the length of gestural sentences and conversations from 6 to 20 months of age and to test effects of both child's and caregiver's prior gesturing behavior on those trajectories. I used qualitative transcripts from observations to illustrate the content and context of gesturing interactions between preverbal children and caregivers.

Growth modeling allows me to describe the shape of development of gesture use over time. For both sentence length and conversation length I began with an unconditional baseline model, then added specifications of child age – first just linear, then linear and quadratic, then linear, quadratic, and quartic, etc – until I found the most parsimonious model that explained the most variance just by using child age. After establishing the shape of growth, I added variables for prior and current child and caregiver gesturing behavior, examining their main effects and testing their interactions with child age.

Results

Sentences

The quantitative coding and transcripts created from the videos revealed that infants do indeed combine different gestures to create gestural sentences. Infants begin to form 2-gesture sentences as early as 9 months, but do so more consistently around 11 months. Examples of 2-gesture sentences from the transcripts are the following:

Female, 11.8 months:
Time 00:02:01: *point* (index finger extended toward visual focus)
Time 00:02:03: *star* (fingers apart, extending then retracting repeatedly)

Female, 12 months:
Time 00:01:45: *snack/eat* (closed fingers of one hand tapping mouth)
Time 00:01:50: *more* (closed fingers of both hands tapping each other)

Infants begin to create 3-gesture sentences at around 1 year of age, but this stays a rare occurrence compared to 2-gesture sentences. Examples of 3-gesture sentences are the following:

Male, 11.7 months:
Time 00:02:14: *point*
Time 00:02:15: *wave* (fingers together, extending and closing toward palm repeatedly)
Time 00:02:16: *point*

Female, 12.6 months:

Time 00:00:08:	*point*
Time 00:00:10:	*bird* (both arms flapping)
Time 00:00:13:	*where* (palms turned up at shoulders)

Most gestural sentences included a *point*, and many of the 3-gesture sentences involved a sequence in which the first gesture was repeated after a point, such as *bird, point, bird*.

Figure 1 shows a scatterplot of the length of infants' gestural sentences by age. The overlaid trajectory shows the results of Model B in Table 1. This trajectory reveals a steady increase in gestural sentence length between 9 and 15 months, a flattening between 15 and 18 months, followed by another increase after 18 months. Infant age alone explained only 8% of the within- and between-child variance in sentence length, indicating that other child characteristics or experiences may be important predictors.

As seen in Model C, the younger infants were when they entered the classroom, the longer their gestural sentences were; further, infants' early gesturing frequency and variety, measured between 10 and 12 months, was positively associated with later gestural sentence length. These predictors explained 85% of between-child variance in sentence length.

Caregivers' use of gestural sentences had a negative impact on length of children's sentences (Model D). Controlling for caregivers' current gestural frequency and sentence length, caregivers' prior gestural sentence length was negatively related to children's sentence length. Together these predictors explained 8% of within-child variation and 45% of between-child variation in sentence length.

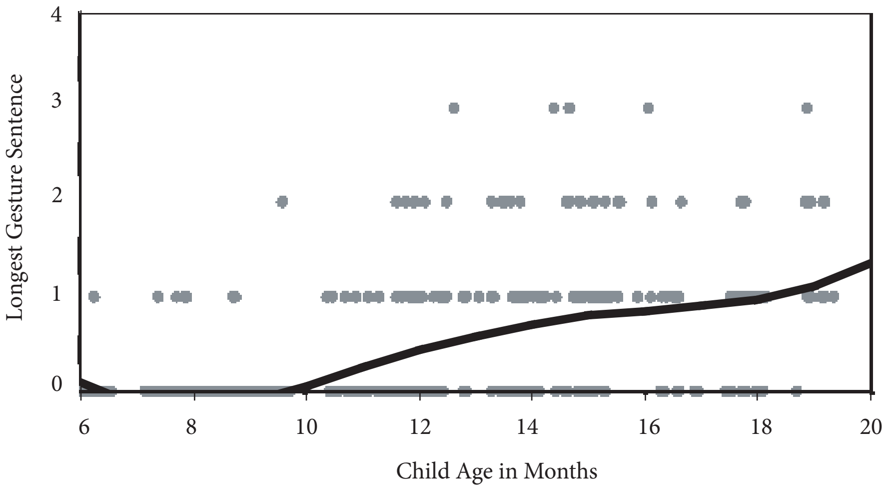


Figure 1. *Scatterplot of infants' longest continuous gestural sentence from 6 to 20 months of age, overlaid with fitted quartic growth model.* NOTE: Height of trajectory is truncated by inclusion of episodes in which children did not gesture.

Table 1. Growth models for the development of children's gesturing sentence length (longest sentence) in a population of 10 infants observed over 8 months

	A. Average means model	B. Average growth	C. Effects of child's early gesturing	D. Effects of caregiver gesturing	E. Effects of child and caregiver gesturing
Fixed Effects					
Initial Status at 6 Months					
Intercept	0.52640***	0.09311	0.61520~	0.05672	0.39730
	(0.07487)	(0.24970)	(0.27960)	(0.24510)	(0.26050)
Age at entry			−0.10180***		-0.06965**
			(0.02721)		(0.02362)
Early gesture frequency			0.31240*		0.28810*
			(0.14410)		(0.12060)
Early gesture variety			0.36780~		0.33720~
			(0.21710)		(0.18550)
Growth over time					
Linear (AGE)		−0.35300~	−0.37370~	−0.26250	−0.31340
		(0.20590)	(0.20360)	(0.19740)	(0.19380)
Quadratic $(AGE)^2$		0.13340*	0.14160*	0.11150*	0.12670*
		(0.0585)0	(0.05776)	(0.05604)	(0.54950)
Cubic $(AGE)^3$		−0.01351*	−0.01456*	−0.01219*	−0.01387*
		(0.00643)	(0.00634)	(0.00614)	(0.00603)
Quartic $(Age)^4$		0.00045~	0.00049*	0.00042~	0.00049*
		(0.00024)	(0.00024)	(0.00023)	(0.00022)
Time-varying effects of caregiver gesturing					
Caregiver prior sentence length				−0.06911**	−0.06934**
				(0.02560)	(0.02559)
Caregiver current sentence length				−0.07743	−0.08294
				(0.06157)	(0.06115)
Caregiver current gesture frequency				.05645***	0.05539***
				(0.01118)	(0.01095)
Variance Components					
Within-child	0.4524***	0.3487***	0.3488***	0.3201***	0.3200***
Between-child	0.0412*	0.0694*	0.0108	0.0385*	0.0042
Fit Statistics: −2LL	754.1	666.1	653.3	631.0	618.6

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Conversations

As early as 7 months, infants responded to a caregiver gesture with their own gesture; however, at this age the infants' gestures were the same as the caregivers' gestures. Therefore, they could be motoric imitations rather than conceptual replies to the caregivers' conversational turns. As early as 10 months, but much more regularly at 11 months, infants replied to caregivers' gestures with different gestures related to the same topic. For example, a caregiver gestures "*more, snack*", and the child replies with the "*all done*" gesture.

Episodes with 4 turns are an important marker of true conversation because in these the infant has replied at least once to the caregivers' gestures and has sustained the conceptual focus through at least 3 turns. Infants engaged in 4-turn gestural conversations as early as 11 months.

Below are examples of conversations between children and caregivers. Horizontal alignment of words, gestures, and actions within each turn represent their sequence.

1. *Discussing another's feelings.* Ellen (11.7 months) is sitting on the classroom floor, another child (Brian) is crying nearby.

Ellen: *waves* <at Brian>
looking at Brian

Caregiver: "Ellen, you are waving at Brian."

Ellen: *cry/sad* (traces finger from eye down cheek)
looks at caregiver, then looks at Brian,

Caregiver: "Yes, Ellen, Brian is crying."
cry/sad

Ellen: *snack/eat* (fingers of one hand tap mouth)
looks at caregiver, at Brian, back at caregiver

Caregiver: "Ellen, you think Brian is hungry?"
snack/eat

Ellen: *sleepy/nap* (palms together, under cheek)
looks at caregiver

Caregiver: "Ellen, you're thinking Brian is ready for his nap?"
sleepy/nap

Ellen: *sleepy/nap*
crawls away from caregiver and Brian

2. *Finding comfort after mom's departure*. Cindy (13.3 months) is in her caregiver's arms; her mother has just left the classroom.

Cindy: *point*
Looks toward the door, extends arm toward door

Caregiver: "Your mom went out that door when she went bye-bye."
wave

Cindy: *wave*
looking at door

Caregiver: "That's right. Mom went bye-bye."
mother (fist running along jaw) *wave*
"You'll see her later at Popsicle Time"
Popsicle (fist tapping chin)

Cindy: *mom*
looks at caregiver

Caregiver: "Yah, you'll see mom at Pops time, later at Pops time."
Mom *Popsicle* *later*

3. *Negotiating play and snack.* Tony (13.5 months) sits at the snack table with his caregiver.

Caregiver: "Do you want more crackers, or are you all done?"
more *all done*
"Do you want more?"
more

Tony: looks outside, then back to caregiver
outside (fingers in loose claw, twisting at wrist)

Caregiver: "You can go play outside when you're all done eating."
play (thumb and pinky extended, middle fingers closed, rotating wrist)
"Do you want more, or all you all done?"
more *all done* (palms down, waving back and forth in front of torso)

Tony: looks at caregiver

Caregiver: "Do you want more snack?"
more

Tony: *all done*
looking at caregiver

Caregiver: "O.k., you're all done. Let's clean up so we can go play."
all done

4. *Clarifying which fish.* Ellen (18.9 months) sits in the book-reading area with her caregiver.

Ellen: *fish* (lips puckered, smacking together)
looks at caregiver

Caregiver: "Do you want to go see the fish in the fish tank?"
fish *point* <across room at tank> *fish*

Ellen: *more* (fingers of each hand together, tapping)
looking at caregiver

Caregiver: "You want more. More of what, Ellen?"
more

Ellen: *point* *yes* (head nods)
looks toward books on the floor

Caregiver: "Oh, you want to read the fish book again?"
"Where is the book?"
where book (palms together, opening out)

Ellen: *point* <at pile of books> *yes*
looks at caregiver, looks back at books

Figure 2 is a scatterplot of the gestural conversation length infants engaged in between 6 and 20 months of age; the overlaid trajectory shows the results of Model B in Table 2. The maximum gesture conversation length observed was 16 turns, though most conversations were less than half that length. Children's age explained 27% of variation in conversation length, indicating that other child or caregiver factors may also predict a dyad's gestural conversation length.

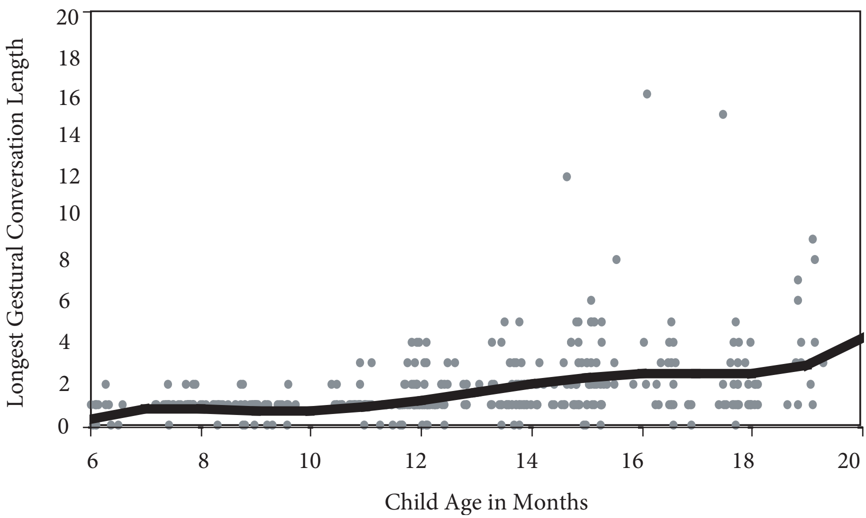


Figure 2. Length of gestural conversation between caregivers and infants from 6 to 20 months of age, overlaid with fitted quintic growth model.

Table 2. Growth models for development of gesturing conversations between children and caregivers for10 infants from 6 to 20 months of age

	A. Average means model	B. Average growth	C. Effects of child's early gesturing	D. Effects of child sentence length	E. Effects of caregiver sentence length
Fixed Effects					
Initial Status at 6 Months					
Intercept	1.61710*** (0.17660)	0.17400 (0.41550)	1.0757 (0.63850)	0.5686~ (0.2753)	−0.1646 (0.3749)
Age at entry			−0.1568* (0.0730)		
Early gesture frequency			0.7072~ (0.3727)		
Growth over time					
Linear (AGE)		1.04250* (0.48070)	1.0355* (0.4793)	0.6930* (0.3511)	0.7373~ (0.4204)
Quadratic $(AGE)^2$		−0.56710* (0.21040)	−0.5616** (0.2098)	−0.3689* (0.1541)	−0.3554~ (0.1846)
Cubic $(AGE)^3$		0.12000** (0.03892)	0.1189** (0.0388)	0.07414** (0.02859)	0.07221* (0.0345)
Quartic $(AGE)^4$		−0.01013** (0.00317)	−0.01003** (0.00317)	−0.00617 (0.00234)	−0.00597* (0.00280)
Quintic $(AGE)^5$		0.00030** (0.00009)	0.00030** (0.00009)	0.00018** (0.00007)	0.00017* (0.00008)
Time-varying effects of child gesturing					
Current gesture frequency				0.3318*** (0.0198)	
Current average sentence length				0.1678* (0.0705)	
Time-varying effects of caregiver gesturing					
Caregiver current gesture frequency					0.09696*** (0.00908)
Caregiver current avereage sentence length					−0.01449 (0.04810)
Variance Components					
Within-child	1.0124***	0.6473***	0.6471***	0.3611***	0.4934***
Between-child	0.2772*	0.2997*	0.1767*	0.0244~	0.2004*
Fit Statistics: – 2LL	1054.7	897.6	892.7	670.3	798.1

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

As seen in Model C of Table 2, the younger children were when they entered the infant classroom, the longer their later gesturing conversations were. Also, children's early symbolic gesture frequency predicted longer conversations. These two variables accounted for 41% of between-child variance in conversation length. Further, controlling for infants' current gesture frequency, their sentence length was also positively related to conversation length (Model D). In contrast, caregivers' gestural sentence length was unrelated to dyad conversation length when controlling for caregivers' gesturing frequency (Model E).

Discussion and conclusion

Preverbal infants are capable of combining gestures to represent and communicate complex ideas. Infants' early symbolic repertoires predict their later ability to combine symbolic representations in the gestural mode. While adult modeling of symbolic gestures (as measured by gesturing frequency) supports children's gestural combinations, caregivers' own combinations actually suppress children's sentence length. It is as if when adults combine many gestures in sequence, the infants could not get a gesture in edgewise.

Infants can also use gestures to converse with adults who are using both words and gestures. It appears that the earlier children are exposed to gestures and the more representational skills they exhibit through gesture (early gesturing frequency and longer gestural sentences), the more they are able to engage in conceptually focused gestural turn-taking, or conversation. However, aside from a natural association between adults' gesturing frequency and the length of conversations, adults' gesturing behavior does not affect dyad conversations.

Future studies should examine the relationship between gestural combinations and the gesture-word combinations documented by Goldin-Meadow and colleagues (Iverson & Goldin-Meadow 2005, Ozçaliskan & Goldin-Meadow 2005) as spoken language emerges. Since use of symbolic gestures with children is associated with earlier vocabulary production (Goodwyn, Acredolo, & Brown 2000), we may hypothesize that symbolic gesture use predicts children's earlier use of gesture-gesture and gesture-word combinations. This should be tested experimentally. Further, since children's gesture--speech combinations elicit more complex language from adults (Goldin-Meadow, Goodrich, Sauer, & Iverson 2007), we may ask whether children's gesture-gesture combinations also elicit more responsive language from adults. This may in part explain the relationship between symbolic gesture use and advanced language development.

In conclusion, given a rich gesture environment, infants can create gestural sentences and converse in the gestural mode. They make use of gestures to negotiate requests, describe observations, and even discuss abstract concepts such as future events and the internal states of others. Children's symbolic gestures reveal the representational and communicative capacities that do not wait for words.

References

Acredolo, L. and Goodwyn, S. 1988. "Symbolic gesturing in normal infants." *Child Development* 59: 450–466.

Bates, E., Bretherton, I., Snyder, L., Shore, C. and Volterra, V. 1980. "Vocal and gestural symbols at 13 months." *Merrill Palmer Quarterly* 26: 407–423.

Caselli, M. C. 1983. "Communication to language: Deaf children's and hearing children's development compared." *Sign Language Studies* 39: 113–144.

Crais, E., Douglas, D. and Campbell, C.C. 2004. "The intersection of the development of gestures and intentionality." *Journal of Speech, Language, and Hearing Research* 47: 678–694.

Cohen, J. 1960. "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement* 20: 37–46.

Goldin-Meadow, S. and Butcher, C. 2003. "Pointing toward two-word speech in young children." In *Pointing: Where Language, Culture, and Cognition Meet*, K. Sotaro (ed), 85–107. Mahwah, NJ: Lawrence Erlbaum.

Goldin-Meadow, S., Goodrich, W., Sauer, E. and Iverson, J. 2007. "Young children use their hands to tell their mothers what to say." *Developmental Science* 10: 778–785.

Goodwyn, S. W., Acredolo, L. P. and Brown, C. A. 2000. "Impact of symbolic gesturing on early language development." *Journal of Nonverbal Behavior, Special Issue: Gesture and Development* 24: 81–103.

Iverson, J. and Goldin-Meadow, S. 2005. "Gesture paves the way for language development." *Psychological Science* 16: 367–371.

Jones, S. E. and Zimmerman, D. H. 2003. "A child's point and the achievement of intentionality." *Gesture* 3: 155–185.

Kaye, K. and Charney, R. 1981. "Conversational asymmetry between mothers and children." *Journal of Child Language* 8: 35–49.

Liszkowski, U., Carpenter, M., Henning, A., Striano, T. and Tomasello, M. 2004. "Twelve-month-olds point to share attention and interest." *Developmental Science* 7: 297–307.

Liszkowski, U., Carpenter, M., Striano, T. and Tomasello, M. 2006. "12-and 18-month-olds point to provide information for others." *Journal of Cognition and Development* 7: 173–187.

Murray, L. and Trevarthen, C. 1986. "The infant's role in mother–infant communication." *Journal of Child Language* 13: 15–29.

Ozcaliskan, S. and Goldin-Meadow, S. 2005. "Do parents lead their children by the hand?" *Journal of Child Language* 32: 481–505.

Rutter, D. R. and Durkin, K. 1987. "Turn-taking in mother-infant interaction: An examination of vocalizations and gaze." *Developmental Psychology* 23: 54–61.

Singer, J. D. and Willett, J. B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Wagner, M. 2006. *First Steps to Communication: A Pragmatic Approach*. Tübingen, Germany: Narr Francke Attempto Verlag

Werner, H. and Kaplan, B. 1963. *Symbol Formation*. Oxford, England: Wiley.

CHAPTER 9

Giving a nod to social cognition

Developmental constraints on the emergence of conventional gestures and infant signs

Maria Fusaro[1] and Claire D. Vallotton[2]
Harvard Graduate School of Education[1] and Michigan State University[2]

Developmental researchers recognize that multiple component skills and social processes underlie children's communication. Infants' gestures have catalyzed consideration of non-verbal behaviors as markers of early communicative and social competence. The current study examines infant sign and conventional gesture production to inform debate on developmental and contextual constraints on communicative competence. Based on observations over eight months, we describe the emergence timing of gestures and signs in ten infants' spontaneous behavior. We test whether two features of gestures and signs, relative frequency of caregiver use and motoric complexity, explain variation in emergence timing. We find that while these features may constrain whether infants produce particular gestures or signs, additional explanatory mechanisms must account for the late emergence of some conventional gestures.

Introduction

Children develop a communicative repertoire that broadens and becomes increasingly more complex. Developmental researchers from social-pragmatic and dynamic systems perspectives recognize that multiple component skills, as well as systematic features of the social context, underlie children's early communication attempts (Bruner 1975, Fogel & Thelen 1987). Pre-linguistic infants' intentional use of actions as proto-declaratives and proto-imperatives highlights the need to look for precursors to communication in children's non-verbal behavior (Bates, Camaioni, & Volterra 1975). Thus far, developmental studies of gesture have emphasized *pointing* as marking a breakthrough in intentional communication toward the end of the first year of life (Carpenter, Nagell, Tomasello, Butterworth, & Moore 1998), likely reflecting infants' emerging understanding of others as intentional agents (Tomasello, Carpenter, & Liszkowski 2007; Crais, Douglas, & Campbell 2004). However, infants can use other conventional gestures and infant signs to communicate. In particular, infant signs, described below, afford a unique perspective for studies of early communicative competence as children can use these body actions to refer systematically to objects, people, and events before they do so using verbal language (Acredolo & Goodwyn 1988).

Children's use of infant signs and conventional gestures can inform debate on the developmental and contextual constraints on communicative competence. In this chapter, we move this agenda forward by describing the timing of emergence of conventional gestures and infant signs in the spontaneous behavior of ten infants. While the timing of emergence of conventional gestures has been reported in prior research in language development (Fenson, Dale, Reznick, et al. 1994), their emergence in relation to infant signs has not yet been explored. We describe features of gestures and signs that infants produce, namely, the frequency of use in the social context, and their motoric complexity. Based on this analysis, we argue that additional elements of gesture and sign use, including social-cognitive demands, must be considered in an explanation of their emergence in children's communication.

Defining conventional gestures and infant signs

Conventional gestures include those body movements used to convey a locally agreed-upon meaning. We focus on four such gestures used in many cultures – *pointing*, *waving* the hand in greeting, *nodding* the head "yes," and *shaking* the head "no." On a popular parental-report measure of child language production, these gestures are referred to as "first communicative gestures" (Fenson et. al 1994). They are also referred to as emblems, which follow standards of form and carry meaning that can be "read" from the movement (McNeill 1998). Cultures have unique collections of conventional gestures and vary in the richness of their repertoire and in social norms for display. Other conventional gestures in the United States include *thumbs up* (indicating success) and putting the index fingers to the lips (requesting silence).

In contrast to conventional gestures, which are ubiquitous in a given culture, infant signs (or symbolic gestures) can be introduced into a local setting, such as a home or childcare center. This informal communication system includes requesting *more* by tapping together the grouped fingers of both hands, and representing *ball* by motioning up and down with the palm of the hand, as if bouncing a ball. By one year of age, infants can use these signs to label objects and to communicate requests and observations (Acredolo & Goodwyn 1988, Goodwyn & Acredolo 1993). Each of these hand, arm, and mouth motions carry semantic meaning and are used systematically in association with the same concept over time. Yet, they are informal; their specific form may vary between families or childcare centers. Caregivers can introduce signs from existing programs, such as the Baby Signs® Program. However, caregivers and preverbal children invent some signs specific to their communicative needs (Acredolo & Goodwyn 1988). Infant signs lack the formal properties of sign languages used by deaf populations, notably syntax.[1] They also differ from both formal sign language and conventional gestures in that ultimately, infant signs are replaced by verbal language.

The current study

The use of infant signs, as well as conventional gestures, in some childcare centers creates a unique means for examining the development of communication skills during the transition from preverbal to verbal language. The current study uses observations from a larger study of children's gesture and sign use (Vallotton 2008), conducted in a child care center that has used infant signs since 1990. The first goal of the current study is to provide descriptive information about the use of particular signs among a group of infants exposed to them in child care. The second is to examine the features of conventional gestures and infant signs, and features of the broader context (e.g., gesturing input), to consider sources of variation in their timing of emergence in children's communicative repertoire. Our research questions include the following:

1. Which infant signs do hearing children in a sign-rich environment learn and use, and at what ages are these signs produced?
2. Do frequency of input and motoric features of conventional gestures and infant signs explain why some gestures and signs are produced early, some later, and some not at all?

In the next sections, we consider the potential roles of input frequency and motoric complexity in explaining variation in the timing of emergence of gestures and signs.

Input frequency

In early verbal language acquisition, the density of maternal speech is a reliable predictor of variation in the child's vocabulary size (Huttenlocher, Haight, Bryk, Selzer & Lyons 1991). If caregiver modeling of gestures and signs supports children's use of these communicative acts, then more frequent exposure to certain gestures and signs may explain their earlier emergence in spontaneous communication. Two related predictions follow: (1) the average emergence age for particular gestures and signs will be earlier among those occurring most frequently in caregivers' communication, and (2) because conventional gestures are part of the child's broader social milieu, they will emerge earlier than infant signs in children's communication.

Motoric complexity

The motoric demands of most gestures and signs should presumably be light as cumbersome movements requiring extensive practice would hardly be useful in real-time

1. The term "home-sign" is also not used here as it has been associated with signs invented by deaf children who lack most language input.

communication. Nonetheless, infant signs vary in form, with some requiring the placement of one hand in a static position and others bimanual coordination of two different movements. The first gestures and signs children use may be those that are motorically easiest to produce. The classic view of motor development is that infants gain control of the body from the head down (cephalocaudal), from the midline of the body outward (proximal-distal), and from large muscle groups to smaller ones (Appleton, Clifton, & Goldberg 1975; Gesell 1946). Control of the head and trunk precedes infants' facility to use the hands to reach for an object (Bertenthal & von Hofsten 1998). Further, controlled complex actions likely emerge from the coordination, or mapping, of single actions (Fischer & Bidell 2006). If motor complexity explains variation in emergence timing, those gestures and signs involving only the head, such as *head shaking* and *head nodding,* would be among the earliest to emerge. Also, the earliest manual gestures and signs to emerge would be those that involve gross motor movements, those comprised of only one action, and among bimanual gestures, those for which both hands perform simultaneous and symmetrical movements rather than separate or complementary movements (Corbetta & Thelen 1996).

Methods

Sample

Participants were 7 female and 3 male infants and 24 non-parental caregivers in an infant classroom at a university laboratory school. Infants spent 3 to 12 hours per week in the classroom. They were between 5.5 and 11.0 months of age at the beginning of the observation and between 14 and 19.5 months at the end. Caregivers were 22 university students and 2 hired teachers (22 female and 2 male). Students cared for children as part of a child development internship for a minimum of 3 and maximum of 9 months.

Exposure to infant signs

Caregivers were taught to use infant signs in conjunction with words through explicit instruction by their supervisors. Caregivers were given a list and descriptions of signs to be used; the same list was sent home with parents, though home-based use was not reinforced. Posters were placed around the classroom as reminders for caregivers to use the signs. Table 1 provides brief descriptions of each of the four conventional gestures and 66 infant signs that caregivers produced during the observation period. Though children spent approximately the same amount of time in the classroom, overall exposure to gesturing was not uniform. Infants were never instructed or required to use signs; they learned them only through informal caregiver modeling.

Table 1. Descriptions of conventional gestures and infant signs, features of caregiver and infant production, and complexity of associated motor actions

	Gesture Description	Caregiver frequency		*N* infants producing gesture or sign	Avg. emergence age (mos.)	Motor complexity (1–4)
		N	% of total (*N* = 2788)			
Conventional Gestures						
Point	Single extended finger, without touching referent.	636	22.81%	10	10.65	1
Wave	Open palm, waving side-to-side at wrist. OR Fingers vertical, opening and closing together.	148	5.31%	9	12.03	1
Yes	Up-down head nodding at the neck.	26	0.93%	3	15.57	NA
No	Head shaking side-to-side.	10	0.36%	7	14.18	NA
Infant Signs						
Snack	Fingers of one hand grouped, tapping mouth.	301	10.80%	4	13.35	1
More	Grouped fingers of both hands tapping together.	215	7.71%	5	11.10	2
All done	Hands open, palms down, waving back and forth.	165	5.92%	4	14.44	2
Hear	Open palm over ear.	146	5.24%	2	15.10	1
Where	Palms of hands up next to shoulders.	119	4.27%	6	13.45	2
Sit	Index and middle finger of both hands tapping one on top of other (making an X).	115	4.12%			3
Play	Pinky and thumb extended, hand rotating.	91	3.26%	5	13.88	1
Bottle	Loose fist to mouth.	80	2.87%	3	13.94	1
Ball	Up-down palm motion, as if bouncing ball.	68	2.44%	4	15.07	1
Outside	Open palm twisting, as if opening door.	67	2.40%	6	12.43	1
Parent	Open palm, thumb tapping between forehead and chin (left side).	46	1.65%	3	14.32	1
Fish	Smacking pursed lips.	39	1.40%	3	13.85	NA
Star	Fingers of one or both hands vertical and wiggling.	37	1.33%	8	12.56	1
Sad	Drawing forefinger down cheek, as if tracing tear.	33	1.18%	3	14.91	1
Book	Palms opening together.	33	1.18%			2
Later	Rotating right thumb in open left hand.	32	1.15%	1		3
Bib	Pat chest, indicating someone else's bib.	30	1.08%	1		1
Duck	Fingers to thumb, opening and closing.	29	1.04%	3	14.31	1
See	Finger pointing to eye; OR finger extending from eye forward.	27	0.97%	1		1
Hat	One hand tapping top of head.	27	0.97%	1		1

	Gesture Description	Caregiver frequency N	Caregiver frequency % of total (N = 2788)	N infants producing gesture or sign	Avg. emergence age (mos.)	Motor complexity (1–4)
Popsicle Time	Tapping back of palm to chin.	26	0.93%	2	14.95	1
Wait	Right fist tapping left palm.	21	0.75%	1		3
Spider	Index fingers rubbed together as in "Itsy, Bitsy Spider."	19	0.68%	1		2
Sleep	Folded hands, laid against cheek.	18	0.65%	1		3
Horse	Hand stroking face as if petting nose of horse.	17	0.61%			1
Camera	One hand in half-circle shape framing eyes, one or two fingers moving down as if pushing a button.	13	0.47%			3
Elephant	Back of hand to nose.	13	0.47%			1
Phone	Fist of one hand to ear.	11	0.39%	1		1
Happy	Open hands, palms outward, framing face.	10	0.36%	1		2
Swim	Palm flat, perpendicular to body, weaving back/forth.	9	0.32%			3
Necklace	Fingers of both hands grasp, move up and over head, then down meeting in front of neck.	8	0.29%			4
Bird	Arms or hands fluttering.	7	0.25%	4	11.17	2
Rocking Horse	Torso rocking forward and back.	7	0.25%			NA
Music	Torso swaying side to side.	6	0.22%	5	13.83	NA
Bunny	Wrinkling nose.	6	0.22%	1		NA
Diaper	Patting hip.	6	0.22%			1
Wash	Two hands running over one another.	6	0.22%			4
Heart	Two hands drawing heart on chest.	5	0.18%			4
Slide	One hand swoops over and down in front of torso.	5	0.18%			3
Baby	Arms folded at chest, rocking baby.	4	0.14%			4
Hair	One hand stroking head as if brushing hair.	4	0.14%			1
Tiger	Claw-shaped hand swiping near face.	4	0.14%			1
Juice	Index finger to cheek.	3	0.11%	1		1
Cow	Hand over head, thumb and pinky pointing up.	3	0.11%	1		1
Cat	Hand gently stroking opposite forearm.	3	0.11%			3
Cleanup	Palm down, circular motion.	3	0.11%			1
Gentle	One hand stroking other hand.	3	0.11%			3
Loud; Water; Big; Car; Eyeglasses; Giraffe; Lotion; Roll; Butterfly; Frog; Laugh; Milk; Open; Pig; Rain; Smile; Talk; Tall; Train		<3	<.10% each	Loud = 1 Water = 1		

Videotaping procedures

Over an 8.5 month data collection period, each infant was observed an average of 40 times in 5-minute episodes. Infants were videotaped during normal program routines in their childcare classroom; approximately half of the recordings were made during snack-time and half during free-play. On average, each infant was filmed a total of 200 minutes, approximately 1% (0.93) of their total 360 hours in the classroom over the data collection period. The gesture and sign use measured included entirely spontaneous behavior. Communication was not elicited by the researchers; normal classroom routines in this gesture-rich environment served as interaction contexts which might be natural elicitors of gesture and sign.

Coding

All gesturing behavior was coded from the video-recorded episodes in real time unless there was a technical problem (such as an obscured camera view) that rendered the behavior of the child or caregiver unrecognizable. For each observed gesture, coding captured which gesture was performed, who performed it, and when it occurred within the episode. Coders were university students trained to recognize all gestures and signs by learning to perform them from written and visually demonstrated instructions and from seeing examples on training videos. Inter-coder reliability was assessed using Cohen's Kappa (Bakeman & Gottman 1987). Coders obtained a Kappa of .75 or above on 5 consecutive episodes before beginning to code independently. Upon reassessment of 15% of all tapes, coders achieved Kappa scores of .83 and above.

Results

Infants' use of gesture and sign

To address the first research question, we examined the set of all gestures and signs produced across infants and the average age at which each item was first observed. In addition to the four conventional gestures, caregivers produced 66 infant signs at least three times each.[2] Table 1 presents the number of infants (of 10) observed using each item. Each infant was observed *pointing*. Seven or more children produced *waves* and *head shakes,* while three produced *head nods*. Two or more infants were observed using a subset of 17 signs. Those produced by the greatest number of children included *star* (n = 8), *outside* (n = 6), and *where* (n = 6). Several infant signs were observed

2. Infant signs used once or twice by caregivers are listed in the bottom row of Table 1. Given their infrequency (each comprising <.1% of caregivers' total gesture and sign use), they are excluded from the analysis.

among a smaller number of children, with an additional subset of 12 signs each produced by one child.

The signs that two or more children learned and used were typically relevant to the activities and objects in the classroom context. They represented animals (e.g., *bird, fish, duck*) desired objects (e.g., *snack, ball*) and activities (*play*, *outside*: for going outside, *star*: for singing a song about stars). Less concrete signs, representing concepts beyond the here and now, included *where, parent*, and *Popsicle Time* (a center-specific event at the end of the day when children and parents sit and eat popsicles).

We examined the average age at which each of the 21 gestures and signs produced by at least two children was first observed in children's spontaneous communication. In Figure 1, mean emergence ages are marked with diamonds, with error bars representing the standard deviation around the average. Overall, the emergence ages for conventional gestures (top part of figure) were more disparate than those for infant signs (bottom part of figure). Most signs were first observed between 11 and 15 months of age, though individual variability was apparent.

To determine which gestures and signs were particularly early or late to emerge, we calculated the average emergence age across the combined set of 21 items as a benchmark. This value accounts for the emergence age for *pointing* for each child observed *pointing*, the emergence age for *waving* for each child observed *waving*, and so on for the remaining items. This benchmark age was 13.14 months ($SD = 2.63$) and is represented as a vertical dashed line on Figure 1. Analyzed in separate categories, conventional gestures emerged on average at 12.44 months ($SD = 2.68$) and infant signs at 13.43 months ($SD = 2.57$).

The emergence ages for individual gestures and signs were compared to the benchmark (13.14), using one-sample t-tests (two-tailed, .05 level of significance). The average emergence age for *pointing* was significantly earlier than the benchmark ($t = -2.66$, $p = .026$). In contrast, the average emergence age for *head nods* and *head shakes* were significantly later than the benchmark (*nod*: $t = 4.83$, $p = .040$; *shake*: $t = 2.98$, $p = .025$). This pattern is aligned with previous studies suggesting that *head shakes* and *head nods* are typically observed later than *pointing* and *waving* based on parental report and researcher observation (Crais, Douglas, & Campbell 2004; Fenson et al. 1994). Infant signs showed fewer systematic differences in emergence timing. *Popsicle time* emerged slightly later than the benchmark ($t = 8.23$, $p = .077$). While additional signs were produced relatively early and late, wide individual differences were apparent and use by only a subset of children limited the statistical power needed to detect significant differences.

Accounting for variability in emergence

To address the second research question, we examined each of the gestures and signs according to the frequency with which caregivers used them and their motor complexity. In the following sections, we present evidence relevant to these sources of variation in emergence timing.

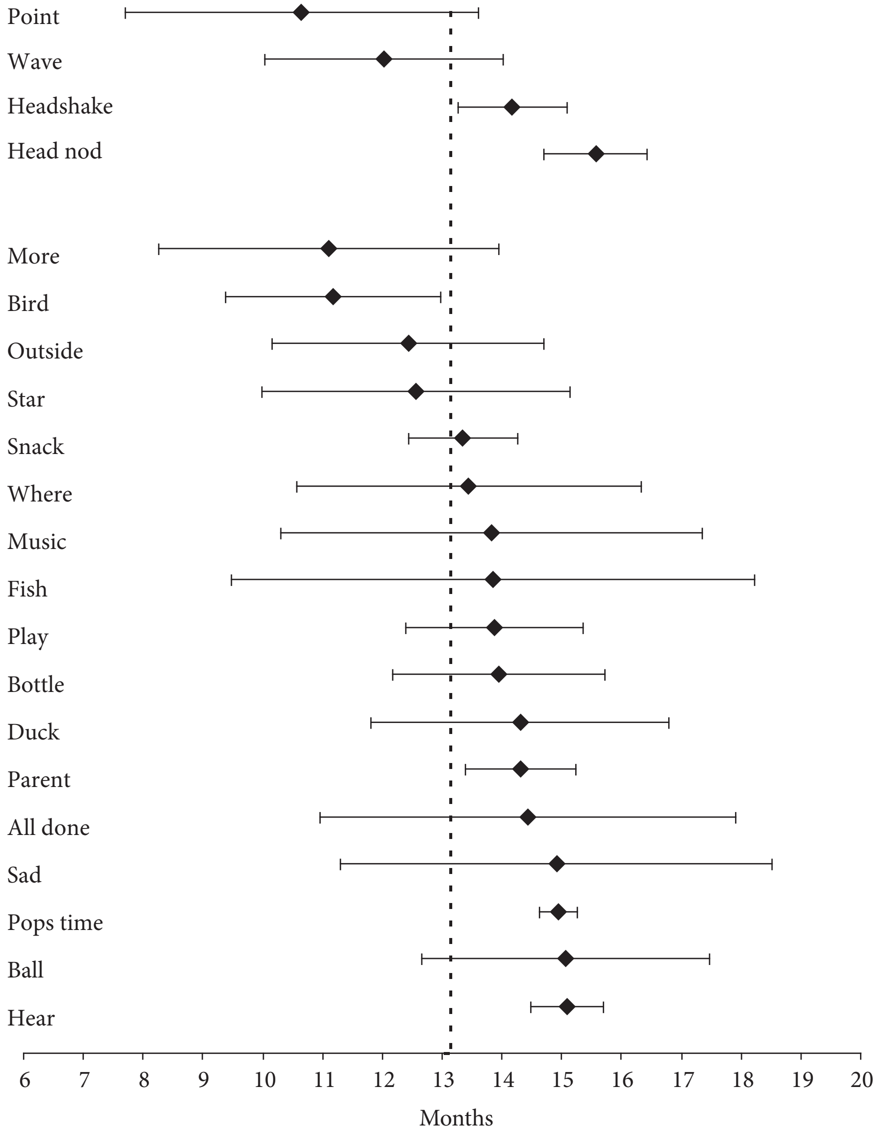


Figure 1. Average emergence age (in months) of conventional gestures and infant signs produced by two or more infants.

Frequency of input

Table 1 presents the number of times caregivers used each gesture and sign and the percentage of the total set of caregiver behaviors represented by each action. By far, the most commonly produced action was *pointing*, comprising 22.81% of caregivers' non-verbal repertoire. The next most frequent items were the infant signs for *snack* (10.80%) and *more* (7.71%), reflecting, in part, the snack-time context in which half of the observations were made.

Correlation analyses were conducted to examine the relationship between the frequency of input (% of instances) and the average emergence age for individual gestures and signs produced by two or more children. Initial analyses revealed a negative correlation such that higher frequency gestures were earlier to emerge (r = -.521, p = .015). However, this relationship was largely driven by *pointing*, which was an outlier in terms of caregiver frequency and was particularly early to emerge. Setting aside *pointing*, the relationship between frequency and emergence age was still negative but weaker and non-significant (r = -.257, p = .27).

A frequency-based explanation for variation in emergence timing suggests that conventional gestures, which are ubiquitous beyond the child care center, will emerge earlier than infant signs. We examined the ages at which each infant was first observed using any infant sign and any conventional gesture. On average, infants produced their first conventional gesture at 10.04 months (SD = 2.94), a few weeks earlier than their first infant sign at 10.41 months (SD = 2.02). A paired-samples t-test (two-tailed, .05 level of significance), confirmed that this difference was not statistically significant (t = .34, *ns*). This difference was also insignificant when first pointing gestures were excluded from analysis. Further, the sample was evenly divided among those who were observed using a conventional gesture before using an infant sign (n = 5) and those first observed using an infant sign (n = 4), with one infant observed using each at the same observation session. The same results were obtained when analyzing only those infants for whom observations began prior to 8 months of age (n = 6). Thus, use of conventional gestures in the broader social milieu does not appear to lead to earlier emergence of these gestures, compared to infant signs, in children's communication.

We next examined whether the signs children produced were particularly high frequency in caregiver input. Among infant signs used by two or more children (n = 17), the average caregiver frequency, expressed as percent of all instances, was 3.11% (SD = 2.88). The average frequency for never-produced signs (n = 18) was .51% (SD = .94). A two-tailed t-test revealed that the average caregiver frequency was higher among signs the children used than among signs they did not use (t = 3.55, p = .002, equal variances not assumed). Signs produced by only one child (n = 12) tended to be infrequent, comprising an average of .62% (SD = .38) of the caregivers' repertoire. Thus, frequency of caregiver input appears to be associated with whether or not children produced a given infant sign, but not with the variability in age of emergence.

Motor complexity

Controlled head and torso movements develop earlier than finer movements of the hands and fingers. Thus, among the gestures and signs modeled by caregivers, the lowest levels of motoric complexity apply to *music* and *rocking horse*, and to *head nodding* and *head shaking* as they require only gross motor movement of the torso or head. The emergence age for *music* was widely variable among the five children who produced it and was not significantly different from the benchmark average. No children were observed using the *rocking horse* sign. The *head nod* and *head shake* emerged later than most manual infant signs, suggesting that motor complexity alone cannot explain their late emergence.

We categorized each manual gesture and sign that caregivers produced, based on motor complexity, following Dennis and colleagues (1982) and Daniloff and Vergara (1984), who analyzed the motor demands of formal sign languages. Their seven categories of increasing complexity were collapsed into four (Table 2), yielding groups of gestures and signs large enough for comparison. The categorization of each action is included in Table 1.

The majority of manual gestures and signs that caregivers produced were relatively simple, unilateral without crossing the midline of the body (Category 1). However, caregiver gestures and signs covered the full range of motor complexity levels. Infants produced a subset of items from the first two categories. Those signs requiring the third motor complexity level were each used by only one child. No children were observed using the signs that were the most motorically challenging (Category 4).

Average emergence ages by motoric category were calculated using the same method used for the overall benchmark. Only those items used by two or more children were used in these calculations. For the simplest manual items (Category 1), the average emergence age was 12.99 months (*SD* = 2.52). For Category 2, average emergence age was 12.56 months (*SD* = 2.96). The difference between these averages was

Table 2. Coding of manual gestures and signs according to motoric features, from category 1 (simplest) to 4 (most complex)

		Number of gesture and sign types produced	
Category	**Feature**	**Caregivers***	**Infants****
1	Unilateral: not crossing midline	25	13
2	Bilateral mirror movements: not crossing midline	7	4
3	Bilateral: not crossing midline Unilateral across midline Bilateral: one base, one mover	9	0
4	Bilateral: both movers Bilateral crossing midline: both hands cross	4	0

* Includes gestures and signs produced by caregivers 3 or more times.

** Includes gestures and signs produced by two or more infants.

not significant (t = .628, *ns*). This similarity in emergence ages does not support the motor complexity hypothesis for explaining variation in the timing of emergence for manual signs. However, motor complexity may be related to whether a particular sign is ever observed in infants' spontaneous communication.

Discussion

Children's production of conventional gestures and infant signs provides a window into early communicative competencies. Previous research has identified the end of the first year of life as typical for entry into non-verbal, intentional communication. We found that infants learned and spontaneously produced a large subset of signs modeled by caregivers. Their signs referred to concepts relevant to the childcare context, including a small subset used to represent abstract ideas. In this sample, infants were first observed using infant signs in their spontaneous communication when they were approximately 10 months of age. Thus, infant signs provided a means for these children to refer to specific objects and events before their first birthday.

Measuring spontaneous communication

Gesture researchers must choose between measuring spontaneous and elicited non-verbal behavior. This study focused on spontaneous production because of our interest in understanding children's naturalistic use of gestures and signs. This design eliminates the possibility that observed behaviors are disconnected from real-world behavior. A limitation of this approach, however, is that the observed actions include only those that were relevant to the user's communicative goals. Thus, we cannot rule out that children might have produced additional signs had they been prompted to do so. While the current study provides insight into infants' spontaneous gesture and sign use, future studies might complement this approach by examining children's production of signs in response to explicit elicitation.

Variation in emergence timing

Although their average emergence age was similar to that of infant signs, conventional gestures showed a more pronounced differentiation in emergence timing; *pointing* was early to emerge, while *head shaking* and *head nodding* were relatively late. In this sample, no specific infant signs were systematically early or late to emerge in infants' spontaneous communication.

We found mixed evidence that the frequency of caregivers' use of particular gestures and signs is related to emergence timing in infants' communication. In line with this account, *pointing* was heavily represented in caregivers' communication and was early to emerge. However, when we excluded *pointing* as an outlier, the negative relationship between frequency and emergence age was no longer significant. Nonetheless, infants were more likely to produce those signs that were relatively more frequent in caregivers' communication. Thus, input frequency may predict whether infants use a particular sign, but not when it emerges, at least during the infancy period represented in this study.

In terms of motor complexity, a similar pattern was found; complexity appeared to be related to whether, but not when, particular signs emerged in children's communication. All of the gestures and signs produced by two or more children required gross movements of the head or torso or relatively simple movements of the arms and hands; only three of the more complex bi-manual signs were ever observed, in each case by one infant. As discussed in more detail below, motor demands also do not explain the relatively late emergence of *head nodding* and *head shaking*.

Integrating component skills and context

In this chapter, we suggested that there are multiple constraints on the timing of emergence of conventional gestures and infant signs such as features of the communicative context and of the gestures. This analysis suggests that a multi-faceted account is needed to explain whether and when infants spontaneously use particular gestures and signs. For instance, the signs for *bird* and *music* were relatively infrequent in input, comprising only .25% and .22% of caregivers' total observed repertoire, respectively. However, their relative motor simplicity (flapping the arms in synchrony; swaying the torso) might facilitate their inclusion in children's communication. Further, children's interest in communicating about these topics, such as referring to birds at the bird-feeder or requesting music, also contributes to their occurrence in the child's spontaneous repertoire.

The integrated roles of caregiver input, motoric complexity, and context are also reflected in the absence of particular signs in children's behavior. For instance, *sit* accounted for over 4% of caregivers' total gesturing (over 100 observations in our recordings), yet no infants produced it. This absence may be accounted for by the need to coordinate the placement of fingers of both hands (motoric complexity) or by the sign's irrelevance to the child's communicative goals; caregivers have concerns for classroom management and cleanliness that infants do not share, which may explain why some signs were not observed in children's communication (e.g., *sit*, *wash*, *wipe nose*).

Late emergence of head gestures

This descriptive study replicates previous reports that *head nodding* and *head shaking* emerge later than *pointing* and *waving* in children's communication (Fenson et al. 1994; Crais, Douglas & Cambell 2004). We also found that *nodding* and *shaking* are relatively late to emerge compared to many infant signs. This pattern is intriguing, given that these gestures are motorically simple and are modeled both in and out of the infants' child care context, whereas infant signs require manual activity and are largely confined to the classroom setting. Some other factors besides frequency of input and motoric complexity must explain their late emergence. One such factor may be the social-cognitive complexity associated with their use.

Developmentalists studying pragmatics in verbal language have argued that development in children's understanding of social interaction contributes to advances in the pragmatic sophistication of their communication (Ninio & Snow 1996). Increasing social understanding is reflected in the broadening range of communicative acts that children learn to control in speech. Guidetti (2005) similarly argues that the child's developing ability to adapt and respond to adult dialogue may explain the increasing frequency of agreement and refusal messages children produce with words, head nods, and head shakes between one and three years of age. It is possible that gesture and sign follow a similar progression, such that children control a broadening set of forms that serve an expanding range of communicative goals. Those gestures and signs used to perform the simplest communicative acts should emerge earlier than those serving more complex functions.

Using *head nods* and *head shakes* may be more socially and conceptually complex than using gestures and signs refer to or to request a tangible entity or event. *Nods* and *shakes* are given in reply to another person's offer, suggestion, or question, and convey agreement and refusal messages. Infants have the option of responding to an offer or question by performing a relevant behavior, such as pushing away a refused object or showing excitement when a caregiver offers to repeat an interesting activity. Intentionally conveying agreement or refusal messages, whether in words or gestures, may thus reflect a breakthrough in children's ability to respond to others' messages using conventional modes of communication.

References

Acredolo, L. and Goodwyn, S. 1988. "Symbolic gesturing in normal infants." *Child Development* 59: 450–466.

Appleton, T. R., Clifton, R. K. and Goldberg, S. 1975. "The development of behavioral competence in infancy." *Review of Child Development Research* 4: 101–185.

Bakeman, R. and Gottman, J. M. 1987. "Applying observational methods: A systematic view." In *Handbook of Infant Development* (2nd ed.), J. D. Osofsky (ed), 818– 854. New York: Wiley.

Bates, E., Camaioni, L. and Volterra, V. 1975. "The acquisition of performatives prior to speech." *Merrill-Palmer Quarterly* 21: 205–226.

Bertenthal, B. and von Hofsten, C. 1998. "Eye, head and trunk control: The foundation for manual development." *Neuroscience and Biobehavioral Reviews* 22 (4): 515–520.

Bruner, J. S. 1975. "From communication to language: A psychological perspective." *Cognition* 3: 255–287.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G. and Moore, C. 1998. "Social cognition, joint attention, and communicative competence from 9 to 15 months of age." *Monographs of the Society for Research in Child Development* 63 (4): i-174.

Corbetta, D. and Thelen, E. 1996. "The developmental origins of bimanual coordination: A dynamic perspective." *Journal of Experimental Psychology: Human Perception and Performance* 22: 502–522.

Crais, E., Douglas, D. D. and Campbell, C. C. 2004. "The intersection of the development of gestures and intentionality." *Journal of Speech, Language, and Hearing Research* 47: 678–694.

Daniloff, J. K., and Vergara, D. 1984. Comparison between the motoric constraints for Amer-Ind and ASL sign formation. *Journal of Speech, Language and Hearing Research, 27,* 76–88.

Dennis, R., Reichle, J., Williams, W. and Vogelsberg, T. 1982. "Motor factors influencing the selection of vocabulary for sign production programs." *Journal of the Association for the Severely Handicapped* 7 (1): 20–32.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. and Pethick, S. J. 1994. "Variability in early communicative development." *Monographs of the Society for Research in Child Development* 59 (5).

Fischer, K. W. and Bidell, T. R. 2006. "Dynamic development in action and thought." In *Handbook of Child Psychology: Theoretical Models of Human Development* (6th ed., Vol. 1.), W. Damon and R. M. Lerner (eds), 313–399. New York: Wiley.

Fogel, A. and Thelen, E. 1987. "Development of early expressive and communicative action: Reinterpreting the evidence from a dynamic systems perspective." *Developmental Psychology* 23 (6): 747–761.

Gesell, A. 1946. "The ontogenesis of infant behavior." In *Manual of Child Psychology*, L. Carmichael (ed), 295–331. New York: Wiley.

Goodwyn, S. W. and Acredolo, L. P. 1993. "Symbolic gesture versus word: Is there a modality advantage for onset of symbol use?" *Child Development* 64 (3): 688–701.

Guidetti, M. 2005. "Yes or no? How young French children combine gestures and speech to agree and refuse." *Journal of Child Language* 32: 911–924.

Huttenlocher, J., Haight, W., Bryk, A., Selzer, M. and Lyons, T. 1991. "Early vocabulary growth: Relation to language input and gender." *Developmental Psychology* 27: 236–248.

McNeill, D. 1998. "Speech and gesture integration." In *The Nature and Functions of Gesture in Children's Communication*, J. M. Iverson, S. Goldin-Meadow, et al. (eds), 11–27. San Francisco: Jossey-Bass.

Ninio, A. and Snow, C. E. 1996. *Pragmatic Development*. Boulder, CO: Westview Press.

Tomasello, M., Carpenter, M. and Liszkowski, U. 2007. "A new look at infant pointing." *Child Development* 78 (3): 705–722.

Vallotton, C. D. 2008. "Signs of emotion: what can preverbal children "say" about internal states?" *Infant Mental Health Journal* 29 (3): 234–258.

CHAPTER 10

Sensitivity of maternal gesture to interlocutor and context

Maria Zammit* and Graham Schafer
University of Reading

Child-directed communication may be systematically modified either because (1) it scaffolds language learning (the 'Facilitative Strategy Hypothesis') or (2) as a consequence of the semantic simplicity of interactions with children (the 'Interactional Artefact Hypothesis'). To compare these hypotheses, we compared maternal gestural production in dialogue with adults and children. We also examined the sensitivity of gestural production to children's concurrent linguistic level. Twenty-nine mothers and their 16–24-month-olds were video-recorded during a free play session, and during picture and word description tasks. In interaction with children, maternal gestural repertoires were limited, typically comprising concrete deictic and representational gestures; abstract emphatic gestures were rare. Maternal gesture and children's current vocabulary were positively correlated. Thus, maternal gestural modification appears to scaffold word learning, supporting the Facilitative Strategy Hypothesis.

Child-directed speech is systematically modified in comparison with adult-directed speech (Snow 1972). Child-directed action and gesture are also modified relative to adult-directed communication (Brand, Baldwin, & Ashburn 2002; Shatz 1982). There are two influential explanations for modification of child-directed communication (CDC). First, modification in CDC may scaffold linguistic development (Barrett, Harris, & Chasin 1991; Hampson & Nelson 1993; Shatz, 1982). According to this account, the relative simplicity and redundancy in CDC aids in parsing information and resolving ambiguity; child-directed actions facilitate infant attention, thereby enhancing learning and comprehension (Brand, Baldwin, & Ashburn, 2002; Iverson, Capirci, Longobardi, & Caselli 1999). Such modifications, in which adults adjust communication to the level of their interlocutor, we henceforth refer to as the Facilitative Strategy Hypothesis, or FSH. Second, perhaps child-directed speech is concrete,

* to whom correspondence should be addressed. Current address is Faculty of Health, Leeds Metropolitan University, Civic Quarter, Leeds, UK

brief and less complex than adult-directed speech because adults speak to children about a smaller range of subjects and in a less abstract way than to adults. The relative simplicity of child-directed speech is an artefact of the "...*narrow set of semantic relations typically expressed...*" in speech to children; the "...*apparent...simplicity of CDS (child-directed speech) is best understood as an artefact of its semantic simplicity* ..." (Pine 1994: 17). This view is henceforth referred to as the Interactional Artefact Hypothesis, or IAH. Both CDC and IAH can be extended to gestural interaction.

In this chapter, we set out to establish (1) if maternal child-directed gestures are modified relative to adult-directed gestures both within and across contexts and (2) whether such modification is sensitive to the size of children's vocabularies. The two different views of modification make different predictions. FSH predicts that mothers adjust communicative behavior when the child requires support; thus, communication should be sensitive to both the context and the child. In contrast, IAH predicts that maternal communicative style is sensitive to the semantic context only, and not to the child.

Several studies have investigated mothers' use of gesture in interaction with infants (Gutmann & Turnure 1979; Iverson et al. 1999; Namy, Acredolo, & Goodwyn 2000; O'Neill, Bard, Linnell, & Fluck 2005; Rowe, Pan, & Snow 2003; Schmidt 1996; Shatz 1982), but no published study has compared child-directed gesture with adult-directed gesture within-subjects. We, therefore, set out to study, within-subjects, mothers' use of gesture in three contexts and with two different interlocutors.

Only one previous, unpublished, study directly compared child- and adult-directed communication (Bekken 1989). Bekken observed triadic communication between mothers, 18-month-old daughters and an unfamiliar female adult. Mothers produced around twice the number of adult-directed as child-directed gestures, usually in the form of speech-gesture combinations ('speech-gesture acts', henceforth SGA). However, although mothers gestured more frequently to adults than to children, there were no reliable differences in the relative *proportion* of speech-gesture acts to speech alone acts directed to adults versus children because adult-directed speech was also more frequent than child-directed speech. It is additionally possible, however, that the observed similar proportion of SGA rates to children and to adults is simply specific to triadic rather than dyadic interactions.

The majority of research investigating child-directed gesture (CDG) has observed mother-child interaction during free play (Iverson et al. 1999, Bekken 1989). However, O'Neill et al. (2005) found considerably higher maternal CDG rates during a structured counting task and free play session with 20-month-olds than is typically observed during free play sessions. This finding offers some support for the FSH over the IAH. It further suggests that observing mother-child interactions during a single context may limit the scope of the findings. Our task therefore employed dyadic interaction in three distinct contexts: (1) a word description task, in which adults talked about a topic presented to them in the form of a single word; and (2) a picture description task, in which a topic was presented as a single image; and (3) unstructured interaction.

The first two tasks might be expected to elicit differential amounts of gesture because concrete referents elicit more gesture than do abstract topics (Feyereisen & Havard 1999). In contrast, free interaction has no concrete referent or specific instruction, thus potentially inhibiting all gesture types. Mothers repeated all three contexts with an adult and a child interlocutor.

FSH and IAH make differential predictions about maternal sensitivity to children's vocabulary. Under FSH, differences in maternal gesture to children might be expected to occur as a function of the child's vocabulary as mothers 'tune' their support to children's lexical knowledge. Under IAH, because the semantic context is held constant between high and low vocabulary interlocutors, we would expect little or no difference in maternal gesture as a function of the child's vocabulary.

Method

Participants

Participants were 29 British, white, mother-child dyads, recruited when children were aged 16–24 months. All mothers were married or living with partners and were aged between 20 and 40 years. Families were middle class, scoring at least 3.5 or above on the Socio-economic status coding scheme in the Life Events and Difficulties schedule (Brown & Harris 1978). All mothers were educated to British 'A' level standard or above, scoring at least 2 on the Educational status coding scheme (Brown & Harris 1978). Twenty-two mothers were full-time caregivers. Table 1 gives the ages and productive vocabulary scores grouped by a median split on high versus low productive vocabulary score (see below).

A different unfamiliar adult interlocutor participated with each mother so that, like each child, they were unfamiliar with the experiment. Adult interlocutors were gender matched to the child.

Table 1. Description of children in sample

Vocabulary score	Gender	N	Age		Productive vocabulary score	
			Mean (SD)	Range	Mean (SD)	Range
Low	Male	10	18.8 (3.1)	16–24	7.1 (7.9)	0–22
	Female	4	16.6 (0.5)	16–24	16.2 (8.1)	7–24
High	Male	8	22.1 (3.2)	15–25	133.8 (102.9)	26–289
	Female	7	21.1 (1.4)	20–23	176.7 (192.7)	34–564

Questionnaires

No standardised parent reported measure of British children's gestural production existed at the start of the research. Therefore, we developed a parent-completed checklist of communicative development. The verbal section of the checklist had previously been administered to British children (Tan & Schafer 2005), while the gestural section was adapted and extended from the MacArthur Communicative Development Inventory: Words and Gestures (Fenson et al. 1994).

Stimuli

A bank of forty nouns familiar to British 15-month-olds was compiled from existing data (Hamilton, Plunkett, & Schafer 2000), being: aeroplane, apple, ball, balloon, banana, bath, bed, bib, boat, book, camera, car, cat, clock, comb, crayons, cup, slide, dog, doll, duck, elephant, fish, flower, hat, jumper, keys, lollipop, pool, phone, rolling pin, shoe, spoon, star, swing, teddy, toothbrush, tree, and umbrella.

Procedure

Mothers visited the laboratory on two occasions, one to three weeks apart, interacting with their child in one session, and with the adult interlocutor in the other. The order of child-adult sessions was randomly assigned and counterbalanced. During each visit, dyads were video-recorded in two structured tasks (word and picture description) and an unstructured free interaction. During the structured tasks, dyads viewed 10 pictures or words randomly selected from the 40-item bank, each projected individually onto the wall of the experimental room for 20s. Random selection of words and pictures typically resulted in different items appearing in each structured task. Mothers had been previously instructed to talk to their interlocutor about each item until they were signalled to stop. At no point was reference made to gesture.

Coding and analysis

The videotaped observations were coded using the scheme of O'Neill et al. (2005). A single speech utterance comprised any verbalisation followed by (a) a silence, (b) a change in conversational turns, or (c) a change in intonation pattern. Each utterance was further classified in one of two exclusive categories: as *speech alone* (SA) or, when a gesture was enacted in temporal overlap with an utterance, as a *speech-gesture act* (SGA). A gesture comprised a hand, arm or body movement preceded and followed by a clear pause or relaxation of hand position. Gesture without speech was never observed.

Coding of maternal gesture by type

The videotaped observations were analysed and all occurrences of maternal gestures were coded, including deictic, representational and emphatic gestures. *Deictic gestures* (e.g. a point to an object) indicated the existence of an object, person, or occurrence of an event. *Representational gestures* referred in non-arbitrary (iconic) fashion to objects, locations, individuals or events. Such gestures describe an attribute or action of an object and differ from deictic gestures in that their meaning is consistent across situations. *Emphatic gestures* (or beat gestures) highlight aspects of discourse structure and/or the content of accompanying speech. They are non-representational, have no specific semantic content or precise referent, and are not linked to a specific hand shape or facial expression. These emphatic or beat gestures are rarely observed during adult-child speech (Iverson et al. 1999).

Results

In order to tease apart the two proposed explanations for maternal adjustment of communication to children (i.e., FSH versus IAH), we first examine maternal production of speech and gesture as a function of interlocutor and context. We then explore variability in maternal gesture *type* production as a function of interlocutor and context. Finally, we examine the relation between maternal gesture production and children's current vocabulary.

Proportional gesture rates

Table 2 presents the mean proportion of maternal speech alone and speech-gesture acts as a function of context and interlocutor, with child-directed communication further grouped by the median split on high versus low vocabulary scores. The

Table 2. Mean proportion of speech alone and speech gesture acts as a function of interlocutor and context

Context	Communicative act	Child directed		Adult directed
		Low Vocabulary	High Vocabulary	
		Mean % (SD)	Mean % (SD)	Mean % (SD)
Free Interaction	*Speech Alone*	84.5 (9.6)	82.5 (9.7)	42.9 (31.2)
	Speech-Gesture	15.5 (9.7)	17.5 (9.7)	57.1 (31.2)
Word description	*Speech Alone*	82.8 (11.3)	79.7 (12.6)	53.9 (28.3)
	Speech-Gesture	17.2 (11.3)	20.3 (12.6)	46.1 (28.3)
Picture description	*Speech Alone*	77.7 (10.3)	80.7 (10.9)	64.4 (23.7)
	Speech-Gesture	22.3 (10.3)	19.3 (10.9)	35.6 (234)

majority of child-directed communicative acts consisted of speech alone, irrespective of vocabulary group or context. The majority of adult-directed communicative acts were speech alone during structured tasks, in contrast with the free interaction where the majority of adult-directed communicative acts were speech-gesture acts (SGA). Maternal volubility was unaffected by children's vocabulary status (high versus low) (*t*s < 1), so for this analysis we collapsed data across vocabulary groups. (The lack of sensitivity of maternal volubility to the children's vocabulary status would offer tentative support to the IAH over the FSH.) Proportion of speech-gesture acts produced by mothers to children and adults are shown in Figure 1.

A 2x3 repeated measures ANOVA, with Interlocutor and Context as the independent variables and with the proportion of maternal gesture accounted for by SGA as the dependent variable[1], revealed a robust main effect of Interlocutor $F(1,28) = 42.7$, $p < .001$, a marginal main effect of Context, $F(2,27) = 3.2$, $p = .06$, and a significant Interlocutor X Context interaction $F(2,27) = 6.9$, $p = .004$. To examine this interaction further, we conducted two one-way ANOVAs to examine the effect of Context upon maternal production of SGA with each interlocutor separately. These revealed a robust effect of Context on adult-directed SGA ($F(2,27) = 6.1$, $p = .006$) and a non-significant effect of Context upon child-directed SGA ($F(2,27) = 1.5$, $p = .3$). Bonferroni adjusted pairwise comparisons of adult-directed SGA by context confirmed an increase in SGA during the free interaction versus the structured picture task ($p = .002$), and between the two structured tasks (word versus picture $p = .02$), but no differences during the free interaction versus the structured word task ($p \approx 1$).

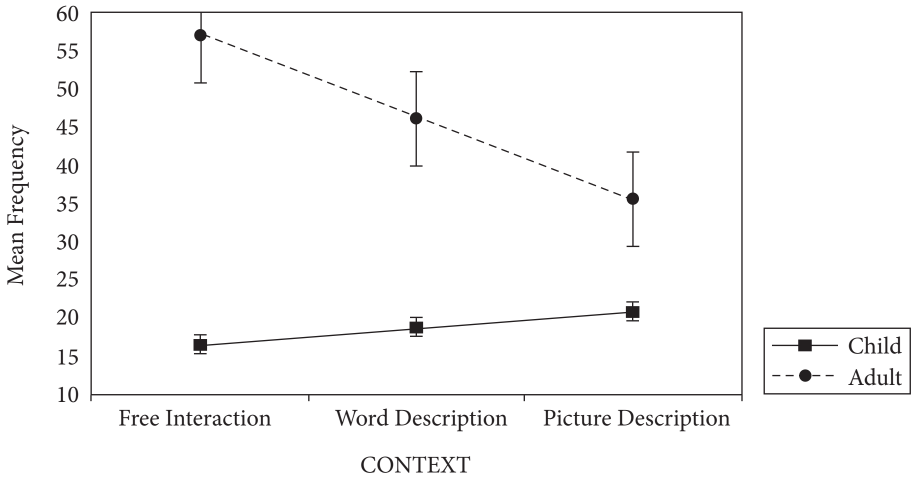


Figure 1. Maternal production of speech-gesture acts as a function of interlocutor and context (standard error bars shown).

1. Because proportional data is not independent, we analyse only maternal productions of speech-gesture acts.

Gesture types

Prior to examination of the rate per minute of maternal gesture production by type, we performed square root transformations to ensure the data were normally distributed. We next performed a series of independent t-tests to check whether children's vocabulary status (high versus low) significantly affected the mean frequency of maternal rate of production of each gesture type (Deictic, Emphatic and Representational) within each context. Table 3 presents the untransformed means and standard deviations of each gesture rate as a function of context and interlocutor, with child-directed communication further grouped by the median split on high versus low vocabulary scores. Maternal gesture production by type within context was again unaffected by children's vocabulary status (high versus low) (*t*s < 1), so we collapsed data across vocabulary groups, comparing child-directed versus adult-directed gesture. Again, the non-sensitivity of maternal volubility to the children's vocabulary status would offer tentative support to the IAH over the FSH.

Table 3. Transformed mean gesture rate by type, interlocutor and context

Gesture	Context	Interlocutor	Vocabulary	M (SD)	N
Deictic	Free interaction	Child	*Low*	2.2 (1.2)	8
			High	2.4 (1.1)	5
		Adult		1.6 (0.7)	13
	Word Description	Child	*Low*	2.9 (1.2)	8
			High	3.6 (1.6)	5
		Adult		1.9 (0.9)	13
	Picture Description	Child	*Low*	4.1 (0.9)	8
			High	3.7 (1.3)	5
		Adult		1.4 (0.4)	13
Representational	Free interaction	Child	*Low*	1.4 (0.5)	2
			High	1.3 (0.4)	3
		Adult		2.0 (1.2)	5
	Noun Description	Child	*Low*	3.6 (1.6)	2
			High	2.3 (1.1)	3
		Adult		2.7 (0.6)	5
	Picture Description	Child	*Low*	2.7 (0.4)	2
			High	1.5 (0.5)	3
		Adult		2.5 (0.3)	5
Emphatic	Free interaction	Child	*Low*	1.0 (n/a)	1
			High	1.8 (0.5)	7
		Adult		4.5 (1.4)	8
	Noun Description	Child	*Low*	2.6 (n/a)	1
			High	1.9 (0.6)	7
		Adult		4.5 (2.1)	8
	Picture Description	Child	*Low*	1.4 (n/a)	1

Deictic gestures

Figure 2 presents maternal deictic gesture production (rate per minute) as a function of interlocutor and context. Inspection of Figure 2 suggests there was more child-directed gesture during the two structured tasks than the free interaction. This contrasts with adult-directed deictic gesture, which appeared to vary little. In a 2x3 ANOVA on these data, there were significant main effects of Interlocutor ($F\ (1,28) = 135.4, p < .001$) and Context ($F\ (2,27) = 7.9, p = .02$) and a reliable Interlocutor X Context interaction ($F\ (2,27) = 11.1, p < .001$). This suggests that context affects maternal production of deictic gesture differently during interaction with children than adults. To examine this interaction, we conducted two one-way ANOVAs to examine the effect of Context upon maternal production of deictic gestures with each interlocutor separately. These revealed a robust effect of Context on child-directed deictic gestures ($F(2,27) = 14.3, p < .001$) and a non-significant effect of Context upon adult-directed deictic gestures ($F(2,27) = .308, p = .7$). Pair-wise comparisons confirmed increased child-directed deictic gesture production during the picture task versus the free interaction ($p < .001$), and during the picture versus the word task ($p < .001$), but no significant differences during the free interaction versus the word task ($p = .3$).

Representational gestures

Figure 3 presents maternal representational gesture production as a function of interlocutor and context. However, the low numbers of mothers who used representational gestures (see Table 3) make these data unsuitable for parametric analysis. We therefore

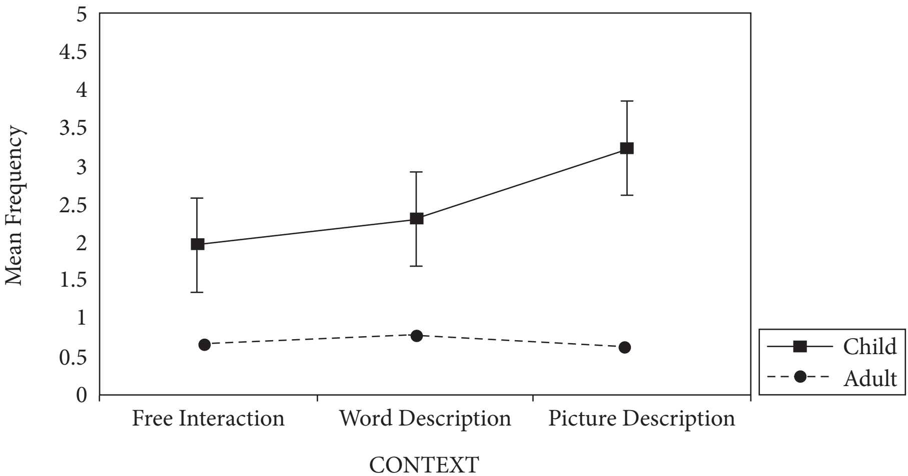


Figure 2. Maternal production of deictic gesture as a function of interlocutor and context (standard error bars shown).

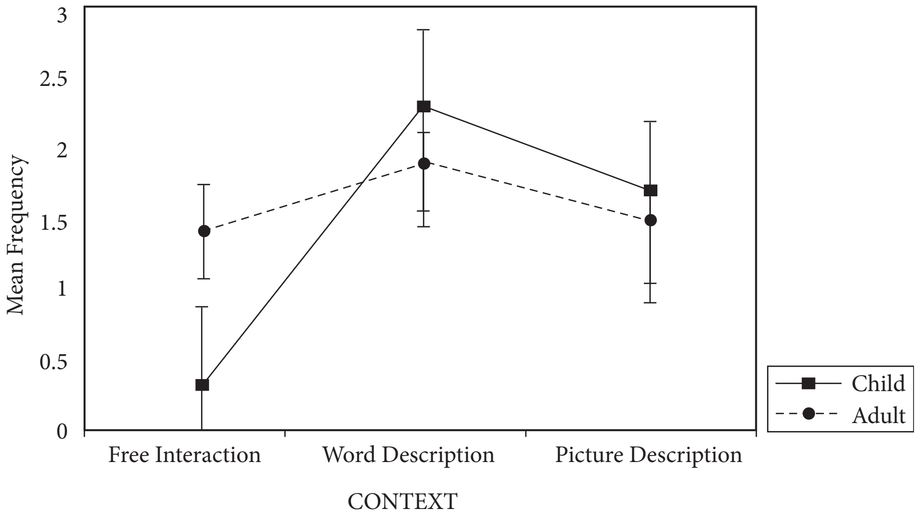


Figure 3. Maternal production of representational gesture as a function of interlocutor and context (standard error bars shown).

conducted two non-parametric Friedman's tests (one for each interlocutor) with maternal representational gesture production as the dependent variable and Context (three levels: free interaction; word task; picture task) as the independent variable. This revealed a significant main effect of Context on child-directed representational gestures χ^2 (2) = 17.1 $p < .001$, but not on adult-directed representational gestures χ^2 (2) = 3.2 $p = .20$, offering support for the IAH.

To further explore the effects of context on representational gesture production with children we performed three post-hoc Wilcoxon signed ranks tests, Bonferroni-corrected for multiple comparisons. During interaction with their child, mothers produced more representational gestures during the word description and picture description tasks than during free interaction (Z = –4.1, $p < .005$ in both cases), with a non- significant difference between the word and picture task (Z = 1.9, $p = .18$).

A series of Bonferroni-corrected Wilcoxon signed ranks tests enabled exploration of the effects of interlocutor on representational gesture production within each context. Mothers produced significantly fewer child-directed than adult-directed representational gestures during the free interaction Z = 3.4, $p = .003$. However, no significant differences in child-directed versus adult-directed representational gestures were observed during either of the two structured tasks: word Z = .79, n.s.; picture Z = 1.7, n.s.. These findings, in conjunction with examination of Figure 3, suggest that the effects of interlocutor and context were not independent, but rather interact together to influence maternal productions of representational gesture. In particular, it appears that the free interaction was the only context in which mothers modified their use of representational gesture, and they do so only for children.

Emphatic gestures

Figure 4 presents maternal emphatic gesture production as a function of interlocutor and context. Emphatic gesture production was not subject to analysis because of low numbers of mothers who used emphatic gestures (see Table 3), making the data unsuitable for parametric analysis. However, inspection of Figure 4 indicates child-directed emphatic gestures were both rare and relatively unaffected by context, while adult-directed emphatic gesture varied across contexts, being reduced in the presence of a visible referent.

Gesture and children's current communicative ability

Our final set of analyses focused on the relation between maternal speech and gesture and children's reported vocabulary status. Under the FSH maternal gesture is predicted to be sensitive to child's current vocabulary, while the IAH would predict that it is not. Table 4 presents correlations between maternal gesture across contexts and children's reported communicative status. Frequency of maternal Speech Alone significantly and positively correlated with children's productive vocabulary size (r = .42). Frequency of maternal speech-gesture acts (SGA) was not correlated with any measure of child communicative ability. However, a more fine-grained analysis reveals that (1) frequency of maternal emphatic gesture production was positively correlated with children's gestural repertoire (r = .39); (2) frequency of maternal deictic gesture production was positively correlated with the number of words children were reported to understand but not yet say (r = .41); (3) frequency of maternal representational gesture

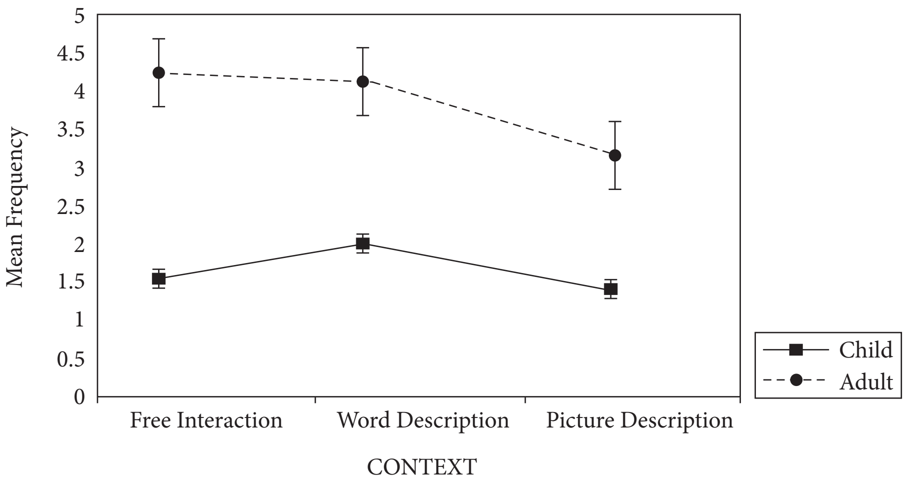


Figure 4. Maternal production of emphatic gesture as a function of interlocutor and context (standard error bars shown).

Table 4. Pearson's r correlations between maternal and child speech and gesture measures across tasks

	Mother				
	Communicative acts		Gesture types		
Child vocabulary measures	Speech alone	Speech & gesture	Emphatic	Deictic	Representational
Understood, not yet said	−.19	−.04	.22	.41**	−.01
Productive vocabulary	.42**	.02	.25	−.05	−.37*
Receptive vocabulary[?]	.34	.04	.22	.12	−.35***
Gestural repertoire	.29	−.04	.39**	.04	−.17
Speech-gesture combinations	.29	−.14	.25	−.23	−.28

Note *p = .05 **p < .05. ***p < .001

[?] Receptive vocabulary is the total number of words understood and produced by children

production was negatively correlated with children's productive vocabulary (r = −.37), and near significantly negatively correlated with children's receptive vocabulary (r = −.35).

Discussion

Modification of maternal communication

In line with previous research (Iverson et al. 1999, O'Neill et al. 2005), mothers in our study reliably produced fewer child-directed than adult-directed speech-gesture acts (SGA). Child-directed SGA accounted for around 20% of maternal communicative acts, contrasting with 50% for adult-directed communication. Adult-directed SGA, unlike child directed communication, was sensitive to context, with structured tasks having an inhibiting effect on adult-directed SGA. This restriction by mothers of SGAs to children compared with adults in the same context is consistent with the Facilitative Strategy hypothesis (FSH) and not with the Interactional Artefact hypothesis (IAH). There was little variation in the proportion of child-directed SGAs across contexts, which, in contrast, appears to offer support to IAH. However, the effect of context was different between child and adult interactions, providing further support for FSH over IAH (because the latter predicts that context should affect communication with both interlocutors equally).

There is a substantial body of literature suggesting that maternal modification of communication with children is a characteristic of the parent, rather than a response to the child's immediate linguistic level (Smolak & Weinraub 1983; Huttenlocher, Height, Bryk & Seltzer 1991; Cohen & Beckwith, 1976). Our findings are intriguing, because they suggest that the amount of communication produced by mothers was influenced by both the context and the conversational partner, albeit to different degrees. Our findings suggest SGAs are influenced by both interlocutor and context. However, interlocutor effects were stronger than context effects: the interlocutor effect was robust ($p < .001$), while the context effect was only marginally significant ($p = .06$). Child-directed SGAs did not vary significantly across tasks, contrasting with adult-directed SGAs, where free interaction > word description ≈ picture description. However, a more interesting picture emerges when gesture types are viewed separately.

Modification of gesture types

Consistent with previous research (Bekken 1989, Iverson et al. 1999, O'Neill et al. 2005, Shatz, 1982) mothers in our study, during interaction with children, employed context dependent, concrete, deictic gestures, particularly point gestures. They rarely used emphatic gestures with children. Mothers used a wider range of gestures during interaction with adults, thus supporting the notion that child-directed gesture is modified relative to adult directed gesture.

We have presented two opposing explanations for such modification. Under FSH such modification scaffolds language learning and is sensitive to context and children's vocabulary status. In contrast, IAH results from the semantic simplicity of child-directed communication and is unaffected by context. We found robust effects of context on deictic gesture along with a reliable Context X Interlocutor interaction, in which context had more effect on interaction with children than interaction with adults. Similarly, context reliably affected child-directed but not adult-directed representational gesture. Emphatic gestures appeared unaffected by context. This differential effect of context on interlocutor offers tentative support for the FSH, rather than the IAH.

We found no differences in the rate or type of gestures mothers produced as a function of the child's current vocabulary group (high versus low). The evidence discussed thus far appears to suggest that while mothers adjust the amount and type of gestures they produce according to the general age of the interlocutor (adult versus child), this adjustment is not closely tied to the child's linguistic ability, consistent with the findings of Smolak and Weinraub (1983).

Sensitivity of gesture to children's current linguistic level?

Maternal production of child directed gesture did not differ as a function of children's current vocabulary group (high versus low). However, the presence of correlations between maternal gestural production and child vocabulary measures would suggest that mothers may use aspects of the child's perceived communicative ability to direct their own communicative attempts. Examination of the correlational data permits more sensitive analysis than does a median split and reveals an apparent effect of child's vocabulary. Children's productive vocabulary was correlated strongly and positively with maternal production of speech alone acts (SAs), but negatively associated with representational gestures. Representational gestures were also negatively associated with children's receptive vocabulary. This indicates that mothers tended to produce fewer representational (descriptive) gestures when children had relatively large receptive vocabularies typically indicating children understood rather more words than they could say. Maternal production of speech and use of descriptive gesture rather than deictic or emphatic gesture related to the size of a child's vocabulary, highlighting the sensitivity of maternal communication to children's current linguistic ability. Furthermore, maternal use of deictic gesture associated strongly and positively with children's vocabulary of words understood but not yet produced. These findings need to be approached with caution, since they represent multiple correlations over the same data set; but they do appear to support FSH over IAH, inasmuch as there are clear relations between maternal communication and child vocabulary. Supporting this conclusion is a body of work indicating that maternal volubility predicts children's vocabulary (Furrow, Nelson, & Benedict 1979; Hoff & Naigles 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons 1991; Rosenthal-Robbins 2003).The links we found between maternal communication and child vocabulary suggest that both maternal production of speech and gesture and representational gesture production may promote vocabulary development in children as would be predicted under the FSH. However, we examined the relation of maternal volubility to children's concurrent productive vocabulary; thus, we are unable to state with certainty from these whether maternal communicative behavior was contingent on children's vocabulary or whether, alternatively, it facilitated vocabulary growth. A prospective longitudinal study goes some way to answering these questions, suggesting that maternal gestural behavior can indeed facilitate learning of individual words (Zammit & Schafer 2011).

Conclusion

This study was, to our knowledge, the first attempt to explore the influence of context on maternal gesture across three contexts varying in the degree of structure. There were several reasons to believe that the structure inherent to each instructional task would influence gesture production. The context effects observed confirm that mothers adjusted gesture according to both the demands of the situation, and – importantly – the needs of the interlocutor. Thus, there was some support for the Facilitative Strategy Hypothesis. The Interactional Artefact Hypothesis, on the other hand, predicts few differences between maternal gesture to children compared with adults during structured interactions, when semantic context is held constant between child and adult interlocutors. The variability of maternal interaction across interlocutors and the correlations between maternal communication and child vocabulary tend to negate this hypothesis.

References

Barrett, M., Harris, M. and Chasin, J. 1991. "Early lexical development and maternal speech: A comparison of children's initial and subsequent uses of words." *Journal of Child Language* 18 (1): 21–40.

Bekken, K. 1989. *Is There Motherese in Gesture?* Unpublished doctoral dissertation, University of Chicago, Chicago.

Brand, R. J., Baldwin, D. A. and Ashburn, L. A. 2002. "Evidence for 'motionese': modifications in mothers' infant- directed action." *Developmental Science* 5 (1): 72–83.

Brown, G. and Harris, T. 1978. *Social Origins of Depression*. London: Tavistock.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. and Pethick, S. J. 1994. *Variability in Early Communicative Development*. Monographs of the Society for Research in Child Development, Serial No. 242, Vol. 59, No. 5.

Feyereisen, P. and Havard, I. 1999. "Mental imagery and production of hand gestures while speaking in younger and older adults." *Journal of Nonverbal Behavior* 23 (2): 153–171.

Furrow, D., Nelson, K. and Benedict, H. 1979. "Mothers' speech to children and syntactic development: Some simple relationships." *Journal of Child Language* 6: 423–442.

Gutmann, A. J. and Turnure, J. E. 1979. "Mothers' production of hand gestures while communicating with their preschool children under various task conditions." *Developmental Psychology* 15 (2): 197–203.

Hamilton, A., Plunkett, K. and Schafer, G. 2000. "Infant vocabulary development assessed with a British communicative development inventory." *Journal of Child Language* 27 (3): 689–705.

Hampson, J. and Nelson, K. 1993. "The relation of maternal language to variation in rate and style of language-acquisition." *Journal of Child Language* 20 (2): 313–342.

Hoff, E. and Naigles, L. 2002. "How children use input to acquire a lexicon." *Child Development* 73 (2): 418–433.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. and Lyons, T. 1991. "Early vocabulary growth: Relation to language input and gender." *Developmental Psychology* 27 (2): 236–248.

Iverson, J. M., Capirci, O., Longobardi, E. and Caselli, M. C. 1999. "Gesturing in mother-child interactions." *Cognitive Development* 14 (1): 57–75.

Namy, L. L., Acredolo, L. and Goodwyn, S. (2000). "Verbal labels and gestural routines in parental communication with young children." *Journal of Nonverbal Behavior* 24 (2): 63–79.

O'Neill, M., Bard, K. A., Linnell, M. and Fluck, M. 2005. "Maternal gesture with 20-month-old infants in two contexts." *Developmental Science* 8 (4): 352–359.

Pine, J. 1994. "The language of primary caregivers." In *Input and Interaction in Language Acquisition*, C. R. Gallaway and B. J. Richards (eds), 15–37. Cambridge: Cambridge University Press.

Rosenthal-Robbins, P. 2003. "Caregivers' contingent comments to 9-month-old infants: Relationships with later language." *Applied Psycholinguistics* 24: 221–234.

Rowe, M., Pan, B. A. and Snow, C. 2003. "Predictors of maternal child-directed speech and gesture in low-income families." Paper presented at the *Biennial Conference of the Society for Research in Child Development*, Tampa, FL.

Schmidt, C. L. 1996. "Scrutinizing reference: How gesture and speech are coordinated in mother-child interaction." *Journal of Child Language* 23 (2): 279–305.

Shatz, M. 1982. "On mechanisms of language acquisition: Can features of the communicative environment account for development?" In *Language Acquisition: The State of the Art*, E. G. Wanner (ed), 102–127. New York: Cambridge University Press.

Snow, C. E. 1972. "Mothers' speech to children learning language." *Child Development* 43: 549–565.
Tan, S. H. and Schafer, G. 2005. "Toddlers' novel word learning: Effects of phonological representation, vocabulary size, and parents' ostensive behavior." *First Language* 25 (2): 131–155.
Zammit, M. and Schafer, G. 2011. "Maternal label and gesture use affects acquisition of specific object names." *Journal of Child Language* 38 (1): 201–221.

CHAPTER 11

The organization of children's pointing stroke endpoints

Mats Andrén
Lund University

The timing of index finger pointing gestures of three Swedish children (recorded longitudinally between 18 and 28 months) was analyzed. 63% of the pointing strokes ended in direct association with the child's own spoken utterance. This is in line with standard descriptions of gesture timing. However, 35% of the pointing strokes were sustained for a longer time – until a response was received from an interlocutor. It is shown here that parents give significantly more elaborated responses when children's pointing strokes are sustained and that the children work actively to achieve this result. The timing of such pointing gestures is thus a matter of interactive coordination between child and interlocutor. Finally, these findings are used as the basis for a discussion of different types of descriptions of gesture timing in the literature and how these may relate to each other.

1. Introduction

In the context of ethology, Hinde (1957: 118) stated that: "The mechanisms underlying behavior are diverse, and a given pattern of behavior may be brought to an end in many different ways. Nothing is gained by grouping all 'causes of endings' under one heading". In this paper, I will try to make very much the same point, but more specifically with respect to children's pointing stroke endpoints. Most research on gesture timing during the last 20 years has been devoted to the formulation of gesture- and speech-production models and models of 'thinking for speaking'. This line of research has yielded important insights into the ways in which gestures are usually coordinated with spoken utterances, especially with regard to the onset of strokes and mainly regarding iconic gestures (cf. Nobe 2000).[1] However, not all aspects of gesture timing are of this kind. The present study focuses on aspects of gesture timing whose logic is

1. Strictly speaking, Nobe (2000) uses the term 'representational gestures', as including 'iconics', 'metaphorics', and 'abstract deictics'.

primarily interactive; that is, it is continuously adjusted during the course of its performance with respect to the behavior of the Other. Also, this study concerns the endpoints of 18–28 month-old children's pointing strokes, rather than the onsets of strokes in the iconic gestures of adults, which have been much more studied.

At issue is how responses from interlocutors vary in relation to two types of pointing stroke durations: (a) strokes that end in direct association with the utterance in which the pointing stroke also started and (b) strokes that end some time after *another utterance* has been delivered (in most cases by the interlocutor). The two main hypotheses are (H1) that parents give more elaborated responses during children's sustained pointing gesture strokes and (H2) that the children fine tune this type of sustained strokes adaptively in relation to the responses they get from their parents: the less elaborated responses they receive, the more they tend to work for such a response.

Similar thoughts about the interactive functions of sustained pointing gestures have been expressed in the literature, based on observations of adults (Sidnell 2005, Clark 2005) and children with Down's syndrome (Wootton 1990). Bavelas (1994: 203, citing personal communication with Adam Kendon in 1988) writes: "when a gesture is held longer than would be needed simply to convey information, it becomes a *kinetically held question*, that is, a request for response from the addressee." The aim of the present study is to make a more systematic evaluation of these claims.

2. Method

2.1 Data

Recordings of three Swedish children from the Strömqvist-Richthoff corpus (Richthoff 2000) were used. Each child, two girls and one boy, was recorded at home at least once a month between the ages of 18 and 28 months as they were interacting with a parent. Throughout the recordings, the participants sat by a table, interacting side-by-side. Activities included book reading, eating, playing with toys, and general conversation.

All instances of index finger pointing performed by the children in the first five minutes of each recording were coded according to the categories described in the following two sections. A total of 393 instances of index finger pointing in the children were found.

2.2 Explicit exclusion of some instances from analysis

There is much variation in children's performance of pointing gestures. Not all of them are performed with an extended index finger and the other fingers curled. Furthermore, in some cases the stroke itself has a movement structure, giving it a kind of inherent temporal extension (in contrast to punctual strokes), and in other cases, the pointing gesture is part of a series of gestures rather than being used on its own. Such "additional" features of pointing gestures can be expected to affect their timing characteristics, but since the aim was to study variation in pointing only along one single dimension, namely differences in parental response when pointing gestures were sustained or not, it was decided to focus strictly on cases where the pointing gesture is as "pure" and prototypical as possible. To be sure, the different ways in which children perform pointing gestures is an interesting topic of systematic study in its own right (Andrén, in preparation), but accounting for the precise timing in relation to all of these features would necessarily involve a much more complex analysis. Therefore, explicit criteria were formulated of when to *exclude* instances of index finger pointing from analysis. Instances were excluded under the following circumstances:

a. The child was simultaneously holding an object in the hand that performed the pointing gesture (cf. Andrén 2010) (n = 7).
b. The pointing stroke also exhibited iconic features such as displaying form or motion (n = 9).
c. The pointing gesture was performed without speech (n = 8).
d. The pointing gesture seemed to be monologic (private) rather than directed to the interlocutor (n = 6).
e. The parent and the child started talking simultaneously (n = 21).
f. The pointing gesture was part of a series of gestures (n = 13).
g. The pointing gesture was very diffuse and did not seem to orient to a specific target (n = 8).
h. The gesture appeared to be combined with haptic exploration of a material being pointed to, using the index finger (n = 5).
i. The pointing gesture was affected by practical problems, such as the child pointing to a book that the parent was simultaneously moving, or the child and the parent collided physically during the action (n = 13).

In all of these cases, the timing characteristics are potentially at least slightly different in terms of the child's coordination of the gesture with his/her own utterance and/or in terms of interactive coordination. After the exclusion of these instances, 303 of the 393 instances remained.

2.3 Coding of the data

The concept of a *stroke endpoint* is used in this paper to refer to the moment in time when a pointing gesture to a certain target eventually turns into a full retraction or into a preparation phase of a new gesture. In cases where there were, for example, repeated tapping of the target during the performance of the pointing gesture, possibly with some pause between some of the taps, this was considered to be a single stroke, rather than being a series of strokes, since it is part of the same overall pointing to a single target. For each index finger pointing gesture, three *types of stroke endpoints* were distinguished:

1. *short*: Endpoint in direct association with the child's own utterance – at the very end of, or during, the utterance.
2. *between turns*: Endpoint shortly after the utterance was finished, but still before the first transition relevance place (TRP) in the next utterance. The first TRP in an utterance corresponds to the first point in the utterance where the turn-so-far may be perceived as a complete turn, although the utterance need not necessarily end at this point (Sacks et al. 1974). The turn-so-far can be a full grammatical clause, but also, for example, a response morpheme.
3. *sustained*: Endpoint after the first TRP in the next utterance, produced either by the parent or the child.

Sustained strokes were considered single gestures even if they exhibited features of *stroke renewal*, such as tapping the target once more after a hold, as long as there was no proper retraction of the pointing gesture in between. This is why the term sustained gestures is used here, in contrast to the more technical and narrow sense of a stroke hold, which would not be said to last across such renewals of a stroke. It should also be noted that the same sustained pointing gesture was sometimes sustained over several further utterances from both of the participants.

Second, all parental responses to the children's utterances that included pointing gestures (sustained or not) were coded for *degrees of response*, arranged on an ordinal scale, ranging from least responsive to most responsive. The categories were defined relative to what the child was talking about and pointing to, and they were defined in the following way:

1. *no response* – Neither responding nor initiating features: when parents did not say anything, or did not perform a certain act if an act was requested, or when the parent indeed did say something, but initiated a new sequence rather than replying to the child's utterance.
2. *minimal response* – Responding, but no initiating features: when the response only contained short response morphemes such as "yes", "mm", "no" or simple repetitions of what the child said (Child: "a ball", Parent: "yes, a ball") mainly serving as an acknowledgement of what the child said, rather than adding any new content.
3. *expanded response* – Both responding and initiating features: when there were not only aspects of acknowledgement in the reply, but also initiating aspects, saying something more or new about what the child was pointing to (Child: "there", Parent: "I wouldn't wanna taste that") or when the parent performs an act in compliance with a child's request for such an act.

This is, of course, a rather harsh simplification of the intricacies of interactive coordination of response and initiative in dialogue. For example, some utterances, such as *wh*-questions, are more response demanding than others. Nevertheless, it was judged sufficiently detailed to provide a foundation for testing the hypotheses. For a considerably richer treatment of initiative and response in dialogue, see Linell and Gustavsson (1987).

Third, in cases where the children produced further utterances during a *sustained* stroke, often with parental responses "inserted" in between, all these subsequent child utterances were coded using a distinction between three levels of *communicative effort*. This distinction was intended to capture the type of effort the child puts into the gesture stroke itself as well as other means of drawing attention to the multimodal utterance as a whole, such as using a stronger voice than in the child's previous utterance. The three levels are defined as:

1. *plain hold*: a continuous hold of the gesture with a new spoken utterance which was essentially a repetition of the previous utterance at the same level of intensity.
2. *renewed stroke*: a renewal of the stroke such as tapping the target again, but without a full retraction, repeating a similar utterance at the same level of intensity as in the previous utterance.
3. *upgraded renewal*: not just renewing the gesture and repeating the previous utterance as in the previous category, but also adding intensifiers such as performing the gesture in a more intense and salient way, using a stronger voice, turning to the parent to establish mutual gaze, or providing a markedly more elaborated version of the previous utterance.

3. Analysis

3.1 The existence of two main types of timing

Most of the analyzed pointing stroke endpoints were of the *short* type (63%), ending in direct association with the child's own utterance in line with standard descriptions of gesture timing. However, there was also a substantial amount of stroke endpoints of the *sustained* type (35%), where the stroke was sustained until at least one more utterance has been delivered (90% of those utterances come from a parent). The *between turns* type of endpoints was very rare in comparison (2%), which means that the children used two almost categorically distinct ways of placing the endpoint of the pointing stroke either within their own utterance or after a next utterance has been delivered, but seldom in between. Moreover, 5 of the 7 instances of the *between turns* type had a stroke where the target was tapped repeatedly with the index finger. This implies that the *between turns* type may often occur in strokes that have an inherent temporal extension due to having a complex movement structure (such as tapping repetition), which may compete to some extent with detailed coordination of gesture and speech. The rest of the analysis concerns only the two most common types of stroke endpoints (*short* and *sustained*). Even though development was not the focus of this study, it may be pointed out that the relative frequencies of *short* strokes and *sustained* pointing strokes for the group as a whole remained constant over the investigated period (correcting for an overall reduced tendency for pointing from around 23 months and onwards). The three children were fairly similar to each other, with one child performing slightly more *sustained* strokes than the other two in the analyzed data.

3.2 Parental responses in relation to short/sustained strokes

According to the first hypothesis (H1), it was expected that *sustained* pointing strokes should be associated with more elaborated responses from the parents than *short* strokes, and vice versa. A Pearson Chi-square test confirmed this hypothesis: χ^2 (df = 2, n = 296) = 22.34 ($p < 0.01$). Raw frequencies are presented in Figure 1. In cases of *sustained* strokes, *expanded response* was significantly more common, whereas *no response* and *minimal response* were significantly less common, compared to when strokes were of the *short* type. When strokes were *short*, the pattern was the opposite (also significant). In sum, children received significantly more response in cases where the stroke was sustained into further utterances. There is only one instance in the data where a child abandons a sustained pointing gesture apparently without having received any response.

It should also be remembered that many means are available for eliciting responses in communication. The claim here is certainly not that the duration of pointing strokes is the primary means for eliciting responses. *Expanded response* was the most common response type in both conditions (*short* and *sustained* stroke, see Figure 1),

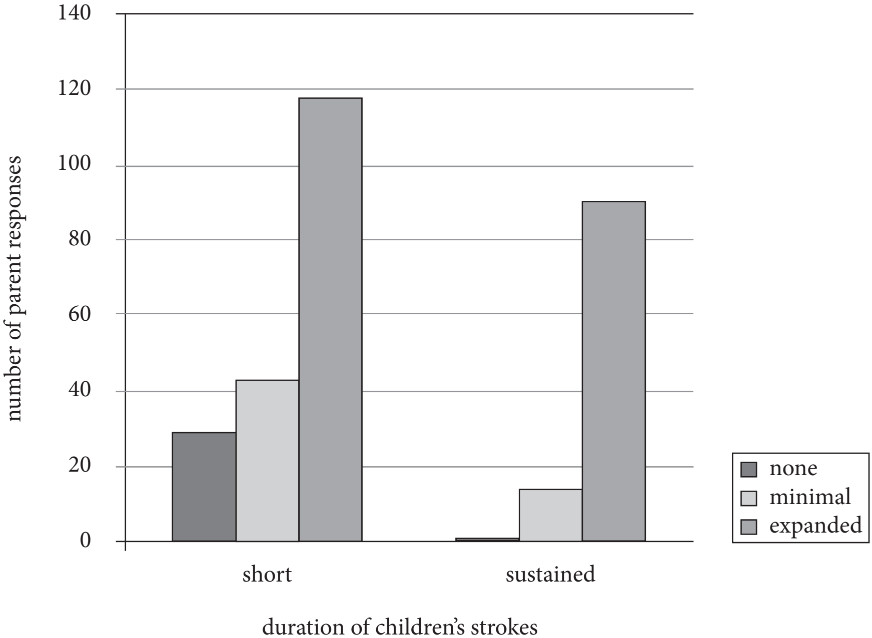


Figure 1. Stroke endpoints and types of response.

and the crucial finding was that parents gave *even more* response when children performed sustained pointing strokes than they did in the context of short strokes.

3.3 The internal dynamics of sustained pointing sequences

In 69% of the cases where a stroke was *sustained*, the stroke ended directly after an adjacent response given by a parent. This, in itself, is evidence in favor of the second hypothesis (H2), that getting a (satisfactory) response from the parent is indeed the primary stopping condition of the sustained strokes. In 31% of the cases even more utterances were exchanged, including a few instances where the stroke was sustained although an *expanded response* had in fact been received. All utterances in such exchanges from both participants were always about the referent being pointed to. It could be argued that the general principle of gestures being coordinated with co-expressive speech remains true here too, even across turns and partly even across speakers.

To further test the second hypothesis, children's reactions to their parents' responses were investigated in cases where the parent responded during a sustained child pointing gesture. More specifically, if the child kept the pointing gesture for at least one more of his/her own utterances after a parent's response, the nature of the child's pointing gesture as it was performed with the new utterance was coded according to communicative effort (*plain hold*, *renewed stroke*, or *upgraded renewal*). There was a strong negative relation (Spearman Rank = –0.70, $p < 0.01$) between communicative effort and the degree of response from the parent, which is strong evidence in favor of the second hypothesis. The less response the children received to their sustained pointing gestures, the more communicative effort they mobilized in the next turn, and vice versa.

4. Discussion and conclusions

According to the findings of the present study, parents gave significantly more response when children performed *sustained* index finger pointing gestures. Furthermore, the children were shown to orient to the content of these parental responses. This was visible in two ways. First, in most cases the children immediately withdrew their sustained pointing gestures when a parental response was given, and second, in cases where there were no such immediate withdrawal, there was a significant inverse relationship between the degree of response from the parent and the communicative effort invested by the child in the child's subsequent utterance(s). That is, the less response a parent gave to a child utterance with a sustained pointing gesture, the more likely the children were to upgrade their "demand" for a response through various sorts of intensifying resources. In short, the children did not perform such sustained pointing gestures at random. They seemed instead to be part of the children's established and typified repertoires of methods for eliciting parental responses. The fact that the children were less satisfied with minimal responses than with extended responses is interesting since it means that the "goal" of their pointing was not only to achieve intersubjectively shared reference to certain targets (joint attention), in which case simple acknowledgements would have been sufficient. The children rather seemed to aim for receiving various sorts of evaluations and comments on the target being pointed to, i.e. an active form of *social referencing*. It remains unclear when the use of these different types of pointing stroke endpoints emerge, since they are already present in the first observations here, where the children are 18 months old. As pointed out in the analysis, the relative frequencies of *short* and *sustained* strokes remained constant, for the group as a whole, during the period studied here (correcting for the fact that there was an overall decline of pointing gestures around 23 months).

Since sustained pointing gestures and the responses from the parents are happening simultaneously, it may not be appropriate to interpret one as cause and the other as effect. As was shown in this study, both parent and child orient to such sustained pointing gestures actively and mutually. This can be seen as a demonstration of the utility of multimodal resources in dialogue since gesture and speech do not interfere with each other in the way that simultaneous speech does. As mentioned in the introduction, adults sometimes make use of sustained gestures too for the very same general purposes (and in several different cultures). However, it would be an overgeneralization to conclude that all kinds of sustained gestures are to be interpreted as requests for response under all circumstances. For example, at the final moment of a theatre act actors commonly freeze and sustain postures and gestures, but in the turn-taking system used in this setting the audience is supposed to wait until *after* a sustained gesture is retracted before providing a response in the form of applauses (Broth 2002). In conversational interaction, though, gestures sustained in this way may well serve similar functions across most contexts. A common denominator between the theatre example and the conversational situations studied here seems to be that a sustained gesture is markedly "in play", whatever that implies in a given context.

Regarding the general theoretical issue of gesture timing, it is interesting to note that different researchers use different terms to talk about such phenomena. In interactionally oriented research gesture timing is often described as a kind of skillful *achievement* or in terms of *recipient design* (Kendon 2004, Goodwin 2000), whereas psychologically oriented research tend to use the term *synchrony* (McNeill 2005). Although interactional and psychological interests need not be mutually exclusive, it is clear that these differences in terminology highlight very different aspects of gesture timing. The term synchrony tends to evoke descriptions of gesture timing primarily as a matter of neural mechanisms in relation to utterance formulation, whereas the terms achievement and recipient design tend to highlight continuous processes of mutual orientation that are not understandable without reference to two or more co-present bodies, and their contextual embedding in a shared field of activity. Whereas psychological research on gesture timing has mainly focused on onsets of strokes (and preparations) as indications of mental activity such as "motor planning", the interactional perspective seems to be the one that has recognized the importance of stroke endpoints as orderly, visibly recognizable and interactionally consequential (Wootton 1990, Sidnell 2005, Clark 2005, Bavelas 1994, Goodwin 2000).

A crucial difference seems to lie in what is considered the 'starting point' of behavior (explicitly and/or implicitly). Does action originate in the mind, or in situations? Obvsiously, the question is ill-formed, since this is not best understood as an either-or question. Still, the emphasis of acts – especially communicative acts – as creations of the mind in psychological research has a tendency to downplay how action is systematically sensitive and responsive to the contingencies of social situations. Interactional accounts of action are often more explicitly geared at treating action in a way that relates it to the situation and previous acts; that is, treating action as a kind of ongoing balance act between initiating and responsive/responding aspects of action. In line with this description, Linell (1998: 211) writes about a change of emphasis from intentionality to responsibility within dialogical frameworks. In this vein, I would like to argue that it is important to think about agency of action in a way that pays proper attention *both* to its initiating and responsive qualities. It seems clear that the sustained pointing gestures described in this paper are not delivered as readymade wholes, as they also exhibit features of responsiveness with respect to the behavior of the Other during their very performance. To be clear, the point of this argument is certainly not to claim that one or the other approach to language and communication is 'wrong' *per se* or that there are approaches that do not have limitations. It is also questionable to push talk about perspectives too far as if they were uniform. The point is rather that the perspectives tend to *complement* each other due to their relatively systematic highlighting of different aspects – in this case regarding gesture timing. Still, I would like to conclude by suggesting that there exists no moment in time such that the process of communication is entirely your own (cf. Schutz 1962).

5. Acknowledgements

Thanks to Gale Stam, Jordan Zlatev, Simon Harrison and Ulrika Nettelbladt for valuable comments on previous drafts of this paper. I would also like to acknowledge the Centre for Cognitive Semiotics (CCS) at Lund University for providing a productive interdisciplinary environment for research on the relations between different semiotic resources.

References

Andrén, M. 2010 *Children's Gestures from 10 to 30 months*. Ph.D. Thesis. Lund University, Centre for Languages and Literature.

Andrén, M. in preparation. Forms of pointing in young children.
Broth, M. 2002. *Agents Secrets. Le Public dans la Construction Interactive de la Représentation Théâtrale*. PhD thesis, Uppsala University.
Bavelas, J. B. 1994. "Gestures as Part of Speech: Methodological Implications." *Research on Language and Social Interaction* 27 (3): 201–221.
Clark, H. H. 2005. "Coordinating with each other in a material world." *Discourse Studies* 7 (4–5): 507–525.
Goodwin, C. 2000. "Action and embodiment within situated human interaction." *Journal of Pragmatics* 32 (10): 1489–1522.
Hinde, R. 1957. "Consequences and goals: Some issues raised by dr. A. Kortlandt's paper. on 'aspects and prospects of the concept of instinct'." *British Journal of Animal Behaviour* 5 (4): 116–118.
Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
Linell, P. and Gustavsson, L. 1987. *Initiativ Och Respons: Om Dialogens Dynamik, Dominans Och Koherens.* Linköping University: Department of Communication Studies.
Linell, P. 1998. *Approaching Dialogue*. Amsterdam: John Benjamins.
McNeill, D. 2005. *Gesture and Thought*. Chicago: Chicago University Press.
Nobe, S. 2000. "Where do most spontaneous representational gestures actually occur with respect to speech?" In *Language and Gesture*, D. McNeill (ed), 186–198. Chicago: Chicago University Press.
Richthoff, U. 2000. *En Svensk Barnspråkskorpus: Uppbyggnad Och Analyser.* PhD thesis. Göteborg University.
Sacks, H., Schegloff, E. A. and Jefferson, G. 1974. "A simplest systematics for the organization of turn-taking for conversation." *Language* 50 (4): 696–735.
Schutz, A. 1962. *Collected Papers I: The Problem of Social Reality*. The Hague: Martinus Nijhoff.
Sidnell, J. 2005. "Gesture in the pursuit and display of recognition: A Caribbean case study." *Semiotica* 156 (1): 55–87.
Wootton, A. J. 1990. "Pointing and interaction initiation: The behaviour of young children with Down's syndrome when looking at books." *Journal of Child Language* 17 (3): 565–589.

CHAPTER 12

Is there an iconic gesture spurt at 26 months?

Şeyda Özçalışkan[1] and Susan Goldin-Meadow[2]
Georgia State University[1] and University of Chicago[2]

Previous research has shown that children understand the iconicity of a gesture at 26 months. Here we ask when children begin to display an appreciation of iconicity in the gestures they produce. We observed spontaneous gesture in 40 children interacting with their parents from 14 to 34 months of age and found that children increased their production of iconic gestures over time. At 26 months, they not only produced significantly more iconic gestures (tokens) than at any previous time point, but they also conveyed significantly more different meanings with those iconic gestures (types). We found similar increases in the iconic gestures that the children's parents produced, suggesting that parents either were sensitive to changes in their children's iconic gestures or perhaps were responsible for those changes. Overall, the results suggest that the 26-month age period is a turning point for children's grasp of the iconicity of a symbol.

1. Introduction

One of the milestones of early language learning is mastering the ability to map a symbol onto a referent. Iconicity (the resemblance between a symbol and a referent) could play an important role in this mapping process. Previous research has shown that, although children can associate iconic gestures with objects at 18 months, it is not until 26 months of age that they truly understand the iconic relation between gesture and object (Namy & Waxman 1998; Namy 2001; Namy, Campbell, & Tomasello 2004), a step that may be important in understanding symbols. We ask here whether this sensitivity to iconicity can be found in children's production, as well as their comprehension, of gestures. In other words, do children increase their spontaneous production of iconic gestures during the period when they have been shown to increase their understanding of iconic gestures? And if so, can this productive surge be traced back to the gestural input children receive from their parents? More specifically, we ask whether the gestures that parents produce have the potential to play a role in children's production of iconic gestures.

2. Children's early iconic gesture production and comprehension

Young children rely on gesture to communicate before they produce their first words (Bates 1976, Bates, Benigni, Bretherton, Camaioni & Volterra 1979; Greenfield & Smith 1976). Children's earliest gestures, produced around 10 months, are *deictic* gestures, gestures whose referential meaning is given entirely by the context and not by the form of the gesture; for example pointing at a bottle to indicate a BOTTLE (Bates 1976). While deictic gestures are the most commonly used gesture type at the early ages, other types of gestures, most notably *iconic* gestures, can also be found in children's early gesture repertoires. Children use iconic gestures to convey actions or attributes associated with an object; for example flapping arms to depict a bird FLYING or holding cupped hands in the air to depict the ROUNDNESS of a ball (Acredolo & Goodwyn 1985). There is, in fact, evidence that very young children can produce a range of iconic gestures – known as *baby signs* – that indicate actions or attributes associated with objects when those gestures are deliberately taught to them by their parents (touching their index fingers to their thumbs and rotating them to convey a spider CRAWLING, raising and extending arms to indicate BIG; Acredolo & Goodwyn 1985, 1988, 2002; Goodwyn, Acredolo & Brown 2000).

Although there has been ample research on the iconic gestures explicitly taught to children, we know much less about the iconic gestures that children spontaneously produce on their own. Previous work suggests that the incidence of spontaneous iconic gestures is rare, accounting for roughly 1 to 5% of the gestures that young children produce (Iverson, Capirci & Caselli 1994; Nicoladis, Mayberry & Genesee 1999; Özçalışkan & Goldin-Meadow 2005a, 2009).[1] But why are iconic gestures so infrequent in children's early gesture repertoires? One explanation could be that iconic gestures are conceptually harder than deictic gestures, as they convey relational information rather than merely pointing out objects and people in the world (Özçalışkan, Gentner & Goldin-Meadow 2011). Indeed, there is evidence that, early in development, children find it difficult to grasp the relation between an iconic gesture and its intended referent (Namy & Waxman 1998, Namy 2001, Namy et al. 2004). At 18 months, children are equally likely to associate an iconic gesture (e.g., hopping two fingers up and down to represent the rabbit's ears as it hops) or an arbitrary gesture (holding a hand shaped in an arbitrary configuration to represent a rabbit) with an object, thus displaying a lack of sensitivity to iconicity. But by 26 months, they are more likely to associate an iconic gesture than an arbitrary gesture with an object (Namy et al. 2004). These findings suggest that, prior to 26 months, children merely associate gestures with objects and thus do not treat gestures as symbols. At 26 months,

1. Interestingly, iconic gestures constitute a relatively large proportion of the gestures produced in adult-adult interactions, accounting for roughly 30% of the gestures produced (McNeill 1992).

however, they begin to discover the iconic relation between gesture and object, which may herald important changes in their understanding of symbols.

3. Do comprehension and production of iconic gestures go hand-in-hand?

Children's sensitivity to iconicity, as measured by gesture comprehension, thus appears to be a late emerging skill, beginning at approximately 26 months. The question we ask here is whether we see evidence of iconicity in the gestures that children produce at around the same time. It is possible that children undergo a shift at 26 months, not only in their *understanding* of iconicity in gesture, but also in their *production* of iconicity in gesture. If so, we should see marked increases in both the number and the diversity of iconic gestures that children produce in their everyday interactions at around 26 months.

We investigated this possibility in a sample of 40 typically developing American children (22 girls, 18 boys), who were being raised as monolingual English speakers. The children were videotaped for 90 minutes in their homes every four months from 14 to 34 months of age while interacting with their parents in their everyday routines (see Özçalışkan & Goldin-Meadow 2005b for details on the sample).

We first looked at changes in children's overall use of gesture over the six observation sessions, from 14 to 34 months. During this time, children increased their gesture production, $F(5, 170) = 10.82$, $p < .001$, from an average of 54 (SD = 36) gestures at 14 months to 122 (SD = 112) gestures at 34 months, with production peaking at 26 months, M = 131 (SD = 90); row 1, Table 1). During this time, children produced three different types of gestures: Deictic gestures (points or hold-ups) were used to indicate objects, people or places. Conventional gestures were forms prescribed by the culture to convey particular meanings (e.g., nodding the head to mean YES, shaking the head sideways to mean NO). Iconic gestures were spontaneously generated forms used to convey actions or attribute meanings (e.g., moving the fist forcefully in air to indicate HIT, holding the hand above head to indicate TALL). The children produced these three types of gestures at different rates, $F(2, 78) = 115.71$, $p < .001$, and iconic gestures were produced significantly less often than either conventional or deictic gestures, p's < .001, *Bonferroni* (Figure 1A).

Nonetheless, as can be seen in Table 1, children increased production of iconic gestures over time, $F(5,170) = 6.56$, $p < .001$ (Figure 2A, solid line). The children also increased production of deictic gestures, $F(5,170) = 11.63$, $p < .001$, but their production of conventional gestures remained flat during this time period, $F(5,170) = 0.6$, *ns* (Table 1, Rows 2, 3). Importantly, abrupt increases were found at different moments for iconic vs. deictic gestures: at 26 months for iconic gestures, $p < .05$, *LSD*; at 18 months for deictic gestures, $p < .001$, *Bonferroni*. The number of children who produced iconic gestures also increased over time. There were only 3 children (8%) who produced iconic gestures at 14 months, but this number increased to 70% (28/40) at

Table 1. Summary of children's and parents' gesture production by child age[a]

	14-months	18-months	22-months	26-months	30-months	34-months
Children						
Mean number of gesture tokens (SD)	54 (36)	91 (64)	119 (75)	131 (90)	123 (68)	112 (65)
Mean number of deictic gestures (SD)	32 (24)	67 (57)	93 (65)	102 (82)	92 (51)	85 (51)
Mean number of conventional gestures (SD)	22 (20)	24 (19)	23 (20)	26 (23)	24 (21)	23 (24)
Mean number of iconic gestures (SD)	<1 (2)	1 (1)	1 (2)	4 (7)	4 (7)	4 (4)
Parents						
Mean number of gesture tokens (SD)	102 (82)	95 (67)	105 (90)	123 (117)	119 (102)	97 (94)
Mean number of deictic gestures (SD)	72 (68)	67 (53)	74 (75)	81 (81)	82 (75)	64 (61)
Mean number of conventional gestures (SD)	28 (21)	26 (20)	28 (26)	38 (40)	32 (38)	28 (36)
Mean number of iconic gestures (SD)	2 (5)	2 (3)	2 (3)	5 (6)	5 (8)	5 (7)

a. SD = standard deviation; the numbers are rounded up to the closest whole number. Each child-parent dyad was observed for approximately 90 minutes at each observation session.

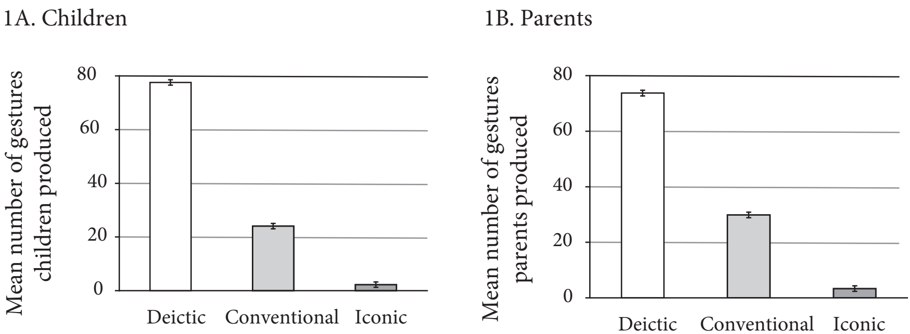


Figure 1. Mean number of deictic (white bars), conventional (gray bars) and iconic gestures (black bars) produced by children (Panel A) and their parents (Panel B) across child ages 14 to 34 months.

26 months and, by 34 months, 98% (39/40) of the children in our sample had produced at least one instance of an iconic gesture in their communications.[2] The pattern was the same for the *types* of meanings children conveyed in their iconic gestures. Children conveyed more different meanings in their iconic gestures with increasing age, $F(5,170) = 9.32$, $p < .001$, with a significant increase again at 26 months of age, $p < .05$, *LSD* (Figure 2A, dotted line).[3]

The children in our study thus displayed a reliable increase in the number and types of iconic gestures they spontaneously produced at 26 months, the age at which children have been shown to be sensitive to iconicity in their comprehension of gesture. As such, 26 months might be a turning point in terms of children's grasp of the iconicity of a symbol.

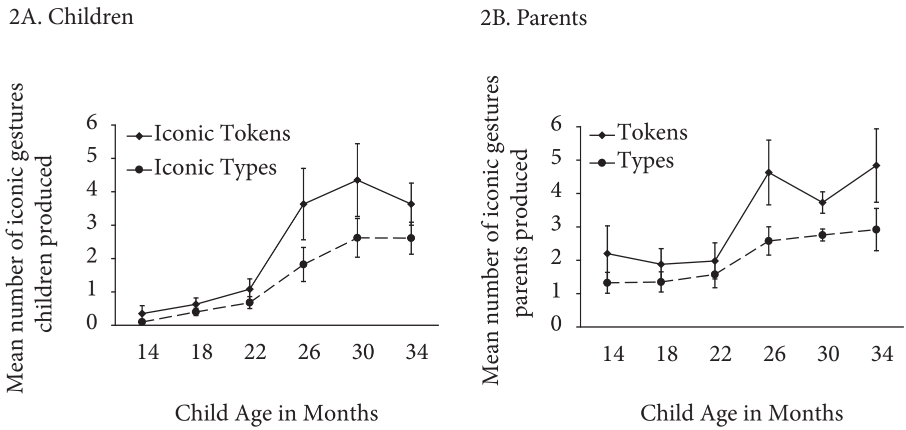


Figure 2. Mean number of iconic gesture tokens (solid lines) and iconic gesture types (dotted lines) produced by children (Panel A) and their parents (Panel B).

4. How do parents use iconic gestures when they talk to their children?

Our analysis shows a steep increase in the token and type frequencies of iconic gestures children produce at 26 months. But why do children show such an increase in their use of iconic gestures around this age? One possibility is that changes in children's use of iconic gestures reflect cognitive or communicative changes within the child. An

2. In contrast to the few children (N = 3) who produced iconic gestures at 14 months, all 40 children produced both *deictic* and *conventional* gestures at 14 months.

3. We used LSD rather than Bonferroni as a posthoc statistic because the change between 22 and 26 months in iconic gesture production (both tokens and types) was a planned comparison, and therefore did not require the Bonferroni correction for unplanned multiple comparisons.

alternative possibility is that changes in children's gestures reflect changes in the gestures that their parents use with them (cf. Rowe & Goldin-Meadow 2009).

At early ages children in the United States spend much of their time with adults, typically their parents. And parents gesture frequently when they speak to their children (Özçalışkan & Goldin-Meadow 2005a; Rowe, Özçalışkan & Goldin-Meadow 2008). There is, moreover, evidence that, as early as 12 months of age, typically developing children understand the deictic gestures that others produce. For example, one-year-old children can easily follow an adult's pointing gesture to a target object (Butterworth & Grover 1988). There is also evidence that parents modify their gestures to accommodate the communicative needs of their young children (Iverson et al. 1999, Özçalışkan & Goldin-Meadow 2005a). For example, parents produce fewer and simpler gestures (e.g., points at concrete objects) when they address young children as opposed to an adult (Bekken 1989). In addition to deictic gestures, parents also provide models for iconic gestures. In fact, the gestures that adults produce during interactive routines with their children can be the basis for the iconic gestures that children first produce (Acredolo & Goodwyn 1985). For example, Goodwyn and colleagues (2000) found that when parents were instructed to use iconic gestures along with their words, their children developed a larger repertoire of iconic gestures (see also LeBarton & Goldin-Meadow 2010).

To explore the effect that parent gesture might have on the child's production of iconic gestures, we analyzed the gestures produced by the parents of the 40 children in our sample during the same time period. The parents represented a heterogeneous mix in terms of both ethnic background and family income. The mother was the primary caregiver in 35 of the 40 families; the father was the primary caregiver in two families; both parents shared the caregiver role in another three families.

We first looked at parents' overall use of gesture over the six observation sessions, from child age 14 to 34 months. Like the children, parents gestured frequently in their interactions with their children. However, unlike the children, the rate at which parents gestured remained essentially unchanged across the entire observation period, $F(5, 170) = 1.12$, *ns*). As can be seen in Table 1 (row 5), parents produced gestures at average rates that ranged between $M = 97$ ($SD = 94$) and $M = 123$ ($SD = 117$) across the six observation sessions, with a peak in production at the 26-month period. At 14 months, children produced significantly fewer gestures than their parents, $M_{child} = 53.83$ ($SD = 36.12$) vs. $M_{parent} = 101.85$ ($SD = 81.63$), $t(39) = 3.74$, $p < .001$. However, by 18 months, the children had caught up, $M_{child} = 91.27$ ($SD = 64.46$) vs. $M_{parent} = 94.97$ ($SD = 67.19$), $t(39) = 0.32$, *ns*, and gestured as frequently as their parents throughout the remainder of the observations, with no reliable differences between them.

Parents also used the same three types of gestures as their children used (*deictic, conventional, iconic*) and, like their children, produced them at significantly different rates, $F(2, 78) = 78.68$, $p < .001$ (Figure 1B). Like the children, parents produced iconic gestures significantly less often than either conventional or deictic gestures, p's $< .001$, *Bonferroni*. Indeed, parents' overall use of each gesture type was almost identical to their children's use, with no reliable differences for either deictic, $t(39) = .57$, *ns*, conventional, $t(39) = 1.58$, *ns*, or iconic, $t(39) = 1.77$, *ns*, gestures.

Parents showed no change in their production of either *deictic*, $F(5,170) = 0.68$, *ns*, or *conventional*, $F(5,170) = 0.18$, *ns*, gestures over the six observation sessions (see Table 1). Interestingly, however, they increased their production of iconic gestures over time, $F(5,170) = 2.70$, $p = .02$, with a significant increase at the 26-month observation session, *LSD*, $p = .01$ (Figure 2B, solid line), thus mirroring the pattern we observed in the children. Parents were also similar to the children with respect to change in types of iconic gestures. They conveyed more different meanings in their iconic gestures with increasing child age, $F(5,170) = 9.32$, $p < .001$, with a marginally significant increase at 26 months, *LSD*, $p = .09$ (Figure 2B, dotted line).

One important difference between parents and children was that, unlike their children, the parents produced iconic gestures at the first observation session: 50% (20/40) of parents produced at least one iconic gesture at the 14-month session, in contrast to only 8% (3/40) of children. Parents did, however, increase their use of iconic gestures over time, from a mean number of 2.20 (SD = 5.22) gestures at 14 months to 4.63 (SD = 5.97) at 26 months; by 26 months all but two of the 40 parents (98%) had produced at least one instance of an iconic gesture in their communications with their children, compared to 28 (70%) of the 40 children. Moreover, there was a positive correlation between parents and children for both mean number of iconic tokens, Spearman's rho = .36, $p < .05$, and mean number of iconic types, Spearman's rho = .30, $p = .06$, across the six observation sessions.

5. Types of meanings conveyed in child and parent iconic gestures

Children and parents produced approximately the same number of iconic gestures, and they displayed a marked increase at the 26-month age period in both tokens and types of iconic gestures. To further explore the relation between the iconic gestures produced by parent and child, we asked whether the particular meanings that the children conveyed in their iconic gestures overlapped with the meanings that their parents conveyed. To do so, we characterized the meaning of the iconic gestures in three different ways.

In the first analysis, we categorized iconic gestures according to whether the form of the gesture depicted an *action* associated with an object (e.g., flapping arms to convey FLYING) or a perceptual *attribute* characteristic of an object (e.g., pinching fingers to indicate SMALL SIZE). As has been reported in the literature (Acredolo & Goodwyn 1988), we found that the children used iconic gestures to convey action information more than the static perceptual attributes of an object (77% vs. 23%). We also found the same pattern for parents (76% vs. 24%; Figure 3). There were no reliable differences between parent and child in the numbers of action, $M_{parent} = 13.95$ [SD = 16.56] vs. $M_{child} = 9.38$ [SD = 8.48], $t(39) = 1.69$, *ns*, or attribute, $M_{parent} = 4.63$ [SD = 5.41] vs. $M_{child} = 3.58$ [SD = 5.09] vs. $t(39) = 1.22$, *ns*, gestures produced.

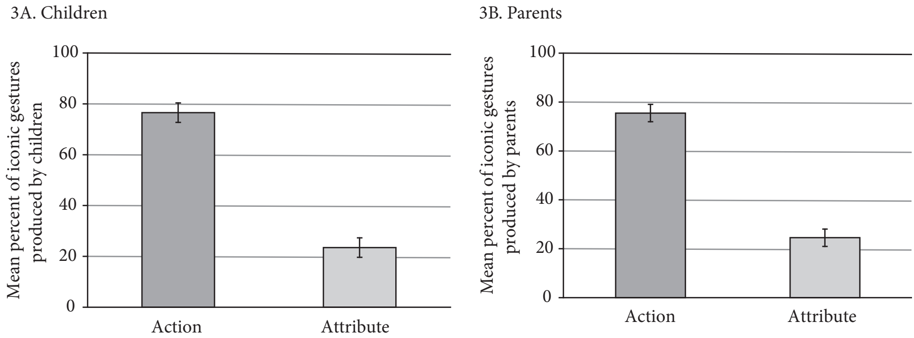


Figure 3. Mean percent of iconic gestures conveying information about the actions (black bars) or attributes (gray bars) associated with an object produced by children (Panel A) and their parents (Panel B).

In the second analysis, we categorized iconic gestures according to whether the gesturer assumed the point of view of an object, animal or person (*character viewpoint*, e.g., the gesturer turned her whole body in circles to represent a mixer); whether the gesturer used her hand to represent an object, animal or person (*observer viewpoint*, e.g., the gesturer used a V-shaped hand to represent rabbit ears); or whether the gesturer used her hand to outline the shape or trajectory of an object, animal or person (*trace gestures*, e.g., the gesturer traced a circle in air to represent the circular path a horse follows). McNeill (1992) reports a predominance of character viewpoint in children's early iconic gestures. As can be seen in Figure 4A, we replicated this pattern in our sample. There was a significant effect of *viewpoint* in children's early iconic gestures, $F(2,78) = 22.71$, $p < .001$: children used *character* viewpoint gestures (66%) significantly more often than *observer* viewpoint gestures (24%), $p < .001$, which, in turn, were used more often than *trace* gestures (10%), $p < .05$. Parents, however, displayed a different pattern. Unlike the children, the parents did not differ in their use of different viewpoints, Figure 4B, $F(2,78) = 0.97$, *ns*. Parents used *character* (36%), *observer* (28%), and *trace* (36%) gestures equally often. When compared to their children, parents produced significantly more *observer* gestures, $t(39) = 2.37$, $p = .02$, and *trace* gestures, $t(39) = 5.42$, $p < .001$.

In the third analysis, we classified each iconic gesture according to the particular meaning conveyed (RUNNING, THROWING, BIG, SMALL). We then examined the overlap of meaning glosses for parent and child in each dyad. Based on previous work suggesting that children learn iconic gestures in interactive routines with parents (Acredolo & Goodwyn 1993), we expected that many of the child's iconic gestures could be found in the parent's gestural repertoire. Surprisingly, however, we found minimal overlap between child and parent iconic gestures. The proportion of meanings found in the children's iconic gestures that were also found in their parents' gestures was under 20% throughout the observation sessions (Figure 5).[4]

4. This percentage could not be calculated at 14 months as only three children produced iconic gestures during this session.

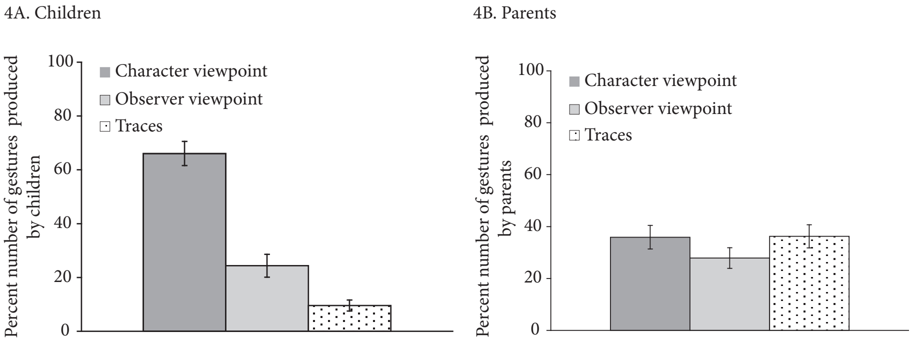


Figure 4. Mean percent of character viewpoints (black bars), observer viewpoints (gray bars) and traces conveyed in the iconic gestures of children (Panel A) and their parents (Panel B).

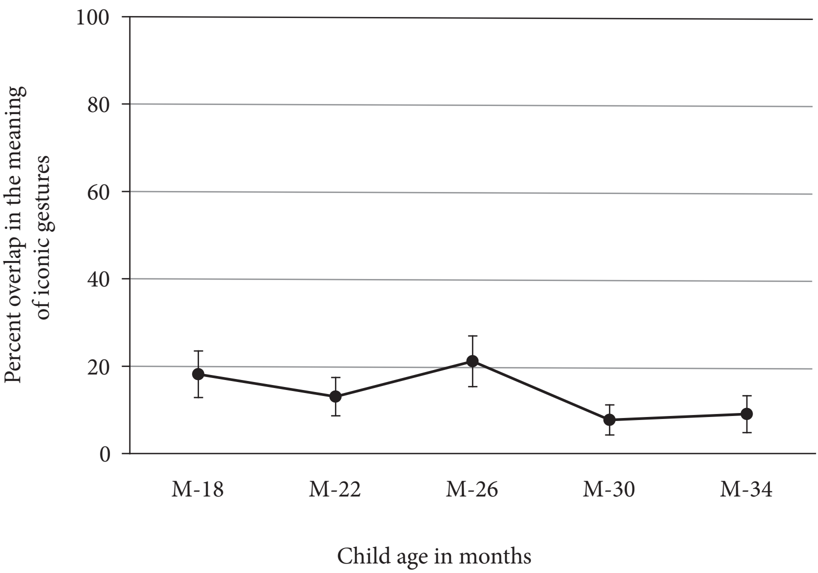


Figure 5. Mean percentage of meanings conveyed in a child's iconic gestures that were also found in the parent's gestures during that session.

5. Conclusions

Previous research has found that understanding the iconicity of a gesture is a relatively late achievement, beginning as late as 26 months (Namy & Waxman 1998, Namy 2001, Namy et al. 2004). Here we explored whether 26 months marks a similar turning point in children's production of iconic gestures. We found that children did indeed display a significant increase at 26 months in the number and types of iconic gestures that they produced during spontaneous interactions with their parents.

Why do we see a surge in iconic gestures at 26 months? One possible explanation for the relatively late occurrence of iconic gestures is that children are modeling their gestures after their parents' gestures. The parents in our study not only produced the same number and types of iconic gestures as their children, they also displayed an increase in iconic gestures at the same time as their children. However, the fact that parents produced iconic gestures during the earliest sessions means that children had a model for iconic gestures at 14 months but didn't appear to use it until 26 months, suggesting that the newly found interest in iconic gestures may have come from the children rather than the parents. Moreover, although both parents and children showed a similar pattern with respect to action and attribute iconic gestures (Figure 3), there was very little overlap in the particular iconic gestures that parents and children produced (Figure 5) and children showed a different distribution of *character* vs. *observer* gestures than their parents (Figure 4). These differences lend weight to the hypothesis that the increase in iconic gestures in parents at child-age 26 months reflects, rather than causes, the increase in iconic gestures in children.

The relatively late occurrence of iconic gestures, particularly in relation to deictic gestures, may stem from the fact that the mapping between symbol and referent is less straightforward, and therefore more cognitively demanding, for iconic gestures than for deictic gestures. Deictic gestures map onto the perceptual world in a direct way; they are used to indicate objects, people or locations that are perceptually cohesive and easily parsed out of the scene. In contrast, iconic gestures select their referents from a diffuse set of relational concepts, and may depend on the language one speaks (Kita & Özyürek 2003, Özçalışkan et al. 2011). In fact, deictic gestures routinely precede children's first nouns (Iverson & Goldin-Meadow 2005), whereas iconic gestures conveying action meanings typically follow children's first verbs (Özçalışkan et al. 2011), further reinforcing the idea that iconic gestures might be conceptually harder than deictic gestures. Iconic gestures may emerge as an outcome of related spoken language achievements, rather than being a precursor to such abilities.

In summary, we have shown that children begin to produce iconic gestures in earnest at around 26 months of age. Although parents also increase their production of iconic gestures at this same time, there is reason to believe that their gestures reflect, rather than cause, changes in the child. Indeed, the fact that children begin to produce iconic gestures at just the moment that they seem to understand iconicity in gesture suggests that this moment may be a turning point in the child's grasp of the iconicity of a symbol.

Acknowledgements

We thank K.Schonwald and J.Voigt for their administrative and technical support and the project research assistants, K.Brasky, E.Croft, K.Duboc, Becky Free, J.Griffin, S. Gripshover, C.Meanwell, E.Mellum, M.Nikolas, J.Oberholtzer, L.Rissman, L. Schneidman, B.Seibel, K.Uttich, and J.Wallman, for their help in collecting and transcribing the data. The research presented in this paper was supported by grant #PO1 HD406-05 to Goldin-Meadow.

References

Acredolo, L. P. and Goodwyn, S. W. 1985. "Symbolic gesturing in language development." *Human Development* 28: 40–49.

Acredolo, L. P. and Goodwyn, S. W. 1988. "Symbolic gesturing in normal infants." *Child Development* 59: 450–466.

Acredolo, L. P. and Goodwyn, S. W. 2002. *Baby Signs*. New York: McGraw Hill.

Bates, E. 1976. *Language and Context*. New York: Academic Press.

Bates, E., Benigni, L., Bretherton, I., Camanioni, L. and Volterra, V. 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.

Bekken, K. 1989. *Is There Motherese in Gesture?* Unpublished doctoral dissertation. The University of Chicago.

Buttterworth, G. and Grover, L. 1988. "The origins of referential communication in human infancy." In *Thought without Language*, L. Wiskrantz (ed), 5–24. Oxford: Clarendon Press.

Goodwyn, S. W., Acredolo, L. P. and Brown, C.A. 2000. "Impact of symbolic gesturing in early language development." *Journal of Nonverbal Behavior* 24 (2): 81–103.

Greenfield, P. and Smith, J. 1976. *The Structure of Communication in Early Language Development*. New York: Academic Press.

Iverson, J. M., Capirci, O. and Caselli, M. C. 1994. "From communication to language in two modalities." *Cognitive Development* 9: 23–43.

Iverson, J. M., Capirci, O., Longobardi, E. and Caselli, M. C. 1999. "Gesturing in mother-child interactions." *Cognitive Development* 14: 57–75.

Iverson, J. M. and Goldin-Meadow, S. 2005. "Gesture paves the way for language development." *Psychological Science* 16: 368–371.

Kita, S., and Özyürek, A. 2003. "What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking." *Journal of Memory and Language* 48 (1): 16–32.

LeBarton, E. S. and Goldin-Meadow, S. 2010. "Experimentally-induced increases in early gesture lead to increases in spoken vocabulary." Under review.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Language and Thought*. Chicago: University of Chicago Press.

Namy, L. L. 2001. "What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories." *Infancy* 2: 73–86.

Namy, L. L., Campbell, A. L. and Tomasello, M. 2004. "The changing role of iconicity in non-verbal symbol learning: A U-Shaped trajectory in the acquisition of arbitrary gestures." *Journal of Cognition and Development* 5 (1): 37–57.

Namy, L. L. and Waxman, S. 1998. "Words and symbolic gestures: Infants' interpretations of different forms of symbolic reference." *Child Development* 69: 295–308.

Nicoladis, E., Mayberry, R. I. and Genesee, F. 1999. "Gesture and early bilingual development." *Developmental Psychology* 35 (2): 514–526.

Özçalışkan, Ş. and Goldin-Meadow, S. 2005a. "Do parents lead their children by the hand?" *Journal of Child Language* 32 (3): 481–505.

Özçalışkan, Ş. and Goldin-Meadow, S. 2005b. "Gesture is at the cutting edge of early language development." *Cognition* 96 (3): B101–B113.

Özçalışkan, Ş. and Goldin-Meadow, S. 2009. "When gesture-speech combinations do and do not index linguistic change." *Language and Cognitive Processes* 28 (24): 190–217.

Özçalışkan, Ş., Gentner, D. and Goldin-Meadow, S. 2010. "Do iconic gestures pave the way for children's early verbs?" Under review.

Rowe, M., Özçalışkan, Ş. and Goldin-Meadow, S. 2008. "Early gesture production and comprehension provide a helping hand for later vocabulary skills." *First Language* 28 (2): 185–203.

Rowe, M. and Goldin-Meadow, S. 2009. "Differences in early gesture explain SES disparities in child vocabulary size at school entry." *Science* 323: 951–953.

CHAPTER 13

The development of spatial perspective in the description of large-scale environments*

Kazuki Sekine
Japan Society for the Promotion of Science/National Institute of Informatics

This research investigated developmental changes in children's representations of large-scale environments as reflected in spontaneous gestures and speech produced during route descriptions. Four-, five-, and six-year-olds ($N = 122$) described the route from their nursery school to their own homes. Analysis of the children's gestures showed that some 5- and 6-year-olds produced gestures that represented survey mapping, and they were categorized as a *survey group*. Children who did not produce such gestures were categorized as a *route group*. A comparison of the two groups revealed no significant differences in speech indices, with the exception that the survey group showed significantly fewer right/left terms. As for gesture, the survey group produced more gestures than the route group. These results imply that an initial form of survey-map representation is acquired beginning at late preschool age.

Keywords: Spontaneous gesture, Speech, Mental representation, Spatial cognition, Preschool children

1. Introduction

In this chapter I focus on gesture and speech spontaneously produced in route descriptions. My goal is to provide evidence of the development of large-scale spatial representations (i.e., mental models in preschool-aged children). I approach this question indirectly by examining the spatial perspective that one takes when one describes a specific environment.

---

* Gesture and speech transcription conventions in this chapter: Square brackets show the starting point and ending point of the motion of the children's hands; **Boldface** marks the stroke phrase of the gesture phrase; Underlining indicates a motionless hold phase; "*" shows self interruption; ":" in speech indicates elongated phonation. Abbreviations that are used in the interlinear gloss are indicated below: ACC accusative, FP final particle, INJ interjection, IMP imperative, NOM nominative. PST past, QUOT quotative.

In earlier studies, researchers have speculated on the representation of large-scale environments, by treating linguistic descriptions alone as indices of representations (Taylor & Tversky 1992). Because previous studies of route description have focused mainly on linguistic descriptions, there has been little analysis of gestures, even though some researchers have pointed out the importance of gestures in route descriptions (e.g., Bühler 1934, Klein 1982, Piaget et al. 1960, Schegloff 1984). In most studies, adult subjects memorize maps or navigate a variety of scales of environments and are required to write route descriptions from memory. Taylor, Naylor, and Chechile (1999) found that when people were offered a bird's-eye perspective in advance, such as a large-scale environmental map, most descriptions were written from a *survey perspective*, viewing the environment from a fixed, single viewpoint. In contrast, when required to describe the route after searching the environment, most descriptions were written from a *route perspective,* taking the form of an imaginary journey. These descriptions contained information about temporal and spatial contiguity, as if the speaker were mentally taking the listeners on a specific route through an imaginary walk (Klein 1982, Linde & Labov 1975). In this route perspective, the speaker's reference point, which is termed an *origo* by Bühler (1934), seemed to be constantly shifting as the description went on.[1]

These studies revealed that the factors influencing the choice of perspective are not only the scale of the environment, but also its configuration, such as the number of paths and the relative size of landmarks (Taylor & Tversky 1996), or the learning goal (i.e., learning the layout vs. the fastest routes), and the source of spatial information in the environment (by studying a map vs. by navigating) (Taylor et al. 1999). Thus, studies have revealed so far that multiple factors can affect the choice of spatial perspective. However considering the fact that adults who showed a survey map perspective were carrying out a task which was most likely to induce that perspective (Taylor et al. 1999), the following question needs to be asked: In a natural setting, when children are not required to learn the environment from a map or use a specific learning goal, what perspective tends to be chosen and when does the survey perspective appear in children's descriptions?

Studies of the development of children's spatial representations, as Graf (2006) pointed out, have shown that, although the acquisition of large-scale spatial knowledge is especially salient in middle and late childhood (cf. Presson, 1987, Allen & Ondracek 1995), its components might be acquired earlier as primitive spatial structures. Traditionally, it is thought that large-scale representations develop from a route map to survey map types (Shemyakin 1962) and that survey map representations are acquired by accumulating partly local-networked landmarks (Hart & Moore 1973). Given that

1. Some researchers categorize an intermediate form of perspective, between route and survey, which has been called a "gaze viewpoint" (Linde & Labov 1975, Taylor & Tversky 1996, Tversky et al. 1999). Because a gaze perspective is rarely produced in the description of large-scale environments (Taylor & Tversky 1996), it is excluded from analysis in the present study.

the coordinated survey map representation appears around 8 or 9 years old as suggested in previous studies, it can be considered that an initial form of the survey map perspective is available to younger children, perhaps those of preschool age when landmarks begin to be learned. However, little is known about the developmental transition from route map to survey map representations.

An approach to this question may be found in research methods used to study spatial knowledge. Sekine (2009) suggested that although recognition methods, including maps, aerial photographs, and environmental models keep stimuli constant, these presentation modes may influence subjects' responses by providing cues about the perspective in advance. Production methods, such as verbal descriptions and sketch maps, are also problematic because they are limited by the skills children possess. Thus, because these methods force participants to transform their spatial expressions into information on a two-dimensional plane and they require drawing skills and verbal competence, results may lead us to underestimate children's spatial knowledge or performance.

To overcome the methodological problems described above, the present study focused on spontaneous gestures as an index of spatial perspective. Importantly, gestures can be projected in three-dimensional space and producing gestures is easy, even for preschoolers (Doherty-Sneddon & Kent 1996). Recently some studies have indicated that gestures are a useful index for approaching spatial representations (Emmorey, Tversky, & Taylor 2000; Sekine, 2009). For example, Emmorey et al. (2000) examined which point of view was taken when a speaker recalled a spatial image by observing gestures that the speaker produced. They asked participants to explain landmarks on a map and divided their verbal responses into two perspective types: route and survey maps. They reported that speakers who assumed a route map point of view produced many gestures in three-dimensional space, as if depicting an actual scene that they might experience in the environment. In contrast, speakers who took a survey map point of view tended to produce gestures in two-dimensional space, as if they were using their frontal space as a desktop or a blackboard.

However, studies of children's spatial perspective so far have mainly focused on linguistic aspects of utterances and paid little attention to gestures (Blades & Medlicott 1992, Gauvain & Rogoff 1989). In fact, in verbal route descriptions even 9-year-olds fail to show survey descriptions (Gauvain & Rogoff 1989). Studies focusing on gestures have revealed that people communicating spatial information tend to produce more of the *spontaneous gestures* (hereinafter referred to as a 'gesture') that co-occur with speech (McNeill 1992) than those communicating non-spatial information (Rauscher, Krauss, & Chen 1996). More importantly, gestures are a primary vehicle for conveying spatial information, especially in young children. Consequently, it is worth examining ontogenetic changes of spatial perspective from the standpoint of gesture research. Therefore, this study examines the development of spatial perspectives in preschool age children by looking at how children use gestures in route descriptions.

2. Method

2.1 Participants

As shown in Table 1, the study lasted for three years, and a total of 122 children participated. The participants consisted of 36 4-year-olds, 36 5-year-olds, and 50 6-year-olds. 15 of the 4-year-olds who began participating in 2004 continued to participate until 2006, and 19 of the 5-year-olds who participated in 2004 continued to participate until 2005. Other children participated just once in the study. Because the main purpose of the study was to examine general tendencies in the development of spatial perspective, longitudinal analysis was not conducted. All participants were native Japanese speakers from middle-class families who attended a public nursery school in Tokyo, Japan. The children's main commute was classified as either *walking*, *cycling*, or *by car*. No relationship was found between route descriptions and direct distance to the nursery school from home, or the type of commute, and these factors are not subsequently mentioned.

2.2 Procedure

In order to obtain the children's route descriptions from the nursery school to their home, interviews were conducted individually in a quiet spare room of the nursery school. Before entering the room for the interview, the experimenter and the child confirmed the location and direction of the gate from a window in the corridor immediately outside of the room. An armless chair was placed in the room, facing the nursery school gate, and a video camera was positioned at a 45 degree angle from the child. The child sat down in the chair and the experimenter sat facing the child. After ensuring the child's knowledge of the direction of the nursery school gate, the experimenter asked, 'How do you go back to your home from the gate of nursery school?' All children interviewed responded to the question by the second prompt. All interviews were recorded by a camcorder. (See Sekine (2009) for the detailed procedure and the experimental setting.)

Table 1. Number of participants and the average month in each experimental year

Year	2004	2005	2006
6 years	15 (13 boys, 2 girls), 77 mo	20 (11 boys, 9 girls), 76 mo	15 (8 boys, 7 girls), 78 mo
5 years	19 (9 boys, 10 girls), 64 mo	17 (9 boys, 8 girls), 65 mo	
4 years	19 (8 boys, 11 girls), 53 mo	17 (6 boys, 11 girls), 51 mo	

2.3 Coding of gestures and spatial perspective

All narratives were transcribed verbatim by a native speaker, and then the *total number of gestures* and the *frequency of gestures per second* were calculated. Following Iverson's criteria (1999), hand movements were classified as gestures only when they had an identifiable beginning and a clear end, and they were synchronized with speech. Spatial perspective was calculated using the same index used in Emmorey et al.'s (2000) study. These researchers argued that, for a speaker who takes a survey map perspective, target locations or landmarks are drawn by gestures on a two-dimensional plane, as if tracing a maze. In the present study, the children's gestures were categorized as *survey map gestures* if they met the following two criteria: the gestures were produced on a two dimensional plane, either horizontal or vertical; and the gestures were used to set up the nursery school as a starting point in the gesture space. Gestures that did not meet these criteria were categorized as *route map gestures*.

3. Results

3.1 Spatial perspective in gestures

First, none of the children changed perspective during their route descriptions. As shown in Table 2, 115 children produced route map gestures during their entire descriptions, and seven produced survey map gestures. In what follows, I refer to these as the 'route group' and 'survey group', respectively. The youngest children in the survey group were 5-year-olds, and out of the total of seven, five were boys and two girls. Because no 4-years-olds produced survey map gestures, they are excluded from the following analyses.

3.2 Mean scores of speech and gestural measures in route and survey group

Table 3 shows the *average age*, *mean amount of total speaking time* (time spent on the route or survey description), *total number of morphemes* (excluding fillers such as 'uh', 'ah', hesitations, and speech errors), *number of landmarks*, such as a park, river, or

Table 2. Number of children in survey and route group

	Survey group			Route group		
	Boys	Girls	Total	Boys	Girls	Total
6 years	3	1	4	29	17	46
5 years	2	1	3	16	17	33
4 years	0	0	0	14	22	36

Table 3. Average age, speech, and gestural performance (SD)

	Survey group *(N = 7)*	Route group *(N = 79)*	T value
Average age (month)	70.1 (9.6)	72.1 (7.1)	0.66
Total speaking time (second)	40.2 (27.5)	37.8 (31.6)	0.2
Total number of morphemes	59.1 (46.9)	50.5 (38.7)	0.55
Number of landmarks	1.7 (1.1)	3.1 (3.2)	1.12
Number of left/right terms	0.1 (0.4)	1.1 (2)	3.72***
Total number of gestures	17.7 (6.6)	10.8 (8.9)	2.01*
Frequency of gestures (per second)	0.5 (0.2)	0.3 (0.2)	2.54**

*** p<.001 ** p<.01 * p<.05

hospital (with the exception of the nursery school and child's own home), *number of left/right terms, total number of gestures* and *frequency of gestures per second* for both groups. A t-test, comparing the mean scores of these indices, revealed that the route group produced significantly more left/right terms than the survey group (t = (84) 3.72, $p < .001$) and that the total number and frequency of gestures in the survey group were significantly greater than those in the route group (t = (84) 2.54, $p < .001$, t = (84) 2.07, $p < .001$, respectively). These results suggest that the children in the survey group tend to depend on gestures to describe their route. Perhaps they rely on gestures to indicate directions because they are lacking left/right terms.

3.3 Describing the starting point in the route group

To better understand how children in both groups describe their route, it is important to look at how speech and gesture interact in more detail. Let us examine the following examples by focusing on the starting point of their routes.

Most children in the route group started their descriptions with the direction of movement or the motion taken immediately after leaving the gate, without explicitly depicting the gate or location of the nursery school in gesture and/or speech. This might be due to the setting in which the route description was collected, where the location of the gate was shared between the child and the experimenter. Figure 1 shows a typical explanation of the starting point in the route group.

Child A, as shown in Figure 1, starts describing her route using the motion verb *deru* 'to get out'. Her utterance implies that the gate is the origin of the motion, but neither the gate itself nor its location is mentioned verbally. The gate is orthogonal to the slope, which inclines to the left. Based on the angle of her gesture and the accompanying speech in Figure 1, it appears that she transposes an origo to the right outside the gate, and indicates the direction of the bottom of the slope from there. Child B ((a) in Figure 2) illustrates a second means used by the route group to express the starting point. He indicates the direction of the gate by pointing from the room to the actual

(a) orite kou (go down the slope do like this)

de-te:	[[**si***]	[saka o	o(a)**ri-te**]	[**kou**	yat-te]	[massugu	**it**-te]	soide
get.out-and	si*	slope ACC	go.down-and	like.this	do-and	straight	go-and	then

"(I/you) get out, go down the slope, and go straight like this and then"

Figure 1. A description of the starting point in the route group (girl A, 5 years old).

(a) [[**kou** **it**-**te**][**kou** **it**-**te**] [saka o ori-te
this.way go-and this.way go-and slope ACC go.down-and
"(I/you) go this way, go this way, and go down the slope, and"

(a) kou itte kou itte (go like this, go like this)

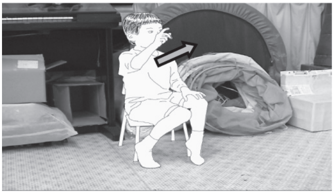
Figure 2. A description of the starting point in the route group (boy B, 4 years old).

environment, that is, without transposing an origo. This child begins by saying 'doing like this and doing like this' while pointing in the actual direction. Like child A, he does not mention the gate or the nursery school itself.

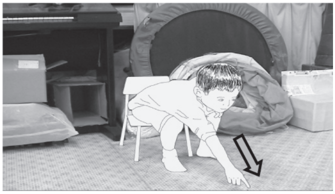
Thus, children in the route group tend to start their explanations with the movement they will take right after they leave the gate, without mentioning starting points in the route. Because describing motion from a route map perspective makes the location of the starting point obvious, children do not need to mention it explicitly. This is a way of explaining the starting point which is mainly seen in the route group. Common gesture characteristics of the route group are that (1) gestures are produced in three-dimensional space with depth and (2) the starting point is not assigned in gesture space.

3.4 Describing the starting point in the survey group

Let us observe a description of a starting point in the survey group (Figure 3). Child C ((b) in Figure 3) sets up his nursery school as a starting point in the gesture space by pointing to the ground while using a demonstrative *koko* 'here'. Interestingly, most children in the survey group, like child C, used words such as *omou* 'to suppose' to make their listener understand that that particular point in their frontal space signified the location of the nursery school. This suggests that children in the survey group notice that their listener has a different perspective from theirs and that they know how to share their perspective with the listener. The fact that they use gesture and speech to make the listener assume a particular point in gesture space as a specific landmark on their route implies that they are aware of the need to make their listener understand what the gesture or the gesture space stand for in order to share perspective with their listener.



(a) anone (well)



(b) koko ga (here is)

(c) matigaeta (made mistake)

[[(a)anone ii (b)**koko ga** hoikuen to omotte yo]
INJ INJ here NOM nursery.school QUOT suppose.IMP FP
"Well, (are you) ready? Suppose that here is the nursery school."

[hoikuen to omotte yo site: koko **kara** **de**-te]
nursery.school QUOT suppose.IMP FP and here from go.out-and
"Suppose as the nursery school and, (I/you) go out from here,"

[(c)**matigae** **ta**]
make.mistake PST
"I made (a) mistake."

[**koko kara**][hoikuen **o** **de-te** sorekara] [maga* **koko** **kara** **to** **ato**
here from nursery.school ACC go.out-and then turn* here from and then
"(I/you) go out the nursery school from here, and turn*, from here and then"

Figure 3. A description of a starting point in the survey group (boy C, 6 years old).

In the survey group, other expressive behaviors showing a deliberate use of the gesture space were observed. For example, child C tried to point in the actual direction of the gate as he used the discourse marker 'well, are you ready?' at the beginning of the route description ((a) in Figure 3). But, before he finished indicating the external environment he moved to the depiction of the gate in his gesture space. This stagnation of a gesture might reflect the speaker's hesitation to choose a perspective and implies that the speaker has multiple descriptive strategies or mental models of the large-scale environment. In addition, it was observed that child C erased a part of the route that had been already depicted by wiping the floor ((c) in Figure 3). This erasing gesture was never observed in the route group. Child C seems to be conscious that the listener might make use of his gestures depicting the route on the floor as an important informational source.

These observations suggest that children in the survey group can symbolically assign a starting point or landmarks in a two-dimension plane of the gesture space and that they try to share it with their listeners who can simultaneously overview the route that the speaker depicts. An implication is that some children in the survey group purposefully choose the survey perspective to describe the route using multiple descriptive strategies. These are considered characteristics of the survey group.

4. Discussion

In this study I investigated the spatial perspective assumed by preschool children as reflected in gestures and speech produced in route descriptions. The study revealed that some children produce survey map gestures, and this implies that children can begin to take a survey perspective from late preschool age.

Comparing the characteristic descriptions of the survey group with those of the route group, I found that, although there is no difference in the average age of the groups, children in the survey group produced fewer left-right terms and a greater number and frequency of gestures than children in the route group. These results indicate that children in the survey group tend to describe the direction of movement mainly through gesture.

Studies of spatial cognition have suggested that survey map representations – which systematically coordinate landmarks in the environment from a single perspective – are acquired around the middle grades of school-aged children (i.e., at 8 or 9 years old). The results of the present study suggest that an understanding of the environment from a bird's-eye viewpoint is available from as early as 5 years of age and that an initial form of survey map representations begins to appear by that period. In contrast to some children in the route group who point directly to their actual route (Sekine 2009), children in the survey group tend to set up the nursery school as the starting point in gesture space and make use of such space symbolically. The symbolic use of space would underlie the survey map representation.

Why are survey map gestures produced? Let us consider factors that influence the appearance of such gestures. First, we consider the lack of directional indicators, such as left-right terms. Children in the survey group most likely avoided left-right terms because of a difficulty indicating left-right with respect to their own bodies. Instead, they chose a strategy in which they depicted the route directly on the floor.

Second, children in the survey group might have a greater ability to adjust their route descriptions according to the listener's knowledge of the route. Generally, when preschoolers describe their route, they express it either by pointing directly to the actual environment or by depicting a view that they can see when they actually walk in the environment (Sekine 2009). However, children in the survey group use the two-dimensional space which lies between themselves and their listener. Children in the survey group might have the ability to speculate that depicting the environment from a survey viewpoint would be a better means of communication for the listener, rather than depicting it from an egocentric perspective, because a description using survey map gestures makes the route visible and sharable between them.

A third factor concerns the characteristics of play the children prefer. I attended this nursery school once a week for six years as a volunteer to support the teacher, so I was familiar with the children who participated in the study, their teachers, and the return routes to their homes. Observing the play preferences of the children who participated in the study for several years, I found that, although this is an anecdote, all children in the survey group were more likely to play with toys such as mini cars or railway models, which induce children to take a bird's eye viewpoint with respect to the miniature models. Play preferences in daily life might influence the way children express or understand their environment.

Considering these factors as influences on the production of survey map gestures, when children attain preschool age, they may start acquiring both a 'spatial perspective' – taking a view which is spatially different from that taken in the here-and-now – and also a 'social perspective' – adjusting their means of expression according to the knowledge status of their listener–. Further studies are needed to examine to what extent the three factors have an influence on the acquisition of survey map perspectives and on how those factors interrelate. In parallel with this, it would also be necessary to investigate the development of meta-communicative abilities, including how the deliberate use of a gesture space or a descriptive strategy is related to changes in large-scale representations. At the same time, studies are needed to reveal the consistency or variability of perspectives taken by each individual child.

By focusing on spontaneous gestures, this study suggests that a survey map perspective, which has been believed to be acquired around the middle grades in school age children, is already starting to be acquired from a late preschool age. The study suggests that spontaneous gestures can be a useful index for understanding a speaker's spatial knowledge or perspective.

Acknowledgements

I am grateful to Elena Levy for her valuable suggestions and comments on earlier drafts.

References

Allen, G. L. and Ondracek, P. J. 1995. "Age-sensitive cognitive abilities related to children's acquisition of spatial knowledge." *Developmental Psychology* 31 (6): 125–135.

Blades, M. and Medlicott, L. 1992. "Developmental differences in the ability to give route directions from a map." *Journal of Environmental Psychology* 12 (2): 175–185.

Bühler, K. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer. (Reprinted by Gustav Fischer Verlag, Stuttgart, 1965)

Doherty-Sneddon., G. and Kent, G. 1996. "Visual signals and the communication abilities of children." *Journal of Child Psychology and Psychiatry* 37 (8): 949–959.

Emmorey, K., Tversky, B. and Taylor, H. A. 2000. "Using space to describe space: Perspective in speech, sign, and gesture." *Spatial Cognition and Computation* 2 (3): 157–180.

Gauvain, M. and Rogoff, B. 1989. "Ways of speaking about space: The development of children's skill in communicating spatial knowledge." *Cognitive Development* 4 (3): 295–307.

Graf, E-M. 2006. *The Ontogenetic Development of Literal and Metaphorical Space in Language.* Gunter Narr Verlag Tubingen.

Hart, R. A. and Moore, G. T. 1973. "The development of spatial cognition: A review." In *Image and Environment*, R. M. Downs. and D. Stea (eds), 246–288. Chicago: Aldine Publishing.

Iverson, J. M. 1999. "How to get to the cafeteria: Gesture and speech in blind and sighted children's spatial descriptions." *Developmental Psychology* 35 (4): 1132–1142.

Klein, W. 1982. "Local deixis in route directions." In *Speech, Place, and Action*, R. J. Jarvella and W. Klein (eds), 161–182. Chichester: Wiley.

Linde, C. and Labov, W. 1975. "Spatial networks for the study of language and thought." *Language*: *Journal of the Linguistic Society of America* 51: 924–939.

McNeill, D. 1992. *Hand and Mind*. Chicago: The University of Chicago Press.

Piaget, J., Inhelder, B. and Szeminska, A. 1960. *The Child's Conception of Geometry*. London: Routledge and Kagan Paul.

Presson, C. C. 1987. "The development of spatial cognition: Secondary uses of spatial information." In *Contemporary Topics in Developmental Psychology*, N. Eisenberg (ed), 77–112. New York: Wiley.

Rauscher, F. H., Krauss, R. M. and Chen, Y. 1996. "Gesture, speech, and lexical access: The role of lexical movements in speech production." *Psychological Science* 7 (4): 226–231.

Schegloff, E. A. 1984. "On some gestures' relation to talk." In *Structures of Social Action; Studies in Conversation Analysis*, J. M. Atkinson and J. Heritage (eds), 266–296. Cambridge: Cambridge University Press.

Sekine, K. 2009. "Changes in frame of reference use across the preschool years: A longitudinal study of the gestures and speech produced during route descriptions." *Language and Cognitive Processes* 24 (2): 218–238.

Shemyakin, F.N. 1962. "Orientation in space." In *Psychological Science in the U.S.S.R. Vol. 1*, B. G. Anayev et al. (eds), 186–255. Washington, DC: Office of Technical Services.

Taylor, H. A. and Tversky, B. 1992. "Spatial mental models derived from survey and route descriptions." *Journal of Memory and Language* 31 (2): 261–292.
Taylor, H. A., Naylor, S. J. and Chechile, N. A. 1999. "Goal-specific influences on the representation of spatial perspective." *Memory and Cognition* 27 (2): 309–319.
Taylor, H. A. and Tversky, B. 1996. "Perspective in spatial description." *Journal of Memory and Language* 35 (3): 371–391.

CHAPTER 14

Learning to use gesture in narratives

Developmental trends in formal and semantic gesture competence

Olga Capirci,[1] Carla Cristilli,[2] Valerio De Angelis,[1] and Maria Graziano[3]
ISTC- CNR, Rome- Italy,[1] Università degli Studi di Napoli "L'Orientale",[2] SESA-Università degli Studi "Suor Orsola Benincasa"[3]

This study analyses the way in which children develop their competence in the formal and semantic aspects of gesture. The analysis is focused upon the use of representational gestures in a narrative context. A group of 30 Italian children from 4 to 10 years was videotaped while telling a video cartoon to an adult. Gestures were coded according to the parameters used in Sign Languages analysis and analysed in terms of the acquisition of their properties, the accuracy of their execution and correctness in content representation.
It investigated also the development of the symbolic competence in relation both to the use of some of these parameters and to the representational strategies adopted.

Results indicate a developmental trend in all the phenomena investigated and point out some formal similarities between gesture and Sign Languages.

Keywords: co-speech gesture development, representational gestures, gesture and sl compositional parameters, italian pre-school and school age children

Introduction

In the last decades, an increasing number of scholars have shown the relevant role played by gesture in the psychological-cognitive processing of content and in the construction of discourse (Kendon 1985, 2004; McNeill 1992, 2005 to name a few).

The tight link recognized between speech and gesture in both processes has led Kendon (2004) to speak of a *speech-gesture ensemble* and McNeill (1992, 2000, 2005) to consider them as two aspects of the same underlying thought process.

Recent findings on the neurophysiology of the motor system have provided a neural basis to this claim (Gallese et al. 1996, Rizzolati et al. 1996, Umiltà et al. 2001, Kohler et al. 2002), demonstrating that hand and mouth movements overlap in a broad frontal-parietal network. This network, called 'mirror neuron system,' would be activated during both perception and production of familiar and meaningful manual gestures and mouth movements (Rizzolatti & Arbib 1998), thus creating a direct link between the sender and a receiver of a message and making the observing and doing something like manifestations of a single communicative faculty, rather than two separate abilities. On the basis of these assumptions Rizzolatti and Arbib (1998) suggest that the mirror neuron mechanism represents the basic mechanism from which language evolved. Nevertheless, if gesture and speech are intimately and remotely connected, they still constitute two different forms of content processing and expression. To the analytic, compositional, conceptual and standardized form of speech, McNeill (1992, 2000, 2005) contrasted the synthetic, holistic, imagistic and idiosyncratic one of gesture.

Yet Calbris (1990), adopting a semiotic approach, identified a variety of handshapes, movement patterns and planes of their execution, suggesting that each of these parameters presents some semantic consistency.

Pettenati et al. (2010) explored the form of representational gestures produced by children (age range 24–37 months) asked to label pictures in words and analysed them with the parameters used to describe deaf children's signs. Results of this study show that gestures representing a given picture exhibit similarities in many of the parameters across children and that these parameters are similar to those described for early signs.

Showing that gestures, like sign languages, have a compositional structure, these works give us the possibility of rethinking McNeill's thesis on their global and holistic nature. Kendon (1985, 2004), moreover, shows that even co-verbal gestures have an internal structure that differentiates them from any kind of physical activity: they are characterized by an 'excursion' (movement away from and to a rest position); a '*stroke*' (the peak of the excursion recognized by naive subjects as what the movement actually 'does' and is 'meant for'); a well 'boundedness' (gestures tend to have clear onsets and offsets).

As for the close and profound link between speech and gesture, an important contribution to their understanding has been given by studies on their developmental aspects. These studies have demonstrated that this link becomes evident from early language development: gesture and speech emerge at about the same time, refer to the same broad set of referents and serve similar communicative functions. In addition, changes in gesture use predict the onset of first words and the emergence of early syntax (Butcher & Goldin-Meadow 2000; Capirci; et al. 1996, 2002; Goldin-Meadow & Butcher, 2003).

In some earlier developmental works, gestures were primarily explored as relevant features of the 'prelinguistic' stage, as behaviors preceding and preparing the emergence of language (substantially identified with speech). In these studies, behaviors such as playing with objects were considered gestures (Bates et al. 1979), thus linking gesture to cognitive skills separated from language but developing together with it within the same time frame and representing sort of 'cognitive precursors' of it.

More recent research supports the view that there is a remarkable continuity between prelinguistic and linguistic development and that the symbolic skills, most evident in linguistic productions, are inextricably linked to and co-evolve with more general representational abilities.

Around one year of age, words and gestures appear to encode similar meanings and go through a similar decontextualization process: both gestures and words are initially strictly related to the actions children perform with objects or with their own bodies. On the basis of these observations, it has been supposed that speech and gesture output systems draw on underlying brain mechanisms common to both language and motor functions (Iverson & Thelen 1999). In the following months, when the verbal system begins to emerge as the primary mode of linguistic communication, gesture shifts from a position of relative communicative equivalence in relation to speech to one of a support system integrated with it.

Recently, some scholars have been devoting their attention to older children, looking at the way in which they come to integrate speech and gesture in more complex tasks, like narratives. The development of narrative competence is a slow process founded on the evolution of psychological-cognitive capacities and on the acquisition of linguistic and textual devices and strategies (Stein & Glenn 1979, Peterson & McCabe 1983, Berman & Slobin 1994, Karmiloff-Smith, 1985).

In a multimodal perspective, Cassell & McNeill (1991) and McNeill (1992) observed the way in which children's gestures are functionally related to the categories of voice (C-VPT/O-VPT) and perspective (inside/outside). Studying gesture in narrative, Kita (2000) and Kita & Wood (2006) showed that children's bodies, as a representational medium, become more and more flexible and that gesture space becomes more and more symbolically distanced from the physical one.

Colletta (2004) analysed spontaneous narratives by 6- to 11-year-old French children, showing that, from 9 years on, narratives gain in linguistic complexity and children use more gestures to represent events and characters.

A recent Italian work from Capirci, Cristilli and collaborators (Capirci et al. 2008) underlines how the nature of the gestures produced during a narrative task changes with age. The study of 40 children (20 aged 5 and 20 aged 9) video-recorded while narrating a cartoon previously shown to them, examined different levels of analysis: syntactic, textual, pragmatic, narrative and gestural. The latter level showed gestures with a referential function (representational and deictic) distinguished from those with a 'pragmatic' one ('pragmatic gestures' refer to characteristics of an utterance meaning which are not part of its referential meaning or propositional content: Kendon 2004). Besides an expected improvement in syntactic, textual and narrative competences, results demonstrated a parallel development in the gestural modality: it was observed that gestures with a referential function (particularly deictic) decrease in favor of the pragmatics and that amongst these, older children produce mostly gestures with a narrative-textual function (*discursive* and *parsing*).

In the present study we aimed at investigating the developmental trends in formal and semantic gesture competence in a narrative context. In particular, focusing on representational gestures, we devoted our attention to the way in which children learn to: (a) exploit the motor-physical potentiality of gesture to express contents; (b) use these motor-physical components as elements of a system that, like any semiotic one, requires that they be accurately performed in relation to their formal properties; (c) use each significant component of gesture to represent referents in a semantically correct way. For the analysis of gesture components we utilized the formational parameters adopted in Sign Language studies: handshapes, movements, hand orientation and place of articulation. This gave us also the possibility of comparing their use by our children with that observed in deaf children exposed to SLs (Boyes-Braem 1975; Meier et al. 2008; Clibbens 1998; Karnopp 2002; Morgan, Barrett-Jones & Stoneham 2007).

Moreover we analysed the representational strategies used by children, considering them from the point of view of the level of abstractness they reveal. The development of the symbolic capacity was investigated also in relation to the way in which children used some gesture components, like the place of their execution.

Method

Participants

Thirty developmentally typical children took part in this research. The children were divided into three groups: group I, mean age 4 (preschool age); group II, mean age 6.5 and group III, mean age 8.7 (school age). All the children were right-handed.

Procedure and task

In order to analyze the narrative abilities of the groups, all the children were video-recorded while telling an adult a short video cartoon story they had watched twice. Both the adult and the setting were familiar to them. The short video cartoon belongs

Table 1.

Groups	Age (Range)	Sex	
		Male	Female
I	4 (3.03–5.08)	4	6
II	6.5 (5.11–7.06)	2	8
III	8.7 (7.07–10.05)	4	6

to ‘Pingu’, a TV series. It lasts 4 minutes and contains no proper words but only some vocalizations. It shows a penguin family (parents and two different aged children) while getting ready for Christmas: the mother makes some biscuits while the children watch the preparation; the parents decorate the Christmas tree outside the igloo while inside the children wrap their presents, and in the end, they all open them under the Christmas tree.

Coding

In order to evaluate the length of children’s narratives, we considered the total number of clauses produced by the three groups, whereas to assess how many and how frequently representational gestures were produced during the narratives, we considered the total number of their occurrences in the three groups and the percentage of gestures per clause.

As for the motor aspects of gestural production, we first considered whether the gestures were produced with one or two hands. In the first case, we transcribed which was involved; in the second case, we analyzed the symmetry between the two hands. Gestures were then coded according to the same parameters used to analyze Sign Languages: handshapes, place of articulation, hand orientation and movement.

To observe the way in which children learn to use gestures in a formally appropriate way, we formulated the concept of ‘formal accuracy’ scored in relation to three parameters. This analysis was based on a free adaptation of the ‘Scale of Gestuality’ proposed by Kendon, who considered it as a scale of gradient properties making some movements ‘more gestural’ than others[1]. The parameters we analyzed are: well boundedness, clearness of the stroke execution, shared space. Each was scored on a scale from 0 to 2.

The well boundedness was scored as follows: 0 = without a clear start and a clear end; 1 = only one of the two is clear; 2 = both are clear; NC (not classified) for consecutive gestures.

The formal clearness of the stroke was scored in relation to the gesture configuration and movement: 0 = both parameters are not clear; 1 = only one of the two is clear; 2 = both are clear.

The space of gesture execution was scored as follows: 0 = not visible by the listener; 1 = peripheral space; 2 = shared space.

We coded the representational correctness on the basis of the semantic pertinence of the gesture components (place, configuration and movement) in relation to the corresponding aspects of the referent (its location, its shape and size, the type and direction of the action). We scored it as follows: 0 = none of them is pertinent; 1 = only one is pertinent; 2 = only two are pertinent; 3 = all three are pertinent.

1. Our classification was based on proposals presented by Kendon in a Seminar given in the Department of Psychology, University of Rome “La Sapienza” on 6 November 2006.

Finally, we analyzed the strategies used in the gestural representation of the referent conceiving a scale going from the highest to the lowest degree of concreteness and motivation. The categories, partially corresponding to those adopted by other scholars (Müller 1998, Streeck 2008) are mime, manipulation, hand becoming an object, shape depiction and/or delimitation (of objects contours), symbolic-conventional representation.

These are the types of gestures we coded according to these categories:

Mime: the gestures children produced with the whole body or only with hands and arms, miming a situation in a holistic manner and identifying themselves in the character (similar to Müller's 'the hands imitates' and Streeck's 'mimesis'); *manipulation*: the gestures by which children represented an object reproducing the shape the hand assumes while seizing it (like Streeck's 'handling'); *hand becoming an object*: the child's identification of a part of his/her hand with the object represented (similar to Müller's 'the hand portrays'); *shape depiction*: the gestures by which the children represented an object depicting its shape (like Müller's 'the hand draws' and Streeck's 'drawing'); *delimitation*: the gestures by which children represented an object delimiting its contours in the air (like Streeck's 'delimiting'); *symbolic-conventional representation*: the gestures used to express in a symbolic and conventional way some more abstract contents like spatial and temporal relationships.

Results

Initially, we analyzed the total number of clauses produced by the three groups of children. The results show that it increases considerably with age: pre-school children (group I) produce 227 clauses, while school children (groups II and III) produce respectively 394 and 395 clauses.

The total number of representational gestures produced by the three groups increases between pre-school and school children: 103 produced by group I, 188 and 178 produced respectively by group II and group III. However, looking at the proportion of representational gestures in the clauses, we found that it is very similar for the three groups: 45% in group I, 48% in group II, 45% in group III.

Analyzing the use of one or two hands, we found that while the first two groups produced almost half of the gestures with one hand and half with two (group I, 47% and 53%; group II, 48% and 52%), group III produced two- handed gestures in a higher proportion (41% one hand, 59% two hands).

For the gestures produced with one hand, we observed which one was involved. The results show a strong preference for the use of the right hand in all the groups: 98% in group I, 82% in group II and 83% in group III. However, a slight increase emerges in the use of the left hand in the two older groups, going from the 2% in group I to the 18% and 17% in group II and III, respectively.

In the case of bi-manual gestures, we analyzed if the two hands were symmetrical (with same handshape and/or movement) or asymmetrical (different handshapes and/

or movements). The great majority of gestures are produced with symmetric hands by all the three groups of children: 98% of gestures in group I, 94% in group II and 87.5% in group III. Thus, the 'symmetry condition' of Sign Languages is respected [2].

Formational parameters

Gestures were coded according to the parameters used in Sign Language analysis (Stokoe 1960; Volterra 1987, 2004).

Hand-shapes

Table 2 shows the percentage of hand configurations mostly produced by the three groups of children, whereas Figure 1 shows the distribution of the different configurations in the three groups.[3]

These six handshapes account for the 84% of the total hand configurations used in the entire sample of 30 children's gestures. These handshapes constitute the basic ones in Sign Languages, and they are the most frequently used by children exposed to these languages (Boyes-Braem 1975; Meier et al. 2008; Clibbens 1998; Karnopp 2002; Morgan, Barrett-Jones & Stoneham 2007).

Looking at the distribution of the different configurations in the three groups in Figure 1, we can see that '5' is the most used by all of them; however, it is interesting to note that the use of this configuration decreases with age, while there is a gradual

Table 2.

Configurations	Percentages of the 3 Groups
5	47.12
B	14.63
A	8.02
C	5.91
L	5.62
T	2.53
Tot.	**83.83**

2. The 'symmetry condition' states that, when two hands move without touching each other, the movement and the configurational features of the sign must be the same or symmetrical for the two hands. Pettenati et al. 2010)

3. The symbols used for representing handshapes are the same adopted in SLs literature. They correspond to numbers or to alphabet letters. Different symbols can be used to represent the same handshapes by each SL: counting and finger spelling vary according to culture.

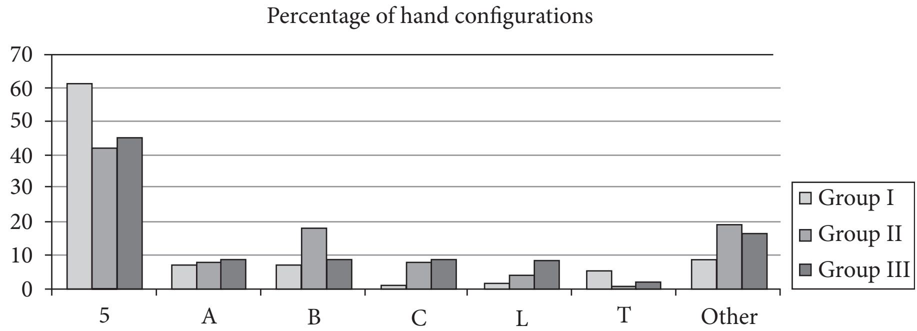


Figure 1.

increase in the use of the 'C' and 'L' ones, which were almost absent in the first group. Figure 1 also shows that the proportion of the other configurations increases with age, especially comparing the first group and the other two. In these other categories the configurations mostly used by the older children, even though not in a significant way, were the '3' and 'F.'

Place of articulation

The place of articulation was coded as 'Not involved hand' when the gesture was produced on the not involved hand; 'Body' when it was produced on different parts of the body (head, trunk, shoulder, etc) not necessarily with direct contact, 'Neutral space' when it was produced in the space in front of the children's body.

As it is shown in Figure 2, neutral space is the most frequent location used by all the children's groups and the non-dominant hand the less frequently used. Nevertheless, group I used a very high proportion of body locations, whereas the use of 'not dominant hand' location gradually increases with age.

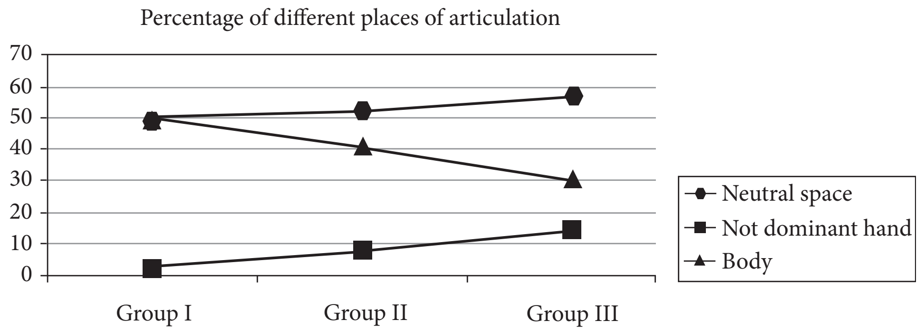


Figure 2.

Hand orientation

Figure 3 shows the different types of palm orientation used by the three groups.

In all three groups the palm is more frequently oriented 'up or down,' but we noted a developmental trend in the use of the right/left orientation from group I to groups II and III.

Movement

We analyzed the movement direction of children's gestures.

As we can see in Figure 4, with age there is a clear shift from the "up/down" to the "in front/behind" direction.

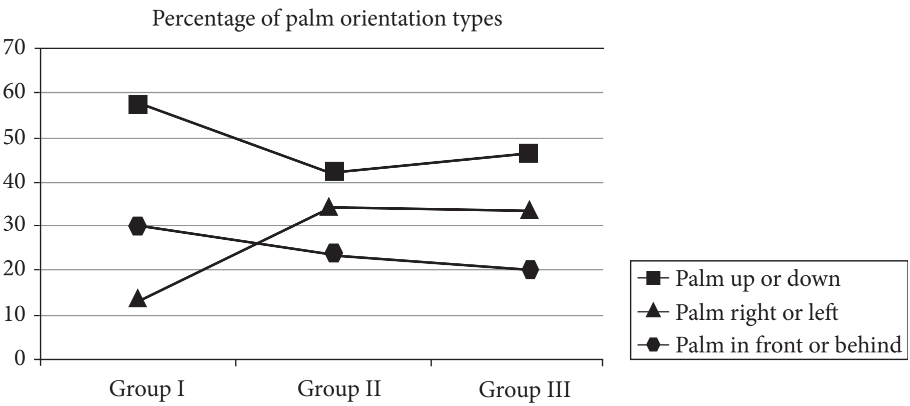


Figure 3.

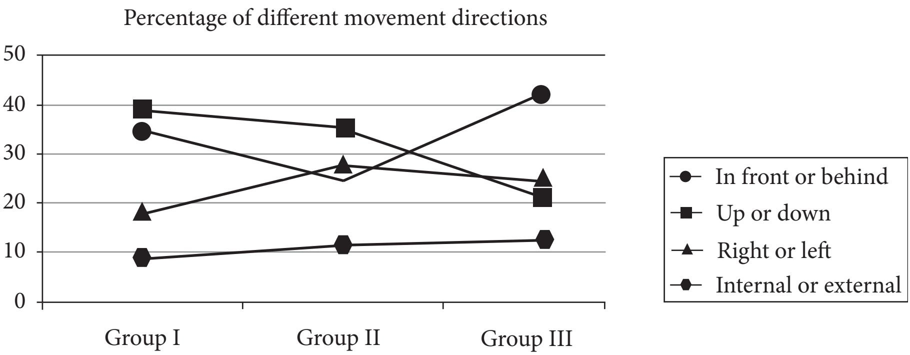


Figure 4.

Formal accuracy

As Table 3 shows, we calculated the mean score of the three groups for each of the three parameters considered.

It can be observed that the formal accuracy of execution increases with age, in particular from the group I to the groups II and III. Only the parameters of the 'shared space' appear to be already mastered by Group I

Representational correctness

For this parameter, we calculated the mean score of the three groups.

We noted an increase in the gestural representational correctness from the group I to the groups II and III as shown in Table 4.

Representational strategies

Figure 5 shows the percentages of representational strategies[4] used by the three groups of children.

The data demonstrate that the 'manipulation' strategy is the most used by all three groups and decreases with age like the 'mime' strategy, (particularly from Group I to Groups II and III) whereas 'depiction of shape/delimitation' and 'hand-becomes-objects' proportionally increase. The symbolic- conventional strategy is the less used by the three groups of children.

Table 3.

	Group I	**Group II**	**Group III**
Well boundedness	1.4	1.5	1.5
Formal clearness of the stroke	1.5	1.7	1.7
Shared space	1.9	1.9	1.9
Total Formal accuracy of execution	1.6	1.7	1.7

Table 4.

	Group I	**Group II**	**Group III**
Representative	2.4	2.5	2.6

4. The label S/C refers to the *symbolic-conventional* strategy

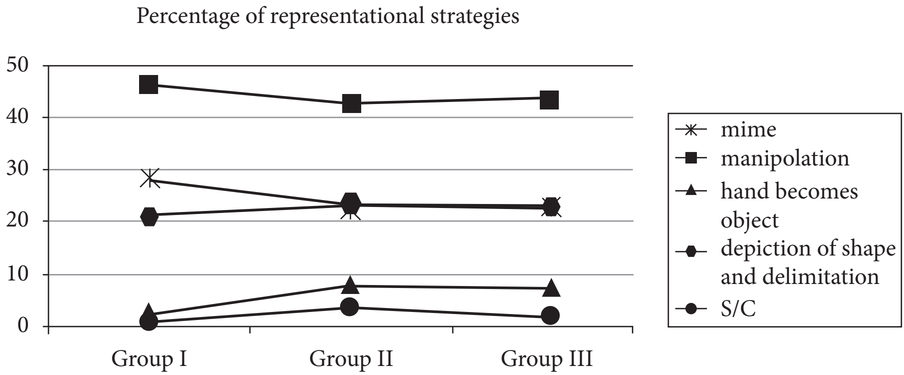


Figure 5.

Discussion and conclusion

One of our aims was to investigate whether and how it is possible to find a gradual mastering of the gestural form of expression as it happens in the mastering of the linguistic one. Our hypothesis of a parallelism between the development of the linguistic and gestural competence has been confirmed by the results which emerged in relation to all the phenomena investigated, starting with the parallelism between the increasing number of clauses and gestures in relation to age.

As for the analysis carried out on the formational parameters, it has shown a development of both the formal and the semantic aspects of gesture. The greater use that older children made of different hand orientations and locations testifies to an increasing ability to exploit the expressive resources of gesture. Also the results on the accuracy of gesture execution and the correct representation of their referents reveal a clear developmental trend. Whereas the former phenomenon testifies to an increasing mastering of the formal properties of gesture, the latter indicates that children have to learn how to use the gesture expressive components for representing in a proper way the aspects of the referents that gesture can codify.

The correlation between the acquisition of the formal and the semantic aspects of gesture compositional parameters demonstrates not only the children's increasing control of the semiotic properties of the gesture code, like for the linguistic code, but also that gesturing, like Sign Languages, constitutes an analytical and compositional system of expression.

Moreover, our research has showed that the motor constraints observed in the production of first signs by deaf children (Conlin et al. 2000) operate also in the way in which hearing toddlers use their gestures (Ann 1996). Such a result would support the notion of a continuum between gestures and signs rather than a clear boundary between non linguistic and linguistic systems (Pettenati et al. 2010).

The results obtained for children's use of the space of gesture execution, also indicate an increasing symbolic competence (McNeill 2000, 2005; Kita 2006). Indeed, as we saw, older children made a lesser use of their body in favor of their non-dominant hand, thus showing an increasing ability to move from a more concrete to a more abstract way of representing the referents related to those designated by the dominant hand.

A developmental trend in the acquisition of the symbolic competence emerged also in the analysis of the gesture representational strategies, which showed a gradual shift from the use of the most concrete and motivated (mime and manipulation) to that of the most abstract and conventional one (hand becomes object and shape depiction/delimitation). Gestural movement becomes less and less like real action in the physical world and becomes representationally more flexible: hand movement can represent something else than hand movement (Kita 2006).

The results of our analysis demonstrate that gesture, like speech and Sign Languages, has formal and semantic properties children have to acquire to develop their communicative competence. While the different semiotic identity between gesture and speech has been leading developmental researchers to investigate the way in which children learn to exploit their different expressive potentialities in order to integrate them into the multimodality of communication, the semiotic affinities between gesture and Sign Language can give scholars the possibility of investigating the similarities between the principles on which their representation of reality and the internal structure of their units are based.

References

Ann, J. 1996. "On the relation between ease of articulation and frequency of occurrence of handshapes in two sign languages." *Lingua* 98: 19–41

Bates, E., Benigni, L., Bretherton, I., Camaioni, L. and Volterra, V. 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.

Berman, R. A. and Slobin, D. I. 1994. *Relating Events in Narrative: A Crosslinguistic and Developmental Study*. Hillsdale (NJ): Erlbaum.

Boyes Braem, P. 1975. "The acquisition of handshape in ASL." Salk Institute Working Paper. The Salk Institute, La Jolla, CA.

Butcher, C. and Goldin-Meadow, S. 2000. "Gesture and the transition from one- to two-word speech: when hand and mouth come together." In *Language and Gesture*, D. McNeill (ed), 235–257. Cambridge: Cambridge University Press.

Calbris, G. 1990. *The Semiotics of French Gesture*. Bloomington: Indiana University Press.

Capirci, O., Iverson, J. M., Pizzuto, E. and Volterra, V. 1996. "Gestures and words during the transition to two-word speech." *Journal of Child Language* 23: 645–673.

Capirci, O., Caselli, M. C., Iverson, J. M., Pizzuto, E. and Volterra, V. 2002. "Gesture and the nature of language in infancy: the role of gesture as transitional device en route to two-word speech." In *The Study of Sign Languages: Essays in Honor of William C. Stokoe*, D. Armstrong, M. Karchmer and J. V. Van Cleeve (eds), 213–246. Washington, DC: Gallaudet University Press.

Capirci, O., Cristilli, C., De Angelis, V. and Ricci, S. 2008. "The relationship between speech and gesture in the development of textual and narrative competence." Paper presented at Multimod International Conference 2009. Toulouse, University of Toulouse II, 9–11 July, 2009.

Cassell, J. and McNeill, D. 1991. "Gesture and the poetic of prose." *Poetics Today* 12 (3): 375–404.

Clibbens, J. 1998. "Research on the acquisition of British Sign Language: Current issues." *Deafness and Education* 22: 10–15.

Colletta, J. M. 2004. *Le Développement de la Parole chez L'enfant Agé de 6 à 11 Ans. Corps, Langage et Cognition*. Sprimont: Mardaga.

Conlin, K. E., Mirus, G. R., Mauk, C. and Meier, R. P. 2000. "The acquisition of first signs: place, handshape and movement." In *Language Acquisition by Eye*, C. Chamberlain, J. P. Morford and R. I. Mayberry (eds), 51–69. Mahwah, NJ: Erlbaum.

Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. 1996. "Action recognition in the premotor cortex." *Brain* 119: 593–609.

Goldin-Meadow, S. and Butcher, C. 2003. "Pointing toward two-word speech in children." In *Pointing: Where Language, Culture and Cognition Meet*, S. Kita (ed), 85–107. Cambridge: Cambridge University Press.

Iverson, J. M. and Thelen, E. 1999. "Hand, mouth, and brain: the dynamic emergence of speech and gesture." *Journal of Consciousness Studies* 6: 19–40.

Karmiloff-Smith, A. 1985. "Language and cognitive processes from a developmental perspective." *Language and Cognitive Processes* 1: 61–85.

Karnopp, L. B. 2002. "Phonology acquisition in Brazilian Sign Languages." In *Directions in Sign Language Acquisition*, G. Morgan and B. Woll (eds), 29–53. Amsterdam: John Benjamins.

Kendon, A. 1985. "Some uses of gesture." In *Perspectives on Silence*, D. Tannen and M. Saville-Troike (eds), 215–234. Norwood, NJ: Ablex Publishing Corporation.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.

Kita, S. 2000. "Why do people gesture?" *Cognitive Studies* 7 (1): 9–21.

Kita, S. 2000. "How representational gestures help speaking." In *Language and Gesture*, D. McNeill (ed), 162–185. Cambridge: Cambridge University Press.

Kita, S. and Wood, C. 2006. "Semiotic characteristics of speech-accompanying gestures in children and adults' narrative." Paper presented at the Current Developments in Gesture Research. Hatfield, UK, June 30th 2006.

Kohler, E., Keysers, C. Umiltà, M. A., Fogassi, L., Gallese, V. and Rizzolatti, G. (2002). "Hearing sounds, understanding actions: action representation in mirror neurons." *Science* 297: 846–848.

McNeill, D. 1992. *Hand and Mind*. Chicago: The University of Chicago Press.

McNeill, D. 2000. *Language and Gesture*. Cambridge: Cambridge University Press.

McNeill, D. 2005. *Gesture and Thought*. Chicago: The University of Chicago Press.

Meier, R. P., Mauk, C., Cheek, A. and Moreland, C. 2008. "The form of children's early signs: Iconic or motor determinants?" *Language Learning and Development* 4: 63–98.

Morgan, G., Barrett-Jones, S. and Stoneham, H. 2007. "The first signs of language: Phonological development in British Sign Language.î *Applied Psycholinguistics* 28: 3–22.

Müller, C. 1998. "Iconicity and gesture." In *Oralité et Gestualité: Communication Multimodale, Interaction*, S. Santi, I. Guaïtella, C. Cavé, and G. Konopczynski (eds), 321–328. Paris: L'Harmattan.

Peterson, C. and McCabe, A. 1983. *Developmental Psycholinguistics: Three Ways of Looking at a Child's Narrative*. New York: Plenum Press.

Pettenati, P., Stefanini, S. and Volterra, V. 2010. “Motoric characteristics of representational gestures produced by young children in a naming task.” *Journal of Child Language* 37 (4): 887–911.

Rizzolatti, G. and Arbib, M. A. 1998. “Language within our grasp.” *Trends in Neuroscience* 21: 188–194.

Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. 1996. “Premotor cortex and the recognition of motor actions.” *Cognitive Brain Research* 3: 131–141.

Stein, N. L. and Glenn, C. G. 1979. “An analysis of story comprehension in elementary school children. “In *New Directions in Discourse Comprehension: Vol. 2: Advances in Discourse Processes*, R. O. Freedle (ed), 53–120. Norwood NJ: Ablex.

Stokoe, W. C. 1960. “Sign Language Structure: An outline of the visual communication systems of the American deaf.” Studies in linguistics. Occasional Papers 8. University of Buffalo.

Streek, J. 2008. “Depicting by gesture.” *Gesture* 8 (3): 285–301.

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. 2001. “I know what you are doing: a neurophysiological study.” *Neuron* 31: 155–165.

Volterra, V. 1987. *La Lingua Italiana Dei Segni*. Bologna: Il Mulino (New Edition 2004)..

CHAPTER 15

The changing role of gesture form and function in a picture book interaction between a child with autism and his support teacher

Hannah Sowden,[1] Mick Perkins[2] and Judy Clegg[3]
Newcastle University,[1] Sheffield University[2] and Sheffield University[3]

Autism is a developmental disorder which impacts on the social, communicative and cognitive abilities of the child. The development of both language and gesture is delayed. Previous research indicates that deictic gestures predominate over representative gestures in this population. This paper presents a case study, Nathan, aged 2:6 years interacting with his support teacher, Joanne. The five minute interaction comprises three distinct phases. In the first phase Joanne engages Nathan's attention by means of deictic gestures, the second phase shows an increase in iconic gesture, and in the final phase Nathan actively contributes to the interaction both verbally and gesturally. We conclude that Nathan is skilled at understanding and using deictic gestures, at imitating representative gestures and can collaboratively engage in interactions. This study indicates that children with autism may combine communicative modalities with more complexity than previously thought.

Introduction

Gesture, which occurs without conscious thought in everyday conversation, is a vital part of communication. We assume that it is an intentional action accompanying speech, usually performed by the hand or arm in the area of the upper torso in face to face interaction. Most people are adept at differentiating between intentional communicative movements and other movements such as mannerisms and fidgeting (Arendsen et al. 2007, Kendon 1978 cited in Kendon 2004). Although primarily hand and arm movements, a gesture will be performed by the part of the body which rewards economy of effort with successful communication. The head can be used to mark assent or dissent, or it can be used as a means of indicating by tilting and jerking actions. Facial expressions and the direct manipulation of objects are excluded from this definition of gesture.

There are several different forms of gesture (McNeill 1992), the most common of these being deictic, iconic, and beat gestures and emblems. Deictic gestures are used to identify a referent, most commonly by pointing and can be concrete or abstract. Concrete deictic gestures indicate a real world object or location. Abstract pointing is used to delineate a spatial area to represent the topic of conversation, for example, a downwards point referring to *here and now* compared to a backwards point referring to some event in the past. Iconic gestures bear some semantic relation to the speech they accompany. They represent a concrete real-world action or object, such as imitating the action of unscrewing a jar lid. Emblems are culturally specific (Morris 1979) and have a precise, paraphrasable meaning, resulting in the possibility of autonomous use. Some examples are the thumbs up sign for "OK" and the thumb rubbing on the tips of the fingers of a curled hand for "money" (Kendon 2004). Beats are rhythmic, repetitive and rapid movements, such as a flick of the fingers. They coincide with stressed syllables (Krauss et al. 2000) and are used, amongst other things, to emphasize speech.

Gesture features prominently in children's early attempts to communicate. Early gestures include requesting or showing objects and pointing (Blake & Dolgoy 1993, Blake et al. 2005). Gesture also plays a role in the support of first words and in the transition to two word speech (Butcher & Goldin-Meadow 2000, Capirci et al. 2005, Iverson et al. 1994, McEachern & Haynes, 2004, Pizzuto & Capobianco, 2005, Iverson et al. 2008).

Whilst much is known about this period of gesture development in typically developing children, less is known about children with communication disorders. Thus far gesture has been studied in relation to Down's Syndrome (Iverson et al. 2003), Williams Syndrome (Bello et al. 2004), Specific Language Impairment (Fex & Månsson 1998) and stuttering (Mayberry & Jacques 2000). This paper presents a case study, examining the gestures of a child with Autistic Spectrum Disorder (ASD).

Autistic spectrum disorder

ASD is a developmental disorder affecting social and communicative abilities in the child. ASD has been researched extensively since its identification in the 1940s, almost simultaneously by Kanner (1943) and Asperger (1944). Estimates for the prevalence of ASD vary from 30 to 60 cases per 10,000 (Rutter 2005). Diagnosis is made on the basis of behavioral information (American Psychiatric Association 1994, World Health Organization 1992). In order to be diagnosed with ASD, children must show impairment in social development, communication and imagination. Currently, reliable diagnosis is possible by the age of two to three years, often signalled by a delay in the development of language.

The social impairment is commonly attributed to deficits in the development of a theory of mind; that is the ability to understand that others may hold different beliefs to oneself. The ability to infer another's thoughts forms the bedrock of our social interactions. Other features of ASD such as the desire for sameness, need for routine, and repetitive and stereotypical actions have been related to deficits in executive functions.

These are higher order cognitive processes which determine priorities, plan actions, and control the ability to switch between tasks. A third proposal for explaining the nature of autism is based on the notion of "weak central coherence" (Frith 2003, Happé & Boot 2008), which claims that people with ASD have a processing style biased towards fine detail rather than global information. This may account for generalisation difficulties and preoccupation with often seemingly irrelevant details.

Gesture in autism

Much of the focus of research into gesture production in children with ASD has been on deictic gestures, which form the majority of the gestural repertoire of children with ASD (Sowden 2008). Pointing gestures are commonly used in two ways: the child either points at an object in order to request it (imperative pointing) or points at an object to comment on it and share experiences (declarative pointing). Declarative pointing is impaired in children with ASD, but imperative pointing is not (Loveland & Landry 1986, Baron-Cohen 1989, Camaioni et al. 1997, Camaioni et al. 2003, Stone et al. 1997). Stone and colleagues (1997) also reported a preference for contact over distal gestures and less use of eye gaze and vocalisations when commenting compared with typically developing children.

The difficulty of declarative pointing for children with ASD is part of the general impoverishment of joint attention skills and as such has been linked to theory of mind deficits (Baron-Cohen 1995, Tomasello & Camaioni 1997). Alternatively Stone et al. (1997) suggest that the impairment lies with the ability to monitor, rather than direct, the attention of another. This invokes the executive functions account as children cannot shift attention between the referent and the interlocutor. This also explains the preference for contact gestures, as monitoring becomes irrelevant with direct manipulation of another's hand. Regardless of the underlying cause, the lack of joint attentional behaviors and declarative pointing is so well attested in young children with ASD that it has been used as part of an early clinical marker of autism (Baron-Cohen et al. 1992), with promising results (Baron-Cohen et al. 1996, Charman et al. 2001).

In comparison with the interest in deictic gesture, other gesture forms have been less extensively researched. Emblems and iconics are known to be limited in both quantity and quality (Wetherby et al. 2004, Wetherby et al. 1998, Stone & Caro-Martinez 1990). Emblems have been reported in studies where the primary focus was deictic gesture (Camaioni et al. 1997, Camaioni et al. 2003, Stone et al. 1997), but these are infrequent, restricted to certain individuals and are mainly imitative and learnt during social routines.

In summary, children with ASD have predominantly deictic gestures in their repertoire. However, the declarative function, as realised through pointing and showing gestures, is an area of difficulty. In addition, children with ASD seem to show preference for contact gestures, such as manipulating the adult's hand and touching objects. Their use of emblems is limited and appears restricted to those learnt by imitation of social routines such as waving, nodding and shaking the head. Iconic gestures have not yet been demonstrated to be spontaneously produced.

Aims of this study

This paper presents a case study of a single interaction between Nathan, a child with ASD, and his support teacher, Joanne. The interaction was based around an animal picture book, which was used to stimulate talk about a wide range of animals. During the discussion both Joanne and Nathan used a variety of gestures to support their spoken word. The video-recording of the interaction was analyzed to investigate the following areas: (a) gesture forms, (b) discourse functions of the gestures and (c) the dynamic nature of gesture form and function in the co-construction of the interaction between Joanne and Nathan.

Method

This case study was taken from a larger longitudinal project which followed eight children for up to a year during their attendance at "Explorers", a first intervention programme aimed at facilitating socialisation and communication skills through naturalistic behavior-based intervention. Children attended three mornings a week, with each session lasting two and a half hours. Six to eight children attended at any one time, with four staff members per session. To be eligible for the project the children needed to have been diagnosed with ASD by a clinical psychologist and to have accepted a place on the Explorers programme.

Participants

The participants in the interaction discussed below are Nathan, 2:4 years old at the time of recording, and Joanne, an experienced and full time member of the Explorers team. Nathan had been attending the programme for one month when the recording took place. Pseudonyms are used throughout.

Although regular assessment is a part of the Explorers programme, standardised assessments are not used. Therefore, a profile of Nathan's core skills will be presented. The three assessments which form the basis of this profile are:

- *The Socialisation Checklist* (National Health Service 2006). This has been developed in-house and assesses the child's communication and behavior
- *The Living Language Detailed Profile* (Locke & Beech 1991). This covers physical, social and linguistic development.
- *The Surrey Speech Language and Communication Profile* (McGregor & Cave 1996). This provides a detailed assessment of the linguistic abilities of the child.

Further details of these assessments, including the scoring system and the contribution of each assessment to the overall profile are given in Appendix A.

Table 1. Nathan's profile of core skills

Socialisation	Communication			Independence	Physical skills
	Receptive	Expressive	Interaction		
Moderate	Severe	Severe	Severe	Moderate	Mild

As shown in Table 1, Nathan's socialisation impairment is moderate as measured by the Socialisation Checklist and the Living Language Profile. His independence and self-help are also moderately impaired. Physical skills and hand-eye co-ordination are good, with only a mild or no impairment. Nathan has the most difficulty with communication, and this severe impairment affects both receptive and expressive abilities. Nathan was beginning to use words productively at the time of the recording. Nathan also makes extensive use of immediate echolalia. In summary, Nathan has moderate autism, which impacts most severely on his communication skills.

Procedure

Nathan was video-recorded for twenty minutes once every two weeks throughout his time in the Explorers programme. The interaction forming this case study occurred towards the end of the first recording, in the final hour of the Explorers session. The interaction with the picture book was initiated by Joanne and lasted for approximately five minutes.

Analysis of data: the video footage was transcribed in detail and analysed qualitatively. Gestures for both Joanne and Nathan were identified and classified according to form. Analysis was based on principles of Conversation Analysis (CA). This approach assumes that interaction is dynamic and is co-constructed by the interacting participants. No predetermined categories are used in the data analysis; instead, salient behaviors of potential interest are examined in their sequential context, and any replications are further scrutinised as possible evidence of more general patterns (Hutchby & Wooffitt 1998). By studying only that which is directly observable, CA is inherently empirical and analysis is driven by the data. Transcription conventions are given in Appendix B.

Results

Three extracts from the interaction will be presented in chronological order and analysed in terms of the areas of investigation: identification of gesture forms, linking gesture form and function, and the dynamic role of gesture in the co-construction of the interaction. Links between each extract will be brought out in the discussion.

Establishing attention: deictic gestures

In this first extract (Extract 1) Joanne is reading the first page of the book to Nathan. It is a rhyme about four puppies. The extract begins part way through.

Extract 1:

1. JOA: three little puppies
 pointing at picture of three puppies in turn
2. JOA: what could I do?
 moving Nathan's hand from the writing
3. JOA: I took the black one home then there were
 pointing at black puppy
4. (.) one (.) two
 pointing at each puppy
5. *Nathan imitates points at the puppies*
6. JOA: two little puppies playing in the sun
 pointing at each puppy
7. *Nathan imitates points*
8. JOA: I took the grey one home
 tapping grey puppy moves tapping to last puppy
9. JOA: then there was?
 holds finger up
10. *Nathan looks at Joanne then plays with flap*
11. NAT: one
 Looks round at noise then back and sees her finger
12. JOA: one
13. NAT: o[ne] [one]
 Imitates finger held up
14. JOA: [one] little [pup]py looking very sad
 pointing at single puppy *pointing at text*
15. *Nathan copies point to puppy*
16. JOA: I took it home and then there were none (.)
 finger traces writing
17. look

The predominant gesture form throughout this extract is the deictic gesture. Joanne makes use of a range of deictic forms: pointing at the pictures accompanying the text (lines 1, 3, 4, 6 and 14), rapidly repeated pointing which makes contact with the book (line 8) and tracing the relevant text (lines 14 and 16). Nathan copies many of these gestures and also points at the pictures in the book (lines 5, 7, and 15).

In this first extract, Joanne is working hard to establish Nathan's attention on the book. They have just started to work together and are collaboratively negotiating how the book should be used. Joanne is using deictic gestures to indicate pictures to Nathan which become relevant as the text progresses. She is using the book in a focused and literal way without deviation from the written text. The deictic gestures serve primarily to support the story revealed through the reading (lines 1, 3, 4, 6 and 14). The tracing gestures (lines 14 and 16) explicitly demonstrate the link between the text and Joanne's expectation of Nathan's attention.

Despite these efforts, it is doubtful whether Nathan is either attentive to the text or understands it. He demonstrates a willingness to interact through imitation of Joanne's gestures (lines 5, 7, 13 and 15), but for Nathan, the book has a less central role. He does not wait for the accompanying text, but indicates each picture in the sequential order in which it appears on the page. Thus, a difference can be perceived; Joanne links the pictures to the text, whereas Nathan focuses on their sequential relationship. Consequences of this difference can be traced in the interaction: Nathan indicates the latter puppies in advance of Joanne (lines 5 and 7) and begins to lose concentration when he has indicated all the pictures (lines 8 and 9).

Joanne responds to his inattention (line 8) by changing the form of the gesture from a single point to rapid multiple taps. This strategy engages Nathan's attention sufficiently for Joanne to use an iconic gesture to represent the idea of "one" puppy. Nathan copies this gesture and its verbal accompaniment. The extract ends with joint attention on opening the flap at the end of the rhyme.

To summarise, this extract contains predominantly deictic gesture which Joanne uses to direct Nathan's attention to different parts of the story. Nathan does demonstrate some awareness of joint attention, directing his attention where Joanne indicates through her gestures. However, he does not attempt to direct her attention and is easily distracted by the flap in the book and noises in the room (lines 10–11).

Talking about animals: Emblems and iconic gestures

The second extract is taken from approximately half way through the interaction. Joanne and Nathan have started a new page in the book. At the top of the page is a picture of a rabbit, underneath which is a rabbit hutch concealing a sleeping rabbit. The rabbit can be viewed by opening a flap. The pictures are accompanied by relevant text.

Joanne no longer relies on the text to talk about the rabbits; instead, the pictures provide a stimulus for a more wide ranging discussion. Without the rigid adherence to the text the necessity for deictic gesture is reduced; Joanne only points in lines 1, 3, and 8. Nathan effortlessly follows the conversation without the additional support of multiple deictic gestures and maintains joint attention throughout the majority of the extract.

Extract 2

1.	JOA:	what do you see?
		Pointing at picture in book
2.	NAT:	*you see?*
3.	JOA:	a: (.) rabbit hop (.) hop (.) hop (.) hop
		Tapping picture *two fingers together move up and down across page*
4.	NAT:	hop [hop (.) hop (.) hop (.) hop (.) hop]
		Moving flat hand up and down in time
5.	JOA:	[hop (.) hop (.) hop (.) hop (.) hop] (.) they're
		two fingers together move up and down across page
6.		hopping (0.2) who's inside?
		opening flap
		(music starts from toy behind Nathan)
7.	NAT:	who's inside?
8.	JOA:	ssh asleep (.) Nathan (.) sssh sleeping
		Finger on lips *point at picture In book* *head on hands and mime sleep*
9		shhh asleep
		Finger on lips then closes flap

Joanne repeatedly introduces and elaborates on a new topic in a similar fashion throughout the book. First, she employs an attention-directing expression (lines 1 and 6). This is often accompanied with a deictic gesture which helps to direct Nathan's attention to the relevant part of the page (lines 1, 3 and 8). Nathan often signals his attention by imitating part of Joanne's immediately prior turn (lines 2 and 7).

After successfully directing Nathan's attention, Joanne provides a verbal label for the animal, in this case "rabbit" (line 3), then represents a characteristic of that animal by means of gesture. For the first rabbit, this is an iconic gesture representing the way the rabbit moves (lines 3 and 5). For the second rabbit, she indicates sleep by the emblematic gesture of resting her head on her joined hands (line 8) and requesting Nathan to be quiet by placing her finger on her lips, an emblem gesture where the right index finger is vertically extended from a fist hand shape and placed across the centre of the lips (lines 8 and 9).

To summarise this sequence, Joanne introduces a new topic firstly by directing Nathan's attention with a deictic gesture and attention-directing expression, secondly by providing a verbal label, and thirdly by giving a gestural description of the animal. This results in fewer deictic gestures. Instead the interaction is dominated by symbolic gestures. As before, Nathan imitates the gestures (line 4) but does not spontaneously produce any iconic gestures himself.

Co-constructing interaction: Mixing gestures

The final extract (Extract 3) is taken from the penultimate page of the book. At the top of the page are four members of the big cat family, including a lion. Underneath are three penguins. As before there is accompanying text, but this is ignored by both Joanne and Nathan. This extract is notable for the increasingly extensive role that Nathan assumes in the interaction as he begins to introduce topics and drive the interaction himself.

Extract 3:

1.	JOA	look
2.	NAT	meow
		points to the cat picture then looks at Joanne
3.	JOA	meow (.) are they cats?
		nods and points to cats then looks back to Nathan
4.	NAT	they're cats
5.	JOA	this one's a li[on]
		points to picture of lion
6.	NAT	[mmm]
		points to picture of penguin
7.	JOA	ra:
		hands as claws pouncing
8.		*Nathan points to penguin on far side of picture*
9.	JOA	l[ion]
		points to lion
10.	NAT	[ra:]
		maintains pointing but looks at Joanne then back to picture
11.	JOA	penguin
		points to penguin looking at Nathan
12.	NAT	penguin
		points to 3rd penguin
13.	JOA	this one goes
		picks up Nathan's hand and moves it to lion picture
14.	JOA	this one goes ra:
		taps lion picture *hands as claws*
15.	NAT	ra[:]
		Imitates and looks at her
16.	JOA	[ra:]
		Hands as claws then turns page

In this extract Nathan actively seeks interaction and has the necessary joint attention skills to ensure that his attempts are successful. Nathan introduces topics (lines 2 and 10) and sustains them over several turns. At the start of the extract, Nathan directs Joanne to the cats by combining a verbal label (meow) with a pointing gesture. Nathan looks at Joanne, checking to see that she has followed his cues before returning his gaze to the picture (lines 2–3). Joanne accepts the topic, and Nathan maintains it over another turn (line 4).

In line 6 Nathan attempts to introduce the penguins with a vocalisation and pointing gesture. However, he does not look at Joanne, and she chooses instead to elaborate on the lion by means of an iconic pouncing gesture (fingers spread and hooked representing the lion's claws). This topical misalignment continues through lines 8–10. Joanne elaborates on the initial lion topic through the naming sequence which has been described previously. First, she combines a deictic gesture with an attention-directing expression (line 3). This is followed by a label "lion" and another deictic gesture (line 5), before elaborating with the iconic pouncing gesture (line 7) and finally repeating the label (line 9). During this sequence Nathan attempts to reintroduce penguins by pointing at a different penguin picture (line 8), but again does not look at Joanne.

By lines 10 and 11 Joanne has completed the naming sequence and is ready to pursue the penguin topic, whereas Nathan attempts to respond to Joanne's lion sequence by imitating her roaring whilst maintaining the penguin point and also looking at her (line 10). Joanne accepts these cues and labels the penguin in line 11. Nathan confirms with his next turn (line 12) before Joanne firmly re-establishes the lion topic by physically moving Nathan's hand to the lion picture (line 13). On establishing the lion topic, Joanne once more elaborates with the iconic pouncing gesture. Nathan copies her and they turn the page together.

Discussion

The aim of this study was to illuminate three different aspects of gesture use by conducting a qualitative analysis of a picture book interaction between a child with autism and his support teacher. These were firstly to identify what gesture forms were used by both Joanne and Nathan, secondly to identify gesture functions used by both and to investigate links between gesture form and function, and finally to investigate the changing role the gesture plays in the unfolding interaction. Each of these aspects will be discussed in turn.

As may be expected Joanne makes use of a range of different gesture forms. Near the beginning of the interaction deictic gestures dominate. Several different forms of deictic gesture were observed including proximal pointing, repetitive pointing, tracing words and physically moving Nathan's hand to the relevant picture. In addition to deictic gesture, Joanne also employed iconic gestures and some emblems in the later extracts.

In line with previous findings, Nathan's repertoire of gestures was more restricted, consisting predominantly of deictic gestures. Specifically, in the first two extracts he spontaneously produced very few gestures, and in most cases gestures were an immediate imitation of Joanne's. Although there was no evidence of spontaneous use of iconic gestures or emblems by Nathan, he imitated these in all three extracts. It is not currently clear how this apparent lack of imagistic gestures may reflect the wider impairments associated with autism, nor whether the proposed cognitive accounts underlying the impairments may be extended to fit this pattern of gesture use.

In terms of the second area of interest, that of gesture functions, the analysis confirmed the dynamic nature of functions throughout the interaction. Gesture consistently supported and reinforced speech, but changed with regard to the intention of the speech it accompanied. In the first extract Joanne deliberately restricted herself to the text in the book. This dictated a sequential ordering for the discussion of pictures and proscribed the duration of each discussion. In effect gesture punctuated the speech, highlighting each picture as it became relevant to the unfolding story. Gesture supported the narrative structure, rather than carrying propositional content. As the interaction progressed, Joanne became less reliant on the text of the book, and gesture increasingly conveyed semantic content. This is particularly evident in Extract 2, where gesture demonstrates how rabbits move, that a rabbit is asleep and that quietness is needed to avoid waking the rabbit. Although gesture is an ancillary system to speech, it can support both narrative structure and propositional content.

The functions of Nathan's gesture are very different. Initially he does not seek to direct Joanne's attention, only doing so towards the end of the interaction. Instead his gesture could be considered to perform a "back-channelling" role – i.e., filling his conversational turn and signalling his engagement with Joanne. A parallel emerges between the use of immediate echolalia in speech and immediate imitation of gesture in the manual modality.

There are several links observed between gesture form and function. Joanne initially uses deictic gesture to situate the text in relation to the page in the hope of facilitating Nathan's ability to attend to the book and follow the speech. The provision of additional support enables Nathan to engage with Joanne and gradually enter into sustained joint attention. Once achieved, a qualitative change can be discerned in the interaction. Joanne eschews the text as a means of providing access to the book for Nathan, and thus pictures constitute the basis of a freer interacting style. This in turn leads to an increase in iconic and emblem gestures as animals are described more fully and gesture takes on more propositional content. Due, in part, to his ability to enter joint attention Nathan is able to cope with these increased demands, demonstrating his engagement through gestural imitation. The final extract reflects a further phase of the interaction, with both Joanne and Nathan combining deictic and iconic gesture to introduce and sustain different topics of conversation.

Conclusion

Although it is not possible to generalise from a single case study, the data discussed here have revealed that gesture form and function are intricately linked and arise as a consequence of the nature of the interaction and engagement of interlocutors. Whilst confirming previous findings regarding gesture form in the communication of children with autism – for example the predominance of deictic over iconic gesture use, this study has also indicated that gesture use in this population may be more complex and varied than previously thought.

References

American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington D.C.: Author.

Arendsen, J., van Doorn, A. and de Ridder, H. 2007. "When and how well do people see the onset of gestures?" *Gesture* 7 (3): 305–342.

Asperger, H. 1944. "Die "Autistischen Psychopathen" im Kindesalter." *Archiv fur Psychiatrie und Nervenkrankheiten* 117: 76–136.

Baron-Cohen, S. 1989. "Perceptual role taking and protodeclarative pointing in autism." *British Journal of Developmental Psychology* 7: 113–27.

Baron-Cohen, S. 1995. *Mindblindness: An Essay on Autism and the Theory of Mind*. Cambridge, Mass: MIT Press.

Baron-Cohen, S., Allen, J. and Gillberg, C. 1992. "Can autism be detected at 18 months? The needle, the haystack and the CHAT." *British Journal of Psychiatry* 161: 839–43.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., Drew, A. and Charman, T. 1996. "Psychological markers in the detection of autism in infancy in a large population." *British Journal Of Psychiatry* 168 (2): 158–163.

Bello, A., Capirci, O. and Volterra, V. 2004. "Lexical production in children with Williams syndrome: spontaneous use of gesture in a naming task." *Neuropsychologia* 42 (2): 201–213.

Blake, J. and Dolgoy, S. J. 1993. "Gestural development and its relation to cognition during the transition to language." *Journal of Nonverbal Behavior* 17 (2): 87–102.

Blake, J., Vitale, G., Osborne, P. and Olshansky, E. 2005. "A cross-cultural comparison of communicative gestures in human infants during the transition to language." *Gesture* 5 (1/2): 201–218.

Butcher, C. and Goldin-Meadow, S. 2000. "Gesture and the transition from one to two word speech: when hand and mouth come together." In *Language and Gesture*, D. McNeill (ed), 235–258. Cambridge: Cambridge University Press.

Camaioni, L., Perucchini, P., Muratori, F. and Milone, A. 1997. "Brief report: A longitudinal examination of the communicative gestures deficit in young children with autism." *Journal of Autism and Developmental Disorders* 27 (6): 715–725.

Camaioni, L., Perucchini, P., Muratori, F., Parrini, B. and Cesari, A. 2003. "The communicative use of pointing in autism: developmental profile and factors related to change." *European Psychiatry* 18 (1): 6–12.

Capirci, O., Contaldo, A., Caselli, M. and Volterra, V. 2005. "From action to language through gesture: A longitudinal perspective" *Gesture* 5 (1/2): 155–177.

Charman, T., Baron-Cohen, S., Baird, G., Cox, A., Wheelwright, S., Swettenham, J. and Drew, A. 2001. "Commentary: The modified checklist for autism in toddlers." *Journal Of Autism And Developmental Disorders,* 31, 145–148.

Fex, B. and Månsson, A.-C. 1998. "The use of gestures as a compensatory strategy in adults with acquired aphasia compared to children with specific language impairment (SLI)." *Journal of Neurolinguistics* 11 (1–2): 191.

Frith, U. 2003. *Autism: explaining the enigma*. Oxford: Blackwell.

Happé, F. G. E. and Booth, R. D. L. 2008. "The power of the positive: Revisiting weak coherence in autism spectrum disorders." *Quarterly Journal Of Experimental Psychology* 61: 50–63.

Hutchby, I. and Wooffit, R. 1998. *Conversation Analysis: Principles, Practices and Applications.* Cambridge: Polity Press.

Iverson, J. M., Capirci, O. and Caselli, M. 1994. "From communication to language in two modalities." *Cognitive Development* 9: 23–43.

Iverson, J. M., Longobardi, E. and Caselli, M. 2003. "Relationship between gestures and words in children with Down's syndrome and typically developing children in the early stages of communicative development." *International Journal of Language & Communication Disorders* 38 (2): 179–197.

Iverson, J. M., Capirci, O., Volterra, V. and Goldin-Meadow, S. 2008. "Learning to talk in gesture-rich world: Early communication in Italian vs. American children." *First Language* 28: 164–181.

Kanner, L. 1943. "Autistic disturbances of affective contact." *Nervous Child* 2: 217–50.

Kendon, A. 1978. "Differential perception and attentional frame: two problems for investigation." *Semiotica* 24: 305–315.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Krauss, R., Chen, Y. and Gottesman, R. 2000. "Lexical gestures and lexical access: a process model." In *Language and Gesture*, D. McNeill (ed), 261–283. Cambridge: Cambridge University Press.

Locke, A. and Beech, M. 1991. *Living Language*. London: Nelson Publishing Company.

Loveland, K. A. and Landry, S. H. 1986. "Joint attention and language in autism developmental language delay." *Journal of Autism and Developmental Disorders* 16 (3): 335–349.

Mayberry, R. and Jaques, J. 2000. Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In *Language and Gesture*, D. McNeill (ed), 199–214. Cambridge: Cambridge University Press.

McEachern, D. and Haynes, W. O. 2004. "Gesture-Speech Combinations as a Transition to Multiword Utterances." *American Journal of Speech-Language Pathology* 13: 227–235.

McGregor, T. and Cave, D. 1996. "Collaboration across six trusts." *Bulletin of the Royal College of Speech and Language Therapists* 533: 6.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Morris, D., Collett, P., Marsh, P. and O'Shaughnessy, M. 1979. *Gestures: Their Origins and Distribution*. London: Cape.

National Health Service ASD Team. 2006. *Checklist for Providing Baseline of Levels for Socialisation/ Communication Development (Early Years) Non School Environment*. Sheffield: Ryegate Children's Centre.

Pizzuto, E. and Capobianco, M. 2005. "The link (and differences) between deixis and symbols in children's early gestural-vocal system." *Gesture* 5 (1/2): 179–200.

Rutter, M. 2005. "Incidence of autism spectrum disorders: Changes over time and their meaning." *Acta Paediatrica* 94 (1): 2–15.

Sowden, H. 2008. *Investigating Gesture in Children with Autism: Development, Input and Interaction.* Unpublished doctoral dissertation. Sheffield, University of Sheffield.

Stone, W. L. and Caro-Martinez, L. M. 1990. "Naturalistic observations of spontaneous communication in autistic children." *Journal of Autism and Developmental Disorders* 20: 437–453.

Stone, W. L., Ousley, O. Y., Yoder, P., Hogan, K. and Hepburn, S. 1997. "Nonverbal communication in two- and three-year-old children with autism." *Journal of Autism and Developmental Disorders* 27 (6): 677–696.

Tomasello, M. and Camaioni, L. 1997. "A comparison of the gestural communication of apes and human infants." *Human Development* 40 (1): 7–24.

Wetherby, A. M., Prizant, B. M. and Hutchinson, T. 1998. "Communicative, social/affective, and symbolic profiles of young children with autism and pervasive developmental disorders" *American Journal of Speech-Language Pathology* 7 (2): 79–91.

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H. and Lord, C. 2004. "Early indicators of autism spectrum disorders in the second year of life." *Journal of Autism and Developmental Disorders* 34 (5): 473–493.

World Health Organisation. 1992. *International Classification of Diseases: Diagnostic Criteria for Research* (10th ed.). Geneva, Switzerland: Author.

Acknowledgements

Thanks go to the illustrator, Nick Coupland.

Appendix A: Nathan's profile of core skills

The table below shows which sub-sections of each assessment contributed to the four major skill classes (Socialisation, Communication, Independence and Physical Skills).

Scoring criteria for each assessment:

A classification for severity of impairment was made for each assessment based on the following criteria:

- Socialisation Checklist and Surrey Speech, Language and Communication calculated by taking range of possible scores and dividing by three:
- Severe = lowest range, Moderate = mid range, Mild = highest range
  Living Language Detailed Profile
  calculated by length of observed delay:
  severe = 12+ months, moderate = 6–12 months, mild = 0–6 months

Table 2. sub-sections of the assessments

	Major skill classes					
		Communication				
Assessment	**Socialisation**	**Receptive**	**Expressive**	**Interaction**	**Independence**	**Physical skills**
Socialisation Checklist	Adaptability to rules and routines Socialisation	Communication			Learning independence	n/a
Living Language	Play and social development	Listening and understanding	Expressive	n/a	Self help and independence	Physical skills Eye and hand co-ordination
Surrey Speech, Language and Communication Profile	Behavior	Receptive	Expressive speech production	Impact Interaction	n/a	n/a

Appendix B: Transcription conventions

JAC	identifies the speaker
:	lengthened vowel
?	rising intonation
!	exclamation
(.)	pause
(0.4)	timed pause
[went up]	over lapping speech
{coughing}	meta-linguistic information
CAPITAL	emphasis
Italic	movement, actions or gesture

PART III

Second language effects on gesture

CHAPTER 16

A cross-linguistic study of verbal and gestural descriptions in French and Japanese monolingual and bilingual children

Meghan Zvaigzne, Yuriko Oshima-Takane, Fred Genesee[1] and Makiko Hirakawa[2]
McGill University[1] and Bunkyo University[2]

This study investigated whether the presence of mimetics (sound-symbolic words) in language influences children's verbal and gestural descriptions by comparing monolingual and bilingual speakers of Japanese and French. Mimetics are present in Japanese, but not French (Kita 2008). 4 to 6-year-old children described motion and object characteristics to an experimenter during a referential communication task. Verbal descriptions were coded as precise or imprecise and produced with or without mimetics and/or iconic gestures. Mimetics and gestures were used frequently in Japanese, particularly for motion descriptions. Bilinguals patterned like monolinguals, except when speaking Japanese they used more imprecise descriptions and fewer mimetics. This shows that presence of mimetics in language and frequent exposure to them promotes their use in conjunction with gestures.

Keywords: Iconic gestures, verbal description, cross-linguistic comparison, bilinguals.

1. Introduction

Children's and adults' speech is frequently accompanied by spontaneous hand and arm movements, called co-speech gestures (Mayberry & Nicoladis 2000, McNeill 1992). McNeill (1992) postulated that speech and gesture are closely related and that both are integral to understanding the speaker's message. Gestures, in particular, often convey more precisely the imagistic components of a message. Hence, gestures relate semantically to the speech they accompany, and they may or may not express the same information. One type of semantically related gestures is iconic gestures. These gestures refer to concrete things like events or objects (e.g. moving a hand continuously in a circle while saying *he is rolling*). The present study investigated whether the precision of children's verbal descriptions and their use of iconic gestures were influenced by (1) the specific language spoken, (2) whether the speaker was monolingual or bilingual, and (3) the type of information described.

Languages vary in the extent to which they contain highly imagistic words (e.g. onomatopoeia). For example, Kita (2001) reports that the use of mimetics, sound symbolic words, is frequent in Japanese (*giongo/gitaigo*). Mimetics are a class of words that vividly encode information about physiological, psychological, and affective states (e.g. heavy, tired, negativity) and events (e.g. repetition, manner of movement) experienced via all sensory modalities (e.g. vision, touch). Kita found that Japanese-speaking adults produced iconic gestures with a mimetic 95% of the time. Allen et al. (2007) found that Japanese-speaking children and adults used mimetics, but did not examine gesture use. The present study investigated whether the availability of mimetics in a language can contribute to children's verbal descriptiveness and use of iconic gestures by contrasting two languages (Japanese and French) that differ in this respect. Japanese has many mimetics, whereas French has few words (onomatopoeias) that could be considered to have mimetic properties (Kita 2001, 2008).

In this study, we also wanted to compare monolingual and bilingual children's verbal and gestural descriptions. With bilinguals, it is possible to compare performance in two languages within the same individual while controlling for cognitive ability and cultural experience (Nicoladis 2002). We investigated whether French-Japanese bilinguals would speak and gesture like French and Japanese monolinguals when using each language. If bilinguals follow language-specific patterns, we would have further support that the properties of one's language influence how one uses speech and gesture to describe things. Furthermore, by comparing bilinguals and monolinguals, we can examine whether language ability relates to verbal descriptiveness and iconic gesture use. The language ability of bilinguals may differ from that of monolinguals insofar as they may have smaller vocabularies, often due to their reduced exposure to one or both of their languages. This could result in less descriptive speech by bilinguals. Gestures may thus be used to compensate for lower language ability, and bilinguals might be expected to use more gestures than monolinguals, at least in their less proficient language (Nicoladis 2007).

People's descriptions may also be affected by what they are describing and, in particular, how they describe animated motion events (Kita & Özyürek 2003; McNeill & Duncan 2000; Özyürek, Kita, Allen, Furman, & Brown 2005; Stam 2006, 2008). These researchers found that the information expressed in gestures often mirrored that expressed in speech (i.e. path or manner of movement), but gesture sometimes conveyed additional information (e.g. path, manner, direction). No systematic cross-linguistic studies have examined speech and gesture use for object descriptions. However, we know that English-speaking children and adults gesture about an object's shape, size, and position (Church & Goldin-Meadow 1986, Holler & Beattie 2003, Riseborough 1982). Moreover, the information in gesture does not always match that conveyed in speech (Church & Goldin-Meadow 1986). In the present study, we compared descriptions of motions and objects by French and Japanese speakers to investigate whether the type of information being described would affect the children's verbal and gestural descriptions. Descriptions were elicited using a referential communication task (RCT) where children described the difference between two animated animal cartoons to an experimenter. The cartoons differed in one characteristic: the manner of the animal's movement (motion characteristic), or the shape or size of the animal (object characteristic).

We hypothesized that iconic gestures would accompany Japanese verbal descriptions more frequently than French verbal descriptions because Japanese speakers frequently use mimetics, while French has few such words (Kita 2001, 2008). Descriptions by monolinguals and bilinguals were expected to differ only if the groups differed in language proficiency to a degree that would influence performance on the RCT. If bilinguals could not verbally describe the scene characteristics, they might compensate with increased use of gestures (Gullberg 1998). Furthermore, it was expected that the dynamic nature of motion events would result in higher gesture use when children described motions compared to objects.

2. Methods

2.1 Participants

Eleven French-Japanese bilingual (7 male, mean age 5:8, range 4:2 to 6:7), 12 French monolingual (3 male, mean age 5:0, range 4:1 to 6:7), and 12 Japanese monolingual (4 male, mean age 5:4, range 5:0 to 5:10) children participated. Four of the bilingual children were French-dominant, four were Japanese-dominant and three were balanced according to their vocabulary size in each language as assessed with the Expressive One-Word Picture Vocabulary Test (Academic Therapy Publications Inc., 2000). The bilinguals were recruited from a Japanese language school and Japanese culture center in Montréal, Canada. The French monolinguals were recruited from a participant database of families living in the greater Montréal area, and the Japanese monolinguals were recruited from a daycare in Tokyo, Japan. To be included in the study, the bilingual children's exposure to French and Japanese had to total to 90%, and they had to be able to perform the RCT when using each language. The monolinguals had to have been exposed to their respective language at least 90% of the time. All monolinguals met this criterion, but due to difficulties in finding age appropriate French-Japanese bilinguals, we included two bilingual children who were exposed to French and Japanese for a total of 70% to 80% of the time (and thus had exposure to another language 20% to 30% of the time).

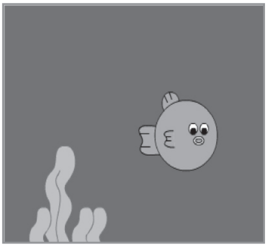
2.2 Materials and apparatus

The RCT used to elicit verbal and gestural descriptions consisted of eight pairs of animated cartoons that differed on one scene characteristic related to the animal depicted in the cartoon. Four pairs differed with respect to the animal's motion characteristics (manner of movement) and four differed with respect to the animal's object characteristics (shape, size). See Figure 1 for example cartoon pairs and Table 1 for the scene characteristics of each cartoon pair. Two sets of cartoons were created because the bilinguals performed the task twice (once in each language). One set was used for all French sessions (monolinguals and bilinguals), and the second was used for all Japanese sessions. The animals and backgrounds differed in each set, but the scene characteristics remained the same. Three practice pairs were created to give the children experience with each type of scene characteristic.

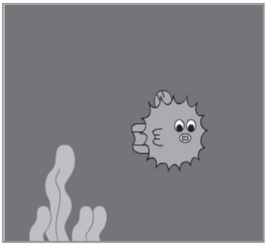
During the experiment, the child and experimenter sat facing each other at a small table. The experimenter viewed the cartoons on a Dell Inspiron laptop, which was connected to an LCD ViewSonic monitor on which the child viewed the cartoons. A Javascript program displayed the animated cartoons side by side on the screens, and a yellow star was placed above the target cartoon. The child was instructed to describe the scene characteristics so that the experimenter could guess the target cartoon. The animations played repeatedly until the experimenter "guessed" by pressing a key to indicate her choice.

a. Wings still

b. Wings flap



c. Smooth



d. Spiky

Figure 1. Still image examples of cartoon pairs depicting the animals' (a, b) motion characteristics and (c, d) object characteristics.

The experiment also had two visibility conditions. In half the trials, the child and experimenter could see each other (visible), and in the other half they could not see each other (non-visible) because a cardboard wall was placed between the child and experimenter. Since we found that the children gestured in both visibility conditions (see Zvaigzne, Oshima-Takane, Groleau, Nakamura, & Genesee 2008), we collapsed our data across both visibility conditions for the purposes of this paper.

The children's expressive vocabulary level was assessed using the Expressive One-Word Picture Vocabulary Test (EOWPVT, Academic Therapy Publications Inc., 2000). The children were shown pictures of objects or activities, and they had to name the objects or actions. This test was created for and normed with English-speaking children in the United States; thus, we modified the administration and scoring for our participants. First, the test began at the 3-year-old level for everyone in case the bilinguals had less vocabulary than the same-age monolinguals. Second, we omitted 18 items during scoring because they were perceived to be culturally specific (e.g. windmill). Third, raw scores were calculated by summing the number of correct items from Item 10 onward until the child failed five consecutive items. There are no norms available for French- or Japanese-speaking children; thus, our analyses are based on raw scores. The EOWPVT was administered according to test guidelines except for the changes described.

A Language Environment Questionnaire was completed by the children's parents. This questionnaire asked for demographic information and language experience (e.g. exposure to French, Japanese, and other languages in various settings).

2.3 Procedure

All participants were tested individually. The French monolinguals and bilinguals were tested in a large playroom at a university laboratory. The Japanese monolinguals were tested in a small room at their daycare. Monolinguals had one session; bilinguals had

Table 1. Scene Characteristic differences of the cartoon pairs. Characteristic differences listed first were those of the target cartoons

Scene characteristic	Characteristic difference	Animal for French	Animal for Japanese
Motion characteristic	Flapping wings, still wings	Bird	Butterfly
	Swing, jump	Monkey	Squirrel
	Rolling, sliding	Dog	Pig
	Jumping, running	Frog	Rabbit
Object characteristic	Spiky, smooth	Fish	Lizard
	Square, round	Bug	Turtle
	Fluffy, smooth	Cat	Dog
	Fat, thin	Bird	Mouse

one French and one Japanese session, scheduled one to three weeks apart. The order of language was counterbalanced across participants. French monolinguals and bilinguals completed the RCT, followed by the EOWPVT. The task order was reversed for Japanese monolinguals. All experimental sessions were video-recorded.

The experimenter described the RCT as a guessing game. Using the practice trials, the experimenter explained that they would see two cartoons side by side which were exactly the same except for one difference (scene characteristic). The child had to find the difference and give the experimenter clues so she could guess which cartoon had the star. Children who had difficulty were encouraged with questions unrelated to the scene characteristic (e.g. *are they the same color?).* To keep the children motivated, they received stickers throughout the task. After the practice trials, eight test trials were presented in total with four trials in each visibility condition. The order of the visibility conditions was counterbalanced across participants. In addition, the order of the first and second sets of four test trials was counterbalanced across participants.

2.4 Coding

Native or near native speakers of French and Japanese transcribed the children's and experimenter's speech verbatim in CHAT format for French (MacWhinney 2000) and JCHAT format for Japanese (Oshima-Takane, MacWhinney, Sirai, Miyata, & Naka 1998). The CHAT and JCHAT formats are used in the CHILDES system for producing computerized transcripts of speech that can be analyzed by various CHILDES programs. Children's mean length of utterance (MLU) in words and morphemes was calculated using the CHILDES MLU program (MacWhinney 2000). One-word answers to experimenter questions (e.g. *yes, no, okay*), utterances containing unintelligible speech, and speech that was erroneously or unintentionally repeated within utterances (e.g. *he he he looks square*) were excluded from the MLU analyses.

The children's speech and gestures were coded together by native or near native speakers of French and Japanese, and then a second native or near native speaker verified the original coding. Each clause of a response where the child described (or attempted to describe) the scene characteristics was coded. In the transcription and coding, we did not mark where pauses occurred within an utterance; therefore, the gestures produced with utterances may have been produced with speech or during pauses. There were a few instances where gestures were not produced during an utterance, and these were excluded from the analyses.

The key words in each response were coded as *precise*, *imprecise*, or *other*. A *precise* response included clear, descriptive, and appropriate words to specify the scene characteristics (e.g. *has spikes; jumping*). Responses were coded as *imprecise* if they lacked clear descriptive words. Most often, these were responses such as *it looks like this, it goes like this.* Essentially, *imprecise* descriptions were not understood by the experimenter. A description was coded as *other* if (1) no clear descriptive words were used (e.g. *like a real bug)*, (2) a negative descriptor was used (e.g. *not jumping*), (3) it was not easily classifiable, or (4) the child described something other than the target characteristic. The key words were also coded for word type (e.g. verb, adjective) to determine the frequency of word and mimetic use. Mimetics were only used in Japanese. For example, *pyonpyon* was used to describe jumping and *gizagiza* was used to describe spiky.

All gestures produced by the children were coded (e.g. iconic, pointing), but only iconics were analyzed because they convey information about scene characteristics. Children's responses were coded as produced *with* or *without* gesture.

The verbal description and gesture codes were combined to produce four dependent variables: precise description without iconic gesture, precise description with iconic gesture, imprecise description without iconic gesture, and imprecise description with iconic gesture. The frequency of responses in each category was calculated separately for motion and object characteristics, and language for the bilinguals (French, Japanese). To control for variability in children's talkativeness, proportions were calculated by dividing response frequencies by the total number of responses the child gave for a particular scene characteristic and language.

3. Results

The means and standard deviations for the children's raw scores on the EOWPVT and their MLU in words and morphemes are shown in Table 2. The bilinguals had significantly lower vocabulary scores in French than the French monolinguals ($t\,(21) = -2.71$, $p < .05$) and significantly lower vocabulary scores in Japanese than the Japanese monolinguals ($t\,(21) = -3.88$, $p < .05$). Figures 2 and 3 summarize the mean proportions of precise and imprecise descriptions of the motion and object characteristics, with and without gestures, for each group.

The French and Japanese monolinguals did not differ in how often they produced precise descriptions without gesture. For precise descriptions with gesture, a marginally significant interaction was found between language and scene characteristic, $F\,(1, 22) = 3.94$, $p = .06$. French monolinguals described motions precisely with gesture

Table 2. Expressive language measures

		Vocabulary score (raw)		MLU in words		MLU in morphemes	
		M	*SD*	*M*	*SD*	*M*	*SD*
Monolingual	French	37.92	10.80	5.92	1.23	6.23	1.35
	Japanese	39.17	11.04	5.12	1.22	7.04	1.66
Bilingual	French	24.91	12.21	7.14	1.66	7.46	1.74
	Japanese	20.00	12.63	4.47	1.65	5.92	2.10

slightly more often than objects, whereas the Japanese monolinguals used these responses significantly more often for motions than objects, t (11) = 2.96, $p < .05$. This finding is likely related to mimetic use, as the Japanese monolinguals produced significantly more mimetics for motion descriptions ($M = .71$) than for object descriptions ($M = .28$, t (11) = 3.91, $p < .05$). The Japanese monolinguals provided significantly more imprecise descriptions without gesture than the French monolinguals, F (1, 22) = 4.79, $p < .05$. The monolinguals did not differ, however, in their use of imprecise descriptions with gesture.

For the French-Japanese bilingual children, the language used influenced their use of precise descriptions without gesture, and this interacted moderately with the scene characteristic being described, F (1, 10) = 3.76, $p = .08$. When the bilinguals spoke French, object characteristics were described precisely without gesture significantly more often than motion characteristics, t (10) = −3.77, $p < .05$. The same was found when the bilinguals used Japanese, though the difference was not significant, $p > .05$. With respect to mimetic and gesture use by the bilinguals when using Japanese, this was higher for motion descriptions ($M = .32$) than for object descriptions ($M = .09$), though this difference did not reach statistical significance, t (10) = 1.83, $p > .05$.

When the French-Japanese bilinguals spoke French, they were similar to the French monolinguals in how frequently they used each type of description. When the bilinguals spoke Japanese, they produced precise descriptions with and without gesture to a similar degree as Japanese monolinguals. However, this was not the case for

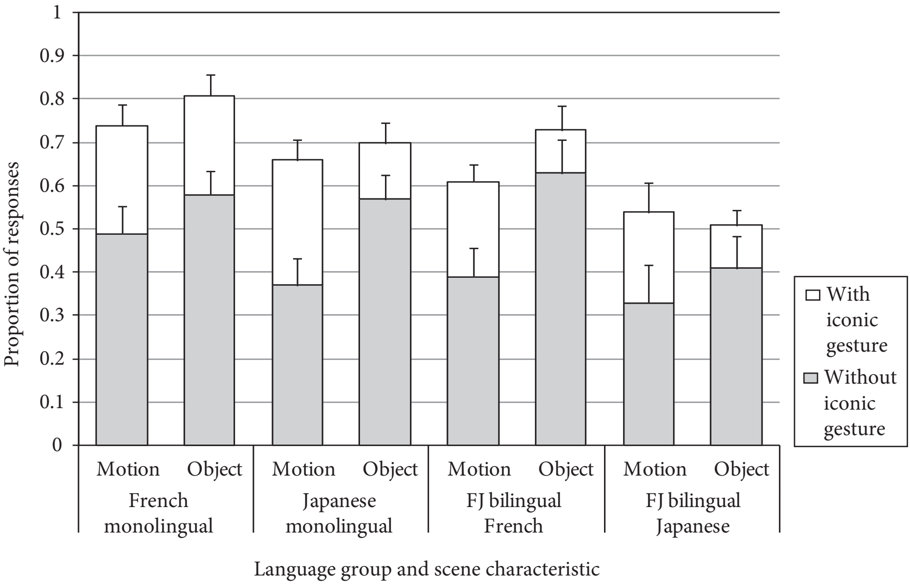


Figure 2. Mean proportions and standard errors of precise responses, with or without iconic gestures, for motion and object characteristics by language group.

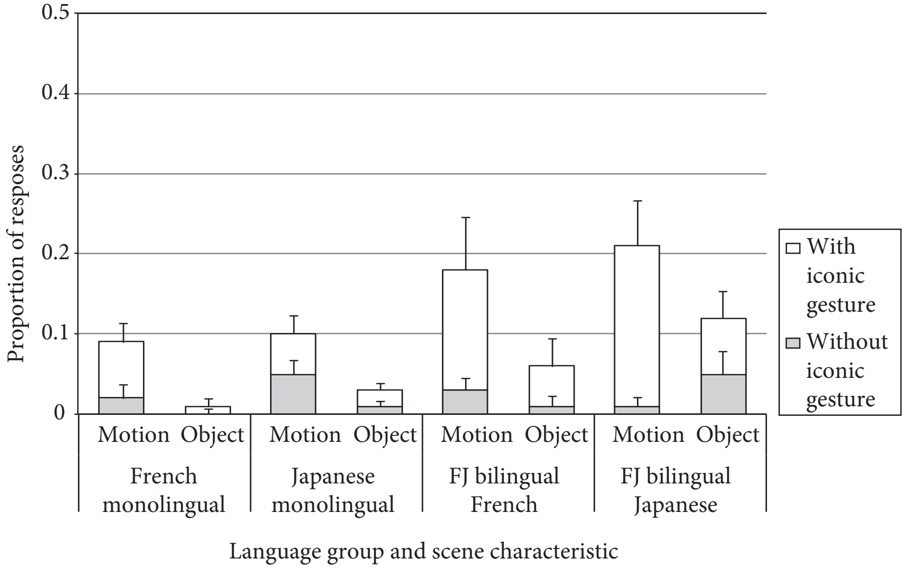


Figure 3. Mean proportions and standard errors of imprecise responses, with or without iconic gestures, for motion and object characteristics by language group.

their use of imprecise descriptions. Language group interacted with scene characteristic for imprecise descriptions without gestures, F (1, 21) = 6.12, $p < .05$. The Japanese monolinguals used these types of descriptions slightly more for motions than objects, while the bilinguals used them more for objects than motions, t (10) = –1.85, $p = .09$. The bilinguals also provided significantly more imprecise descriptions with gestures than the Japanese monolinguals (F (1, 21) = 6.11, $p < .05$), and the bilinguals actually used these types of descriptions more often when describing motions than objects, t (10) = 3.35, $p < .05$. When mimetic use was examined, the Japanese monolinguals and bilinguals speaking Japanese produced similar amounts of precise mimetic descriptions without gesture, but precise mimetic descriptions with gesture were used significantly more often by the monolinguals than the bilinguals, F (1, 21) = 6.38, $p < .05$.

Overall, the monolingual and bilingual children described object characteristics with precise responses and no gestures significantly more often than motion characteristics, *p's* < .05. In contrast, motion characteristics were described with precise responses with gestures and imprecise responses with or without gestures more often than object characteristics were, *p's* < .05.

4. Discussion

In the present study, we investigated the effects of language, language group, and scene characteristic on children's verbal descriptions with and without iconic gestures. We expected greater use of iconic gestures in Japanese than French because of the frequent use of mimetics in Japanese. Both mimetics and iconic gestures can be used to vividly and effectively convey the imagistic and affective nature of objects and events (Kita 2001). There was some evidence for this, but it depended on the scene characteristic being described. That is, both Japanese monolinguals and French-Japanese bilinguals using Japanese produced more gestures with motion than object descriptions when providing precise responses. Furthermore, a moderate to large proportion of the motion descriptions were mimetic in nature (monolinguals 71%; bilinguals 32%).

Our French-Japanese bilinguals had significantly lower vocabulary scores than both monolingual groups, and consequently may have had some difficulty describing the scene characteristics verbally. Despite this, the bilinguals' response patterns were similar to those of the French monolinguals for all response categories and to those of the Japanese monolinguals' for precise responses. Differences were only found for imprecise responses in Japanese. More specifically, the bilinguals produced significantly more imprecise descriptions with iconic gestures for motion characteristics. This might be due to decreased mimetic use by bilinguals when speaking Japanese compared to the Japanese monolinguals. Indeed, when producing precise responses with gesture, the bilinguals used mimetics significantly less often than the Japanese monolinguals. The similarities and differences between the bilinguals and monolinguals could be a result of living in Montréal, a French environment. Exposure to Japanese, including the use of mimetics, is limited for these bilingual children. Bilinguals living in Japan would probably be more similar to Japanese monolinguals in their mimetic use. We are currently conducting a similar study with Japanese-English bilingual children living in Japan to address this issue.

Consistent with our prediction, the scene characteristic being described influenced gesture use such that motion characteristics were accompanied by gestures more than object characteristics. Scene characteristic unexpectedly influenced verbal descriptions without iconic gesture as well. Descriptions of objects tended to be precise while descriptions of motions were often imprecise. Perhaps objects can be described more easily, while the dynamic nature of motion events renders them more difficult to describe verbally. The object characteristics in our study were relatively simple, however, and this issue should be examined further in future research.

In conclusion, we found that the presence of mimetics in Japanese was associated with co-speech gesture use when describing motion events in particular. Moreover, mimetic and gesture use was seen more often in the Japanese monolinguals than the bilinguals, likely due to their limited exposure to and proficiency in Japanese. Future research should examine bilinguals with higher proficiency in Japanese, as well as other bilingual groups to fully understand how and why speakers use iconic gestures with mimetics.

References

Academic Therapy Publications Inc. 2000. *Expressive One-Word Picture Vocabulary Test* (2nd ed.). Novato: Academic Therapy Publications.

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T. and Fujii, M. 2007. "Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish." *Cognition* 102: 16–48.

Church, R. B. and Goldin-Meadow, S. 1986. "The mismatch between gesture and speech as an index of transitional knowledge." *Cognition* 23: 43–71.

Gullberg, M. 1998. *Gesture as a Communication Strategy in Second Language Discourse. A Study of Learners of French and Swedish*. Lund: Lund University Press.

Holler, J. and Beattie, G. 2003. "How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?" *Semiotica* 146: 81–116.

Kita, S. 2001. "Semantic schism and interpretive integration in Japanese sentences with a mimetic: a reply to Tsujimura." *Linguistics* 39: 419–436.

Kita, S. 2008. "World-view of protolanguage speakers as inferred from semantics of sound symbolic words: A case of Japanese mimetics." In *The Origins of Language: Unraveling Evolutionary Forces,* N. Masataka (ed), 25–38. Tokyo: Springer.

Kita, S. and Özyürek, A. 2003. "What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking." *Journal of Memory and Language* 48: 16–32.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah: Lawrence Erlbaum Associates.

Mayberry, R. and Nicoladis, E. 2000. "Gesture reflects language development: Evidence from bilingual children." *Current Directions in Psychological Science* 9: 192–196.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, D. and Duncan, S. D. 2000. "Growthpoints in thinking-for-speaking." In *Language and Gesture*, D. McNeill (ed), 141–161. Cambridge: Cambridge University Press.

Nicoladis, E. 2002. "Some gestures develop in conjunction with spoken language development and others don't: Evidence from bilingual preschoolers." *Journal of Nonverbal Behavior* 26: 241–266.

Nicoladis, E. 2007. "The effect of bilingualism on the use of manual gestures." *Applied Psycholinguistics* 28: 441–454.

Oshima-Takane, Y., MacWhinney, B., Sirai, H., Miyata, S. and Naka, N. (eds). 1998. *CHILDES Manual for Japanese* (2nd ed.). Nagoya: JCHAT Project.

Özyürek, A., Kita, S., Allen, S., Furman, R. and Brown, A. 2005. "How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities." *Gesture* 5: 219–240.

Riseborough, M. G. 1982. "Meaning in movement: An investigation into the interrelationship of physiographic gestures and speech in seven-year-olds." *British Journal of Psychology,* 73: 497–503.

Stam, G. 2006. "Thinking for speaking about motion: L1 and L2 speech and gesture." *International Review of Applied Linguistics* 44: 145–171.

Stam, G. 2008. "What gestures reveal about second language acquisition." In *Gesture: Second Language Acquisition and Classroom Research,* S. G. McCafferty and G. Stam (eds), 231–255. New York: Routledge.

Zvaigzne, M., Oshima-Takane, Y., Groleau, P., Nakamura, K. and Genesee, F. 2008. "The function of children's iconic co-speech gestures: A study with French-Japanese bilinguals and French monolinguals." In *BUCLD 32: Proceedings of the 32nd Annual Boston University Conference on Language Development,* H. Chan, H. Jacob and E. Kapia (eds), 598–609. Boston: Cascadilla Press.

Acknowledgments

This research was supported by a Social Sciences and Humanities Research Council grant to the second and third authors, and a Ministry of Education, Culture, Sports, Science and Technology grant (No. 18520448) to the fourth author. We thank the Montréal Japanese Language Centre, Japanese Canadian Cultural Center of Montréal, Mominoki daycare, and the children and parents who participated. We also thank Alisa Koyama, Andrew Pinchefsky, Azusa Yamaguchi, Carrie Cheung, Chigusa Kurumada, Kayo Nakamura, Kyoko Sato, Galia Pogossova, Hisako Noguchi, Hitomi Mitsuo, Leila Salem, Mari Umeda, Michi Tomoda, Patricia Groleau, Serina Nishioka, Yasuko Senoo, and Yuri Fujisaki who helped with data collection, transcription, and coding, and Marie Lippeveld for helpful comments.

CHAPTER 17

Gesture and language shift on the Uruguayan-Brazilian border

Kendra Newbury
Western Washington University and United States Air Force Academy

The linguistic phenomenon in which a prestige language variety supplants a traditional one, language shift, as well as the related phenomenon of superstratum and substratum interference, leading to mixture, have been widely studied in linguistics. Scholars, however, have not applied these linguistic theories to non-verbal communication, such as gesture. In applying these concepts to the hegemonic displacement in northern Uruguay of the traditional Portuguese variety by the national language, Spanish, this chapter demonstrates that gestural convention is interconnected with the linguistic outcome of language contact among these border bilinguals. Focusing on gestures that are traditionally associated with each language, the results confirm expected generalizations about gesture shift as a parallel phenomenon, while they reveal conclusions about how gesture differs from language, including the absence of gesture-switching and the phenomenon of latency, or rather, the delay in the adoption of culturally-defined paralinguistic forms when a speech community undergoes language shift.

Introduction[1]

Language and emblematic gestures are culturally-bound expressions of communication in that a sign, either as a written word or physical gesture, is represented by its sound, in the former, and by the physical occupation of space, in the later with respect to the image that they evoke. The speaker, then, has two distinct means by which he or she may express the same concept. Given the cultural relationship between these verbal and non-verbal expressions in a speech community, it stands to reason that a shift in language would imply a concomitant shift in gesture. This hypothesis is confirmed, in part, in the northern region of Uruguay along the Uruguayan-Brazilian border. There, three bilingual speech communities in Artigas, Uruguay undergoing language

1. I acknowledge with gratitude Nicole Douglas for assisting in this research by videotaping the informants and compiling the visual data for analysis.

shift exhibit a move away from one language toward another not only in speech, but also in gesture, a phenomenon I characterize as gesture shift.

The present study isolates those gestures that are distinct in Uruguay and Brazil and, in identifying the sociological factors of people who use them, demonstrates that gesture is a related expression of language shift. As speakers move away from the stigmatized Uruguayan-Portuguese (UP) variety toward Spanish, motivated by pressure from the linguistic majority to conform to the Uruguayan standard, they abandon both the UP variety and culturally-related gestures. As gesture shift occurs, the intermediate outcome of this process is gesture variation in the speech community, a situation in which gestures from both cultures are expressed with variable frequency, giving the appearance of a mixed paralinguistic code or, as I describe more fully below, gesture mixture. This interrelated phenomenon exemplified in the transition from Uruguayan Portuguese (UP) to Spanish suggests that gestural paralanguage should be considered an integral part of the study of language contact and shift.

Previous research

Scholars have recognized the relationship between language and gesture among bilinguals or speakers who adopt/have adopted a new language/culture. Previous studies have focused on the areas of folklore, bilingualism, and second-language acquisition. Rickford and Rickford's (1976) study recognizes the expressions of insult and their corresponding gestures for "cut-eye" and "suck-teeth" as African "survivals" in Guyana, the Caribbean, and the United States among black informants, descendants of those who adapted to the New World. Pika, Nicoladis, and Marentette's (2006) study of English-Spanish and French-English bilinguals demonstrates the possibility of transfer among these "hybrid" gesturers. Bilinguals of a high-frequency gesture language exhibit a similar frequency when speaking English, which is a low-frequency gesture language, when compared to their English monolingual counterparts in recounting a cartoon sequence. Gullberg (2006) draws attention to the importance of second-language learners in acquiring the gestural repertoire of a target language. She acknowledges that an L2 gestural system develops in a similar way to language and may offer insights into phenonema of interference and interlanguage. Choi and Lantolf (2008) demonstrate such a gestural interlanguage. L2 learners who have acquired advanced proficiency in the language acquire path gestures of typologically different languages, as found in Stam (2006), but retain manner gestures of their L1 thinking-for-speaking patterns, as demonstrated by the speech-gesture growth point. These studies appear to indicate that the relationship between language and gesture is one in which gesture is often preserved, contributing to a mixture between the language of one culture and the gesture of another. Admixture through preservation can also be evidenced in emblematic gesture shift in a bilingual/bicultural community undergoing language shift over a number of generations. Emblematic gesture shift represents a novel and dynamic area of future research in gesture studies of which there have been no previous investigations.

Contact along the Uruguayan-Brazilian border

Language contact in northern Uruguay is the result of a hegemonic battle between Portugal and Spain for control of the territory between the Cuareim River, along the southern border of Brazil, and the River Plate, which now divides Uruguay and Argentina on the east coast of Latin America. The Portuguese set out to actively colonize the area in the 17th century, with the founding of Colônia do Sacramento in 1680, while Spain's response was the establishment of Montevideo in 1726 (Coolighan and Arteaga 1992: 108). During the following century and a half, the land was the object of contention between the two crowns, each vying for sovereignty. Between 1815 and 1828, the Portuguese and, subsequently, the independent Brazilian nation gained authority over the entire region, naming it the Cisplantine Province (Coolighan and Arteaga 1992: 261–262, 285, 296). It was thereafter established as an independent nation under the Treaty of Montevideo (Coolighan and Arteaga 1992: 292–296, 336). In the following years, the Uruguayan government actively sought to extinguish vestiges of the Portuguese language through education reform, recognizing that education and justice were largely conducted in Portuguese in the northern half of Uruguay (Academia Nacional de Letras 1982: 12, 21–22, 24). Despite these official efforts, rural inhabitants and members of the urban working class along the northern border of Uruguay have maintained their Portuguese inheritance as UP, a regional Portuguese-based variety, owing to limited access to education and speakers' relative isolation from the Spanish-speaking majority (Hensey 1972, Elizaincín 1992, Elizaincín, Behares & Barrios 1987, Carvalho 1998, Douglas 2004).

Today, with the increased importance placed on education and pressure from the majority to abandon the stigmatized UP variety in favor of Spanish for its socioeconomic advantages, UP is converging toward a variety that is highly influenced by Spanish, while an increasing number of UP/Spanish bilinguals are socializing their children solely in Spanish, which is conversely influenced by UP (Newbury in progress). Also spurring the change is a desire by the UP-speaking community to show loyalty to the Uruguayan nation by abandoning the language associated with their Brazilian neighbor (Douglas 2004). The shift towards Spanish results in structurally-mixed varieties by means of two processes: (1) continued superstratum interference of Spanish on the UP variety, as the language of the dominant culture influences the minority language and (2) substratum influence of the heritage variety, UP, on Spanish, as border bilinguals adopt the language of the majority (Newbury in progress). The ensuing level of mixture in both UP and Spanish manifests itself in language variation both within the individual and in the community of which s/he is a member. Admixture can also reveal itself as mixed variants comprised of elements of both language varieties. The amount of variation is relative to predictable social factors, such as gender and age, but most importantly to location, in this case a rural vs. urban setting (Douglas 2004).

Theoretically, language contact and gesture contact could be similar, in principle, with respect to the resultant mixture arising during language/gesture shift. First, language shift often produces intermediate outcomes in which the speaker draws from both languages, having available to them variants from two distinct vocabularies, such as UP *Natal* and Spanish *Navidad* 'Christmas.' With respect to gesture mixture, the speaker would also have both culturally-specific emblematic gestures. Gesture mixture, then, provides the speaker with an alternate variant in the repertoire of non-verbal expression. That is, the alternation would exist within an individual speaker and the speech community, the gesture being binary, either culturally Brazilian or Uruguayan, irrespective of the language with which it co-occurs. Second, the intermediate could be a hybrid variant. For example, from the perspective of speech, a speaker may, through false analogy, hypercorrect the recognizable *-dad* '-ly' morpheme and apply Portuguese phonology to Spanish *Navidad* 'Christmas', yielding UP *Navidade* [navi⍰da⍰i]. In principle, this phenomenon is also possible with gesture. Emblematic gestures can be analyzed with respect to two components, namely their form and motion. In theory, these gestures may comprise, for example, a mixture of the hand posture of one culture and the spatial shape of another or a combination of forms from both cultures accompanied by a particular movement. Where both types of mixture occur in language, both language variation and language hybridization (Douglas 2004), the purpose of this seminal study is to determine whether these two forms of gesture mixture actually occur in an area of gesture contact and shift.

Gesture shift

The consequences of language contact are sensitive to social factors, such as gender, age, socioeconomic status, and urbanization (Labov 2001). Regional variation is closely related to factors of social class and network, whereby urban bilinguals have looser social network ties and exposure to social stratification. Rural communities, conversely, are less socially-stratified, and have a dense network of multiplex relationships (Milroy 1980). In northern Uruguay, urban speakers have more exposure to the prestigious Uruguayan standard, a means by which a speaker can achieve higher status or the perception of such. In rural communities, there is less exposure to standard variants of language and gesture and less pressure to achieve upward mobility through their use. In addition, each community's spatial construct, i.e. whether or not it has a focal point and ease of access to the urban center of Artigas, determines network patterns and, therefore, the nature of the distribution of linguistic and gestural variants.

Method

Both Uruguayan and Brazilian speakers frequently employ emblematic gestures. Brazilians are commonly known to have a gestural repertoire that more closely approximates that of Italians. I observe that Uruguayan gesture in the capital of Montevideo seems to have influence from Italian immigration, given structurally similar gestures that are present. However, Brazilians appear to have more emblematic gestures than Uruguayans. In many cases, these gestures are distinct.

In order to study regional variation of gesture and its relation to language shift, 37 subjects participated in this study, of which 17 served as informants from the urban center of Artigas and two rural villages outside Artigas, Sequeira and Bernabé Rivera. As a point of comparison, 13 subjects represented two control groups. Six speakers from Montevideo, the capital of Uruguay, which has little to no contact with the border community, and seven speakers from Quaraí, the town adjacent to Artigas on the Uruguayan-Brazilian border, with daily contact with Uruguayans, but virtually no influence on the language or culture, were interviewed to confirm the expected standard for the Uruguayan and Brazilian gestures. Also participating were four informants from Buenos Aires, the Argentine capital, in order to demonstrate that the Uruguayan gesture manifested in Montevideo is, in fact, more related to Hispanic than Lusophone culture. Finally, four monolingual subjects from the city of Artigas were interviewed as another contrast group for the results gathered from the study of bilinguals in the region.

The study focuses on the responses from speakers of the three distinct communities. Artigas, Uruguay is the department capital and the second largest city on the northern border with a population of 40,249 (Instituto Nacional de Estadística 1998: 7). Informants live in a peripheral neighborhood where UP continues to be spoken. Sequeira is a rural village, with a population of 878 (Instituto Nacional de Estadística 1998: 43), stretching along a highway 85 kilometers within the Uruguayan border south of Artigas. Bernabé Rivera is an isolated and tight-knit rural community of 421 inhabitants located along the border (Instituto Nacional de Estadística 1998: 42). Since it has no local access to Brazil, villagers travel the 70 kilometers northeast to Artigas, a good portion of which is not paved (Douglas 2004: 43). UP is the predominant variety in the rural communities. All communities have access to Brazilian television, but access to Uruguayan television is limited to the city of Artigas (Douglas 2004: 42).

The informants from the three distinct Uruguayan speech communities were UP/Spanish bilinguals from the working class, with an equal or nearly equal number of men and women participating from each community. The interviews were conducted by me, a bilingual speaker of both standard Spanish and Brazilian Portuguese. The study began with 75 concepts that might elicit emblematic gestures based on knowledge of gesture in both cultures. Of these, three gestures revealed themselves to be the most salient, representing those that elicited immediate and unambiguous responses. The interviews were conducted spontaneously in a variety of natural settings. Culturally bilingual informants were asked in standard Spanish for the gesture that would express the cue that was given verbally. At the end of the interview of the 75 concepts, I gave the cues in Portuguese and repeated the experiment, focusing on approximately ten of the gestures to determine whether the gestural code changed with the language.

Results

Three quotable gestures were identified as clearly distinct in Uruguay and Brazil with respect to the referent. These are "Let's eat!", "He's stingy!", and "Come here!" The first gesture, "Let's Eat!", or the concept of eating in general, is articulated differently in the two nations. In Uruguay, with fingers held together, mainly the thumb, index, and middle fingers, the speaker gestures toward the mouth. In Brazil, with the palm down and fingers bent at the palm, the speaker flaps all four fingers in a downward fashion near the mouth. Table 1 demonstrates that in Quaraí, on the Brazilian side of the border, Brazilians consistently use the nationally-accepted gesture, while Uruguayans in Montevideo did not use any gesture culturally-related to Brazil.

When comparing the three locations in question, regional variation is apparent. In the urban center of Artigas, informants generally use the Brazilian gesture, with one informant choosing the Uruguayan variant. In Bernabé Rivera, the Brazilian variant is clearly predominant, and in Sequeira, it is the only variant recorded. The result in Artigas and Bernabé Rivera represents a mixed repertoire within the community. The gestures of Brazilian heritage are being maintained in Bernabé Rivera and Sequeira, while there is movement toward the Urugayan counterpart in Artigas.

The second quotable gesture, "He's stingy!", also demonstrates the retention of Brazilian forms in the rural areas and the tendency to adopt the Uruguayan variant in the city. The control groups uniformly produced the expected variants. The Uruguayans from Montevideo expressed this concept by bending the arm and tapping under the elbow with other hand, representing a person who does not bend his hand to reach into his or her pocket, also known as the golden elbow. This gesture, interestingly, is identical to the gesture used in greater Brazil to indicate jealousy. Alternatively, a person may make the movement of walking on one's elbows by bending both arms and mimicking the motion. The Brazilians from Quaraí formed the gesture by raising a fist with a bent elbow, the fingers of the fist facing the body. A variation of this gesture

Table 1. Gesture for "Let's Eat!"

	Quaraí	Montevideo		Artigas	Sequeira	Bernabé Rivera	
Brazilian	7	0		3	7	6	
Uruguayan	0	4	1*	1	0	0	1*
Other	0	2		0	0	0	
No response	0	0		0	0	0	

* Represents a secondary response.

Table 2. Gesture for "He's stingy!"

	Quaraí	Montevideo	Artigas	Sequeira	Bernabé Rivera
Brazilian	7	0	3	4	6
Uruguayan	0	6	1	0	0
Other	0	0	0	3	0
No response	0	0	0	0	0

includes the placement of the thumb between the index and middle fingers. In either case, this gesture represents a person who does not open his hand to give money. The results are shown in Table 2, which are similar to those for "Let's Eat!"

One informant has adopted the Uruguayan variant in the urban center of Artigas, while the Brazilian variant is strongly maintained. While one could expect some resistance to adopting the Uruguayan variant, given that it connotes a different semantic content in Brazil, none of the informants in Uruguay offered that shape and movement to connote the Brazilian conception of jealousy. This suggests that informants were not familiar with this Brazilian gesture, thereby not contributing to a barrier to possibly create what would be equivalent to homonyms, or rather, the same gesture conjuring the image of two different referents, stinginess and jealousy. Instead, the community has two gestures to connote one referent, serving as gestural synonyms.

In the final example, the emblematic gesture for "Come here!" demonstrates a similar tendency to shift away from the Brazilian form. In this case, there is an added complication: the two Brazilian gestures can be represented in one of two ways, the second of which is equivalent to the Uruguayan expression. In Uruguay, the hand is held in front of the body with the palm up, with the speaker moving one or all fingers toward him or herself. While this movement is also common in Brazil, Brazilians alternatively extend the arm with the palm down, moving the fingers toward the body. Therefore, any palm-down gestures are considered, here, as definitively Brazilian, while a palm-up emblem could be either. In Table 3, below, the Brazilian control group did use the palm-up alternative, although more often as a second demonstration of the gesture (represented by superscript ^{U} in the chart below). The Uruguayans in Montevideo never employed the palm-down variant. This was also collectively the case among Argentinean informants in Buenos Aires.

Table 3. Gesture for "Come here!"

	Quaraí		Montevideo	Artigas	Sequeira		Bernabé Rivera	
Down	1	2^{U}	0	0	3	1^{U}	5	1^{U}
Up	6		6	4	4		1	
*	0		0	0	0		0	
Ø	0		0	0	0		0	

In Artigas, the palm-up emblem was used exclusively with no expression of this Brazilian alternative. In Bernabé Rivera, the palm-down gesture was predominant, followed by Sequeira, demonstrating slightly less use of this form. The shift away from the alternate form seems to be apparent in Artigas. The reason for the more rapid change could be that the Brazilian palm-up alternative is equivalent to the Uruguayan expression, as if they were homonyms. The support for an existing gesture facilitates its increased adoption to the point that it equates with usage in Montevideo.

Discussion

Location and urbanization, which is related to social class in Uruguay, was a main factor in gesture use, and the frequency of each variant in each of these three communities clearly demonstrates gesture shift in progress. These results demonstrate that, in the case of these three emblematic gestures, there appears to be a similar trend: the tendency for a speaker to maintain Brazilian gestures in the rural areas, and the onset of abandonment of those culturally-specific gestures in Artigas, the urban center. With respect to the difference between Sequeira and Bernabé Rivera, the absence of the Brazilian emblems is more frequent in the former community than the latter. Language contact research that showed that Sequeira, a town whose physical space is not conducive to strong network ties, as is Bernabé Rivera, and which despite its distance from Artigas is closer by way of greater highway access, shows a greater tendency to adopt the Spanish standard among its speakers (Douglas 2004). The result of this gestural study parallels that of this sociolinguistic research conducted in the same area.

What we are witnessing, therefore, is the intermediate outcome of language shift, mixture in both verbal and non-verbal communication. With respect to language contact and shift, the intermediate outcome is a mixture of both UP and Spanish by way of alternate forms within an individual's speech, and variation in the community. With the progression toward abandonment of the stigmatized UP variety, UP speakers incorporate and use increasingly more Spanish variants, particularly in urban Artigas (Douglas 2004: 323). At the same time, bilinguals who speak Spanish also produce mixture at all three levels, form-internal, speaker-internal, and community-internal, owing to substratum influence, primarily at the structural level (Newbury in progress).

In gesture shift, a continuum between Montevideo and Quaraí is developing, with more rapid progression in the urban center than in the rural areas as more speakers adopt the Uruguayan variants. Alternation within the individual, exemplified by those cases in which the informant gave a second alternative to the question, is the basis for variation in the speech community as each individual chooses his or her primary form. Given the experiment conducted with respect to language choice and gesture, the repertoire appears to belong to a single code as the gestures performed remained the same notwithstanding a change in the language in which the cue was provided by the interviewer.

Given that the Brazilian gesture, when distinct, is predominant, it is evident that the base is Brazilian as is the UP vernacular. I propose that the speech community in northern Uruguay has one set of heritage gestures related to their Brazilian colonial past, evidenced by the maintenance of Brazilian gestures in the two rural villages. However, through contact with members of the majority culture, or rather, superstratum interference, that set has expanded to include alternate gestures similar to a borrowing or loan. The Uruguayan imitation remains, therefore, in free variation with the Brazilian norm. The process of maintenance or shift is dependent upon the speech community. The urban community is more likely to have gesture mixture and eventual shift, given the higher frequency of Spanish monolingualism, prestige and loyalty, greater social stratification, as well as looser network ties. Sequeira and Bernabé Rivera, despite these villages being well inside Uruguay, exhibit a higher frequency of Brazilian gestures and will resist shift longer as there is less contact with the dominant culture, such as the media or daily contact with monolingual Uruguayans, and therefore, less access to Uruguayan gestural input and contexts in which they are used. These rural inhabitants may be slow to shift, owing to the lack of social stratification and dense network ties that provide a model or incentive for upward mobility, thereby perpetuating the status quo.

While there is a parallel relationship between language shift and gesture shift in this community, it appears that it is occurring at different rates, as evidenced by the frequency of each gestural expression when comparing regional use among urban and rural speakers who are fully proficient in Spanish. Most speakers seem unaware of the difference between gestures, both being equivalent, and therefore, the gesture does not hold the same value of identity and prestige as speech. There appears to be no stigma attached to Brazilian gestures, as there is with the spoken UP variety, perhaps owing to a speaker's perception of the association between verbal and non-verbal communication. If a speaker is already engaged in Spanish, the gesture is automatically associated with the language and, therefore, the gesture does not carry social stigma. There is, therefore, no motivation to abandon the gestures of Brazilian origin.

For this reason, gesture shift lags behind its language counterpart. Shift toward the Spanish standard has been occurring over the last century and a half, with the majority of the speakers, primarily those of the upper and middle classes having already become monolingual in Spanish. It is the end of language shift among bilinguals in rural and working-class urban areas that we are witnessing in the Artigas region (Douglas 2004). Yet, it is interesting to note that Spanish monolinguals in Artigas will often produce Brazilian gestures, owing again to the Brazilian linguistic and cultural substratum influence of the region. Therefore, it appears that Brazilian gestures have a wider range of use, not being limited to UP bilinguals alone. However, as in language shift, Uruguayan gestures are likely to become more predominant and may eventually displace Brazilian ones, albeit at a subconscious level, to resemble the output observed among Montevideo informants.

Contrary to the original hypothesis of a hybrid gesture, such as the mixture of the form from one culture with the motion of another, this possibility was never realized in the course of this investigation. The resultant gesture mixture, therefore, does not occur within the gestural mixture itself, but rather is a function of emblematic alternatives available to the speaker or the speech community to express a mental concept. Therefore, unlike language mixture, emblematic gesture mixture is limited to variation within the repertoire of the individual, and consequently, of the community.

While there is no mixture within an individual gesture, there appears to be only one repertoire of gestures in the community from which speakers select and which derive from both cultural traditions. It is not that a speaker is gesturally bilingual, alternating between the gestural codes that are appropriate to the language being spoken, as in code-switching, but rather, one may have at his or her disposal variants from both cultures, functioning much like synonyms within one language. While there was some self-reporting of gestures being either Uruguayan or Brazilian, gesture-switching did not correlate with code-switching from Spanish to UP. This is evidenced by the fact that, toward the end of Spanish-language interviews with bilinguals, informants accommodated the interviewer, who switched to a Portuguese variety, asking the participants to reproduce earlier gestures. While the linguistic code was different, the gestures remained the same. This suggests that while there is some consciousness of the cultural difference between gestures, speakers have just one gestural code, which may include variants from either culture, and are, therefore, not bigestural in the same way that an individual is bilingual. Therefore, gesture mixture is distinct from language mixture in this community undergoing language shift. Speakers do not mix within the gestural expression itself, nor is there evidence that they alternate gestures as they switch between linguistic codes.

Conclusion

The use of gesture closely follows patterns of language behavior in an area undergoing language shift. Gesture study in a bilingual community demonstrates a number of significant conclusions. (1) With respect to location and urbanization, emblematic gesture is socially-stratified in the same way as other aspects of language, a fact which might motivate future study for determining correlations between gesture repertoire and age, gender, social class, and ethnicity. (2) An individual's repertoire in a bilingual community may include competing gestures from both cultures as gestural synonyms without conscious recognition that they are culturally-distinct expressions of the same notion. (3) Phenomena of substratum and superstratum interference are present in Spanish monolingual urban Artiguense speakers using Brazilian gestures and UP bilinguals using Uruguayan gestures, respectively. (4) There is no gesture-switching analogous to code-switching in this population of bilingual speakers undergoing language shift. Bilinguals consistently use one gestural code and do not alternate gesture based on established principles of code-switching even when the spoken code has changed. (5) Gesture mixture manifests itself as variation within the compendium of gestures available to the individual and/or the community, rather than the mixing or hybridization of elements within the gesture itself. (6) Gesture variation, in this study, is a function of location. Urban informants from Artigas and rural informants from Sequeira, a village which is less cohesive but has greater access to the city, show a disposition toward adopting the gestures associated with the prestige culture, while rural informants from insular and tight-knit Bernabé Rivera maintained those of their Brazilian heritage. (7) Gesture shift in this bilingual speech community, though parallel to language shift, exhibits latency, the adoption of non-verbal communication lags behind the adoption of language. (8) Latency can be attributed to the lack of association of gesture to language, given one gestural code. Therefore, culturally-specific gestures are not associated with prestige or stigma, driving the adoption or rejection of a gestural form.

While the study of gesture generally falls outside the scope of linguistic research as a form of non-verbal communication because it does not necessarily co-occur with language, in studying language shift in bilingual Uruguay, gesture and the study of linguistics appear to be closely intertwined. While the study of linguistics encompasses all aspects of linguistic production, including those that are paralinguistic, such as prosody, gesture is a form of non-verbal communication that is often regarded as falling outside the scope of the discipline by linguists, though foremost gesture theorists argue that language and speech are inseparable (McNeill 2005, Kendon 2004). The current study demonstrates the importance of integrating gesture into linguistic studies, as well as the need to understand linguistic background while conducting studies that focus on gesture, particularly in culturally heteronymous communities.

References

Academia Nacional de Letras. 1982. *Estudio sobre el problema idiomático fronterizo*. Montevideo: Comisión para el Estudio del Español en la Zona Fronteriza, Academia Nacional de Letras.

Carvalho, A. M. 1998. *The Social Distribution of Uruguayan Portuguese in a Bilingual Border Town*. Ph.D. Dissertation. University of California at Berkeley. Ann Arbor: UMI.

Choi, S. and Lantolf, J. P. 2008. "Representation and embodiment of meaning in L2 communication: Motion events in the speech and gesture of advanced L2 Korean and L2 English speakers." *Studies in Second Language Acquisition* 30 (2): 191–224.

Coolighan, M. L. and Arteaga J. J. 1992. *Historia del Uruguay: Desde los orígenes hasta nuestro días*. Montevideo: Barreiro y Ramos.

Douglas (now Newbury), K. 2004. *Uruguayan Portuguese in Artigas: Tri-dimensionality of Transitional Local Varieties in Contact with Spanish and Portuguese Standards*. Ph.D. Dissertation. University of Wisconsin, Madison. Ann Arbor: UMI.

Elizaincín, A. 1992. *Dialectos en contacto: Español y portugués en España y América*. Montevideo: Arca.

__, Behares L. and Barrios, G. 1987. *Nos falemo brasilero: Dialectos portugueses en Uruguay*. Montevideo: Amesur.

Hensey, F. 1972. *The Sociolinguistics of the Brazilian-Uruguayan Border*. The Hague: Mouton.

Gullberg, M. 2006. "Some reasons for studying gesture and second language acquisition (Hommage a Adam Kendon)." *International Review of Applied Linguistics in Language Teaching* 44 (2): 103–124.

Instituto Nacional de Estadística. 1998. *VII censo general de población, III de hogares y V de viviendas*. Montevideo: República Oriental del Uruguay, Instituto Nacional de Estadística.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Labov, W. 2001. *Principles of Linguistic Change, Vol. 2: Social Factors.* Language and Society 29. Oxford: Blackwell.

McNeill, D. 2005. *Gesture and Thought.* Chicago: University of Chicago Press.

Milroy, L. 1980. *Language and Social Networks.* Oxford: Blackwell.

Newbury, K. in progress. "The substratal relationship of Portuguese to Uruguayan Spanish among bilinguals in northern Uruguay."

Pika, S., Nicoladis E. and F. Marentette, P. 2006. "A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer?" *Bilingualism: Language and Cognition* 9 (3): 319–327.

Rickford, J. R. and E. Rickford, A. 1976. "Cut-eye and suck-teeth: African words and gestures in New World guise." *The Journal of American Folklore* 89 (353): 294–309.

Stam, G. 2006. *Changes in Patterns of Thinking with Second Language Acquisition*. Ph.D. Dissertation. University of Chicago. Ann Arbor: UMI.

PART IV

Gesture in the classroom and in problem-solving

CHAPTER 18

Seeing the graph vs. *being* the graph

Gesture, engagement and awareness in school mathematics

Susan Gerofsky
University of British Columbia

This study is situated within a body of new work in mathematics education that involves studies of gesture, kinesthetic learning and embodied metaphor and mathematical understandings (for example, Lakoff & Núñez 2000; Nemirovsky & Borba 2003; Goldin-Meadow, Kim & Singer 1999). This chapter reports findings from the first two years of the author's multi-year study exploring variations of secondary students' gestures when asked to describe mathematical graphs. Three diagnostic categories emerged from this data with regard to learners' degree of imaginative engagement and ability to notice mathematically salient features when encountering graphs.

Current work in mathematics education and gesture

Mathematics educators are increasingly working with gesture as a way of revealing unconscious aspects of mathematics learning and teaching, since teachers and learners produce gestures in a largely unconscious way, as a byproduct of communicating and expressing ideas (Núñez 2004; Roth 2001, 2009). Gestures produced by mathematics teachers and learners provide a rich source of data, comparable in scope to that provided by language, which can be read in terms of bodily metaphors, object development in the formation of mathematical concepts, and the relationships among mathematical concepts. Mathematics education researchers have taken up gesture studies in a number of different ways, and for different purposes.

There has been a great deal of interest in studying students' and teachers' gestures, together with accompanying language, to gain access to processes of mathematical concept formation. These studies often involve detailed microanalysis of very short bursts of language and gesturing. In some studies, the focus is on the teacher's gestures and their role in student learning (Alibali & Nathan 2007; Goldin-Meadow, Kim & Singer 1999; Goldin-Meadow, Nusbaum, Kelly & Wagner 2001). In others, researchers focus on student gestures produced while communicating mathematically with peers or teachers (Flevares & Perry 2001; Goldin-Meadow, Cook & Mitchell 2009; Radford et al. 2003; Rasmussen, Stephan & Allen 2004). Through the study of student gestures, individually and in groups, and in conjunction with speech, drawing and writing, researchers aim to tease out an accurate and detailed description of the process of mathematical concept formation. This can include growth points within the individual's development of a concept (McNeill 2000), the spread of a concept through social interaction in group work, and misapprehension of another's concept (Reynolds & Reeve 2002, Maschietto & Bartolini Bussey 2009).

Other studies look more broadly at embodied metaphors and mathematical concepts (Lakoff & Núñez 2000, Edwards 2009). This research, with roots in semiotics, linguistic semantics and cognitive science, takes up the idea that abstract mathematical concepts are necessarily grounded in our physical, embodied experiences of the world – and that, in fact, the historical origins of these abstract concepts always emerge from empirical, sensory observations (Radford 2009, Arzarello et al. 2009, Nemirovsky & Ferrara 2009, Tall 2004). In these studies, particular observed gesture/language conjunctions are used as exemplars of a broader framework of metaphor within a culture or subculture. These mathematical metaphors are placed within a larger structure of cultural metaphors dealing with time, space, size, quantity, pattern and relationship – metaphors that give shape to mathematical abstractions (Núñez 2009, Arzarello & Edwards 2005). While some studies of embodied mathematical metaphor use detailed microanalysis of a particular interaction as illustrative, others treat culturally-typical gestures as semantic elements of language (in the sense of Saussure's idealized *langue*) and use these in a wide-ranging analysis of mathematical metaphor (Saussure 1915/1965).

My own area of interest involves gesture, embodiment and the teaching and learning of mathematical functions and graphing in secondary school mathematics (Gerofsky 2008). Even within this rather specialized area, researchers take a variety of different approaches. Most work with spontaneous gestures produced in conjunction with speech, writing, drawing and the use of manipulatives when students and teachers are engaged in mathematical communication and problem-solving (Tall 2003). Much of the research around the pedagogy of mathematical functions concerns the idea of *covariation* – that is, noticing how the y (vertical, dependent) elements of a function vary when the x (horizontal, independent) elements change (Nemirovsky & Borba 2003, Robutti & Ferrara 2002) . Studies of this kind often use researcher-designed manipulatives or demonstration tools that separate the x and y elements of a two-dimensional function, allowing learners to move and gesture in ways that show their conceptual models for covariation. I have done some work in this area as well, using Etch-a-Sketch commercially-produced drawing toys as manipulatives that separate x and y elements as two different control knobs, which can be operated in a coordinated way by two different people (Gerofsky & Marchand 2006).

However, the research project reported here (called 'Graphs and Gestures' in short form) takes a different approach to the pedagogy of mathematical functions and graphs through gesture that does *not* highlight covariation. This study contrasts with other work on graphing and gesture in three ways. First, I chose not to study spontaneous speech-accompanying gestures, but rather to elicit gestures in response to pictures of the graphs of functions. Second, I made a conscious choice not to study covariation here, but rather to work with the graphs holistically (and thus not to pull apart their x and y elements). Third, the aim of this research was *applied* rather than simply *descriptive*. I wanted to find ways to improve the teaching and learning in this portion of the secondary mathematics curriculum, using elicited gestures to analyze, and then diagnose and affect student learning in a positive way. In other words, the analysis of student gestures of the graphs of functions is a necessary step on the way to designing an improved pedagogy of graphing. In this applied, design-oriented approach, my work has something in common with that of software and learning tool designers based in cognitive science, who use students' spontaneous gestures while problem-solving as data for the design of 'conceptually-ergonomic' software and other learning tools (Abrahamson 2004).

Preamble: Background and purpose of this study

Working as a linguistic/paralinguistic researcher in mathematics education, I became interested in students' use of gesture in communicating about mathematical graphs, as observed in classrooms where I supervised or assessed teachers and in my own teaching practice. I had noticed that most mathematics teachers (including me) produce a lot of gestures when explaining concepts around mathematical functions and their graphs. It is also common for teachers to elicit gestured graphs from their students as a way of checking student understanding of taught concepts.

I had noticed informally that my own gestures when teaching sometimes differed quite dramatically from those of some of my students. For example, when standing in front of a class of high school students seated at tables, I would ask the whole class simultaneously to gesture the shape of the graph for the function $y = 4$. I would be watching to see whether students knew that this function was a straight horizontal line located four units above the x-axis. What I notice was this: that while most students did gesture a horizontal straight-line movement, some placed this line at the level of their nose or forehead, while others placed it lower, at throat, chest or even waist level. What is more, some students made a very small movement, using only hand and wrist and pointing with the tip of an index finger, while others used a flattened hand and their whole arm to make the horizontal line gesture.

My initial observations took into account the fact that students were seated at tables (unlike me, standing at the front of the room) and that some students might be embarrassed about making large physical movements in front of their peers in the midst of math class. Nonetheless, the variation apparent even within these constraints intrigued me. With a bit of introspection, I realized that my own schema for graphing placed the 'origin' (the (0.0) point) of a Cartesian graph at my navel (which implied an interesting link with the term 'origin' for me at least). When I gestured a graph in front of the class, I imagined the *x*-axis at my waist level, and like most teachers, I used large, whole-body gestures to communicate the shape of the graph to the whole class. It was clear to see that some of my students shared my internalized notions of the placement of the *x*-axis and my propensity for large, physical gestures of graphs, while others placed the *x*-axis higher and used smaller gestures within the classroom context.

Moving from these informal observations to an initial exploratory study, I was interested in observing the following features of gestured graphs:

- variations in the placement of the *x*- axis in relation to the gesturer's body, and potential cognitive, cultural and semiotic interpretations of this placement,
- variations in modes of gesturing a symmetrical graph (using one or both hands, and making use of the body's bilateral symmetry or not),
- variations in eye tracking of gestured graphs,
- interpretations of time, acceleration and fictive motion in relation to the *x*- axis shown by gesture,
- effects of school instruction about graphs and functions on the gestures of advanced secondary math students as compared to those of novice learners in the early years of secondary school,
- genres, conventions or schemata of graph gesturing that might emerge from a reasonably large sample of gestured graphs.

In the first exploratory study, I recruited ten faculty colleagues and family members as convenient subjects and asked each to gesture a given assortment of graphs in front of a video camera. Participants were given 17 cards with enlargements of mathematical graphs, taken from a calculus course and chosen for their variety (symmetrical, asymmetrical and asymptotic graphs, graphs chosen for their interesting visual rhythms, and graphs situated mostly above or below the *x*-axis). Subjects were videotaped in individual clinical sessions. Each participant was asked to stand facing a stationary video camera on a tripod; the graph cards were placed face down on a nearby table. Subjects were asked to look at one card at a time and describe the graph using gesture, as if communicating this shape to someone who could see them but not the graph. Participants were encouraged to use vocal sounds and language to describe the graph as well, but told that they should avoid technical mathematical descriptions because these might be accurate enough to inhibit the need for gesture.

Observations from this exploratory study included the following:

- Participants reported that they were not consciously aware of the choices they made in gesturing the given graphs. Although the gestures produced were not spontaneous gestures accompanying speech and were produced deliberately at the

researcher's prompting, subjects were largely unconscious of features of the gestures they made.
- Participants' gestural representations of the graphs varied widely with regard to placement of the axes (especially the *x*-axis), symmetry, acceleration, direction of movement, handedness, and large vs. small kinesthetic engagement.
- Many participants treated the *x*- axis as a representation of time.
- Some participants used metaphors and non-verbal vocal sounds extensively to describe the graphs. These participants also tended to use large kinesthetic motions in their gesturing. Participants whose motions were more constrained also tended to use fewer metaphoric or non-verbal vocalized descriptions.

Graphs & gestures pilot study in schools: Research questions, participants, procedures

Based on results of the exploratory study, the pilot study reported here explored the following emergent research questions:

1. Can 'elicited gestures representing the graphs of mathematical functions' (hence abbreviated as 'graph gestures') be categorized in a meaningful way that captures the spectrum of learners' cognitive approaches to graphs?
2. If so, can the categories of graph gestures be qualitatively correlated with student attentiveness to and engagement with mathematical features of these graphs?

This pilot study was carried out in April and May, 2008 at three Vancouver, Canada public secondary schools: an east-side school (generally low SES), a west-side school (high SES), and a centrally-located mini school that drew from the whole district (mixed SES). In each school, math teachers were asked to find three or four students willing to participate from two grades, Grade 8 (age 13) and Grade 11 (age 16), representing diversity in terms of gender, ethnicity, and math achievement and enthusiasm. I asked the teachers *not* to inform me before the sessions which students were the "top," "average" or "struggling" math students in their estimation. Grade 8 students would be novices, with little exposure to graphing in school mathematics; Grade 11 students would just have completed a year of intensive study of the graphs of functions and relations. I included the teachers of these students in the study to watch for possible transfer effects from teachers to students.

As in the earlier exploratory study, subjects started the first session standing in front of a stationary video camera on a tripod. I had chosen five of the original 17 graphs, selecting those that had elicited the most varied responses in the exploratory study. As before, students were asked to look at one graph card at a time and describe each using gesture, vocal sounds and words, but not technical mathematical descriptions. I asked each participant to do three 'takes' of his/her gesture for each graph card, the third without words. Figure 1 shows the five graphs used as prompts for the graph gestures.

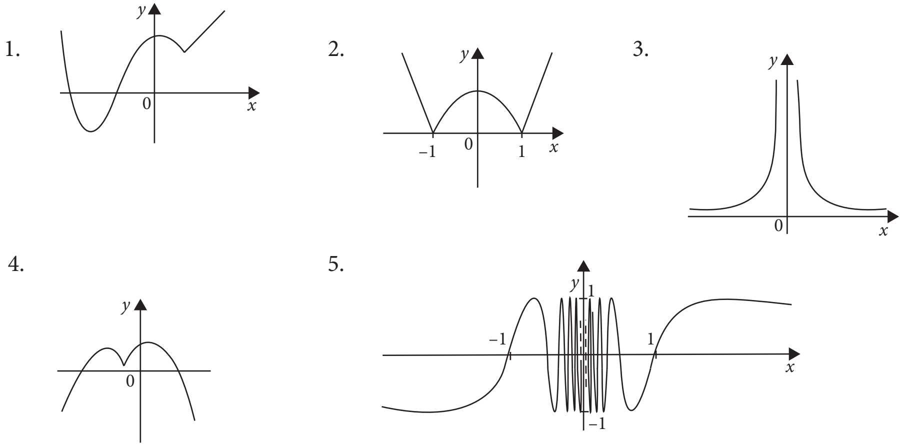


Figure 1. The five graphs used in the second pilot study (in schools).

Table 1.

Session 1 (videotaping gestures at 3 schools)	**M**	**F**	**Totals**
Grade 8 students	5	6	11
Grade 11 students	5	6	11
Teachers	2	2	4
Totals:	**12**	**14**	
Session 2 (re-viewing & discussing tapes)	**M**	**F**	**Totals**
Grade 8 students	3	4	7
Grade 11 students	3	1	4
Teachers	2	2	4
Totals:	**8**	**7**	

I returned to the schools a week later for a second videotaped session, where I conversed with each participant as we watched the video of his/her earlier session. The second session gave me the chance to ask participants what they noticed about their own gestures, share with participants what I had noticed about their gestures, and ask participants for their insights into *why* they had gestured as they had. Table 1 shows numbers of students and teachers who participated in the two phases of data collection by gender and grade.

Results: An emergent diagnostic pattern in graph gestures

My observations of students' graph gestures showed a range of variations in terms of the features of initial interest: placement of the *x*-axis in relation to the body, treatment of the *x*-axis as the time axis, symmetrical or sequential gesturing of graphs, and so on. These individual features clustered in three generic categories, and each participant's collection of gestured graphs fell predominantly into one of these:

1. **An "arm's-length visual model" of the graph (11 of 22 students):** These gestured graphs involved small movements of a finger, hand and arm, without a great deal of larger kinesthetic movement involving the spine. For these students, it was as if they were tracing a small graph on a vertical pane of glass or sheet of paper in front of their upper body, using a finger-tip 'pencil'. Students in this group were the most likely to emphasize accuracy above all. These students would often indicate the locations of the horizontal and vertical axes before they began gesturing and would take pains to place particular numerical values on their 'air graph' and to draw or redraw their gestured graph so that it accurately passed through the correct values.

These students placed the *x*-axis relatively high on the body (at heart, shoulder, throat or nose level) and used a single finger on their dominant hand to make a rather restrained gesture of the graph. Many of these students said in the follow-up interview that they wanted to place the graph where they could *see* it (within their peripheral vision, without moving their heads from a central looking-forward position). Most of these students tracked the line of the imagined graph with eye movement as they traced it with a finger.

This group of students included those who were the slowest to gesture their graphs. In taking pains to make sure their gestures were correct, some of these students moved very slowly, without acceleration, and even made 'erasing' gestures before redoing their gestured graphs. All of the students who did not treat the *x*-axis as the time axis belonged to this group (although many in this group did treat the horizontal axis as a representation of time).

2. **"Being the graph/being in the graph" (9 of 22 students; 4 of 4 teachers):** These gestured graphs involved noticeable movement of the spine and often markedly kinesthetic, whole-body movements. Some students' gestures required them to reach, move off balance or take a step or two. Most of these students used their whole hand or arm, rather than a single finger, to make the gesture, and several used two hands held palm-to-palm, as if preparing to dive into water. One student's gesture was very much a gesture of diving, as if he were following the shape of the graph with his whole body through water.

Students in this group were notable for their bodily, visceral engagement with the shape of the graph. It appeared as though they were 'in' the graph, experiencing the fluctuations of its shape as a movement or journey along a trajectory. Even when gesturing a symmetrical graph with both hands, students in this group would bend, reach and stretch their bodies as if they were touching or riding along the graph.

Most students in this group placed the x-axis relatively low on their bodies (from heart height to waist or hip height). This placement, combined with knee-bending and reaching, allowed them to achieve a whole-body representation of the graph more or less within *reach* (in contrast with the first group, who wanted to have the graph appear within *sight*).

Notably, and particularly for the Grade 8 students in this gesture category, these large gestures were very frequently accompanied by verbal metaphors describing the graphs' shapes in terms of other familiar objects or phenomena. These students were also the most likely to use non-verbal vocal sounds to represent the fictive 'motion' of the graphs. Some of the students in this group produced long strings of metaphors for each graph, which offered contrasting analogies that could function as 'tools for thinking' about different features of the graph and its underlying mathematics. For example, one of the Grade 8 students described Graph 4 as follows:

> **Amber:** This one looks like a round M, or two blobs of jello stuck together, or like when you were in kindergarten and you drew birds, you always draw them as an M. Looks like a kindergarten M or a birdie, like when you draw crows in the sky... Round, two hills again, kind of. And then, two jello blobs, one big, one small, beside each other...yeah. Looks like a 3, but then turned...turned somehow. And then, looks kind of like, there's this little bomb thing, and then it shoots out, shoots outwards.

3. **"Inaccurate, not aware of what counts as salient" (2 of 22 students):** These students had difficulty producing gestures for the graphs, hesitating repeatedly or rushing through the task. Gestural movements did not correspond accurately to the shapes of the graphs, and often large sections of the graph were omitted. Successive 'takes' of the same graph often differed wildly. These students sometimes tried to produce two-handed, symmetrical gestures to represent asymmetrical graphs, produced 'pointy' gestures for rounded curves or vice-versa, and often picked up the graph cards between takes to stare at them at close range. These students produced metaphors for portions of a graph (for example, describing a shape as a 'half pipe', 'hill' or 'checkmark'), but not for its overall shape.

It appeared that these students were encountering two kinds of difficulties: a struggle with perceiving each graph in its overall shape (as a unified entity rather than a collection of parts) and a lack of schemata for identifying and interpreting mathematically salient features of the graphs (relative heights of maxima/minima, axis crossings, discontinuities, symmetries or asymmetries, etc.)

After students' graph gestures were coded and categorized, results were compared with teachers' holistic year-long assessments of student participants as "top", "average" or "struggling" students. The correlation was striking: of the 22 students videotaped, 21 were accurately categorized as top, average or struggling students based solely on the coding of their graph gestures. Students with graph gestures in Category 1 were the average students – hard-working, but depending on rote memorization of formulas and algorithms to get by in math class. Students whose graph gestures fell into Category 2 were the top mathematics students, who brought both accuracy and imagination to their work and were judged by their teachers to have the greatest depth of mathematical understanding. Students whose graph gestures were coded as Category 3 were struggling and at risk of failing mathematics – with the exception of one student. That student consistently gestured only a portion of the graph (for example, only the right-hand side of a symmetrical graph), but was rated as one of the top students in her class by her teacher.

Discussion

I propose the following reasons for the predictive accuracy of these coding categories for graph gestures:

- Category 1 students were precise and followed rules carefully, but often depended on memorization and algorithmic thinking rather than engaging fully with math concepts. These students had learned to value specificity, accuracy and correctness as the principle features that would lead to success in mathematics class and perhaps even the principle features that characterized mathematics as a discipline. An overriding focus on precision and accuracy offered students in the first group a singular, 'one-way' and somewhat rigid approach to a graph (or other mathematical concept). Keeping mathematical ideas 'at arm's length' gave them a sense of control and correctness, but at the cost of full imaginative engagement.
- Category 2 students' visceral, experiential approach to the graphs and multiple metaphors and verbal/kinesthetic/visual representations allowed them multiple potential entry points for sense-making and the creation of more robust mathematical conceptual objects. These students' whole-body engagement offered a way to bring somatic and imagistic imagination into play in their mathematics. Accuracy was not discounted or sacrificed here, but it was not treated by these students as the most salient feature of their exploration of the graphs. Students in the second group showed a conceptually flexible approach to mathematics learning.
- Category 3 students were in urgent need of help in learning to see graphs as whole objects and in bringing attention to those features of graphs considered mathematically salient. Both of the students identified in category 3 were in Grade 8, at the start of their secondary school career. If they were to carry on learning mathematics in mainstream classes, immediate remediation was needed.

Further implications for research and practice

These results have strong implications for the practice of mathematics education in secondary schools. Until recently, the norm for high school mathematics classes was to seat students in rows at individual desks and to encourage them to sit quietly, copying down the teacher's lecture notes, answering the teacher's questions and working silently and individually on examples and homework questions. In fact, despite recent reform efforts, the great majority of North American and other mathematics classes continue to operate within these norms.

Results from this pilot study suggest, however, that mathematics students benefit from being able to engage in an embodied, visceral way with mathematical objects likes graphs through large gestures and kinesthetic whole-body movement. Contrary to earlier classroom norms, this study shows that the students who hold mathematics 'at arm's length' and use the most restrained movements to gesture graphs are less capable of noticing mathematically-salient points than the students who internalize the mathematics and make large gestures. In other words, *being the graph* in a fully-embodied way fosters engagement and attentiveness far more than merely *seeing the graph*.

This suggests that an embodied gestural approach to the teaching of graphs and functions would be helpful in offering a multimodal resource for learners to draw on in their studies. That is not to say that current teaching methods using algebra, word problems, tables of values and drawn diagrams ought to be abandoned – quite the contrary. Rather, these more traditional methods ought to be supplemented by elicited large, close-up gestures, especially in the initial stages of teaching mathematical functions. Gestural work is not sufficient on its own, but when accompanied by focused teaching that helps make salient features of graphs visual, kinesthetic and audible, gesture can play an important role as both a mode of expression and an experiential learning resource.

Two further hypotheses arise from the results of this pilot project:

Hypothesis 1: Gestured graphs can offer the basis for a concise and accurate diagnosis of students' patterns of noticing and engagement in secondary mathematics.

Hypothesis 2: An early intervention that leads all students to gesture graphs closer to that of the most engaged students' gestures can improve students' patterns of noticing and engagement in secondary mathematics.

Ongoing research in the Graphs and Gestures project will test whether whole-group interventions of this kind at an early stage in secondary mathematics education might help draw student attention to salient features of the graphs of mathematical functions several years *before* students are introduced to these functions through the traditional means of mathematical word problems, algebraic equations and tables of ordered pairs. We will experiment with teaching students to 'read graphs with the body' as a primary way of knowing about mathematical functions, with the hope that a groundwork of embodied, gestural mathematics of functions in Grade 8 will become a useful tool and referent for these students when they spend most of the year in their Grade 11 and 12 mathematics classes working with mathematical functions and their graphs.

References

Abrahamson, D. 2004. "Embodied spatial articulation." In *Proceedings of the 26th Meeting of PME-NA, Vol. 2*, D. E. McDougall and J. A. Ross (eds), 79 –797. Toronto: Preney.

Alibali, M. W. and Nathan, M. J. 2007. "Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson." In *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron and S. J. Derry (eds), 349–365. Mahwah, NJ: Erlbaum.

Arzarello, F. and Edwards, L. 2005. "Gesture and the construction of mathematical meaning." In *Proceedings of the 29th Conference of IGPME, Vol. 1,* H. L. Chick and J. L. Vincent (eds), 123–154. Melbourne: PME.

Arzarello, F., Paola, D., Robutti, O. and Sabena, C. 2009. "Gestures as semiotic resources in the mathematics classroom." *Educational Studies in Mathematics* 70 (2): 97–109.

Edwards, L. D. 2009. "Gesture and conceptual integration in mathematical talk." *Educational Studies in Mathematics* 70 (2): 127–141.

Flevares, L. M. and Perry, M. 2001. "How many do you see? The use of nonspoken representations in first-grade mathematics lessons." *Journal of Educational Psychology* 93: 330–345.

Gerofsky, S. 2008. "Gesture as diagnosis and intervention in the pedagogy of graphing." *Proceedings of the 32nd Conference of IGPME*. Morelia, Mexico: PME.

Gerofsky, S. and Marchand, P. 2006. "The body, the senses and mathematics learning." In *Proceedings of the 2006 Meeting of the Canadian Mathematics Education Study Group*, P. Liljedahl (ed). Calgary, Alberta.

Goldin-Meadow, S., Cook, S. W. and Mitchell, Z. A. 2009. "Gesturing gives children new ideas about math." *Psychological Science* 20 (3): 267–272.

Goldin-Meadow, S., Kim, S. and Singer, M. 1999. "What the teacher's hands tell the student's mind about math." *Journal of Educational Psychology* 91 (4): 720–730.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D. and Wagner, S. 2001. "Explaining math: Gesture lightens the load." *Psychological Science* 12 (6): 516–522.

Lakoff, G. and Núñez, R. 2000. *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics into Being*, New York: Basic Books.

Maschietto, M. and Bartolini Bussi, M. G. 2009. "Working with artefacts: Gestures drawings and speech in the construction of the mathematical meaning of the visual pyramid." *Educational Studies in Mathematics* 70 (2): 143–157.

McNeill, D. (ed.) 2000. *Language and Gesture.* Cambridge: Cambridge University Press.

Nemirovsky, R. and Borba, M. 2003. "Perceptual-motor activity and imagination in mathematics learning." In *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th Conference of PME-NA,* N. Pateman, B. Dougherty and J. Zilliox (eds). Honolulu, HI: University of Hawaii.

Nemirovsky, R. and Ferrara, F. 2009. "Mathematical imagination and embodied cognition." *Educational Studies in Mathematics* 70 (2): 159–174.

Núñez R. 2004. "Embodied cognition and the nature of mathematics: Language, gesture, and abstraction." In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, K. Forbus, D. Gentner and T. Regier (eds), 36–37. Mahwah, NJ: Erlbaum.

Núñez, R. 2009. "A fresh look at the foundations of mathematics: Gesture and the psychological reality of conceptual metaphor." In *Metaphor and Gesture*, A. J. Cienki and C. Müller (eds), 93–114. Amsterdam: John Benjamins.

Radford, L. 2009. "Why do gestures matter? Sensuous cognition and the palpability of mathematical meanings." *Educational Studies in Mathematics* 70 (2): 11–126.

Radford, L., Demers, S., Guzmán, J. and Cerulli, M. 2003. "Calculators, graphs, gestures, and the production of meaning." In *Proceedings of the Conference of PME27 –PMENA25, Vol. 4*, N. Pateman, B. Dougherty and J. Zilliox (eds), 55–62. Honolulu, HI: University of Hawaii.

Rasmussen, C., Stephan, M. and Allen, K. 2004. "Classroom mathematical practices and gesturing." *Journal of Mathematical Behavior* 23: 301–323.

Reynolds, F. J. and Reeve, R. A. 2002. "Gesture in collaborative mathematical problem-solving." *Journal of Mathematical Behavior* 20: 447–460.

Robutti, O. and Ferrara, F. 2002. "Approaching algebra through motion experiences." In *Proceedings of the 26th IGPME Conference, Vol. 1*, A. D. Cockburn and E. Nardi (eds), 111–118.

Roth, W-M. 2001. "Gestures: Their role in teaching and learning." *Review of Educational Research* 71: 365–392.

Roth, W-M. and Thom, J. S. 2009. "Bodily experience and mathematical conceptions: From classical views to a phenomenological reconceptualization." *Educational Studies in Mathematics* 70 (2): 175–189.

Saussure, F. (1915/1965). *Course in General Linguistics.* New York: McGraw-Hill.

Tall, D. 2003. "Using technology to support an embodied approach to learning concepts in mathematics." In *História e Tecnologia no Ensino da Matemática, vol. 1*, L. M. Carvalho and L. C. Guimarães (eds), 1–28, Rio de Janeiro, Brasil.

Tall. D. 2004. "Building theories: The three worlds of mathematics." *For the Learning of Mathematics* 24 (1): 29–32.

CHAPTER 19

How gesture use enables intersubjectivity in the classroom

Mitchell J. Nathan and Martha W. Alibali
University of Wisconsin-Madison

In communication it is essential for speaker and listener to establish intersubjectivity, or "common ground." This is especially true in instructional settings where learning depends on successful communication. One way teachers enable intersubjectivity is through the use of gestures. We consider two circumstances in which gestures establish intersubjectivity: (a) making conversational repair and (b) explicitly relating the novel (target) representation to a familiar (source) representation. We also identify two main ways gesture is used in establishing intersubjectivity. *Linking gestures* are sets of attention-guiding gestures (often deictic gestures) that delineate correspondences between familiar and new representations. *Catchments* use recurrent hand shapes or movements to convey similarity and highlight conceptual connections across seemingly different entities.

Communication is an effort to ground meaning in both the cognitive and social realms (Clark 1996). To enable common ground, or *intersubjectivity*, among agents in a social setting, one needs to delineate the common referents for all listeners. One situation in which establishing intersubjectivity is particularly important, though also challenging, is classroom instruction. In this social setting, there are frequent references made to complex ideas, new representations and abstract systems of notation. In such circumstances, intersubjectivity serves both the student and the teacher. For the student, common ground is necessary in order to comprehend the teacher's actions and statements. For the teacher, common ground is necessary in order to connect to students' prior knowledge and experiences, as well as to interpret and assess students' actions and comments, and to appropriately respond to students' questions.

We propose that one way that teachers enable intersubjectivity is through the use of gestures. We consider two different circumstances in which gestures serve this role. In one, the teacher identifies or anticipates misunderstandings and uses gestures to institute conversational repair. In the second, the teacher presents a novel representation.

To imbue the new representation with meaning, the teacher uses gesture to connect the novel (target) representation to a familiar (source) representation.

In addition to identifying these two circumstances, we identify two main ways in which gesture is used for establishing intersubjectivity. First, *linking gestures* are sets of attention-guiding gestures (often deictic gestures) that delineate referential correspondences between the familiar and new representations (Alibali & Nathan 2005, 2007; Nathan 2008). Second, gestural *catchments* (McNeill & Duncan 2000) use repeated features, such as recurrent hand shapes or hand movements, to convey similarity and to highlight conceptual connections across seemingly different entities. The findings presented here suggest a new view of instruction as communication and underscore the central role of gesture for enabling intersubjectivity during instructional communication as a way to foster meaning making and learning.

Intersubjectivity in the classroom

Common ground provides a shared frame of reference within which any interaction unfolds. Its centrality to many theories in the social sciences cannot be overstated. Vygotsky (1986) considered intersubjectivity to be at the heart of learning and of consciousness itself. Schegloff (1992) elevates intersubjectivity, stating that it is "theoretically anterior" to all other considerations in social science because without intersubjectivity, social science stands without reference to the world it purports to identify and describe. Rather than a force that acts on the discourse, intersubjectivity can be regarded as a precondition for discourse itself to occur (Nystrand 1997). When established, intersubjectivity affects listeners' comprehension and subsequent uptake (Wells & Arauz 2006). Even when speakers exhibit divergent – possibly opposing – perspectives, there can be intersubjectivity, as people draw upon commonly held concepts and representations when they articulate their ideas and critique the ideas of others (Nathan, Eilam & Kim 2007).

Classroom learning depends on successful communication between teacher and student and among students. Intersubjectivity is particularly important and challenging when teachers communicate about new representations or concepts. To enable intersubjectivity, one needs to (a) delineate common referents and (b) establish the relations between them. In this chapter, we present two examples to illustrate gesture's role in each of these activities. In the first example, we show how gesture is used in anticipation of a trouble spot during instructional communication. In this case, the teacher utilizes a gestural catchment to provide conversational repair so that all those involved are likely to come away with a common understanding of the referents involved. In the second case, we examine a teacher's use of a novel abstract representation – a matrix – to record, compare and ultimately algebraically model patterns of growth. In this example, the teacher uses a series of attention-guiding deictic gestures, which we call *linking gestures*, to establish the correspondence between values in the matrix and the physical objects to which they refer.

Gestural catchment to provide conversational repair

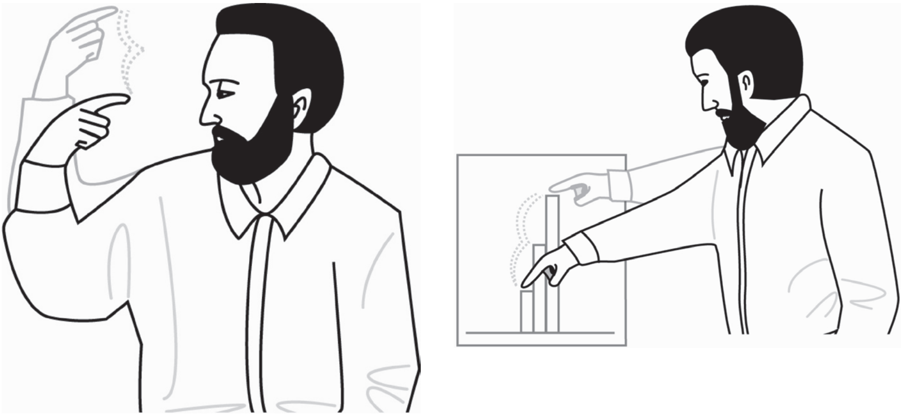
In this example, a 7th grade math teacher starts out sitting on the desktop, speaking and gesturing without any physical referent to his beginning algebra class. He describes the verbal rule the class generated the previous day to describe the growth patterns of a set of tiles – "add the next odd number" – and relates it to the specific values that follow from the rule. Brackets indicate the speech that co-occurs with the gesture in each line. Gesture descriptions are then provided in boldfaced text.

Example 1: The Bar Graph

1 ((Facing the class)) You can notice that [you can add ... the next odd number]
Sequence of hopping gestures
2 so ... ((T turns and walks to the board))
3 we added [three]
Point to graph
4 [we added five to get the next one]
Point to graph
5 [we added seven to get the next one]
Point to graph

In Line 1 the teacher describes a general rule that was used to obtain a series of values and uses a hopping gesture (Figure 1a) as if to depict the pattern exhibited by the corresponding bar graph in thin air. He seems to realize that this may be difficult for the students to understand; he pauses in the midst of line 1, and says, "So ..." in Line 2, stands up and moves to the board, where there is a bar graph the class had prepared the day before. In Lines 3–5 he repeats the hopping gesture (Figure 1b), indicating the tops of the bars, this time using specific referents in speech as well ("we added three", "we added five"). Thus, in this episode, the teacher linked (1) a verbal description of an abstract rule ("you can add ... the next odd number"), (2) a set of concrete values that implement that rule ("3, 5"), and (3) two variations of a graphical representation of the series of bars that result from execution of the rule: a figurative graph in the air and a graph drawn on the board.

For our current purpose, the move at Line 2 from facing the class to referring to the drawing of the graph on the board is of particular interest. It is here that the teacher acts out a kind of replication of his previous action. The first act (Line 1; see Figure 1a) can be interpreted as a sign (an iconic one) in the Peircean sense, following LeBaron and Streeck (2000). The hand in motion is acting out an idea without a perceptually salient context or referent. Of course, it is meaningful to the speaker. We can also infer that it is intentional as well, since he means for these particular gestures to be understood by the listeners.



(a) (b)

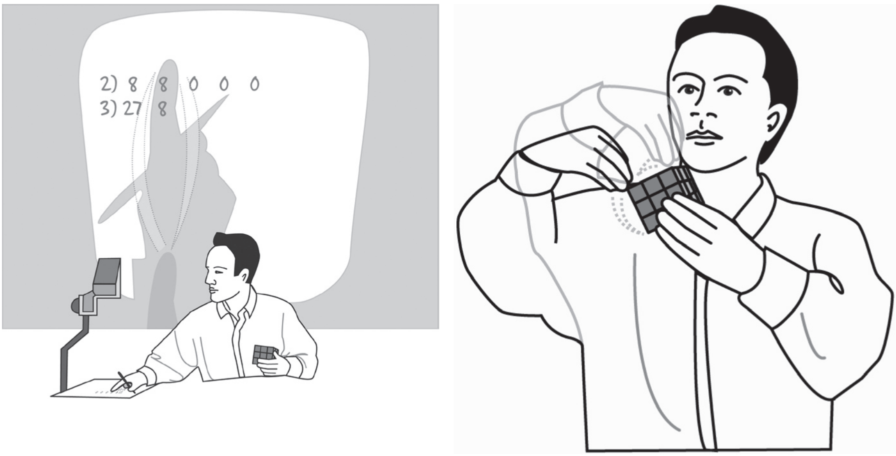
Figure 1. (a) Teacher describes a general rule the class used to obtain a series of values (add the next odd number) using a hopping gesture (a series of points that hop from one location to the next) in space in front of him. (b) He repeats the hopping gesture, this time indicating points at the tops of the bars in a bar graph and using specific referents in speech as well ("we added three", "we added five"). See text for additional details.

However, these are not sufficient conditions for effective instructional communication. Specifically, the teacher's actions are not likely to be understood by the students because the teacher has not satisfied the two conditions for intersubjectivity: (a) He has identified no common referent for the students (i.e., students do not know he is invoking the bar graph), and (b) there is no relation established between such a referent (if there is one) and his sign (i.e., students do not know that each finger position in the sequence is locating the tops of each bar in the graph of the mathematical function). Line 2, then, serves as a self-initiated repair (Schegloff, Jefferson & Sacks 1977) to the instructional conversation. Repairs can take many forms, but they are commonly enacted by repeating or re-voicing the offending material. After recognizing a trouble source (Line 1), the teacher re-voices and employs a salient representation of the graph in the process. He has thereby provided one of the conditions for intersubjectivity – a common referent. The second condition – establishing the relation between sign and referent – is met through the gestural catchment. In a *catchment* (McNeill & Duncan 2000), distinct features of a gesture such as hand shape, location, orientation, trajectory of motion, and so on, are reenacted in order to reinstate the referent of the original gesture. The teacher not only indexes the graph, he also replays the same hopping motion that he used before, and in this way re-invokes his earlier idea from Line 1.

In this short instance, we see how a teacher made a pedagogical move using gesture to address a potential communication failure. The original communicative act was rich with gesture, but the specific form was not one with shared meaning for the speaker and listeners. The teacher's gestures serve to establish intersubjectivity, both by indexing a common and appropriate referent and by establishing the link between the original sign and the intended referent.

Use of linking gestures to enable intersubjectivity for a novel representation

In the next example, an 8th grade mathematics teacher introduces a novel representation to students and shows its usefulness for recording, comparing and modeling the pattern of growth exhibited by cubes of varying side lengths. One aim of the activity is to show students how the growth of the different constituent parts of the cube (the corners, edges, faces, and the total number of blocks) as a function of the side length, follows different mathematical functions (constant, linear, quadratic and cubic functions, respectively). Standing next to an overhead transparency projector while holding a Rubik's Cube in his left hand and pointing to the transparency with his right hand (Figures 2a and 2b), the teacher refers to the values of the number of blocks of each type for each cube in the sequence. When this episode commences, the class has already reviewed the entries for cubes of side length 2 (and volume of 8) and 3 (and volume of 27). His immediate point is to show how the matrix reveals patterns in the data; in this case, the relevant pattern is that no matter the size (above 2), there will always be eight corner blocks.



(a) (b)

Figure 2. (a) While holding a Rubik's cube, the teacher points to the column in the matrix that displays the number of corner cubes as a function of side length (a constant function). (b) The teacher points to the corners of the Rubik's cube.

Example 2: The Cube Problem

1 T: ((Facing the class)) how many of them
2 T: [have three faces painted]?
Points to cube faces.
3 S: Eight
4 T: ((enters '8' in the second row of the second column))
5 T: Eight, and as a matter of fact you should see a pattern right away
6 T: [How about this column]?
Point traces matrix column with 2 entries of '8'.
7 S: eight ... eight
8 T: [((Holding cube))] it's always the [corners, right]?
Beat in silence while holding cube.
Points to corners of cube.
9 T: ((Sets down 3x3x3 Rubik's cube and picks up the 5x5x5 cube))
10 T: and no matter how big the cube gets,
11 T: [there's still always eight corners], right?
Points to corners of cube.
12 T: ((creates table rows 3, 4, 5, and 6))
13 T: ((enters '8' in second column of each row))

In this episode, the teacher is constructing a matrix representation that summarizes for various sizes of cubes, the number of small blocks of each type (one face showing, two faces showing, etc.). The teacher wishes to highlight the fact that, regardless of side length, there is a constant number of "corner" cubes (those that have three faces showing). To make this point, in Line 6 he highlights the column in the matrix that will be filled with 8's when it is completed (Figure 2a). In Line 8, he then points to corners on a small sample Rubik's cube (Figure 2b). To elaborate his point about the universality of this property, he holds up an even larger Rubik's cube in Line 9. He then goes on to mark out additional entries in the matrix (Lines 12–13), which, following his words (Lines 10–11), are expected to all contain the same value in the second column.

The central actions in this episode, for our purposes, occur at Lines 6 and 8–11. Here the teacher uses gesture to link the entries in the matrix and the specific, physical referents on the cube. He establishes the link through a coordinated series of gestures (in this case, pointing actions to the column in the matrix and to the corner cubes). The matrix entries are, of course, the novel representational form. They are highly abstract, not only because of the numerals used to denote the number of blocks, but also because *position* in the matrix denotes the physical referent: The row header is always the side length; the first column always the total number of blocks (volume); and the next column is always the number of corners.

The matrix is a powerful, versatile, but potentially opaque representation. Because of this, intersubjectivity cannot be assumed. Links must be established to connect to students' knowledge and to denote the references of each of the representations. To enable intersubjectivity, both referent and relationship must be established. The teacher accomplishes this using a series of inter-connected pointing gestures between the entries in the matrix and the corresponding components of the example cube he is holding (Figures 2a & 2b). Speech also helps to establish the relationship between the linked referents in this example (Lines 10–11). Thus, during this short episode, the teacher linked (1) a symbolic representation, specifically, the (as yet incomplete) column within the matrix that will represent the constant function, (2) two different physical instantiations of the constant function on two different Rubik's cubes, one small and one large, and (3) a verbal description of the constant function ("no matter how big the cube gets, there's still always eight corners").

Discussion

The two examples presented here highlight the role that gestures play in fostering intersubjectivity during mathematics instruction – laying out and then employing a taken-as-shared set of ideas and representations. In the first example, a teacher uses a gestural catchment to fix an impending communication failure. Gestures help to establish intersubjectivity by indexing a common referent and by establishing the link between the original sign and intended referent. In the second example, the teacher establishes intersubjectivity through a series of inter-connected pointing gestures between entries in the matrix and features of a physical cube. These examples contribute to our understanding of the conditions that exist during socially mediated learning.

Studies of communicative gesture, as with language more generally, unveil a rich and complex set of processes that appear to reflect *both social and individual* aspects of human behavior. There is a lively debate within the literature about the primary role that gesture serves. Some argue that gestures that co-occur with speech serve primarily a self-oriented role, either in facilitating lexical access (e.g., Krauss 1998) or in facilitating the packaging of information in verbalizable form (e.g., Kita 2000). Others (e.g., Kendon 1994, Roth 2003) argue that gesture primarily serves the audience, contributing to the likelihood that ideas will be understood.

A perspective that draws on the role of gestures for intersubjective meaning making suggests a third position (Nathan 2008). According to this view, establishing common ground is paramount to communication, and gestures simultaneously enable individual and communicative (social) functions (Ishino 2007). In serving social functions, gestures guide listener attention, convey substantive information, manage social interactions, and express emphasis (Alibali, Nathan & Fujimori, 2011). Gestures can be used to ground abstract ideas by invoking concrete referents (either physically or as enacted simulations; e.g., Alibali & Nathan 2007, Hostetter & Alibali 2008). They are more frequent when listeners pose questions and exhibit lack of comprehension, or when instructional ideas and representations are novel and more abstract (Alibali & Nathan 2007). At the same time, gestures also serve individual functions .Constructing and maintaining intersubjectivity – especially in educational settings where the goal often is to convey new ideas – makes considerable demands on the processes that mediate speech production. Accessing lexical items and packaging information into syntactic units are integral to formulating fluent speech, and fluency contributes to communicative effectiveness. Thus, by supporting speech production, the self-oriented functions of gesture also contribute to promoting intersubjectivity.

This dual nature of gesture in the personal and social realms is perhaps most striking in example 1, where the teacher uses gestures as a means of conversational repair. Initially (Example 1, Lines 1–2) he expresses a sequence of hops along a line that has no clear referent to the students. He is speaking as he gestures, but the utterance is not coherent and not very descriptive. It is as if he has no immediate words for the ideas he wants to express, though they are readily available to him as simulated actions moving through space (cf. Hostetter & Alibali 2008). His replay of those motions directly on the referent (in this case a drawing of a linearly growing series of entries on a bar graph) provides cohesion (Halliday & Hasan 1976) between the earlier utterance and action and the missing referent and it also helps trigger a more descriptive verbal account of the mathematical idea. In this way gesture appears to have contributed to both the social and individual aims.

In sum, we have argued that teachers use gestures as a tool for enabling intersubjectivity in classroom instruction. We have illustrated that teachers use gestures in service of intersubjectivity when making conversational repairs and when making novel representations meaningful by linking them to other, more familiar representations. We identified two general mechanisms by which gestures are uniquely suited to establish and maintain common ground. In one, gestural catchments (McNeill & Duncan 2000) use repeated features (hand shapes, movements, location) to reinstate conceptual connections across seemingly different entities. In the second, linking gestures guide attention and delineate correspondences between the familiar and new representations (Alibali & Nathan 2007).

In terms of theory, we specified two conditions that we propose need to be met for gestures to enable intersubjectivity. First, the speaker needs to delineate the referents held in common by both speaker and audience, in order to identify taken-as-shared ideas, objects and representations. Second, the speaker must explicitly establish the specific relations that hold between the familiar, common referent and the novel, target representation. Our framework further suggests that, if we wish to understand when and why instructional communication is effective, our analyses must account for teachers' gestures.

References

Alibali, M. W. and Nathan, M. J. 2005. "Teachers use gestures to link multiple representations of mathematical information." In *Gestures and Embodied Action in Teaching, Learning, and*

Practice, N. Enyedy (chair). Symposium presented at the annual meeting of the American Educational Research Association, Montreal.

Alibali, M. W. and Nathan, M. J. 2007. "Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson." In *Video Research in the Learning Sciences*, R. Goldman, R. Pea, B. Barron and S. J. Derry (eds), 349–365. Mahwah, NJ: Erlbaum.

Alibali, M. W., Nathan, M. J. and Fujimori, Y. 2011. "Gestures in the classroom: Whatís the point?" In *Developmental Cognitive Science Goes to School*, N. L. Stein and S. W. Raudenbush (eds), 219–234. New York: Routledge.

Clark, H. H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Halliday, M. A. K., and Hasan, R. 1976. *Cohesion in English*. London: Longman.

Hostetter, A. B. and Alibali, M. W. 2008. "Visible embodiment: Gestures as simulated action." *Psychonomic Bulletin and Review* 15: 495–514.

Ishino, M. 2007. "Intersubjectivity in gestures: The speaker's perspective toward the addressee." In *Gesture and the Dynamic Dimension of Language*, S. D. Duncan, J. Cassell and E. T. Levy (eds), 243–250. Philadelphia: Benjamins.

Kendon, A. 1994. "Do gestures communicate? A review." *Research on Language and Social Interaction* 27: 175–200.

Kita, S. 2000. "How representational gestures help speaking." In *Language and Gesture*, D. McNeill (ed), 162–185. Cambridge, UK: Cambridge University Press.

Krauss, R. M. 1998. "Why do we gesture when we speak?" *Current Directions in Psychological Science* 7: 54–60.

LeBaron, C. and Streeck, J. 2000. "Gestures, knowledge, and the world." In *Language and Gesture*, D. McNeill (ed), 118–138. Cambridge, UK: Cambridge University Press.

McNeill, D. and Duncan, S. 2000. "Growth points in thinking-for-speaking." In *Language and Gesture*, D. McNeill (ed), 141–161. Cambridge, UK: Cambridge University Press.

Nathan, M. J. 2008. "An embodied cognition perspective on symbols, grounding, and instructional gesture." In *Symbols, Embodiment and Meaning: A Debate*, M. DeVega, A. M. Glenberg, and A. C. Graesser (eds), 375–396. Oxford, UK: Oxford University Press.

Nathan, M. J., Eilam, B. and Kim, S. 2007. "To disagree, we must also agree: How intersubjectivity structures and perpetuates discourse in a mathematics classroom." *Journal of the Learning Sciences* 16 (4): 525–565.

Nystrand, M. 1997. *Opening Dialogue: Understanding the Dynamics of Language and Learning in the English Classroom*. M. Nystrand with A. Gamoran, R. Kachur, and C. Prendergast. New York: Teachers College Press.

Roth, W-M. 2003. "Gesture-speech phenomena, learning and development." *Educational Psychologist* 38 (4): 249–263.

Schegloff, E., Jefferson, G. and Sacks, H. 1977. "The preference for self-correction in the organization of repair in conversation." *Language* 53: 361–382.

Schegloff, E. A. 1992. "Repair after next turn: The last structurally provided defense of intersubjectivity in conversation." *American Journal of Sociology* 97: 1295–1345.

Vygotsky, L. S. 1978. *Mind and Society: The Development of Higher Mental Processes*. Cambridge, MA: Harvard University Press.

Wells, G. and Arauz, R. M. 2006. "Dialogue in the classroom." *Journal of the Learning Sciences* 15: 379–428.

CHAPTER 20

Microgenesis of gestures during mental rotation tasks recapitulates ontogenesis

Mingyuan Chu and Sotaro Kita
University of Birmingham

People spontaneously produce gestures when they solve problems or explain their solutions to a problem. In this chapter, we will review and discuss evidence on the role of representational gestures in problem solving. The focus will be on our recent experiments (Chu & Kita 2008), in which we used Shepard-Metzler type of mental rotation tasks to investigate how spontaneous gestures revealed the development of problem solving strategy over the course of the experiment and what role gesture played in the development process. We found that when solving novel problems regarding the physical world, adults go through similar symbolic distancing (Werner & Kaplan 1963) and internalization (Piaget 1968) processes as those that occur during young children's cognitive development and gesture facilitates such processes.

Keywords: gesture, mental rotation, cognitive development, problem solving

Introduction

When we speak, we often spontaneously produce gestures. Gesture and speech are linked at the level of conceptualization, in which a speaker generates prelinguistic thoughts and organizes prelinguistic concepts into suitable units for speaking (Kita 2000; but see, e.g., Krauss, Chen, & Gottesman 2000, for an alternative). One piece of supporting evidence for this view is that individuals produce gestures more frequently when the conceptual complexity of speaking increases but the complexity of other aspects of speaking remains constant (Alibali, Kita & Young 2000; Hostetter, Alibali & Kita 2007; Melinger & Kita 2007; Kita & Davis 2009). According to this view, gesture helps speaking by exploring and organizing spatio-motoric information during the thinking process in preparation for speaking. If gesture is involved in the thinking process for speaking, it is plausible that it also plays a role in the thinking process for other tasks as well. For example, a growing body of evidence has shown that gesture can reveal and shape the thinking processes during problem solving. When adults solve Tower of Hanoi problems[1], gesture and speech mismatch indicates that they are considering alternative strategies (Garber & Goldin-Meadow 2002). In another study, Broaders et al. (2007) showed that encouraging children to gesture can enhance their chance of successfully solving mathematical problem through training. Therefore, gesture is a window into our mind, especially the spatio-motoric thinking process (McNeill 1992, Kita 2000).

Why does gesture play an important role in revealing and shaping the thinking processes in problem solving? One possible explanation can be related to the idea of embodied cognition. That is, our knowledge is deeply rooted in the interaction between our body and the environment, and the simulations of real-world actions and perceptions are the foundation of cognition (Barsalou 1999, Glenberg 1997). Spontaneous gestures, as simulated actions underlying speaking and thinking, may provide insights into the development of individuals' strategies in problem solving and improve individuals' understanding of the problems by providing rich sensori-motor experiences (e.g., Hostetter & Alibali 2008, McNeill 2005).

In developmental psychology, it has long been proposed that children's knowledge of the physical world can be shaped by sensori-motor experiences gained through actions upon physical objects. For example, Piaget (1968) claimed that, young children at the sensori-motor stage (from 0 month to 18 month) of their intelligence development grasp and manipulate every object in their reach distance. By repeatedly acting upon objects, children become able to represent these objects internally. According to Piaget, children's intelligence is rooted in action, and when a certain action can be repeated and generalized, it becomes an internalized action scheme that can be carried out in thought as well as executed materially. Through such an "internalization" process, knowledge gained through the interaction between body and external environment is free from the constraints of the physical world and can be used efficiently to accomplish increasingly complex cognitive tasks. Similarly, Werner and Kaplan (1963) proposed a symbolic distancing process in children's cognitive development. That is, children start out with representations, in which the "symbols" (depicting element) are closely linked to the "referents" (depicted content) both physically and representationally. Over the course of the cognitive development, symbols gradually become separated from referents, and properties of symbols also become independent from properties of referents. Thus, the symbolic distance between the symbols and the referents increases both physically and representationally. For example, children at an early age may seek to gain access to an object by grabbing the adult's hand holding the object, whereas later they become able to point at the object in order to get it (Werner & Kaplan 1963).

1. The Tower of Hanoi is a puzzle that consists of three rods and a number of disks. The puzzle starts with a stack disks of different sizes on one rod with a smaller disk always on top of a larger disk, resulting in a order in which the smallest is on the very top and the largest at the very bottom. The objective of the puzzle is to move the whole stack to another rod, with the following two rules: first, only one disk can be moved at a time; second, a larger disk cannot be placed on top of a smaller one.

Through the "symbolic distancing" process, symbols become self-contained and available to be used freely in thought without anchoring to external referents.

In this chapter, we will review evidence from our recent experiments (Chu & Kita 2008) and other studies in the literature, which investigated how problem solving strategy develops when adults solve problems concerning the physical world and what role gesture plays in the development process. We propose that when adults learn to solve novel problems regarding the physical world, they need to go through similar developmental processes, such as the symbolic distancing (Werner & Kaplan 1963) and the internalization (Piaget 1968) processes that have been observed in children's cognitive development, though it takes much shorter for an adult to go through these processes. More specifically, we hypothesize that adults go through three stages. In the first stage, adults solve the problems by exploring and manipulating, in the form of spontaneous gesture, the stimulus object with their hands. This strategy will provide adults with first-hand sensori-motor experiences about the consequence of the interaction between hand and object. In the second stage, adults still depend on gestures to solve the problems, but the gestural representation becomes *deagentivized*. That is, the agent of the action in the gestural representation disappears, and now the gesturing hand represents the stimulus object and hand movements represent movements of the stimulus object. During the deagentivization process, the symbolic distance between hand and object increases. In the third stage, the knowledge gained through the first two stages becomes *internalized*, and therefore adults become able to solve the problem by pure internal models without the help from overt gestures. Through these stages, the problem solving strategy becomes liberated from constrains of the physical world, and thus can be used more efficiently.

Furthermore, we propose that gesture facilitates the deagentivization process in adults' problem solving. The possible mechanisms underlying the facilitatory role of gesture in the deagentivization process will be discussed in the Section 2.3.

2. Development of problem solving strategy in mental rotation

Since gestures are particularly frequent when people solve problems regarding spatial transformations (Trafton et al. 2006), we used two mental rotation tasks, which are typical spatial transformation tasks, to investigate how the problem solving strategy develops over the course of trials and whether gesture plays a causal role in this strategy change. In the descriptive mental rotation task, participants needed to verbally describe rotation of a three-dimensional object (see Figure 1). In the non-communicative mental rotation task, participants were left alone and asked to choose one of the two mirror three-dimensional objects to match the stimulus object by pressing foot pedals (see Figure 2).

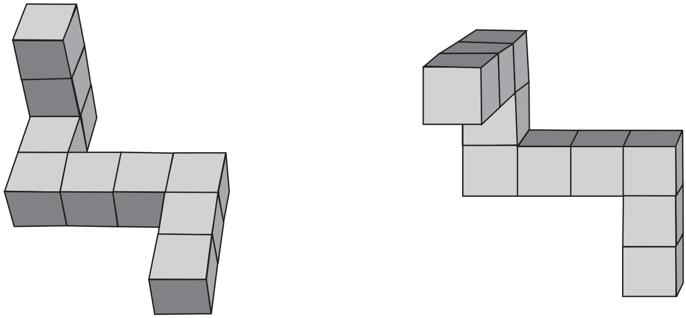


Figure 1. An example of a stimulus in the descriptive mental rotation task.

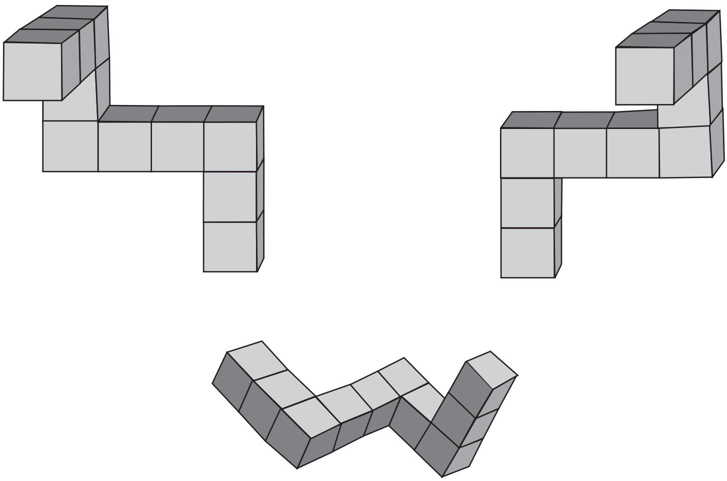


Figure 2. An example of a stimulus in the non-communicative mental rotation task.

We focused our analyses on two types of spontaneous gestures participants produced when they solved the mental rotation tasks: (1) hand-object interaction gestures were those representing the manual exploration and manipulation of the stimulus object. The crucial criterion for this type of gestures was that the participants had to make a grasping or holding hand shape (e.g., the index finger and the thumb were opposed or the two palms were opposed, as if grasping or holding the object). These hand-object interaction gestures reflect the problem solving strategy at the first stage, in which participants gesturally simulated the bodily exploration and manipulation of the stimulus object; (2) object-movement gestures were those depicting the axis, angle, and direction of rotation without any grasping or holding hand shape (e.g., a flat hand representing the object rotated around the wrist or a hand with the extended index finger drew a circle in the air). These object-movement gestures reflected the problem solving strategy at the second stage, in which the representation of the agent disappeared in gestures and the gesturing hand represented the stimulus object.

Furthermore, we also categorized the verbal descriptions of rotation into three types, which were analogous to the distinctions between hand-object interaction gestures (as if an agent manipulated the object) and object-movement gestures (depiction of the object's rotation without an agent). (1) Agent-explicit descriptions highlighted the agent most (e.g., "rotate it clockwise 60 degrees"; "I would rotate it clockwise 60 degrees"). In this type of description, participants used a transitive verb in the active voice. (2) Agent-implicit descriptions only implied the agent implicitly (e.g., "it needs to be rotated clockwise 60 degrees"; "it is rotated clockwise 60 degrees"). In this type of description, participants used a passive form of a transitive verb. (3) Agentless descriptions had no agent at all (e.g., "it rotates clockwise 60 degrees"; "rotate clockwise 60 degrees"[2]; "it is a clockwise rotation 60 degrees"; "clockwise 60 degrees"). In this type of description, participants did not use any transitive verb. Therefore, we have the following deagentivization cline in verbal descriptions of rotation from the most agent salient to the least agent salient: agent-explicit descriptions, agent-implicit descriptions, agentless descriptions.

In addition, the locations of gestures were also coded into near-screen or far-from-screen gestures depending on whether the distance between the hand and the stimulus object in the computer screen was less or more than 20 cm. Therefore, the symbolic distance between hand and object increases as more gestures are performed at the far-from-screen location.

2.1 Evidence for the deagentivization process

According to our theory, adults should start solving the mental rotation task through imaginary manipulation of a stimulus object. Thus, gestures initially should represent somebody holding the object and manipulating it (hand-object interaction gestures). Then the first step in microgenesis of gestures is the deagentivization process, in which the agent disappears from the gestural representation, and gesture becomes more self-contained and detached from the object (rotating-object gestures). Therefore, we investigated the appearance order of the two types of gestures in two time scales: within a single trial, and over the course of the entire experiment in both the descriptive mental rotation task and the non-communicative mental rotation task. In both tasks, we found that hand-object interaction gestures were produced earlier than object-movement gestures were in both time scales (Chu & Kita 2008). Furthermore, in both tasks,

2. One of the reviewers pointed out that "rotate clockwise 60 degrees" can be seen as a case of an ellipsis from "you rotate it clockwise 60 degrees". If that was the case, it should be classified as Agent-implicit description. We think it is unlikely that there is an understood "you" because the participants were instructed to describe the rotation to themselves, although the experimenter was sitting beside the participants.

the location of the gesturing hand became farther away from the stimulus object in the computer screen over the course of the experiment. Finally, we found that in the descriptive mental rotation task, participants' verbal description of rotation and gestural depiction of rotation provided converging evidence for the deagentivization process. Participants who showed a clear sign of deagentivization in gestures (e.g., changing from hand-object interaction gesture to object-movement gesture) also showed a sign of deagentivization in their speech (e.g., changing from agent-explicit description to agentless description), whereas those who did not show a clear sign in gesture tended not to do so in speech either (Chu & Kita 2008).

The above results suggest that, when adults solve mental rotation tasks, they initially imagine manipulating the stimulus object by hands, which can be reflected by the use of hand-object interaction gestures. At this stage, the gestural representation is anchored to the object. As participants become familiar with the stimulus object, the gestural representation is deagentivized in the sense that it becomes self-contained and detached from the stimulus object and the agent of hand movements disappears.

The deagentivization process is compatible with the idea that people go through a schematization process in problem solving, in which irrelevant information is thrown out from the gestural representation of the stimulus object (Schwartz & Black 1996). In this study, participants were presented with line drawings of an interlocking gears system and asked to describe how they decided the rotation direction of the target gear. The authors found that people initially produced gestures representing the rotational movements of each gear. Then, over the course of the experiment, people produced "ticking" gestures which simply marked off each gear without representing the rotational movements of each gear. The authors concluded that people started solving the gear problems by simulating the rotational movement of each gear. Then the dynamics of the gears became faded out from the gestural representation, and people solved the problem by simply counting whether the number of gears was odd or even. During this process, the rotational movement of each gear, which was not directly relevant to the solution of the problem, disappeared in the gestural representation of gears. Similarly, in the deagentivization process, information about the agent, which was not logically necessary for the solution of mental rotation tasks, gradually dropped out of the gestural representation of the stimulus object.

Our finding that the distance between the gesturing hand and the stimulus object increased over the course of the experiment is consistent with the finding in LeBaron and Streeck (2000), in which the authors examined spontaneous gestures produced by a professor in an architecture class. The authors found that the professor initially explored and highlighted a curve shape of a cardboard model by touching and moving his extended index finger along the edge of the curved shape. With increasing frequency over a few minutes time, the professor described the same curved shape by performing similar gestures in mid-air without touching the cardboard, and therefore the physical distance between hand and object increased.

2.2 Evidence for the internalization process

According to our hypothesis, the second step in the microgenesis of the strategy for mental rotation is the internalization process, in which overt gestures are no longer needed, and individuals are able to solve the problem by using internal models as they become more and more familiar with the task. Therefore, we investigated how the rates of hand-object interaction gestures and object-movement gestures changed with the progress of the experiment in both the descriptive mental rotation task and the non-communicative mental rotation task. We found that, in both tasks, the rates of both hand-object interaction gestures and object-movement gestures decreased over the course of the experiment (Chu & Kita 2008). This result suggested that as participants became more experienced in the task, overt gestures were replaced by more efficient internal models.

In the literature of human movement control, it has been suggested that an internal model can accurately predict the sensory consequences of motor commands, and it is essential in performing complex human motions (Wolpert et al. 1995). Individuals appear to use internal models of the physical properties of objects in order to plan and control the grip forces required to stabilize the objects (e.g., Johansson et al. 1992, Flanagan & Wing 1997). Furthermore, in relation to mental rotation tasks, it has been suggested that an internal model becomes decoupled from external motor strategies over the course of the experiment. In a two-task interference study, Wexler et al. (1998) found that although a manual rotation made a mental rotation of the stimulus object faster and more accurate when the two rotations were in the same direction than when they were in opposite directions (the same result was found in Wohlschläger and Wohlschläger (1998)) in the first session of the experiment, the interaction between the directions of the manual and mental rotation disappeared in the second session of the experiment. The authors suggested that an internal model that was not coupled to the external motor strategy developed through practice, and therefore the effect of manual rotation on mental rotation disappeared in the second session of the experiment. In a neuroimaging study, de Lange, Hagoort and Toni (2005) provided supporting evidence for the existence of an internal model in mental rotation tasks, and such an internal model is independent of actual hand movements. The authors found that the dorsal precentral gyrus is responsible for generating internal models for motor plans, whereas the primary motor cortex deals with the actual movement execution.

2.3 Evidence for the facilitatory role of gesture in the deagentivization process

According to our theory, gesture can not only reveal the development of problem solving strategies, but can also play an active role in facilitating the change of strategies. More specifically, we hypothesize that gesture facilitates the deagentivization process when people solve mental rotation problems. In one of our experiment (Exp. 4 in Chu & Kita 2008), we randomly assigned the participants to gesture-allowed and gesture-prohibited groups and compared their verbal description modes in the two groups. Our results showed that the verbal description of rotation overall indicated more deagentivized strategies in the gesture-allowed group than in the gesture-prohibited group. Thus, without the help of gesture, participants in the gesture-prohibited group were less likely to deagentivize their motor strategies. Furthermore, we found that participants were more likely to use deagentivized description modes (e.g., agent-implicit or agentless description) in the first trial in the gesture-allowed group than in the gesture-prohibited group. This suggested that gesture facilitated deagentivization within the first trial even before the verbal description started. Finally, we found that those participants who used agent-explicit description in the first trial were more likely to deagentivize their description mode in the following trials in the gesture-allowed group than in the gesture-prohibited group. These results indicate that gesture facilitates the deagentivization process.

Why can gesture facilitate the deagentivization process? Here, we conjecture two possible mechanisms. First, the flexibility of the gesture execution may provide a problem solver with sudden insights of different strategies. For example, initially people use their gestures to simulate the manipulation of the stimulus object with a holding or grasping hand shape, but sometimes a gesture may be accidentally performed with a more lax flat hand shape. In this case, people may realize that they do not have to solve the problem by simulating the manual action upon the object, and they can use their hand to represent the object itself as well. Such an "accidental discovery" may prompt the transition from hand-object interaction gestures to object-movement gestures, and therefore, the deagentivization process is accomplished.

Another possibility is that gesture may bring out people's implicit knowledge on how to solve the problem by using more efficient strategies. Previous research has shown that when children explain their solutions to problems, their gestures often convey unique information that is not found in the concurrent speech (Goldin-Meadow 2003, Goldin-Meadow, Alibali & Church 1993). For example, children's gestural representations sometimes indicate an implicit awareness of how to solve the problem correctly even though they cannot verbally answer the question correctly (Church & Goldin-Meadow 1986). By gesturing, previously unknown but more efficient strategies might be acted out and noticed. In Broaders et al. (2007), children who failed to solve mathematical equation problems in the pretest were then asked to solve new mathematical equation problems and explain their solutions, with one group encouraged to gesture and the other group told not to gesture during explanation of their solutions. Then, children in both the gesture encouraged group and the gesture prohibited group received the same instructions on how to solve mathematical equation problems. In the posttest phase, children were asked to solve a new set of mathematical equation problems on a paper-and-pencil test. The authors found that children, who were unable to solve the mathematical equation problem in the pretest, added more new and correct problem-solving strategies in their gestures during the manipulation phase when they were told to gesture than when they were not. In addition, the told-to-gesture group solved significantly more questions correctly in the posttest than did the told-not-to-gesture group. The authors suggest that encouraging children to gesture can bring out implicit and correct strategies and subsequently enhances their chance of successfully solving mathematical equation problems through training.

3. Conclusion

In this chapter, we reviewed evidence from our recent experiments (Chu & Kita 2008), in which we investigated the development of problem solving strategies in mental rotation tasks by examining the microgenesis of spontaneous gestures, as well as other studies. Findings from our study and other studies support the claim that when adults solve novel problems regarding the physical world, they go through deagentivization and internalization processes, which are analogous to the cognitive development process in young children. More specifically, we found that people initially solve problems by gesturally simulating the manual manipulation of the stimulus object. At this stage, the strategy is restricted by both the physical properties of the stimulus object and the anatomical constrains of hand and arm. The problem solving strategy then becomes more self-contained and less anchored to the stimulus object. At this stage, the representation of the agent drops out from the problem solving strategy, and the strategy is only restricted by the anatomical limitations of hand and arm. Finally, the problem solving strategy no longer requires overt gestures, and people can solve the problem by internal models. At this point, the problem solving strategy is finally liberated from the restrictions of the physical world that are not essential to the problem and can be used with greater efficiency. Furthermore, we have shown strong evidence that gesture can facilitate the deagentivization process. Therefore, gesture can not only reveal the thinking process in problem solving, but can also play an active causal role in shaping the thinking process.

References

Alibali, M. W., Kita, S. and Young, A. J. 2000. "Gesture and the process of speech production: We think, therefore we gesture." *Language and Cognitive Processes* 15: 593–613.

Barsalou, L. W. 1999. "Perceptual symbol systems." *Behavioral and Brain Sciences* 22: 577–660.

Broaders, S. C., Cook, S. W., Mitchell, Z. A. and Goldin-Meadow, S. 2007. "Making children gesture brings out implicit knowledge and leads to learning." *Journal of Experimental Psychology: General* 136: 539–550.

Chu, M. and Kita, S. 2008. "Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy." *Journal of Experimental Psychology: General* 137: 706–723.

Church, R. B. and Goldin-Meadow, S. 1986. "The mismatch between gesture and speech as an index of transitional knowledge." *Cognition* 23: 43–71.

de Lange, F. P., Hagoort, P. and Toni, I. *2005*. "Neural topography and content of movement representations." *Journal of Cognitive Neuroscience* 17: 97–112.

Flanagan, J. R. and Wing, A. M. 1997. "The role of internal models in motor learning and control: evidence from grip force adjustments during movements of hand-held loads." *Journal of Neuroscience* 17: 1519–1528

Garber, P. and Goldin-Meadow, S. 2002. "Gesture offers insight into problem-solving in adults and children." *Cognitive Science* 26: 817–831.

Glenberg, A. M. 1997. "What memory is for." *Behavioral and Brain Sciences* 20: 1–55.

Goldin-Meadow, S. 2003. *Hearing Gesture: How Our Hands Help Us Think.* Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., Alibali, M. W. and Church, R. B. 1993. "Transitions in concept acquisition: Using the hand to read the mind." *Psychological Review* 100: 279–297.

Hostetter, A. B. and Alibali, M. W. 2008. "Visible embodiment: Gestures as simulated action." *Psychonomic Bulletin and Review* 15: 495–514.

Hostetter, A. B., Alibali, M. W. and Kita, S. 2007. "I see it in my hands' eye: Representational gestures reflect conceptual demands." *Language and Cognitive Processes* 22: 313–336.

Johansson, R. S., Riso, R., Hager, C. and Backstrom, L. 1992. "Somatosensory control of precision grip during unpredictable pulling loads." *Experimental Brain Research* 89: 181–191.

Kita, S. 2000. "How representational gestures help speaking." In *Language and Gesture*, D. McNeill (ed), 162–185. Cambridge, UK: Cambridge University Press.

Kita, S. and Davies, T. S. 2009. Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes* 24 (5): 761–775.

Kita, S., Van Gijn, I. and Van der Hulst, H. 1998. "Movement phases in signs and co-speech gestures, and their transcription by human coders." In *Gesture and Sign Language in Human-Computer Interaction*, I. Wachsmuth and M. Fröhlich (eds), 23–35. Berlin: Springer.

Krauss, R. M., Chen, Y. and Gottesman, R. F. 2000. "Lexical gestures and lexical access: A process model." In *Language and Gesture*, D. McNeill (ed), 261–283. Cambridge, UK: Cambridge University Press.

LeBaron, C. D. and Streeck, J. 2000. "Gestures, knowledge, and the world." In *Language and Gesture*, D. McNeill (ed), 118–138. Cambridge, UK: Cambridge University Press.

McNeill, D. 1992. *Hand and Mind.* Chicago: University of Chicago Press.

McNeill, D. 2005. *Gesture and Thought.* Chicago: University of Chicago Press.

Melinger, A. and Kita, S. 2007. "Conceptualisation load triggers gesture production." *Language and Cognitive Processes* 22: 473–500.

Piaget, J. 1968. *Six Psychological Studies.* NY: Random House.

Schwartz, D. L. and Black, J. B. 1996. "Shuttling between depictive models and abstract rules: Induction and fallback." *Cognitive Science* 20: 457–497.

Trafton, J. G., Trickett, S. B., Stitzlein, C. A., Saner, L., Schunn, C. D. and Kirschenbaum, S. S. 2006. "The relationship between spatial transformations and iconic gestures." *Spatial Cognition and Computation* 6: 1–29.

Werner, H. and Kaplan, B. 1963. *Symbolic Formation. An Organismic-Developmental Approach to Language and the Expression of Thought.* New York: John Wiley and Sons, Inc.

Wexler, M., Kosslyn, S. M. and Berthoz, A. 1998. "Motor processes in mental rotation." *Cognition* 68: 77–94.

Wohlschläger, A. and Wohlschläger, A. 1998. "Mental and manual rotation." *Journal of Experimental Psychology-Human Perception and Performance* 24: 397–412.

Wolpert, D. M., Ghahramani, Z. and Jordan, M. I. 1995. "An internal model for sensorimotor integration." *Science* 269: 1880–1882.

PART V

Gesture aspects of discourse and interaction

CHAPTER 21

Gesture and discourse

How we use our hands to introduce versus refer back

Stephani Foraker
SUNY College at Buffalo

Do speakers use different gestures when first introducing a referent compared to when referring back to the referent? Four adults narrated a story involving two men and several objects. We coded the speech and gestures produced, focusing on the gestures that accompanied nouns or pronouns used to introduce or refer back to referents. The main finding was that gestures with predominantly redundant information (same identity as the spoken referent) occurred more often when introducing a referent in speech, but that gestures with predominantly additional information (different entity than spoken referent, predicate of a referent) occurred more often when referring back in speech. These findings underscore the idea that speakers' gestures can reflect the difference between new and given information in discourse.

Keywords: discourse, information structure, gesture, anaphora, co-reference

1. Introduction

How do the gestures that speakers produce when they talk vary as a function of discourse? McNeill and colleagues have demonstrated that gestures which accompany speech reflect narrative structure as well as meaning, emphasizing that the gestures accompanying speech are a vital part of discourse (McNeill 1985, 1992, 2005; McNeill & Levy 1993; McNeill, Cassell, & McCullough 1994). We first asked whether speakers use different kinds of gestures when they introduce a referent for the first time compared to when they refer back to a referent. Second, we addressed the relationship between a spoken referent and the meaning of an accompanying gesture and its aspects when introducing compared to referring back to a referent.

Discourse consists of repeated reference to the same discourse referents across a series of subsequent utterances, as well as introducing new referents. Typically, repeated reference in speech is established through the use of anaphoric expressions. Anaphors do not describe mental representations of referents directly, but co-refer to or link with antecedent representations that have been previously introduced in the discourse. Anaphors that refer back to a previously introduced referent can take several forms, such as a definite noun phrase, NP, (the cat), a proper name (General Grant), a pronoun (he), or a null anaphor. One explanation of why language users employ an anaphor as a "short-cut" to a previously mentioned referent is the notion of *information structure*. Information structure describes utterance form as a function of the mental states of speakers and listeners, including their current representations of the discourse and the speaker's beliefs about the listener's current representation (Lambrecht 1994). Old referents previously mentioned in the discourse are thought to be more accessible, reflected by the fact that they are often referred to with lighter, less contentful lexical forms, such as a pronoun or null anaphor (Ariel 1990, Chafe 1994, Givón 1983). The concept of old, or recoverable, information aligns most closely with the linguistic notion of *topic*, which at the clause or sentence level is defined roughly as what the utterance or proposition is about (Lambrecht 1994: 15, 118). We will use the terms *given* and *refer back* to indicate this discourse function. New referents that are introduced into a discourse are typically referred to with fuller forms of description, such as definite NPs or names (Ariel 1990, Chafe 1994, Givón 1983). New information, which is unshared or not easily derived from the discourse, is often called the *focus* (Gundel, Hedberg, & Zacharski 1993; Lambrecht 1994: 209). We will use the terms *new* and *introduce* to indicate this discourse function.

Do gestures also participate in the distinction between a referent that is given versus new? Past research by McNeill and Levy (1993; see also McNeill 2000) found that during a narration some gestures were repeated, forming cohesive links across the discourse. Some of these gestures indexed discourse referents, maintaining continuity between gestures by shared location in space, which hand was used, hand shape, or spatial configuration of the two hands. An important aspect of how gestures could encode and track discourse status is McNeill's concept of a catchment: it is "a recurrence of gesture features over a stretch of discourse. It is a kind of thread of consistent visuo-spatial imagery running through a discourse segment that provides a gesture-based window into discourse cohesion" (McNeill 2000: 316). McNeill (2000) posits that such catchments are based on a contrast between old and new information which helps to drive the discourse forward (i.e., communicative dynamism, Firbas 1992: 7). McNeill and Levy (1993) argued that repeated gestures or repeated aspects of gestures helped the speaker track background information, such as which referents were given.

In another experiment, McNeill, Cassell, and McCullough (1994, McNeill 1992: 135–144) exposed participants to videotaped narrations of a cartoon story where the speech and gestures either matched or mismatched the speech. The mismatching gestures included changing the hand used for a referent (as well as spatial location) without any intervening change of location or shift in referent in the actual story (e.g., the left hand on the speaker's left represented Sylvester [speech = "Sylvester"], followed immediately by the right hand on the speaker's right representing Sylvester [speech = "he"]). The participants then retold the story themselves. The researchers found that

30% of the anaphor mismatches from the stimulus video had an effect on the retelling of the story's events.

So and colleagues also found that the location of a gesture is important for tracking a referent during a narration. So, Coppola, Licciardello, and Goldin-Meadow (2005) compared narrations of speakers when they either narrated a story with speech and gesture or with gestures alone. They found that participants used spatial location in their gestures to refer back to previously introduced referents in both cases, although far less often and less consistently when they were also speaking. Using the same narration task, So, Kita, and Goldin-Meadow (2009) also found that the location of speakers' gestures reliably indicated a referent's identity across repeated references, and that this occurred whether the spoken referent was ambiguous or not.

The research to date, then, indicates that gestures can carry information about a spoken referent's identity when referring back to previously mentioned referents. Gesture aspects that appear to carry the referent's identity include which hand is being used, its spatial location, and its hand shape. These aspects seem to be important for consistency of reference, over repeated co-reference. However, it remains an open question as to how gestures might indicate the difference between new versus given referents in a discourse.

The present study examined the gestures that speakers produced during the narration of a story, comparing those that accompanied introducing a new referent versus referring back to a given referent in speech. We first examined whether the type of gesture varied, comparing the prevalence of iconic, metaphoric, and beat gestures. Since referring back typically entails less specification in speech (e.g., pronoun), perhaps speakers will also produce semantically "lighter" gestures, such as more beats than iconics. Indeed, McNeill (1992: 211) and Levy (1984) have found that when referring back in speech, the frequency and complexity of gestures declines as the spoken form is less complex. They have also pointed out that iconics predominate at the narrative level (describing the events of a story), while metaphorics and beats are more frequent at episode boundaries or during meta-narrative and para-narrative speech (McNeill 1992: 214). To our knowledge, though, a straightforward comparison of the gestures accompanying introducing versus referring back on the narrative level has not been reported.

We then compared the different aspects of gestures and their meaning for introducing versus referring back to a referent. Five gesture aspects were coded and analyzed for their relation to the spoken referent: which hand was used, the hand shape, the palm's orientation, the spatial location of the gesture, and the motion of the gesture. Similar to the pattern found in speech, we predicted that the meaning of the different gesture aspects would be redundant with the spoken referent, providing largely the same, semantically full information in speech and gesture. On the other hand, we predicted that fewer gesture aspects would provide redundant information when referring back to a previously mentioned referent in speech. That is, when the referent's identity was given and more accessible, speakers should have less need to specify a referent as fully as when introducing it. We also explored what kind of additional information was available in the different aspects of a gesture that accompanied referring back.

2. Method

We collected narrations from four native-English speaking monolingual adults of college age from the Chicago, USA area (a subset of the So et al., 2005 data set). Following informed consent, they were videotaped while narrating a story involving two men and several objects. They first viewed the 12 scenes of the story together, which lasted 26 seconds. Then the experimenter re-played each scene one by one, and the participant narrated the events for each scene, in order. The narration was produced as a monologue with the experimenter listening passively, rather than as an interactive dialogue, although the experimenter did provide some natural backchannel signals, such as head-nodding.

2.1 Materials

Table 1 provides a description of the stimulus story. The referents considered in this analysis were concrete entities: the officer and the worker were the two animate characters, and the inanimate referents were hat, bench, lunchbox (or lunch), barrel, sandwich, jacket, sink, water taps, water, soap, and bubbles. Referents that were mentioned only once by a participant were not included.

2.2 Data coding

Using Praat (Boersma & Weenink 2007), we first transcribed the speech and annotated it for the two categories of NP (semantically fuller) and pronoun (semantically lighter), and the two discourse functions of introduce and refer back. The first mention of a referent in the whole narration was counted as introducing. All other subsequent mentions of that referent were considered to refer back. Note that this criterion is not as strict as research conducted by Gullberg (2003, 2006), who considered only adjacent clauses, focusing on when the preceding clause contained the referent and the following clause contained the referent as the subject. We chose to include both strict maintenance (in either a parallel or non-parallel sentence slot) as well as cases of returning to a discourse referent following intervening material (see also So et al. 2006, 2009). This choice was informed mainly by our interest in gesture aspects over discourse segments larger than adjacent clauses or sentences. It is an interesting question how the synchronization of gestures tracks with clause-to-clause shifts in discourse, but we have not focused on that finer scale here.

Table 1. Description of the scenes in the stimulus story

Scene	Duration (sec)	Description of each scene
1	2	An officer sits down on a bench beside a barrel with his lunchbox.
2	2	A second man enters the scene from the right. He is taking off his jacket.
3	1	The second man (a worker) salutes to the officer.
4	2	The worker turns on the taps above a sink, which is around the corner from the officer, on the other side of the barrel.
5	2	The officer takes a slice of bread out of his lunchbox and puts it on the barrel top.
6	2	The officer takes some cheese out of his lunchbox and puts that on top of the slice of bread.
7	1	The worker at the sink picks up a bar of soap, which is on the barrel top beside the officer's sandwich fixings.
8	3	The worker washes his hands and face with the bar of soap.
9	2	The worker puts the soap back down on the barrel, but on top of the sandwich fixings (by mistake likely, as he does not seem to look at the barrel top). (This scene has a closer camera angle than other scenes.)
10	3	The officer places a second slice of bread on top of his sandwich fixings, picks it up, and takes a bite.
11	2	The officer stops chewing suddenly with a surprised look on his face.
12	4	The officer is chewing and bubbles are coming out of his mouth.

The speech annotations were then imported into ELAN for accompanying annotation of gestures (Technical Group, Max Planck Institute for Psycholinguistics 2005). We focused on the gestures that overlapped with spoken NPs and pronouns, either in whole or in part. For noun phrases, the gesture had to overlap with the head noun (e.g., *man* in *the man in a suit*); gestures that only overlapped with a verb or predicate (e.g., *in a suit*) were excluded (e.g., Gullberg 2006), as were gestures that occurred fully during a pause (silent or filled).

Gestures were identified based on the stroke and any holds (primarily post-stroke holds), and coded for their type and their physical form. The three types coded were iconic, metaphoric, and beat, following McNeill (1992). Gesture types can often be layered, so the predominant type that the gesture exhibited was used to classify a gesture. Only two deictic points were in the data set, contributed by the same participant, so we did not include them in these analyses. Meta- and para-narrative gestures (McNeill 1992) that did not seem to provide referential or other meaning information about the narrative level were not included in analyses either (see Table 3).

Gesture form was coded by annotating five aspects of the gesture's physical form: which hand was used, hand shape, palm orientation, location, and motion. These aspects were chosen based on past research reviewed in the introduction, as well as an aspect's ability to carry information about a gesture's meaning (McNeill 1992, 2000: 316; Goldin-Meadow 2003; see also Goldin-Meadow, Mylander & Franklin 2007 for these aspects' importance in "home-sign" systems).

1. For *hand used*, we annotated the right and left hands separately, noting symmetrical or asymmetrical for two-handed gestures.
2. For *hand shape*, we noted which American Sign Language hand shape it was the closest to. If the hand shape changed, the beginning and ending shapes were noted.
3. For *palm orientation*, we noted which way the palm faced: up, down, toward or away from the speaker's body, and toward or away from the speaker's vertical center-line. If palm orientation changed, the beginning and ending orientations were noted.
4. For *location*, we annotated two major characteristics. The first was where the gesture occurred in relation to the speaker's body. We used a 9-region matrix, illustrated in Figure 1. The clavicle and hip bones were the horizontal dividers and the two shoulders were the vertical dividers. The center region was also divided into smaller regions for descriptions of locations within this most commonly used region: body landmarks were the sternum (right-left) and the zyphoid process (upper-lower). The matrix was extended into the third dimension with two distances from the body. Near-body was defined as a gesture that fell between touching the body and extending the elbow out to 90°, and far-body was a gesture with the elbow extended past that point. For gestures moving from one region to another, the starting and ending points of the path were noted, as well as any midpoints passed through. The second characteristic of a gesture's location noted was the nature of the location: in neutral space or in a previously defined space. An example of defined space is when a gesture traced the circular outline of a barrel top in neutral space to the speaker's left side, and then a subsequent gesture made use of that outlined region, such as a "put down" movement ending in the barrel top location in space. In such cases, the gestures that shared the (purportedly) defined space had to have a plausible relationship between the meanings of the gestures, tied to the story being narrated.
5. For the *motion* of a gesture, we described the path's shape (e.g., straight, curved), whether it was uni-directional, back-and-forth or in place, the size of the motion, and any descriptive characteristics of the motion's manner (e.g., wavering). For gestures that traced the outline of something, the shape of the outline was noted.

We then interpreted the content or *meaning* provided by each of the gesture aspects. A spoken referent can be accompanied by a gesture that captures information that is *redundant* with the identity of the referent (e.g., *soap* in speech, and a hand shape depicting the shape of a bar of soap in gesture). But a gesture accompanying a referent can, sometimes simultaneously, provide *additional* information which does not simply reinforce the meaning of the spoken referent (e.g., *soap* in speech, and a directional path along which the soap was moved in gesture). For each spoken referent-overlapping

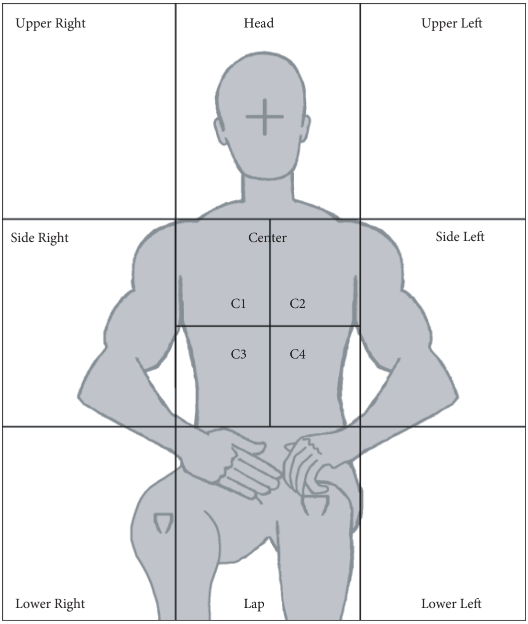


Figure 1. Matrix used to code the location of a gesture. Note that the Right and Left regions extended out as far as the participant reached. Near- and far-body dimensions are not shown; see text for explanation.

gesture pair, we considered whether each of the five physical aspects of a gesture's form (a) provided redundant information about the spoken referent, (b) provided additional information (often about some other referent, but also including a predication of the spoken referent such as an action of or on the spoken referent), or (c) did not appear to provide any information about the spoken referent, other referents, or activity they were involved in. All meta-narrative and para-narrative gestures were excluded for the meaning analyses based on this criterion. Figures 2 and 3 provide examples of this coding scheme. (The fuller spoken context was "There's a guy in some sort of uniform and he's like sitting down to have his lunch. And then, I guess, there's another guy like coming towards him and there's a barrel right next to the police guy.")

The primary way we examined the meaning available in gesture was to classify each gesture as redundant or additional based on the majority of contentful aspects for a particular gesture. Shown in Figures 2 and 3, each physical aspect of a gesture was annotated, and then each aspect's meaning was interpreted in relation to the spoken referent that the gesture overlapped with. Finally, the gesture was classified based on a majority of redundant vs. additional aspects. When there was a tie between the number of aspects, we classified the gesture as additional (this occurred for 5 gestures across the 4 participants).

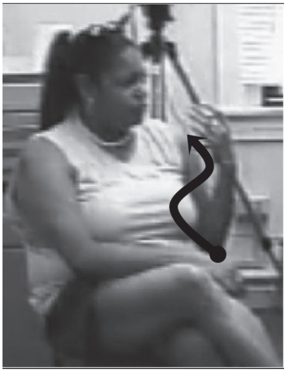
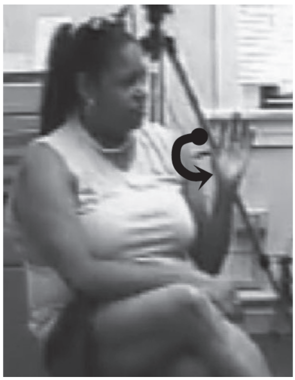
Spoken referent = "another guy" (worker)
Discourse function = introduce the worker
Speech during the gesture = "guy like coming" + pause

Aspect	Form	Meaning	Classify
hand	left	worker	redundant
hand shape	loose 4	n/a	n/a
palm	toward body	n/a	n/a
location	from lap to C2 (through C4), near body, neutral space	worker	redundant
motion	wavering path upward	path of worker entering	additional

Figure 2. Example of the coding scheme for a redundant gesture.

Spoken referent = "him" (officer)
Discourse function = refer back to officer
Speech during the gesture = "towards him"



Aspect	Form	Meaning	Classify
hand	left	worker	additional
hand shape	loose 5	n/a	n/a
palm	toward body rotates to center	(worker) faces officer	redundant
location	C2, near body, defined space	worker	additional
motion	rotation in place	worker turning around	additional

Figure 3. Example of the coding scheme for an additional gesture.

If a particular gesture overlapped with more than one spoken referent, the gesture was classified for each spoken referent separately. Similarly, if more than one gesture overlapped with a particular spoken referent, each gesture was classified separately.

The author annotated and coded all of the speech and gestures of the four narrations. A second coder annotated and coded the speech transcripts, with reliability of 98% for the anaphor form and 99% for discourse function. A third coder annotated gestures for two of the four narrations, coding the gestures for type, physical form, and meaning. Agreement for gesture type was 93% ($n = 83$), for physical form of the gesture aspects 92% ($n = 415$), and for the meaning of gesture aspects 89% ($n = 415$) – redundant aspects were at 95%, with additional and no-information aspects each at 87%.

3. Results

Reported are repeated-measures ANOVAs and paired t-tests, $\alpha = .05$, two-tailed, with participants as the random factor. The results reported below are for tests performed on proportions; tests on arcsine-transformed proportions produced similar results.

First, focusing on the lexical spoken forms, we found that speakers used primarily nouns to introduce an entity ($M = 92.5\%$, $SE = 3.8$), while nouns ($M = 51.4\%$, $SE = 6.3$) and pronouns ($M = 48.6\%$, $SE = 6.3$) were used equally often to refer back. Speakers gestured equally often whether they were introducing a referent ($M = 60.2\%$, $SE = 16$), or referring back (NP: $M = 65.5\%$, $SE = 8$, Pro: $M = 48.4\%$, $SE = 11$). Numerically, each participant gestured the least often when referring back with a pronoun, but this pattern was not significantly different compared to NP cases, $p > .27$.

The first factor was the spoken referent's discourse function: introduce vs. refer back. For the introduce condition, we included only NPs. For the refer back condition, we collapsed across NPs and pronouns since there were no differences between them. Proportions reported below were calculated based on the total number of gestures occurring in the introduce condition separately from the refer back condition.

The first analysis included gesture type as the second factor: iconics, metaphorics, or beats. On average, speakers produced more iconics (63.6%) than metaphorics (18.9%) or beats (17.4%). Comparing these gesture types for introducing vs. referring back showed no significant differences, $ps > .33$. Numerically, beats were more common when referring back (23.0%, $SE = 7$) than introducing (10.5%, $SE = 5$), and iconics were more common when introducing (73.6%, $SE = 13$) than referring back (57.2%, $SE = 14$). Metaphorics showed no difference between introducing (15.9%, $SE = 8$) and referring back (19.8%, $SE = 9$).

The second analysis included gesture meaning as the second factor: redundant vs. additional information. We found evidence that speakers did use their gestures differently in relation to the discourse function of a referent uttered in speech. As shown in Figure 4, when using a noun to introduce a referent, speakers produced gestures that were redundant with the spoken referent 55.4% of the time ($SE = 2.3$) while additional information gestures occurred 44.6% of the time (*SE* = 2.3). In contrast, when using a noun or pronoun to refer back to the referent, speakers produced redundant gestures 37.5% of the time (*SE* = 5.4) and additional ones 62.5% of the time (*SE* = 5.4), significant interaction, $F(1, 3) = 14.73$, $p = .03$. Paired comparisons for redundant vs. additional gestures were marginally different for introducing, $t(3) = 2.36$, $p = .10$ and for referring back, $t(3) = 2.56$, $p = .08$.

We also considered whether this pattern held for *individual referents* within a narration. Not every referent in each participant's narration showed the pattern, but additional gestures outnumbered redundant gestures when referring back at least 80% of the time, for each participant. Also, the pattern held for the main characters (officer and worker), which were animate entities, but was not as strong for the other, inanimate referents. Table 2 provides an illustrative example for one participant, following the officer character. The officer was introduced with a redundant gesture, and referred back to with 4 additional gestures and 3 redundant gestures, as well as cases of not applicable gesture (meta- or para-narrative) and no gesture.

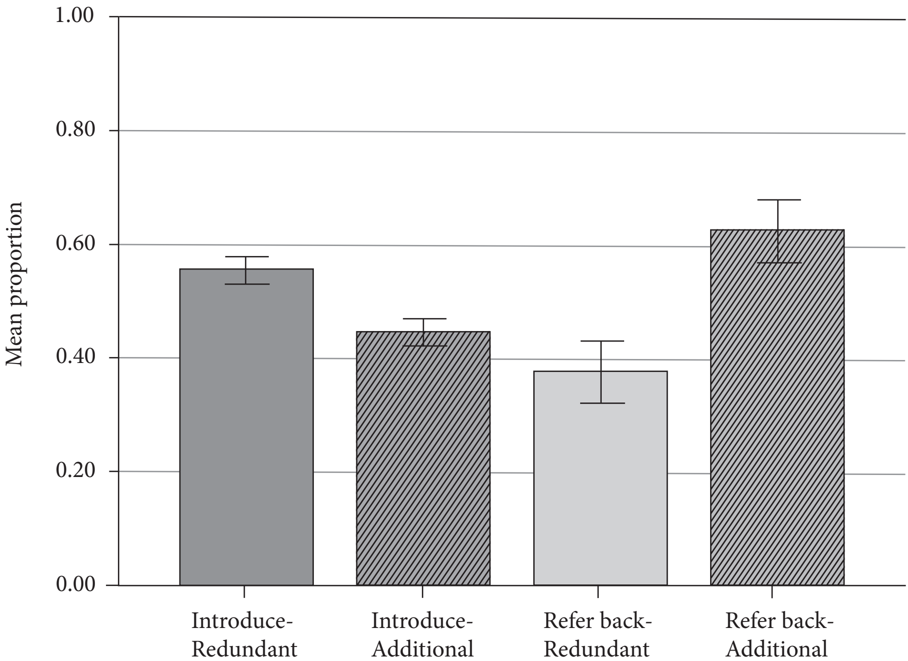


Figure 4. Proportion of redundant (solid bars) vs. additional gestures (striped bars) for introducing (left bars) and referring back (right bars). Error bars represent one standard error of the mean.

Table 2. Gestures that overlap with spoken mention of the officer character produced by one participant, classified as redundant, additional, or not applicable

Order of Mention	Lexical Form	Discourse Function	Gesture No.	Gesture Classification
1	NP	introduce	g1	redundant
2	Pro	refer back		no gesture
3	Pro	refer back	g4	additional
4	NP	refer back	g6	additional
			g7	redundant
5	NP	refer back	g11	not applicable
			g12	redundant
6	NP	refer back	g14	not applicable
			g15	redundant
			g16	not applicable
			g17	additional
7	Pro	refer back		no gesture
8	NP	refer back	g23	not applicable
9	Pro	refer back		no gesture
10	NP	refer back	g32	additional
11	NP	refer back		no gesture
12	Pro	refer back		no gesture
13	Pro	refer back		no gesture
14	NP	refer back		no gesture
15	Pro	refer back		no gesture

When we looked at what kind of additional information was being provided by gesture aspects, we found that the location of another referent or the path of motion (either of the spoken referent or some other referent) provided the most additional information (9% when introducing vs. 28% when referring back), as well as the action of (or on) another referent (12% vs. 17%), followed by a hand shape or outlining indicating another referent (2% vs. 16%). We also found that there were more gesture aspects that did not carry any meaning (redundant or additional) for referring back, particularly when accompanying pronouns (22%) compared to referring back with a NP (9%), or when introducing with a NP (4%).

4. Discussion

The present study found that the type of gestures that speakers produced did not vary when introducing versus referring back to a referent. Iconics were numerically more common when introducing, and beats were more common when referring back, but the differences were not significant. On the other hand, we did find evidence that the gestures speakers produce vary as a function of discourse. When a referent was new, gestures were predominantly redundant with the spoken referent rather than providing additional information beyond the identity of the referent. When a referent was already given, accompanying co-speech gestures provided less redundant information than when introducing – instead, they primarily provided additional information, often about other entities in the discourse, as well as actions of or on the uttered referent. These findings underscore the idea that co-speech gesture is sensitive to co-reference operations and information structure in discourse.

Kita's interface hypothesis (Kita & Özyürek 2003), that gesture and speech generally express the same information, is related to these findings. McNeill (1985, 1992, 2000), too, has stressed the tight synchronization of speech and gesture based on shared meaning. Whether co-expressivity could completely explain the present results is an empirical question for future work. One possibility is that the redundant vs. additional gestures pattern found here might be driven by the propositional content of the speech rather than the given-new status of a referent. For example, it may not be particularly surprising if the gestures that accompanied pronouns were short, not very complex, and less contentful, which could partially account for our results. However, we found that additional information occurred more not only for pronouns that referred back, but also for the more complex NP forms, as well – there were no differences between gestures accompanying pronouns vs. NPs when referring back.

One way to investigate the co-expressivity concern would be to compare the speech segment that a gesture extends over in relation to the proposition that the referent is part of, in speech, and assess whether the kinds of information we found for the motion aspect (including path) would be accounted for. Similarly, if some referent other than the spoken one was indicated by the co-occurring gesture, one would want to know whether that other referent was uttered during the gesture's full extent.

This study is a first step in addressing the question of whether gestures reliably indicate the given-new distinction, as the number of participants reported on in this chapter was rather small, and it would be preferable to examine other stimulus stories with a variety of animate and inanimate referents, and of same and different genders. For example, So et al. (2009) found that when two animate referents were of the same gender, speakers sometimes did not fully specify in speech which character they were referring to (e.g., *he* was ambiguous for referring to one of two males). Interestingly, they found that in such cases, the location of a speaker's gesture did not add any disambiguating information, leaving the spoken referent underspecified in gesture, as well as speech. Since So et al. (2009) focused on the location of a gesture, it would be interesting to know whether the other aspects of a gesture that we have examined in this present study also show no sensitivity to ambiguous reference, or whether it might take multiple aspects of a gesture to add additional, potentially disambiguating information to help specify an ambiguous referent in speech more fully.

Finally, the present results show that speakers' gestures reflect the given-new distinction made in speech, but whether listeners use the information about discourse status that is available in a speaker's gestures is a question that remains to be explored. Because co-speech gestures also play a role in the listener's comprehension, and to some extent are also designed for the listener (Driskell & Radtke 2003, Jacobs & Garnham 2006, Kendon 1994, Özyürek 2002), they may also have the potential to provide listeners with on-line cues about the given-new status of a referent.

References

Ariel, M. 1990. *Accessing Noun-Phrase Antecedents.* London: Routledge.

Boersma, P. and Weenink, D. 2005. *Praat: Doing Phonetics By Computer* (Version 4.5.16) [Computer program]. Retrieved January, 2005, from http://www.praat.org.

Chafe, W. 1994. *Discourse, Consciousness, and Time*. Chicago, IL: University of Chicago Press.

Driskell, J. E. and Radtke, P. H. 2003. "The effect of gesture on speech production and comprehension." *Human Factors* 45: 445–454.

Firbas, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication.* New York, NY: Cambridge University Press.

Givón, T. 1983. "Topic continuity in spoken English." In *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, T. Givón (ed), 343–364. Amsterdam: John Benjamins.

Goldin-Meadow, S. 2003. *Hearing Gestures: How Our Hands Help Us Think.* Cambridge, MA: Harvard University Press.

Gullberg, M. 2003. "Gestures, referents, and anaphoric linkage in learner varieties." In *Information Structure, Linguistic Structure and Dynamics of Language Acquisition*, C. Dimroth and M. Starren (eds), 311–328. Amsterdam: John Benjamins.

Gullberg, M. 2006. "Handling discourse: Gestures, reference tracking, and communication strategies in early L2." *Language Learning* 56: 156–196.

Gundel, J., Hedberg, N. and Zacharski, R. 1993. "Cognitive status and the form of referring expressions in discourse." *Language* 69: 274–307.

Jacobs, N. and Garnham, A. 2007. "The role of conversational hand gestures in a narrative task." *Journal of Memory and Language* 56: 291–303.

Kendon, A. 1994. "Do gestures communicate? A review." *Research on Language and Social Interaction* 27: 175–200.

Kita, S. 2000. "How representational gestures help speaking." In *Language and Gesture*, D. McNeill (ed), 162–185. Cambridge, UK: Cambridge University Press.

Kita, S. and Özyürek, A. 2003. "What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking." *Journal of Memory and Language* 48: 16–32.

Lambrecht, K. 1994. *Information Structure and Sentence Form: Topic, Focus, and The Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.

Levy, E. T. 1984. *Communicating Thematic Structure in Narrative Discourse: The Use of Referring Terms and Gestures.* Ph.D. dissertation, Department of Behavioral Sciences, University of Chicago.

McNeill, D. 1985. "So you think gestures are nonverbal?" *Psychological Review* 92 (3): 350–371.

McNeill, D. 1992. *Hand and Mind: What the Hands Reveal About Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. 2000. "Catchments and contexts: Nonmodular factors in speech and gesture production." In *Language and Gesture*, D. McNeill (ed), 312–328. Cambridge, UK: Cambridge University Press.

McNeill, D. 2005. *Gesture and Thought.* Chicago, IL: University of Chicago Press.

McNeill, D. and Levy, E. 1993. "Cohesion and gesture." *Discourse Processes* 16: 363–387.

Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46,* 688–704.

So, W. C., Coppola, M., Licciardello, V., and Goldin-Meadow, S. (2005). The seeds of spatial grammar in the manual modality. *Cognitive Science, 29,* 1029–1043.

So, W. C., Kita, S., and Goldin-Meadow, S. 2009. Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science,* 33, 115–125.

Technical Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. ELAN – Linguistic Annotator (Version 2.6.3, and later) [Computer program]. First retrieved April, 2006, from http://www.lat-mpi.eu/tools/elan/

Acknowledgments

For helpful comments, I thank attendees of the International Society for Gesture Studies conference in 2007, the gesture community at The University of Chicago, and the editors. I thank Krystal Simms and Rhudwan Nihlawi for their assistance in data annotation. This research was supported by NIH fellowship HD049247 to S. Foraker. Data collection was supported by R01 DC00491 from NIH-NIDCD to S. Goldin-Meadow.

CHAPTER 22

Speakers' use of 'action' and 'entity' gestures with definite and indefinite references

Katie Wilkin[1] and Judith Holler[2]
University of Manchester,[1] and University of Manchester and Max Planck Institute for Psycholinguistics[2]

Common ground is an essential prerequisite for coordination in social interaction, including language use. When referring back to a referent in discourse, this referent is 'given information' and therefore in the interactants' common ground. When a referent is being referred to for the first time, a speaker introduces 'new information'. The analyses reported here are on gestures that accompany such references when they include definite and indefinite grammatical determiners. The main finding from these analyses is that referents referred to by definite and indefinite articles were equally often accompanied by gesture, but speakers tended to accompany definite references with gestures focusing on action information and indefinite references with gestures focusing on entity information. The findings suggest that speakers use speech and gesture *together* to design utterances appropriate for speakers with whom they share common ground.

Keywords: common ground, new and given information, definite and indefinite references, iconic gestures, deictic gestures, entity information, action information, ellipsis

Introduction

One of the central questions gesture researchers have tried to answer in recent years is why we gesture when we speak. This research has led to a greater understanding of the functions of co-speech gestures, and the empirical evidence suggests that they may indeed fulfil a range of quite different functions. For example, co-speech gestures appear to aid the speaker's cognition, such as the processes involved in lexical retrieval (e.g., Pine, Bird & Kirk 2007) or conceptual planning (e.g., Hostetter, Alibali & Kita 2007). Others argue that gestures fulfil communicative functions (e.g., Bavelas & Chovil 2000, Kendon 2004). For example, we know that social context in the form of visibility between speaker and addressee influences gesture rate (Alibali, Heath & Myers 2001; Bavelas, Kenwood, Johnson & Phillips 2002) as well as aspects of gesture form (Bavelas, Gerwing, Sutton & Prevost 2008; Gullberg 2006), and that addressee location can influence speakers' use of gesture space to represent semantic information (Furuyama 2000, Özyürek 2002).

Apart from these overt, physical aspects of the social situation, there is also evidence that more covert processes influence gestural communication, such as the interactants' thinking and understanding. Holler and Beattie (2003b) found that speakers use co-speech gestures to clarify lexical ambiguities for their addressees, both in dialogue-like interactions as well as in more monologue-like narratives. Because verbal ambiguity can be a problem for the addressee but is rarely a problem for the speaker him- or herself, these studies provide evidence that speakers do gesture for their recipient and that they take their addressees' thinking into account when gesturing. Recent research has shown that this conclusion is not restricted to the context of lexical ambiguity but that it generalises to other domains. Some of this research has focused on an aspect fundamental to successful communication, namely the knowledge, beliefs and assumptions interactants mutually share, which has been referred to as 'common ground' (e.g., Clark 1996). Studies examining verbal communication have revealed that common ground leads to more elliptical speech (e.g., Clark & Wilkes-Gibbs 1986, Fussell & Krauss 1989, Isaacs & Clark 1987), amongst other things. Recently, researchers have started to investigate the effects of common ground on gesture use. Gerwing and Bavelas (2004, Study 1) showed that speakers used less complex, precise and informative gestures when they talked to addressees with whom they shared common ground than when talking to addressees with whom they did not share common ground. Similarly, Holler and Stevens (2007) found that speakers encoded less information about the size of entities in gesture when their addressees shared common ground with them regarding this semantic aspect than when they did not. Similarly, Parrill (2010) found that speakers encoded significantly less information about the ground element of an event they were describing when they mutually shared knowledge about this event with their interlocutor than when they did not. Further, findings by Jacobs and Garnham's (2007) suggest that speakers gesture at a lower rate when common ground is built up based on repeated narrations of the same story to the same listener (see also Holler 2003). Taken together, this evidence may lead us to conclude that gestures, like speech, are more elliptical when common ground exists. On the other hand, a study by Holler and Wilkin (2009) revealed that speakers in their common ground condition gestured at a *higher* rate when common ground existed and that they encoded statistically as much semantic information in their gestures in this condition as in the one without common ground.

Several factors could explain the discrepancies between these findings. For example, the studies differed in the way the participants were interacting during the task (free vs. restricted interaction) and in the type of tasks the participants completed (e.g., narratives vs. referential communication tasks). Studies systematically investigating these and other potential factors are currently underway. What we can conclude to date is that common ground appears to influence gestures in a variety of different ways and that the semantic interplay between gesture and speech in this context does not seem to be characterised by one simple pattern. Further research is needed to arrive at a more complete view of how common ground influences communication.

The present study focuses on utterances including definite and indefinite references, and amongst those on references including an indefinite article ('a' or 'an') or a definite one ('the'). Such articles mark information either as 'new' or 'given'. There has been some variation in terms of how new and given information has been defined; in the light of this, Prince (1981) has established three different notions of 'givenness'. This includes the notion of 'givenness' as predictability of a lexical item in its sentential context (based on, for example, Halliday 1967 and Kuno 1972), 'givenness' as saliency in terms of an entity being in the addressee's consciousness (based on Chafe 1976), and 'givenness' as shared knowledge – knowledge the speaker assumes their addressee knows, believes or is able to infer (based on Clark & Haviland 1974). In the present article, we use the latter definition. Consequently, 'new information' is here defined as that which the speaker believes is not yet known by the addressee (i.e., information which is not yet part of the interlocutors' common ground).

Past research has focused on how given information is communicated in discourse and how speakers lexically mark such common ground (e.g., Fetzer & Fischer 2007); however, little research in this area has focused on gesture. One exception is a study by Gerwing and Bavelas (2004, Study 2). This study included an analysis of ten dialogues in which one person had played with a particular toy and described this toy and the actions carried out with it to another person who had not played with or seen the toy. Thus, initial references to features of the toy and its actions were new information, with subsequent information of this kind being given information. Their gestural analysis showed that the accumulating common ground did influence the form of the gestures in that given information was made less salient gesturally and gestures accompanying given information were smaller and less precise. Levy and McNeill (1992) as well as McNeill, Cassell and McCullough (1993) have analysed speakers' verbal and gestural repeated references to the same characters in a story. Their focus was on pointing gestures accompanying initial and subsequent, more attenuated references (mainly in the form of pronouns and zero anaphoras). Pointing gestures were found to occur less frequently with attenuated references (i.e., when the information was given).

The present study compares speakers' gesture use with definite references (e.g., including the lexical marker 'the') and indefinite references (e.g., including the lexical marker 'a/an') in terms of gesture rate and the type of gestures used. The analyses aim to further explore how speakers communicate given and new information in speech and co-speech gesture, going beyond previous research by focusing on grammatical articles (rather than pronouns) and on iconic as well as deictic gestures. Due to the incoherent picture emerging from the previous studies into common ground and gesture, no firm predictions regarding the pattern we may observe can be made.

The data used in the analyses stem from an experiment which was originally designed to manipulate the amount of common ground that exists from the outset of a conversation (common ground based on prior physical co-presence, Clark & Marshall 1981). Participants took part in pairs, with one speaker being allocated the speaker role (and the other the role of the addressee); this person later narrated a story they had seen on video to the addressee participant. In the 'no common ground condition' (NCG), the addressee participant had no knowledge about the story prior to the speaker's narrative. In the common ground condition (CG), the addressee participant watched individual scenes from the video together with the other participant (who then watched the entire video, on their own, prior to narrating the full story to their addressee). For the present analysis we collapse the data from both the 'common ground' and the 'no common ground condition' as they are equally suited to examine common ground that accumulates during the course of a narrative (common ground based on linguistic co-presence, Clark & Marshall 1981). However, we also use the original experimental common ground manipulation as a variable in some of the analyses.

Method

Experimental design

The present study was conducted as an additional analysis on a subset of the data published in Holler and Wilkin (2009).[1] It is based on a between-subjects design with two conditions: the 'common ground' condition (CG), in which participants shared some experimentally induced knowledge about the stimulus material, and the 'no common ground' condition (NCG), in which participants did not share any experimentally induced common ground (other than that which accumulated during the narrative).

Participants

The present analyses are based on fifty-six students (22 female and 34 male) from the University of Manchester who took part in the experiment (all received either payment or experimental credits for their participation). All individuals were right handed (as measured by the Edinburgh Handedness Inventory, Oldfield, 1971) and native English speakers. Each participant was allocated to a same-sex pairing, which was then randomly assigned to one of the two experimental conditions resulting in 14 same-sex pairs in each condition.

1. Only a subset was used in the present analysis because the data in Holler and Wilkin's (2009) study were analysed in two steps, the first one focusing on a smaller subset, at which point the present analysis was conducted.

Materials

A short (about 7 minute long) video was used as the stimulus material. It contained a story in which child and adult human characters were involved in different everyday activities, such as mending a car, grocery shopping, or playing in a barn. From this video, six short scenes (each 2–5 seconds in length) were selected for the common ground manipulation (see Procedure). The participants were filmed in a social observation laboratory including two high definition wall-mounted cameras, each providing the view of one participant, feeding into a dvd recorder in a split-screen format.

Procedure

In both the CG and the NCG conditions, two participants took part at a time, allocated to the roles of speaker and addressee based on their seat choice. The speaker watched the six selected scenes, followed by the whole video. However, in the CG condition, the addressee watched the six scenes together with the speaker (but was absent while the speaker watched the full video). During the following narration phase, the participants sat opposite each other, and the speaker was instructed to tell their addressee what happened in the story as a whole, bearing in mind that (a) their addressee did not know anything about the story (NCG condition), or (b) that their addressee already shared some knowledge about the story with them (CG condition). Addressees were told before the experiment that they would be asked content-related questions at the end. They were also told that they were free to signal their understanding during the narration as they felt appropriate, but that they should not interrupt the speaker to ask questions.

Analysis

Participants' gestural and verbal behavior relating to five of the six selected scenes was included in the analyses. The sixth scene was excluded due to similarity with another part of the video, which made it impossible to decide for certain in all instances which of the two events in the story speakers were referring to.

Speech segmentation

All descriptions of the five target scenes were transcribed verbatim. To identify the respective parts in the narratives, each event was defined in terms of what it comprised semantically (i.e., ideational units, see Butterworth 1975, Holler & Beattie 2002). Only those parts of the narratives were analysed that included semantic information from the five target scenes. The percentage agreement between two independent coders identifying the first and the last word to be considered part of the scene was 87.6%. All discrepancies were resolved through discussion.

Coding for definite and indefinite references based on grammatical determiner

Within the individual speech segments, the following determiners were identified (including both grammatical articles and demonstratives): 'the', 'that', 'a', 'an', and 'this'. 'The' and 'that' were both regarded as lexical markers of common ground (or given information) and were therefore combined in the analyses. We are here not referring to 'that' being used as a demonstrative singling out a referent in physical space (such as when pointing to something) but as a demonstrative in the absence of any nonverbal or physical contextualising cues; an example would be the utterance 'and then that light blue car came along' to refer back to a scene the interlocutors had seen together, or to the car when they had mentioned it beforehand. Similarly, 'a/an', and 'this' were combined as markers of no common ground (or new information); again, we are here referring to the demonstrative 'this' being used without any contextualising information (such as a pointing gesture to an object in the physical surroundings), but, rather, as a general determiner for a referent outside of the common ground, as in 'suddenly, this car comes around the corner' to refer to a car which is not present at that moment. That is, 'the' and 'that' are here classed as definite references and 'a'/'an' and 'this' as indefinite references. While the terms definite and indefinite references also refer to anaphoric expressions (Keysar, Barr, Balin & Paek 1998), the present analysis limits its focus to references including basic definite and indefinite determiners. This means that when a gesture accompanied a part of speech that contained both a grammatical article and an anaphora we used the grammatical article for classification (see Examples 1 and 2). The rationale for this decision was that, while previous analyses have focused on gestures accompanying attenuated references to characters in the form of zero anaphoras or pronouns (e.g., Levy & McNeill 1992; McNeill, Cassell & Levy 1993), in our data the use of pronouns was not that prevalent; instead, most references to the scenes constituting the analytic focus included the entities' grammatical articles (+noun). The present analysis therefore complements those earlier studies.

Gesture coding

Gesture category. Co-speech gestures were identified and categorised according to McNeill's (1992) categorisation scheme by coding them as iconic, metaphoric, deictic (in the present data only abstract deictics occurred), or beats, complemented by Bavelas, Chovil, Lawrie and Wade's (1992) category of interactive gestures[2]. The percentage agreement between two independent judges using these categories classifying

2. The categories of 'beats' and 'interactive gestures' seem to overlap (cf. Bavelas, Chovil, Coates & Roe 1995; Jacobs & Garnham 2007), but during our coding procedure we encountered some gestures that we felt clearly belonged to one and not the other class of gestures, based on the form criteria described by McNeill (1992) and Bavelas et al. (1992); we therefore included two separate categories to capture these gestures.

all gestures co-occurring with references to the five target scenes was 79.9%. Again, all discrepancies were discussed and resolved.

Gesture type. For the second part of the analysis, all iconic and deictic gestures were further classified as 'action gestures' (e.g., an iconic gesture representing someone picking something up; an iconic gesture performed with a single finger moving from left to right to indicate a car driving past), or as 'entity' gestures (e.g., a deictic gesture indicating the presence of an entity, or an iconic gesture representing a whole or part of an object, such as by using the index fingers to outline the square shape of a window). These examples illustrate that the distinction between entity and action gestures is not an absolute one – gestures classed as action gestures included those that were considered to be *primarily* encoding information about an action, but may have included information about entities (such as the narrator's hand carrying out the action representing the character's hand); the rationale for calling these 'action gestures' was that they seemed to *foreground* the action component of the gestural representation. Gestures classed as 'entity gestures' always encoded just entity information. The inter-observer reliability of two independent coders for this binary categorisation was 94.3%. The few disagreements that occurred were subsequently resolved through discussion.

Examples (1) and (2), and the following description, illustrate the coding of one speech segment and its accompanying iconic and deictic hand gestures. The underlined words are the definite and indefinite references based on determiner, and the square brackets mark individual gestures, indicated as subscript preceding the respective gesture and numbered consecutively. The superscript letters within each square bracket indicate whether the gesture primarily encoded action information (A) or entity information (E). If an article type was not accompanied by a gesture, it was coded as having no accompanying gesture (subscript N).

(1) $_{G1}$[the boy ^{E}] $_{G2}$[picks up the piece A] of litter, $_{G3}$[and puts it in the bin ^{A}]

$_{G1}$: abstract deictic gesture pointing towards the right hand side of the gesture space, referring to the boy.

$_{G2}$: iconic gesture showing someone grabbing something which is moved upwards (palm pointing downwards).

$_{G3}$: iconic gesture showing someone holding something enclosed in the hand which moves down and forwards, stopping at about chest height in front of the speaker's body.

(2) $_{N}$The kid... $_{G1}$[picks up a bit of litter ^{A}] off $_{N}$the floor $_{G2}$[and puts it in a ^{A}], $_{G3}$[in a litter bin ^{E}] $_{G4}$[which is a little basket ^{E}] $_{G5}$[attached to a lamppost ^{E}]

$_{G1}$: iconic gesture showing someone grabbing something which is moved upwards (palm pointing downwards).

$_{G2}$: iconic gesture showing someone holding something in the hand which moves down and sideward, stopping at about lap/thigh height to the side of the speaker's body.

$_{G3}$: iconic gesture showing the vertical, straight sides of a small, imaginary, upright container.
$_{G4}$: iconic gesture showing the sides and the base of a small, imaginary, upright container.
$_{G5}$: iconic gesture showing the narrow width, elongated shape, and vertical orientation of an imaginary object

If more than one gesture accompanied a stretch of speech that contained only one article type, then the gesture performed closest in time to the respective determiner (i.e., the gesture with the strongest temporal relation to the word, 'the', 'that', 'a/an', or 'this') was counted. Furthermore, if a part of speech containing an article type had no gestural accompaniment, while a subsequent gesture performed in synchrony with an immediately following part of speech nevertheless appeared semantically related to the preceding speech segment, this gesture was not counted as an accompanying gesture for the former article type but for the one it co-occurred with. Thus, temporal co-occurrence rather than semantic relation was used as the main criterion (although this was equivalent in most cases).

Results

The analyses reported here are based on a corpus of 277 references including the respective grammatical articles. For the statistical analyses, an alpha level of .05 is used throughout (all tests reported are two-tailed).

Definite and indefinite references

Across both conditions, references including definite determiners, 'the' and 'that', were used more frequently (180 times in total) when compared with references including indefinite determiners, 'a/an' and 'this' (97 times in total). This is not surprising since we took into consideration the first time an entity was being referred to, as well as all subsequent references, and speakers tended to refer to some of the entities repeatedly (such as the characters involved in the storyline) – thus establishing exactly the sort of common ground we intended to capture.

Co-speech gestures

Based on our corpus of 277 references and 210 co-speech gestures, we then focused on the proportion of gestures accompanying each reference type (i.e., number of gestures/number of definite references or indefinite references), Table 1. Firstly, the analysis revealed that the same proportional amount of references classed as indefinite was accompanied by gestures as references classed as definite ($z = 0.329$, N-ties = 24, $p = .742$, ns).

Table 1. Overview of average proportions of references classed as definite or indefinite accompanied by gesture or no gesture (in total as well as for individual gesture categories)

Accompaniment	Reference type	
	Definite (the/that)	**Indefinite (a/this)**
No gesture	0.29	0.14
Gesture (all categories combined)	0.71	0.86
Split up by category:		
Iconic	0.86	0.75
Deictic	0.06	0.15
Metaphoric	0.01	0.00
Beats	0.03	0.04
Interactive	0.05	0.06

This pattern held when we considered the individual gesture categories separately, with the exception of iconic gestures, of which a higher proportion accompanied definite references (*Median* = 1, *Range* = 1) than indefinite references (*Median* = .75, *Range* = 1), $z = 2.32$, N-ties = 24, $p = .021$.

'Action' and 'entity' co-speech gestures

Iconic and deictic gestures that accompanied the definite and indefinite references (192 gestures in total) were classified as either 'entity' or 'action' gestures (see Method). The frequencies and percentages can be found in Table 2.

A 2 (gesture type: action vs. entity) x 2 (reference type: definite vs. indefinite) repeated measures ANOVA was carried out and revealed that there was a main effect of reference type (F (1, 27) = 4.50, $p = .043$); out of those references that were accompanied by gesture, more were definite ones than indefinite ones. The main effect of gesture type was not significant (F (1, 27) = 3.16, $p = .087$, ns), meaning that, overall, speakers used as many gestures that focused on actions as gestures that focused on entities. However, the interaction between gesture type and reference type was significant (F (1, 27) = 5.36, $p = .028$), with more 'entity gestures' accompanying indefinite references, and more 'action gestures' accompanying definite references.

Table 2. Average proportions (and frequencies) of definite and indefinite references accompanied by 'action' and 'entity' gestures

Reference type	Action	Entity	Total
Definite ('the'/'that')	**62.71%** (74)	**37.29%** (44)	100% (118)
Indefinite ('a'/'this')	**37.84%** (28)	**62.16%** (46)	100% (74)

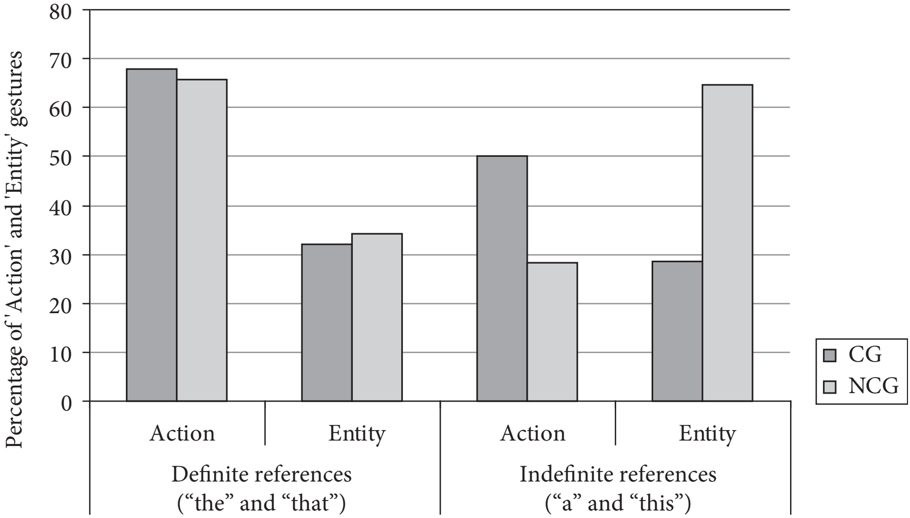


Figure 1. Overview of the mean percentage of 'action' and 'entity' gestures accompanying definite and indefinite references in the two experimental common ground conditions.

Table 3. Average proportions (and frequencies) of 'action' and 'entity' gestures used in the two experimental common ground conditions (common ground and no common ground)

Condition	**Action**	**Entity**	**Total**
CG	**60.9%** (53)	**39.1%** (34)	100% (87)
NCG	**46.7%** (49)	**53.3%** (56)	100% (105)

When considering the experimental common ground manipulation as a third factor with two levels, CG and NCG (see Figure 1), in addition to the effects mentioned above, the statistical analysis revealed a significant interaction between common ground and gesture type (F (1, 26) = 5.16, p = .032), with speakers in the CG condition using mainly 'action' gestures, and speakers in the NCG condition using mainly 'entity' gestures. However, the interaction between the common ground manipulation and reference type was not significant (F (1, 26) = 1.13, p = .297, ns), and neither was the three-way interaction between common ground, reference type and gesture type (F (1, 26) = 2.37, p = .136, ns). Table 3 shows the association between gesture type and experimental condition when considering just those two variables.

Discussion

The present analyses yielded a number of important findings. Firstly, but not surprisingly, we found that speakers used more definite references than indefinite references because they tended to refer to the same characters or objects more than once (and for half of the participants the referents were already in their common ground due to the experimental manipulation). Secondly, the findings show that speakers accompanied these two different types of references with gesture statistically equally often. However, a further analysis revealed that when splitting the amount of gestures up according to different gesture categories, speakers accompanied a higher proportion of definite references with iconic gestures than they did indefinite references. Another analysis distinguished between what we called different 'gesture types', which referred to gestures foregrounding information about entities ('entity gestures') and gestures foregrounding information about actions ('action gestures'). This analysis revealed that 'action gestures' accompanied mainly definite references, and 'entity gestures' mainly indefinite ones. Finally, we found that the manipulation of common ground that exists from the outset of a conversation (that is, common ground based on prior physical co-presence, Clark & Marshall 1981) interacted with gesture type; whereas speakers in the common ground condition used mainly 'action gestures' when referring to those segments of the story constituting common ground, speakers in the no common ground condition used mainly 'entity gestures' with references to the same semantic events. Taken together, these findings suggest that common ground was associated mainly with iconic gestures and action information, and no common ground mainly with abstract deictic gestures and entity information. The main conclusion to be drawn from these findings is that the semantic interplay between gesture and speech is not characterised by a simple, parallel pattern according to which both speech and gesture are more elliptical in the context of common ground. Rather, it appears that speakers employ the two modalities to package the information they intend to convey in a manner most appropriate with respect to the recipient's knowledge status, which can involve more complex representations in gesture even when common ground exists.

This appears to fit the results obtained from an earlier analysis of a similar dataset (Holler & Wilkin 2009). Amongst others, this analysis revealed that speakers gestured at a higher rate (with regard to iconic and deictic gestures) when common ground existed (referring to common ground existing from the outset). Further, their findings showed that, overall, speakers' gestures did not decrease significantly in semantic content when common ground did exist as compared to when it did not. The authors argued that this does not mean that the gestures were not 'recipient designed' (Sacks, Schegloff & Jefferson 1974). Rather, they suggest that the gestures continued to play an important communicational role, but that this role may be different to that of the gestures accompanying the same event descriptions when no common ground existed. The pattern revealed by the present analysis fits this notion; it suggests a shift in semantic focus regarding the gestural representations accompanying references to entities of different information status. The pattern is characterised by more semantically complex gestures accompanying references to information that is in common ground. Although we did not systematically quantify the information contained in the gestures using our 'entity'/'action' distinction, we observed that many 'action gestures' encoded also some entity information, whereas the 'entity gestures' only ever encoded entity information. Of course, the entity gestures could have been encoding information about several entities at once, and action gestures might have been highlighting just one particular dimension of a movement (e.g., direction). Due to this we cannot claim that action gestures *always* contained more information than entity gestures, but a large number of them appeared to do so. The shift in gestural focus observed in the present dataset may be one factor that could explain the lack of a difference in the amount of semantic information represented in gesture found by Holler and Wilkin (2009). The authors speculated that many of the gestures referring to information in common ground were semantically complex instead of elliptical so that they could fulfil a back-up function in case of speakers' uncertainty about specific information being in common ground or not (i.e., with the gestural information compensating for ellipsis in speech in case it is needed). Another possibility they mention is that these fairly complex gestures may assist speakers in focussing their addressees on the correct aspect of their mutually shared knowledge. The fact that, in the present study, when entity information was in the common ground, speakers put less emphasis on the individual characters and more emphasis on the actions carried out by these is agreeable with both of these possible explanations. More research is needed to illuminate this issue further.

Our findings are in line with Foraker and Goldin-Meadow (2007) who found that speakers tend to use gesture to depict the identity of a referent when this referent is newly introduced in speech, but that they used gesture to represent supplementary information about the referent when the referent had already been mentioned.

Our findings may be conceived of as complementing those studies providing evidence of increased ellipsis in gesture in the context of given information. Levy and McNeill (1992) and McNeill, Cassell and Levy (1993) found that pointing gestures (abstract deixis) occurred mainly with initial references to characters in a narration and less frequently with later ones. Once information about the identity of the referent was in the interactants' common ground, they used no gestures with their referring expressions, or the pointing gestures were replaced by other types of gestures (McNeill et al. 1993: 16). Similarly, Gerwing and Bavelas (2004, Study 2) found evidence of a reduction in semantic content in gesture when these gestures were referring to given instead of new information. Although our data show no statistically significant reduction in gesture use when common ground existed, they do show that speakers gesturally emphasise different semantic event aspects (i.e., the actions rather than the entities) and that they used mainly iconic rather than deictic gestures to do so. With regard to depicting entity information, the gestures in our corpus did tend to become more elliptical. However, our data throw a different light on the topic, as we have provided evidence that gestures do not always become more elliptical overall when information is in common ground (although this *can* happen), or that they simply disappear altogether because communication may be conceived of as easier when information is given. Instead, our data suggest that gestures continue to be important for communication and that they combine with speech in a variety of ways to achieve a successful and pragmatically appropriate exchange of information.

In addition to exploring the exact functions co-speech gestures fulfil in this context, future studies will need to establish to what extent the interaction patterns the present and previous studies have revealed are specific to the particular communicative situation examined; that is, the functions of gestures may be specific with respect to whether speakers communicate information that is common ground based on prior physical co-presence, linguistic co-presence, or visual co-presence, for example. Further insights may also be gleaned from more detailed analyses which take into account the structure of individual utterances and the sequence of the gestures accompanying them; after all, the present analyses are based on aggregate data, summarising the occurrences of different gesture types across the discourse, which provides us with merely a first glimpse of what may be going on. Nevertheless, one important conclusion we can draw from the present findings is that the way in which interlocutors communicate information that is in the common ground they share appears to be complex, with partly parallel, partly complementary changes happening to gesture and speech. Thus, only a multi-modal enquiry will be able to provide us with a more complete view of communication in this domain.

References

Alibali, M. W., Heath, D. C. and Myers, H. J. 2001. "Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen." *Journal of Memory and Language* 44: 169–188.

Bavelas, J. B. and Chovil, N. 2000. "Visible acts of meaning. An integrated message model of language use in face-to-face dialogue." *Journal of Language and Social Psychology* 19: 163–194.

Bavelas, J. B., Chovil, N., Coates, L. and Roe, L. 1995. "Gestures specialised for dialogue." *Personality and Social Psychology Bulletin* 21: 394–405.

Bavelas, J. B., Chovil, N., Lawrie, D. A. and Wade, A. 1992. "Interactive gestures." *Discourse Processes* 15: 469–489.

Bavelas, J. B., Gerwing, J., Sutton, C. and Prevost, D. 2008. "Gesturing on the telephone: Independent effects of dialogue and visibility." *Journal of Memory and Language* 58: 495–520.

Bavelas, J. B., Kenwood, C., Johnson, T. and Phillips, B. 2002. "An experimental study of when and how speakers use gestures to communicate." *Gesture* 2: 1–17.

Butterworth, B. L. 1975. "Hesitation and semantic planning in speech." *Journal of Psycholinguistic Research* 4: 75–87.

Chafe, W. L. 1976. "Givenness, contrastiveness, definiteness, subjects, topics, and point of view." In *Subject and Topic*, C. N. Li (ed), 25–55. New York: Academic Press.

Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. and Marshall, C. R. 1981. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, A. K. Joshi, B. Webber, and I. Sag (eds), 10–63. Cambridge: Cambridge University Press.

Clark, H. H. and Wilkes-Gibbs, D. 1986. "Referring as a collaborative process." *Cognition* 22: 1–39.

Fetzer, A. and Fischer, K. (eds.) 2007. *Lexical Markers of Common Grounds*. London: Elsevier.

Foraker, S. and Goldin-Meadow, S. 2007. "Gesture and discourse: How we use our hands to refer back." Talk presented at the 3rd International Society for Gesture Studies Conference: Integrating Gestures, June 18–22, Chicago, IL.

Furuyama, N. 2000. "Interaction between instructor and learner." In *Language and Gesture*, D. McNeill (ed), 99–117. Cambridge: Cambridge University Press.

Fussell, S. R. and Krauss, R. M. 1989. "The effects of intended audience on message production and comprehension: Reference in a common ground framework." *Journal of Experimental Social Psychology* 25: 203–219.

Gerwing, J. and Bavelas, J. B. 2004. "Linguistic influences on gesture's form." *Gesture* 4: 157–195.

Gullberg, M. 2006. "Handling discourse: Gestures, reference tracking, and communication strategies in early L2." *Language Learning* 56: 155–196.

Halliday, M. A. K. 1967. "Notes on transitivity and theme in English." *Journal of Linguistics* 3: 199–244.

Haviland, S. E. and Clark, H. H. 1974. "What's new? Acquiring new information as a process in comprehension." *Journal of Verbal Learning and Verbal Behavior* 13: 512–521.

Holler, J. 2003. *Semantic and Pragmatic Aspects of Representational Gestures: Towards a More Unified Model of Communication*. Unpublished doctoral dissertation, University of Manchester, UK.

Holler, J. and Beattie, G. 2002. "A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk." *Semiotica* 142: 31–69.

Holler, J. and Beattie, G. 2003a. "How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process?" *Semiotica* 114: 86–116.

Holler, J. and Beattie, G. 2003b. "Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?" *Gesture* 3: 127–154.

Holler, J. and Stevens, R. 2007. "An experimental investigation into the effect of common ground on how speakers use gesture and speech to represent size information in referential communication." *Journal of Language and Social Psychology* 26: 4–27.

Holler, J. and Wilkin, K. 2009. "Communicating common ground: how mutually shared knowledge influences the representation of semantic information in speech and gesture in a narrative task." *Language and Cognitive Processes* 24: 267–289.

Hostetter, A., Alibali, M. and Kita, S. 2007. "I see it in my hands' eye: Representational gestures reflect conceptual demands." *Language and Cognitive Processes* 22: 313–336.

Isaacs, E. A. and Clark, H. H. 1987. "References in conversations between experts and novices." *Journal of Experimental Psychology: General* 116: 26–37.

Jacobs, N. and Garnham, A. 2007. "The role of conversational hand gestures in a narrative task." *Journal of Memory and Language* 56: 291–303.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Keysar, B., Barr, D. J., Balin, J. A. and Paek, T. S. 1998. "Definite reference and mutual knowledge: Process models of common ground in comprehension." *Journal of Memory and Language* 39: 1–20.

Kuno, S. 1972. “Functional sentence perspective.” *Linguistic Inquiry* 3: 269–320.
Levy, E. and McNeill, D. 1992. “Speech, gesture, and discourse.” *Discourse Processes* 15: 277–301.
McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
McNeill, D. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
McNeill, D., Cassell, J. and Levy, E. 1993. “Abstract deixis.” *Semiotica* 95: 5–20.
Oldfield, R. C. 1971. “The assessment and analysis of handedness: The Edinburgh inventory.” *Neuropsychologia* 9: 97–113.
Özyürek, A. 2002. “Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures.” *Journal of Memory and Language* 46: 688–704.
Parrill, F. 2010. “The hands are part of the package: gesture, common ground and information packaging.” In *Empirical and Experimental Methods in Cognitive/Functional Research*, J. Newman and S. Rice (eds), 285–302. Stanford: CSLI.
Pine, K. J., Bird, H. and Kirk, E. 2007. “The effects of prohibiting gestures on children's lexical retrieval ability.” *Developmental Science* 10: 747–754.
Prince, E. 1981. “Toward a taxonomy of given-new information.” In *Radical Pragmatics*, P. Cole (ed), 223–255. New York: Academic Press.
Sacks, H., Schegloff, E. A. and Jefferson, G. 1974. “A simplest systematics for the organization of turn-taking in conversation.” *Language* 50: 696–735.

Acknowledgements

We would like to thank Gale Stam and Mika Ishino, as well as independent reviewers, for their valuable comments on an earlier draft of this manuscript. We would also like to thank the Economic and Social Research Council for funding this research (research grant RES-061-23-0135) as well as the participants taking part in our study.

CHAPTER 23

"Voices" and bodies

Investigating *nonverbal parameters of the participation framework*

Claire Maury-Rouan
LPL, CNRS – Université de Provence

According to interactional and dialogic linguistics, utterances may be seen as complex constructions in which alternate "voices", more or less identifiable in reported speech, or less transparent as in polyphony, may be heard. As vocal changes had been established in reported speech, it was hypothesized that shifts in facial expressions, gesture or posture paralleling shifts of footing might also be found. The analysis of videotaped data showed that two distinct formats: reported speech and polyphony in perspective shifts co-occurred with relevant nonverbal cues. Based on the degree of variation of accompanying vocal and visual parameters, three relevant types ('underplayed', 'animated' and 'lively') were found in reported speech. Perspective shifts were found to start with a pause, a shift in the speaker's posture (head tilt) and pitch range variation. Two distinct cases of perspective shifts were found, whether the speaker became (–gaze) or remained (+gaze), possibly indicating quite different mental forms of perspective shifts.

1. Polyphonic "Voices" inside speech

From an interactional point of view, discourse has often been described as the result of a joint construction of utterances. When speaking, we plan and adapt the form and content of our discourse in accordance with an internalized image of our addressees, remembering previous interactions and anticipating the decoding progression and possible reactions of our interlocutors.

Such a conception of speaking has been described from several angles. It is typically present in Bakhtinian linguistics (Bakhtine 1929, Bakhtine 1979, Voloshinov 1968) and in the various more or less Bakhtinian enunciative theories of *polyphony* developed by Ducrot (1984), Authier-Revuz (1982), Vion (1988) or Nölke (1993). In addition, Goffman's (1981) conception of *footing* is based on a quite similar viewpoint. These various theories describe a given speaker as not being the unique source of his discourse. In Ducrot's model, for instance, a given speaker may stage several distinct voices (called "énonciateurs" by Ducrot) within his speech. In a very similar way, Goffman distinguished between three "figures" as being alternately responsible for the words being heard within the speaker's utterance: the "animator" (as the individual voicing the words – the "author" (as the identifiable source of the words, in the case of reported speech, for instance) and the "principal" (as the party who is socially responsible for what is said e.g. *science, religion*).

In a number of cases, production formats rely on clear structure marks, as in reported speech, quotations or speech repairs. But in other formats, the identification of the sources becomes a subtle and highly complex task, for instance, when some kind of speech sounds as though it is reported speech, but without any clear cues to an identifiable "author". Similarly, distinct voices may be heard when several stances are staged, or simply alluded to by a given speaker who animates a fictitious dialogue within his own speech. Such allusions sometimes amount to tenuous traces like the presence of the French particle *quand même* ('yet'). The many degrees in the ways this variety of sources may be displayed, hidden or implied in speech have been been described in Bakhtinian terms as "heterogeneity" under two forms: "shown" or "constitutive" (Authier-Revuz 1982; Vion 1998; Maury-Rouan, Vion & Bertrand 2007). In the same way, Goffman's production format allows a considerable expansion of the various figures animated in the speaker's utterances, thus presenting a highly complex *laminated* speaker, as Goodwin and Goodwin (2004) pointed out.

From a more interactional viewpoint, polyphony may have several advantageous effects. As they invite a variety of voices into their speech (or, in Goffman's words, as they produce constant changes in footing), speakers display a reflexive stance towards their utterances; this may give them the image of being broad-minded and considerate interactants. It also produces softening face work effects as it changes a harsh statement into a softened, open-minded proposition.

In their 2004 paper, Charles and Marjorie Goodwin commented on Goffman's Participation Framework as being exclusively speaker-oriented. In Goffman's model, they pointed out, "speakers and hearers inhabit different worlds," speakers being endowed with "rich cognitive and linguistic capacities" while other participants "are left cognitively and linguistically simple" (being solely classified as "ratified-non ratified addressees/over-hearers/bystanders..), as the process of mutual monitoring is usually neglected in Goffman's study of the participation process, whereas "the process of designing an utterance for a particular type of addressee can shape its structure in powerful ways (…)" as "speakers reconstruct their sentences in their course to make them appropriate to different kinds of recipients" (p. 224).

Furthermore, Goodwin and Goodwin criticized the fact that Goffman's production format usually highlights the sole stream of speech – "over other forms of embodied practice that might also be constitutive of participation in talk". The same blindness could be found in most of the work currently published in the linguistic field of polyphony research and of Bakhtinian studies on interaction and speech, with the exceptions of a few authors like Vincent and Perrin, according to whom "reported speech is woven into a series of non-verbal events" (1999: 291).

2. "Voices" and bodies

This study investigates changes that might occur in speakers' faces, gaze, postures and gestures when such "voices" are being staged. At its present phase, it does not escape Goodwin and Goodwin's reproach of being exclusively focused on speakers, and solely attempts to offer an insight into the "other forms of embodied practices" (Goodwin & Goodwin 2004: 25) that might parallel participation in talk.

A number of studies have investigated the part played by prosody in various cases of reported speech, one of the most frequent (and most clearly identifiable) shifts in footing. In reported speech, Elisabeth Couper-Kuhlen (1998) explained, the unity within a single speaker of the three figures (author, animator and principal) dissolves, and they become independent. Couper-Kuhlen found that a set of phonatory cues helped speakers in identifying such shifts, as she demonstrated that vocal changes are often crucial for coherence in helping interlocutors to detect changes in footing because verbal prefaces (e.g.: "*he said...*") are currently omitted in spontaneous speech, and appropriate shifts in personal deixis (mainly personal pronouns) may be lacking, or ambiguous. The main indication of the staging of an alternate figure inside the animator's speech is then transferred to vocal deixis. This may include notable shifts in four levels of speech characteristics: loudness, pitch, tempo, and voice quality. In addition to identifying deictic functions, Couper-Kuhlen noted that such phonatory voice variations often convey different stances in the way the summoned figures are presented in the animator's discourse, a property confirmed by the reactions of interactants. Consistently, Bertrand (2002) found a correlation between different degrees of the speaker's involvement and overall variations of the F0 parameter.

The presence of a strong link between facial expressions and vocal qualities, and between facial, head or even slight hand movements and changes in vocal pitch and stress has been established through a number of studies, from Fonagy (1983), Bolinger (1985), Ekman (1979) and Ohala (1980). This leads us to hypothesize that shifts might be found in facial expression and possibly in other movements, gaze, gesture or posture, occurring in the same way as the vocal changes that parallel shifts of footing in reported speech. This insight has been reinforced by data from Goodwin (1984) and McClave (1998).

Goodwin (1984: 228–229) described together with changes in voice and intonation, several displays in a narrator's body position (involving clasped or separate hands, elbow position, head movements, gaze, leaning forward) as significant indications "about the nature and extent of her orientation to the conversation itself, as well as displays about the type of talk she is producing, and relevant differences within that talk" including parts of a narrative and reported speech. Relevant action patterns corresponded to the speaker's and the recipients' orientation and to the sequential organization of talk.

In a comprehensive study of head movements, McClave (1998) noted that speakers change their head positions to indicate different discourse levels.

> "Speakers change the orientation of their heads when switching from indirect to direct discourse. This is not simply a realignment or a reorientation toward the interlocutor, as (...) speakers at times break mutual gaze and turn their heads away from the recipient at the beginning of a quote". ". . . .The speaker seems to assume momentarily a new head position to give voice to another's words. Thus a different speaker is conceptualized as occupying a different space. The spatial shift signals the speaker's change in footing from narrator to that of a character in the narration and makes visible the change in narrative structure." (McClave 1998: 369).

3. Corpus and methods

In the present study, two corpuses were investigated for non-verbal correlates of shifts of footing. Corpus n°1 is an 80 minute video recording of an 82 year old lady, Ariane, interviewed about her life story, in the familiar setting of her sitting-room, in the presence of a couple of friends (alternate addressees during the interview) and a cameraman.

The second corpus, filmed in a research lab, recorded two young men (volunteers in the experiment) engaged in casual talk. They had been asked to talk about memories of unusual events in their lives.

To date, 36 out of the 80 minutes of the 'Ariane' Corpus (n°1) have been thoroughly analyzed, and 10 out of the 60 minutes of the 'Young Men' Corpus have been partially analyzed.

With a view to describe the parameters of nonverbal shifts, the verbal data from both corpuses were first analyzed following discourse analysis criteria, with special attention to three polyphonic discourse patterns: reported speech, speech repairs and verbally explicit perspective shift sequences. Relevant voice and gesture parameters were investigated in a later phase.

The nonverbal parameters of perspective shifts and of speech repairs were investigated in the total corpus (46 minutes). The nonverbal accompaniment of reported speech sequences was investigated in Corpus n°1 'Ariane' (36') only.

4. Findings

In two out of the three discourse patterns, reported speech and perspective shift sequences were found to display relevant nonverbal cues. Up to this point, speech repairs do not appear to have been marked by any relevant nonverbal aspects.

4.1 Reported speech

4.1.1 *Prosodic displays*

The prosodic cues of reported speech as described either by Couper-Kuhlen (1998) or Bertrand (2002) were found to be present in the corpus, although unevenly distributed. Three types of distinct reported speech animation seemed to stand out, corresponding to regular patterns in the corpus (even though some of the cases might have been better described in terms of degrees in a scale of prosodic marking).

Type 1: 'Underplayed' reported speech.
In a number of cases, where the presence of reported speech was clearly displayed through verbal *prefaces* ("he told me, she used to say..."), vocal parameters were found to be very tenuous, reduced to subtle tempo variations. No difference as to pitch range, volume, or voice quality was perceptible in the transitions between direct speech preceding the sequence of reported speech and the reported speech sequence itself: the only prosodic cue amounts to a tenuous shift in scansion marking the reported utterance.

Example 1

Example n° 1
Type 1: example (from Ariane: 2:30):
<preface>/reported utterance

... un auteur, dont le nom me reviendra, <qui racontait> "je suis une enfant trouvée"
... a writer, whose name I should remind, <who told > "I am a foundling"

Type 2: 'Moderate' reported speech
In the second group, reported speech was more vividly animated: scansion was more distinctly present, and shifts in volume were perceived, but no changes were heard as to pitch or voice quality.

Type 3: 'Theatrical' reported speech.
In the last type of reported speech, the speaker gave life to "author" figures in a way close to a theatrical performance, not only displaying shifts in tempo and in volume but also adopting vocal quality and pitch changes as if actually impersonating the absent speaker's voice. This third type of reported speech type was found to be the most frequent in the corpus (as shown in Table1).

Example 2
Type n°3: example (from Ariane : 30:01):
(the narrator is telling how a spinster relative refused to be married to a distant cousin):

reported utterance /	<preface	> main speech/reported utterance
"non ! non !"	*< elle lui dit*	*> je la revois/"non ! non ! donnez-le à*
"no ! no !	"<she told him>	I can still picture her/" no ! no ! give him to
Marie-Thérèse!"		
Marie-Thérèse"		

4.1.2 *Visual parameters*

Type 1: In 'underplayed' reported speech, where scansion was the only varying prosodic parameter, the speaker did not display any change in facial expression, trunk or head position; no gesture was produced, and the gaze maintained the same direction as during the previous direct speech sequence. *Underplayed* reported speech thus appeared to be semiotically embedded within the direct speech speech sequence. The facial expression, its qualifying and/or commentary value, and the general body display also matched the main utterance. Were it not for the slight shift noted in scansion, *underplayed* direct reported speech would be in all respects similar to the way syntactically embedded indirect reported speech would be voiced and nonverbally performed. (In such cases, the use of direct reported vs. indirect reported speech might simply serve highlighting functions).

Type 2: In *moderate* reported speech, displaying shifts in volume in addition to distinct scansion, several shifts in visual nonverbal behavior were found: gestures matching the verbal meaning of the utterance were produced, giving more embodied life to the animated figure; consistently with McClave's 1998 findings, shifts in posture – typically head movements – often marked the beginning of the reported speech sequence. Gaze direction conformed to the behavior usually found in speakers (alternating look-away moments and control glances).

Type 3: In type 3, *theatrical* reported speech, together with voice quality shifts, affecting pitch range and vocal quality combined with the other variations in vocal parameters, facial expressions matching the impersonated voice pitch and quality appeared, in addition to posture shifts, head movement and gestures already found in type 2. As in the previous types, gaze direction alternated look-away moments and control glances.

4.2 Perspective shifts

Perspective shift was the other type of discourse pattern to appear with a distinct nonverbal accompaniment. In discourse, perspective shifts are observed when a speaker interrupts the development of a point to take into account an alternate viewpoint. This may be operated in a carefully organized rhetoric way, using adequate syntactic and pragmatic apparatu, as in written argumentation. In spontaneous oral speech, although instances of planned and pragmatically controlled perspective shifts may also be found, a number of them seem unprepared and simply seem to occur: the speaker,

Table 1. Vocal and visual parameters in the various types of reported speech

	Prosodic parameters: variation in:	**Visual parameters variation in:**	**Total number of cases:**	**%**
Type 1: *underplayed*	scansion	–	19	28%
Type 2: *moderate*	scansion, volume	gestures head movements shifts in posture	15	22%
Type 3 *theatrical*	scansion, volume, pitch, vocal quality	gestures head movements shifts in posture facial expressions	33	48%
			68	100%

while uttering a point, as if struck by an alternate viewpoint, stops and starts again wording the different opinion or standpoint. Thus, perspective shifts share some of the aspects of speech repairs, such as dysfluency, pause and restart on a new version, and both behaviors imply a reflexive attitude. However, a distinction seems to be justified, defining speech repairs as a search for better lexical adequacy in wording a consistent point of view as opposed to the contemplating of a distinct opinion or standpoint found in perspective shifts. In terms of polyphony, a distinct voice is heard within the speaker's discourse with the irruption of this new perspective, whether this voice can be identified as that of a former interlocutor, a virtual opponent or an imaginary or collective party in relation to whom the speaker's utterances are produced.

In this type of framework, the animated figure is not always clearly referred to deictically as opposed to most of the cases in which direct reported speech is used; the figure's identity may simply be sketched or inferable from the verbal organization of the utterance revealing the presence of alternating voices.

4.2.1 *Verbal and vocal marks of perspective shifts*

Several cases of such perspective shifts were found in both corpuses, verbally and vocally displayed in a similar way. Verbal displays currently include relevant introductive discourse particles such as the French *mais* ('but, however'), *enfin, bon* ('well', 'actually'), followed by the verbal utterance of the alternate point of view. In other cases, the indication of perspective shifts may be limited to tenuous allusions to alternate stances as in the use of some French adverbs: *finalement*' ('everything considered'*)* or *quand même* ('nevertheless').

On the prosodic level, perspective shifts are frequently marked by variation in pitch range (to a higher pitch range level in our corpuses).

4.2.2 *Nonverbal components of perspective shifts*

Although sharing a set of consistent visual displays, the perspective shifts in the corpus need to be placed in two distinct categories according to the way speakers embody the utterance conveying the reflexive speech act (Table 2). In the first category, the speech stream suddenly stops; after a short pause, the speaker displays a shift in posture, typically consisting of a head movement, as the head is tilted backwards and to the side, and the face is oriented to a different direction the reflexive re-oriented part of the utterance begins. In this first case, the speaker is –gaze during this shift and remains so at least until the verbal utterance is almost completed. Examples 3, 4, and 5 illustrate this category.

Example 3
perspective shift (–gaze): example from Young Men (1:52 to 2:02)
(on *enfin bon* ('actually', 'well') it suddenly has occurred to the speaker that his mother might have been right after all):

non par contre +++ ma mère elle faisait des +++ enfin bon + hé + je comprends hein + mais
(makes faces) (head tilts backwards)
no my mother did some ++++ actually well I understand ok + but

bah euh + j'étais pas propre longtemps hein
um + it took me a long time to get toilet-trained

In this example, the tilt of the head coincides with the utterance of the adverb *quand même* ('nevertheless') which apparently represents (and condensates) a previous dialogue with the speaker's mother's on her child raising methods.

Example 4
perspective shift (–gaze): example from Ariane (6:05 to 6:12)
(the speaker suddenly stops grieving on the hard way she was raised as a child, as it suddenly occurs to her that the Victorian education principles left her mother no choice : *mais elle pouvait pas* – 'but she could'nt'):

Ah c'était + au moins on savait où on allait + mais elle pouvait pas faire autrement je pense

Ah it was + at least we knew where we stood + but she couldn't act differently I guess
(head tilts backwards)

+++ *elle pouvait pas*
+++ she just couldn't

Example 5
perspective shift (–gaze): example from Young Men (2:22 to 2:30)

apparemment c'était radical quoi + mais putain c'était + quand j'y pense je me dis putain
apparently it seemed radically efficient but it was fuckin' + thinking of it I realize

c'était quand même vachement violent quoi
it was nevertheless fuckin' tough though
(head tilts backwards)

In the second category, the interruption of the speech stream is less noticeable; a shift in posture including the tilting backwards of the head is present, but often less distinctly displayed; however, the most striking difference is the presence of a constant gaze towards the interlocutor's face.

Example 6
perspective shift (+gaze): example from Ariane (18:54 to 18:58)

C'était + c'était + c'est vrai que ça a existé ça

It was + it was + it's true that such things actually did exist
(head tilts backwards slightly)

Table 2. Distinct perspective taking displays

	Discourse strategy	**Actually experienced**
Prosodic parameters	Speech is interrupted (briefly) Pitch range shift (frequent)	Speech is interrupted (briefly) Pitch range shift (frequent) (more distinctly than in Discourse strategy)
Postural parameters	Posture Shift Head tilted backwards and to the side Face is re-oriented	Posture Shift Head tilted backwards and to the side Face is re-oriented (more distinctly than in Discourse strategy)
Gaze	+gaze	–gaze

5. Discussion and conclusion

The behavior schemes found in the two distinct types of perspective shifts display head movements very similar to the ones described by McClave (1998) and that we also found in reported speech sequences. Thus in perspective shifts, one might consider, as in reported speech, that "the speaker seems to assume momentarily a new head position to give voice to another's words"(McClave 1998: 369), even though in such cases the different speaker might be described as a mere voice, a stance alluded to by the main speaker's utterance. Nevertheless, such a different speaker is conceptualized by the speaker as occupying a different space so that a spatial shift signals the speaker's change in footing to that of another figure, and makes the change in perspective visible.

According to studies on the relation of gaze direction to speech (Kendon 1967, Lee & Beattie 1998, Goodwin 1981) such a withdrawal of gaze is considered to coincide cognitively with planning phases as opposed to phases of feed-back control. Therefore, it is likely that a speaker, when engaged in thinking-while-speaking on being "struck" by a new idea or by the relevance of an alternate standpoint, will display such a –gaze attitude.

On the other hand, the presence of similar verbal, prosodic and nonverbal parameters, but with a speaker being +gaze, might amount to something more *prepared* and the utterance of the words expressing a reflexive standpoint should in this case be seen as more rhetorical, as a kind of discourse strategy: the staging of an alternate voice is thus used as one of the many ways of avoiding too much assertiveness.

In a –gaze speaker, these same cues should be seen as marking the genuine irruption of another figure's point of view, or as the presence of a voice being heard within the inner debate of a laminated subject.

A parallel might be established between such a partition between rhetorical and experienced heterogeneity in discourse and the functions of facial movements and expressions in human communication. According to Ekman and Friesen (1969) and Ekman (1984), speakers seem to be capable of making a secondary use, for semiotic functions, of the same behaviors that appear spontaneously when subjects experience a given emotion or affect, in the case of smiles or of eyebrow raising for instance. In a similar perspective, rhetorical perspective-taking displays might turn out to be semiotic behaviors based on the mimicry of spontaneous reactions.

References

Authier-Revuz, J. 1982. "Hétérogénéité montrée et hétérogénéité constitutive, élements pour une approche de l'autre dans le discours." *DRLAV* 26: 91–115.

Bakhtine, M. 1929/1963/1970. *La Poétique de Dostoïevsky.* Paris: Seuil.

Bakhtine, M. 1979/1984. *Esthétique de la Création Verbale*. Paris: Gallimard.

Bertrand, R. 2003. "Etre soi avec les mots d'autrui." In *Le Sujet*, J.-M. Merle (ed), 269–280. Paris: Ophrys.

Bertrand, R. and Espesser, R. 2002. "Voice diversity in conversation: a case study." Speech Prosody 1. (11–13 avril 2002). In *Proceedings of Speech Prosody 2002,* B. Bel and I. Marlien (eds), 171–174. Aix en Provence: Laboratoire Parole et Language, .

Bolinger, D. I. 1985. *Intonation and Its Parts*. London: Edward Arnold.

Clift, R. and Holt, E 2006. "Introduction." In *Reporting Talk: Reported Speech in Interaction*, R. Clift and E. Holt (eds), 1–15. Cambridge: Cambridge University Press.

Couper-Kuhlen, E. 1998. “Coherent voicing. On prosody in conversational reported speech.” In *List N°1 Interaction and Linguistic Structures* (Konstanzer Online Publikations System: KOPS Datenbank).

Ducrot, O. 1984. “Esquisse d’une théorie polyphonique de l’énonciation.” In *Le Dire et le Dit,* 171–233. Paris: Minuit.

Ekman, P. 1979. “About brows. Emotional and conversational signs.” In *Human Ethology*, M. von Cranach, K. Foppa, W. Lepenies and D. Ploog (eds), 169–248. Cambridge, Cambridge University Press.

Ekman, P. 1984. “Expression and the nature of emotion.” In *Approaches to Emotion*, K. Scherer and P. Ekman (eds), 319–343. Hillsdale, NJ: Lawrence Erlbaum.

Ekman, P. and Friesen, W. 1969. “The repertoire of nonverbal behavior.” *Semiotica* 1 (1): 49–98.

Fonagy, I. 1983. *La Vive Voix*. Paris: Payot.

Goffman, E. 1981. “Footing.” In *Forms of Talk*, E. Goffman (ed), 124–159. Philadelphia: University of Pennsylvania Press.

Goodwin, C. 1981. *Conversational Organization*. New York: Academic Press.

Goodwin, C. 1984. “Notes on the structure and the organization of participation.” In *Structures of Social Action*, J. M. Atkinson and J. Heritage J. (eds), 225–246. Cambridge: Cambridge University Press.

Goodwin, C and Goodwin, M. H. 2004. “Participation.” In *A Companion to Linguistic Anthropology*, A. Duranti (ed), 222–244. Oxford: Basil Blackwell.

Lee, V. and Beattie, G. 1998. “The rhetorical organization of verbal and nonverbal behavior in emotion talk.” *Semiotica* 120 (1–2): 39–92.

Kendon, A. 1967. “Some functions of gaze direction in interaction.” *Acta Psychologica* 26 (1): 1–47.

Maury-Rouan, C. 2002. “Do Listener’s facial expressions influence speaker’s discourse?” Paper delivered at Gesture: The Living Medium, 1st Conference of the ISGS. Austin, Texas.

Maury-Rouan, C. and Noël-Jorand, M-C. in press. “Dialogism and dialogue : physical vs. imaginary presence of addressee’s influence over discourse parameters.” *Journal of Pragmatics*.

Maury-Rouan, C, Vion, R. and Bertrand, R. 2007. “Voix de discours et positions du sujet. Dimensions énonciatives et prosodiques.” *Cahiers de Praxématique, 49, Numéro Spécial*: *À la recherche des voix dans le Discours*, J. Barberis (ed), 133–158.

McClave, E. 1998. “Cognitive and interactional functions of head movements in conversation.” In *Oralité et Gestualité*: *Communication Multimodale, Interaction*, S. Santi, I. Guaitella, C. Cave and G. Konopczinsky (eds), 365–370. Paris: L’Harmattan.

Nölke, H. 1993. *Le Regard du Locuteur – Pour une Linguistique des Traces Enonciatives*. Paris: Kimè.

Ohala, J. J. 1980. “The acoustic origin of the smile.” *Journal of the Acoustic Society of America* 68: S.33.

Vincent, D. and Perrin, L. 1999. “On narrative vs. non-narrative functions of reported speech: a socio-pragmatic study.” *Journal of Sociolinguistics* 3 (3): 291–313.

Vion, R. 1998. “De l’instabilité des positionnements énonciatifs dans le discours.” In *Pragmatics: Selected Papers from the 6th International Conference of the IPRA, Vol. 2*, J. Verschueren (ed), 577–589.

Voloshinov, V. N. 1986. *Marxism and the Philosophy of Language*. Cambridge: Cambridge University Press.

CHAPTER 24

Gestures in overlap

The situated establishment of speakership

Lorenza Mondada and Florence Oloff
ICAR Research Lab (CNRS, Univ. Lyon2, ENS Lyon)

This paper aims at contributing to the analysis of overlaps in turns-at-talk from both a sequential and a multimodal perspective. Overlaps have been studied within Conversation Analysis by focusing mainly on verbal and vocal resources; taking into account multimodal resources such as gesture, bodily posture, and gaze contributes to a better understanding of participants' orientations to the sequential organization of overlapping talk and their management of speakership.

First, we introduce the way in which overlaps have been studied in Conversation Analysis, mainly by Jefferson (1973, 1983, 2004) and Schegloff (2000, 2002); then we propose possible implications of their multimodal analysis. In order to demonstrate that speakers systematically orient to the overlap onset and resolution we analyze the multimodal conduct of overlapped speakers. Findings show methodical variations in trajectories of overlap resolution: speakers' gestures in overlap display themselves as maintaining or withdrawing their turn, thereby exhibiting the speakership achieved and negotiated during overlap.

1. Overlap as a classical topic for Conversation Analysis

Turns at talk are organized in such a way that addresses the alternation of turns among participants, the distribution of rights and obligations among speakers and the minimization of both gaps and overlaps between turns (Sacks, Schegloff & Jefferson 1974). One of the fundamental features of the turn-taking model is the participants' orientation to the principle of "one party talks at a time". Nevertheless, overlap – that is, simultaneous talk by more than one party at a time – is a recurrent phenomenon in social interaction. This could be seen as a violation of the turn-taking system; however, it is a phenomenon that, although departing from the normative expectations that organize talk-in-interaction, consistently takes them into account: "No matter how much overlapping may be found in the talk [...], the talk appears to be co-constructed by reference to one-party-at-a-time as its targeted design feature [...]" (Schegloff 2000: 3).

Considered in this way, overlap is an interesting phenomenon to look at in order to understand how turn-taking actually works.

If we exclude from this domain of inquiry the cases where activities are organized on the basis of other principles, such as collective choral responses orienting to "all-at-a-time" schisms producing two or more parallel simultaneous conversations (Egbert 1997), we can distinguish two environments for overlaps which have been studied by reference to the principles of turn-taking within one conversation:

a. Environments where overlap is not dealt with as problematic by the participants (cf. Schegloff 2000: 5–6) and results mainly from their monitoring of the current turn organization and from their projection of the Turn-Constructional Unit (TCU) completion at the next Transition-Relevance Place (TRP). In this case, overlap onset exhibits the online analysis done by co-participants of the current turn: overlap occurs at systematic places, such as in terminal or pre-terminal positions, where the projection of the turn's completion is possible. In a similar way, the production of continuers and assessments in overlap orients to the TCU boundaries and is systematically positioned, without claiming the right to take the floor (Schegloff 1982, Goodwin 1986). These overlaps orient to the minimization of their length and to the imminent completion of the overlapped turn.
b. Environments where overlaps are resolved through an "overlap management device" (Schegloff 2000): in this case, more than one speaker is contesting or claiming for a turn space at the same time and overlaps become "problematic" with respect to turn-taking. These overlaps can be sustained and competitive, but do orient to their minimization too, thanks to the management device, which provides for a procedure for negotiating the overlap's end.

Various dimensions can be considered for describing overlaps:

1. The *positions within the ongoing turn* where overlap emerges, exhibiting the real time analysis done by the overlapper of the progressive, incremental organization of the turn;
2. The *positions within overlap itself*: concerning the overlap beginning (*onset*), we can distinguish (cf. Jefferson 1983) a *pre-onset position* (where overlap has not yet taken place but where it can be prevented, generally by taking into consideration vocal or multimodal resources appearing in pre-beginnings of the incipient turn), an *onset position* (where the overlap properly starts), a *post-onset position*, a *post-post-onset or middle position*; concerning the overlap's end (*resolution*), we can distinguish (cf. Schegloff 2000) a *pre-resolution*, a *resolution* and a *post-resolution* position.
3. The *resources* distributed in these positions;
4. The *practices* mobilizing these resources in order to get various interactive jobs done, such as increasing volume and pitch, speeding up or slowing down (Local, Kelly & Wells 1986), repeating, recycling (Schegloff 1987), restarting (Goodwin 1981), all timely organized in relation to the various phases of the overlap, defining its trajectory.

Talk is not only organized step by step, in an incremental way, sensitive to the contingencies of context and of others' interactional conducts; talk is also organized by a *topology of sequentially ordered positions* offering various opportunities to participate. The timed positionings of overlaps show precisely how participants do orient in a finely grained way to these opportunities in order to adequately initiate their interactional moves.

Moreover, the way in which participants mobilize resources and practices for beginning their turns in overlap also exhibits the way in which they consider the emergent process of establishing and accomplishing speakership itself (Mondada 2007). Since overlap is a place where possible competitive turn beginnings are confronted, the way in which overlapping speakers format their turns within the overlap manifests their orientation to their speakership, their rights and obligations as speakers, as well as their stabilization or vulnerability within overlap. Thus, what happens during overlap is interesting to look at for the analysis of the establishment and transformation of speakership.

2. Overlap and multimodality: Simultaneity of talk and simultaneity of gestures

Within Conversation Analysis, overlap is a phenomenon that has been predominantly defined in relationship with talk. In this context, one can wonder about the issues raised by the simultaneous organization of multimodal resources within the sequential organization of talk-and-other-conducts in interaction, especially in environments where overlap occurs. How can video data and multimodal phenomena such as gestures, gazes, facial expressions, bodily postures, contribute to or even transform the view we have of turn-taking and overlaps? Existing work on turn-taking and multimodality has complexified our understanding of the way in which the sequential organization of the turn adjusts to and reflexively integrates multimodal details (see Goodwin 1981). As a consequence, the study of the timed character of social interaction has taken into account both *successive* and *simultaneous* relationships: what characterizes multimodality is that it unfolds in a finely tuned coordination with talk, even if their mutual adjustment within an ordered *Gestalt* is not just a matter of synchronicity (Schegloff 1984; Kendon 2005).

Overlaps concern simultaneous talk; thus, we can ask what happens when other kinds of simultaneities are considered, related to multimodal conducts, gestures, glances, facial expressions, body postures (Oloff 2009). Two possible lines of inquiry can be sketched in this respect. On the one hand, it is possible to explore multimodal conducts during overlapping talk, questioning the way in which gestural and other visual resources contribute (or not) to overlap practices (practices for overlapping, for avoiding overlap, for resolving overlap). This is the line of inquiry adopted here.

On the other hand, one can wonder if multimodal conducts occurring simultaneously with talk could be considered as overlapping it or not. This option is explored by Schmitt (2005), who consequently discusses the concept of "kinetic overlap" on the basis of one example of pre-selection done by a participant in a visible, hyperbolic gestural way during the ongoing talk of a current speaker; the latter notices the former's pre-selection, but selects her only much later on, continuing to speak for a long moment without any perturbation.

The first line is not unrelated to the second. Interestingly, some simultaneous multimodal conducts are oriented to by participants as having an overlapping (or even interrupting) character – as producing some perturbation of talk, similar to the overlapping talk (note that this is *not* the case in the example studied by Schmitt 2005). This can happen when multimodal turn pre-beginnings project imminent (verbal) turn-taking, or when turns are accomplished gesturally (for example providing for the second pair part of an adjacency pair). In these cases, multimodal conduct either projects and prepares (as in the pre-beginnings) or substitutes talk. However, this is *not* the case of all multimodal conducts occurring simultaneously with talk. This leads us to consider overlap primarily as a verbal phenomenon, crucially related to the disruptive potential of more than one person speaking at the same time. Simultaneous talk is produced in a specific way, which does not have the same properties and does not offer the same opportunities as simultaneous multimodal conducts. Moreover, multimodal conducts co-occurring with talk and coordinated with it do not occur in a synchronic way with it, but within a different temporality (mostly *preceding* talk). This feature produces an interesting shift of turn boundaries: in the same way that turns can have gestural pre-beginnings that produce a flexible left boundary of the turn, they can also have gestural post-completions that expand turns in a multimodal way (Mondada 2007). We can predict that these positions can be variously oriented to by co-participants, either as belonging to the turn or not and either as overlapping it or not.

In this paper, we focus on the participants' multimodal conduct during overlap, on the basis of video recordings of everyday activities in interaction. Considering that "hand gesturing is largely, if not entirely, a speaker's phenomenon" (Schegloff 1984: 271), we can wonder what happens in overlap: how are overlapping and overlapped speakers gesticulating? Do gestures follow the same trajectories as talk, for example, exhibiting perturbations when talk is perturbated during overlap or in its aftermath? What does the analysis of gestures during overlap reveal about the way in which participants orient to, recognize and define speakers' rights and obligations?

Schegloff indicates three exceptions to the match between gestures and speakership observed above – all dealing with overlaps in one way or another:

1. The first one concerns incipient speakers. In pre-beginnings, not-yet-speakers or imminent speakers produce gestures that contribute to their transformation from recipients to newly established speakers (see Mondada 2007).

2. The second one concerns "nonspeakers" trying to tell something without interrupting, who orient to the very possibility of organizing a multimodal course of action simultaneously to talk without disrupting it.
3. "A third type of exception occurs when a current speaker is interrupted, and yields to the interrupter. Such at-that-moment nonspeakers may hold a gesture that was in progress at the point of interruption to show that they consider their turn still in progress and intend to resume after the interruption." (Schegloff 1984: 271). This is the type we investigate in this paper. In the example analyzed by Schegloff (1984: 272), the speaker's gesture is frozen throughout overlap and is remobilized when he resumes the turn. We will show that this is one among a range of possibilities. Their variety highlights the embodied conception participants have of speakership as it is locally defined, achieved and sustained.

In this paper, we focus on the trajectory of the ongoing speaker's gestures as she is overlapped by others. On the basis of a video recorded dinner conversation among friends, we analyze a collection of cases showing how these gestures display the participant's orientations to the local definition and recognition of speakership. Four configurations are described:

a. Current speaker continues to gesture during overlap (3.),
b. Current speaker continues to gesture but shows some gesture perturbations (4.),
c. Current speaker holds/freezes gestures before continuing them (5.),
d. Current speaker abandons her gesture (6.).

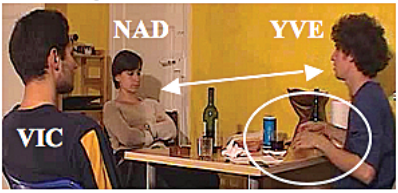
In the first three cases the speaker maintains her speakership; in the last one she looses it.

3. Current speaker continues to gesture during overlap

In what follows, we provide a systematic analysis of various gestures' trajectories during overlapping talk. The first case we focus on displays the continuity of gestures during overlaps, showing how the speaker can maintain an orientation to his status as current speaker, even when he is overlapped by others. The excerpt is taken from a dinner conversation between friends, Victor (VIC) being the host receiving Nadine (NAD) and Yves (YVE) for dinner. Yves has just graduated from an art school where he studied cinema and tells about his experience as a film director. Transcript conventions are explained at the end of this paper.

(1) (PM_150_024010_extrS)

```
1     YVE   comment dire °ou° (.)   °en tout° cas *mettre* en
            how to say °or° (.) °well in any° case*to bring*
      yve                                         *....*shakes->
2           en:::. en avant: (.) °euh° *ce  disposi*+tif,
            to the:::. fore: (.)  °er° *this   devi*+ce,
      yve                              *....body pos*
                                                 gaze NAD+--->1.7
            left hand---------------------------------->1.4
3           [(c'qu'] étaient) qui  #*'taient des&
            [(that] were)     which#* were
4           [((noise of lighter))]
      yve   ------------------------*2h in front of body-->
      fig                           #fig1
5     YVE   &(min*uscules) caméras,#
            &(min*uscule) cameras, #
      yve   -----*percussion gesture-->l. 10
      fig                          #fig2
```



#fig1

#fig2

```
6           [*°donc euh]+::[:.° (y a des)  #par+ties:,+où&
            [*°so er]+::   [:.° (there are)#par+ts:,  +where&
7     NAD   [*mm,hm.]+     [(de            #sem+blant)+de::]
            [*mm,hm.]+     [(to            #sim+ilar) +to:::]
      nad    *nods
      yve    *......turns body left------------------------>>
            gaze NAD-+,,,,,,,,,,,,,,,,,,,,,,,+.....+gaze NAD
      fig                                  #fig3
8     YVE     &voilà]#:,+où y a (où) j' passe d'une
            &exactly]#:,+where there are (where) i pass from one
      yve   -------,,,,,+gaze table & hands---------------->>
      fig                #fig4
```

#fig3

#fig4

```
9     YVE   [#caméra à] *#l'au:tre,  *.h j'arrive&
            [#camera to]*#the other, *.h i'm coming&
10    NAD   [#°ouais enfin.°]*#
            [#°yeah well.°]*#
      yve   percussion---*...Rhand up*---------->>
      fig    #fig5        #fig6
```

#fig5

#fig6

Line 1, Yves is talking to Nadine and Victor about the technical aspects of one of the filmsettings he worked on during his studies. After having mentioned the "minuscule cameras" (l. 5), Yves continues his explanation, but is overlapped by Nadine (l. 6–7), who produces first an acknowledgement and then starts a longer turn, but drops out after a long vowel stretch on "to:::" at the end of line 7. After a short account responsive to Nadine ("exactly", l. 8), Yves continues his turn, describing a scene where he as a director had to switch between the cameras. A last short overlap between Yves and Nadine occurs at line 10, where Nadine is not projecting more to come.

From "bring" (l. 1) on, Yves shakes this hand in front of his head during the development of his TCU, looking at the table. He then changes the position of his body leaning backwards (l. 2), letting slide his left elbow from the table and moving this hand downwards. He is now looking at Nadine who is sitting opposite to him. His right hand lets go of the lighter he was holding until this moment (l. 4), freeing this hand for the gesture to come. Then, Yves holds both hands in front of him in the air, palm down (l. 4, Figure 1), freezing them for a moment. From line 5 on, he starts moving both hands in a percussion gesture in front of his body (Figure 2).

After the first overlap with Nadine (7), Yves begins to turn his torso to the left, maintaining this slightly different orientation until the end of the excerpt. He also turns his head in the same direction, not looking at Nadine anymore on the second overlap onset ("(to)", l. 7), but at his gesturing hands (Figure 3). By doing this, he turns away from the overlapping speaker Nadine and her left hand gesture (see the white circle on Figure 3–4) during the second overlap, now following his own talk's trajectory. He briefly gazes in her direction during the second overlap, turning his head again back to his hands at the end (l. 8, Figure 4). Interestingly, he lifts his hands during the second overlap when his gaze reaches Nadine again (l. 7, gaze towards Nadine), continuing the percussion gesture in an amplified way, being performed at a higher level in front of his body. However, his movements don't signal a perturbation as he continues his percussion gesture during Nadine's overlapping talk, neither changing the rhythm, tempo or amplitude nor his bodily orientation (see Figure 5: overlap onset of the third overlap, Figure 6: end of third overlap). After the last overlap with Nadine, Yves continues gesturing, bringing his hands to a slightly different position when he begins a new TCU.

In this excerpt, Yves is developing a multi-unit-turn and maintains his status as a current speaker during Nadine's overlapping talk. He doesn't glance at her but instead slightly turns away while she is attempting to develop her talk (second overlap). The continuity of his gesture shows no visible perturbation due to her talk. Yves continues his talk and his gesturing after the last overlap, while Nadine withdraws from the turn (l. 10, her turn format and prosody manifest a drop out).

Here, the current speaker organizes his gesture's trajectory in a way that ongoing gesticulations are held when other participants overlap his turn. In this way, even if Yves orients toward Nadine's overlapping talk by shortly glancing in her direction, by responding to what she said (see his responsive "exactly", l. 8) and by repeating a previous segment he produced in overlap ("where", l. 6, repeated after "exactly", l. 8), his gestures are carried out continuously. Thus, Yves continues his turn at a multi-dimensional level: he keeps talking and gesticulating, both in a continuous form and without hitches, displaying the continuity of his speakership.

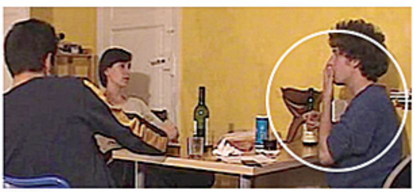
4. Current speaker continues with some perturbations

While in the first case the overlapped speaker goes on gesticulating in a continuous way, in other cases his gestures can display some perturbations. In the next excerpt, Yves is the current speaker, overlapped first by Victor, then by Nadine:

(2) (PM_150_023945_extrF)

```
1    YVE   s'tout que moi j'avais huit caméRAs,
           'specially as i had eight cameRAs,
2          (0.5)
3    NAD   [hm.hm,]
4    YVE   [donc la] galère t'imagin:es, bon c'tait:
           [so the] pain you ima:gine, well that was:
5          c'est: c'tait un choix, dès le début,
           that's: that was a choice, from the start,
6          (0.6)
7    VIC   ouais [c'est (c- huit) caméras.]=
           yeah  [it's (i- eight) cameras.]=
8    YVE         [xxx ausSI,]=
                 [xxx toO,]=
9    NAD   =(e+n ne-)+ en mê[me temps?]
           =(a+t th-)+ at th[e same time?]
10   YVE                     [c't dire que]:. voilà:,
                             [that is]:. exactly:,
     yve      +.....+gaze NAD----------------->l.13
11   YVE   [t'as huit caméRAS*:,    et do[nc]#(.)&
           [you've eight cameRAS*:, and s[o] #(.)&
12   VIC   [ouais:, (°et donc.*et pis do[nc.°)]#
           [yeah:, (°and so.*  and so  th[en.°)]#
13   NAD                               [ou]#ais:,&
                                       [ye]#ah:,&
     yve                      *...lifts 2hands----->
           gaze NAD---------------------------->l.17
     fig                                    #fig7
14   NAD   &[(et alors). ch:]oisir #*en*sui[#te, (d)e::.]
           &[(and so) ch:]oose          th[#en (the::.)]
15   YVE   &[°euh-.°*je-,]                 [#faut qu'
           &[°er-.°* i-,]                  [#i have to
     yve   --------*.....Lhand chin-*--*,,percussion-->
     fig                        #fig8    #fig9
```

#fig7



#fig8

```
16   YVE   &j'ch]oisi*[ss:e,] (.) [*t]#'vois, [+faut que j'choisisse&
           &   ch]oo:*[se,]   (.) [*y]#'see,  [+i have to choose&
17   NAD          [de] la  mer[*:-]#          [+la meilleure:.]
                  [of] the ber[*:-]#          [+the best:]
     yve   ----------*suspension---*percussion gesture-------->l.19
           gaze NAD---------------------------+,,gaze away--->l.19
     fig                               #fig10
```

#fig9

#fig10

```
18   YVE   °(c'ui qui)°] (chalon)# cha- è:- chaque caméra,# à
            °(the one)°]  (eacho)# ea-  e:-  each camera, #at a
     fig                          #fig11                  #fig12
19         tel   + moment.+* [(.) (bien)]*
           given+ moment.+* [(.) (good)]*
20   NAD                     [ouais oua]*is. c'est énorme
                             [yeah ye]*ah. that's tremendous
     yve   -----+,,,,,,,,+gaze NAD --->>
           percussion----*freeze hands--*,,,,-->>
```

#fig11

#fig12

Yves is still talking about the film set and explains the difficulties of using eight cameras at the same time (l. 1, 4–5). Despite the rising intonation on “start” (l. 5) which projects more to come, Yves does not go on immediately, so Victor starts talking after the pause (l. 6–7). But Yves continues his comment about the cameras (l. 8), overlapping part of Nadine’s question to him (l. 9), asking if he was using the eight cameras at the same time. He produces a short second pair part directly after the overlap (“exactly:.”, l. 10) and continues with a longer explanation. He is overlapped first by Victor (l. 12) and then by Nadine, who seeks probably to link up with what Yves has said (see “(and so)”, l. 14). Despite some perturbations, Yves completes his turn and Nadine drops out (l. 17). She self-selects again only when Yves has completed his turn (l. 20).

How do participants multimodally resolve these overlaps? We observe that Yves’s short explanation (l. 10) is still designed for Nadine (cf. his gaze Figure 7). While Victor is overlapping his talk (l. 12), Yves begins to lift both hands, he then makes a stroke gesture on the last item of his TCU (Figure 7) and freezes both hands for a while. During the next overlap, this time with Nadine, Yves lifts his left hand and touches his chin (l. 14–15, Figure 8). He lowers his hand quickly afterwards and starts a percussion gesture with both hands before the next overlap onset (l. 15, Figure 9), thus anticipating a possible completion of Nadine's turn ("(and so) choose then", l. 14). But Nadine continues her simultaneous talk, and shortly before the end of this overlap (l. 16), Yves nearly suspends his gesturing again, briefly freezing both hands in front of his torso before continuing his percussion gesture as he continues his turn (Figure 10). Interestingly, Yves does not only suspend his gesture, but also his turn for a beat, time for Nadine to say in the clear "the ber:-", (l. 16–17). The continuation of Yves's turn again displays his orientation to a possible completion of Nadine's turn ("of the ber:-" could be a complement of "choose", completing Yves' syntactical construction). Nadine finally drops out after the next overlap ("the best", l. 17), while Yves continues his percussion movement until the end of his TCU (l. 19, Figure 11–12).

In this excerpt, Yves tries to maintain his status as a current speaker. The perturbation due to Nadine's overlapping talk is not only visible in the suspension of his current TCU (l. 15), but also in the suspension of his gesture. Yves briefly stops, transforming his gesture in a self-touching gesture, and then continues it while continuing his turn. Another perturbation is visible as he is not only suspending his turn and then repeating it ("i have to choose", l. 16), but also slightly freezing the movement of his hands, taking up his percussion gesture again before recycling what Nadine has previously overlapped. Whereas at the beginning Yves is clearly oriented to Nadine (continuously looking at her until line 16), he turns his head slightly to the left when recycling his syntactical construction (l. 16) and starts looking at the table while continuing his turn (Figure 11–12). This head movement matches the part of his turn where no more overlap and no more gestural perturbation occur, signalling Yves's attention to the trajectory of his own talk (and not to Nadine's). He only reorients to her at the end of his TCU, bringing his head back to its former position (l. 19–20), indicating that he gives her the floor.

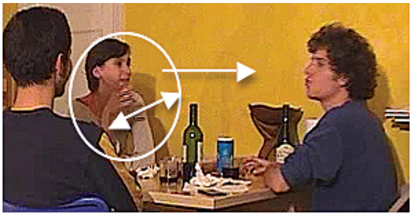
This excerpt shows the current speaker being overlapped by two co-participants who contribute to his talk – since Yves integrates materials from Victor's and Nadine's overlapping turns. These recyclings display his responsiveness to their contributions. During the overlaps, Yves continues his ongoing turn as well as his ongoing gestures, but exhibits some perturbations while taking into account his co-participants' concurrent actions. Once the simultaneous talk is resolved, the speaker continues his gesturing without any perturbation.

5. Current speaker suspends his gesture and then continues

Overlaps can occasion some perturbations both in current speaker's talk and gesture, displaying her sensitiveness to others' conducts. These perturbations can lead to a suspension of the gesture, as in the following excerpt: Nadine's turn will be overlapped several times, leading her to suspend her gesture before continuing it:

(3) (PM_150_025258_extrG)

```
1    NAD   &ça peut être chouEtte un truc avec] des
           &this can be    grEAt something with]
     nad   >>gaze in front---------------------------->
2    NAD   +*mouveMENTS:,+euh* qui *fe+raient[:, euh .h]
           +*MOVEments:,+ er *which*would+be:[er, .h]
3    VIC                                   [°styl]*és.°+
                                           [°styl]*ish.°+
     yve   >gaze nad------------------------------------>
     nad   +..gaze vic---+            +..gaze yve--------+
           *..opens palm-----*beat*open-------------*..->
4          +(0.2)
     nad   +gaze vic-->
           -shakes right hand-->
5    NAD   ouais, pas naturEl:, m:ais à la fois
           yeah, not nAtural:,  b:ut at the same time
     nad   >---gaze vic------------------------->
6          nat[ur*+el,      MAIS#:        EUH::.]
           nat[ur*+al,       BUT#:         ER::.]
7    YVE      [ça veut*+dire, que#tu vOIS i' y a]+une*scène
              [that means,*+that#you sEE there's]+an*incre-
     yve   --nad------+...gaze vic--------------+,,,,,
     nad   >-gaze vic-+...gaze yve--------------------->
           -----------*--lateral mouvements----------*...->
     fig                      #fig13
8    YVE   incroyAble da[ns ba-     *#barry +lyn+dON[:,
           dible scEne i[n  ba-     *#barry +lyn+dON[:,
9    VIC                [ouais voilà*# c'est+ça,+ °x[x° &
                        [yeah  right*#that's+it,+ °x[x° &
10   NAD                                            [tu&
                                                    [you&
     yve   >gaze in front------------------+..gaze vic-->
     nad   >---gaze yve-----------------------+.gaze vic->
           ..right hand to chin.....*touches chin---------->
     fig                        #fig14
```



#fig13

#fig14

```
11   VIC   &+*x quand #c'est dansé, +*c'est pas non*#] plus&
           &+*x when   #it's danced,+*it's   not *#]natural&
12   NAD   &+*ferAIs la#    difféREN+*CE,    entre:*#]
           &+*would make a# dIFFeren+*ce,  between:*#]
     yve   -+...gaze nad------------------------------------>
     nad   ---vic-------------------+,,,, looks in front-->
           -*..opens palm ----------*,,,,,,,,,,,,,*chin-->
     fig              #fig15                          #fig16
13   VIC   &na[ture]l. °mais:°
           &ei[the]r.  °but:°
14   NAD      [ouais.]
              [yeah.]
```

#fig15

#fig16

```
15   NAD     [*entre    +#la   DAN:SE    [*ET     LE*:::. *]
             [*between  +#the  DAN:CE    [*AND    THE*:::. *]
16   YVE     [*mais enCO+#RE, c'est PAS::[*c'est pAs*trOp:*]
             [*and   STI+#LL, it's NOT:: [*it's  nOt*too: *]
     nad                +...gaze yve----------------------->
             -*...open palm lateral mouv.-----------*.throat>>
     yve                                   *---nods---------*
     im.                #fig17
17   YVE     c'est pas trop: (ouais parce que,)+#par exemple,&
             it's  not too:  (yeah because,)   +#for example,&
     yve     ---gaze nad-----------------------+,,,,,
     im                                        #fig18
```

#fig17

#fig18

In this fragment, the three friends are talking about how actors can move in a film, opposing dancing vs. non-dancing, but still choreographed, movements. After having positively assessed Yves's explanation (l. 1), Nadine goes on with a detailed description of the movements. During this turn, her right hand moves progressively into action, first with a short beating gesture with the open palm (l. 2). When Victor suggests a possible completion ("stylish", l. 3), Nadine suspends her turn, looks at him and starts performing a small shaking gesture with her right hand, thereby projecting the refusal of his suggestion (l. 4). After a minimal acknowledgment ("yeah", l. 5), she gets back to her suspended turn and starts performing a lateral movement with the right hand (l. 6, Figure 13). Anticipating her possible completion, Yves self-selects in overlap (l. 7); both speakers upgrade to a competitive volume some syllables later (Schegloff 2000). In this context, Nadine progressively drops out of her turn: first she directs her gaze to Yves, then she withdraws from the turn, and finally she shortly retracts her gesturing right hand, slightly touching her chin (Figure 14).

Yves announces the description of an "incredible scene" in the movie "Barry Lyndon" (l. 7–8). While he is still talking, Victor responds to Nadine's previous turn (l. 9). As Victor and Nadine engage in mutual gaze during this overlap, Yves drops out of his turn (l. 8). At a possible completion of Victor's turn ("that's it", l. 9), Nadine self-selects (l. 10) and brings her hand into an open palm vertical position (Figure 15), which would enable her to easily carry out again the previous lateral movement (cf. Figure 13). Despite the incompleteness of her turn ("between:", l. 12), Nadine retracts her hand again (Figure 16) and stops talking – although projecting a continuation (see her gaze withdrawal from Victor, Figure 15–16, and her minimal acknowledgement, l. 14).

Both Yves and Nadine orient to the next transition-relevance place and self-select in overlap (l. 15–16). While Yves is looking at Nadine, Nadine visibly continues her suspended previous turn: she repeats the connector "between" (cf. l. 12) and immediately brings her hand back into the vertical open palm position (Figure 17). Although she considerably raises her voice during the overlap, she doesn't complete the second complement ("the dance and the::.", l. 15). Shortly before withdrawing from the turn, her right hand stops moving laterally and is retracted to her throat (Figure 18). The participants treat the turn as complete for all practical purposes: Yves starts nodding on the negation (end of l. 16), visibly responding to her, and Nadine seems to orient to the completeness of her turn, adopting a recipient's posture (cf. Figure 18), while Yves gets back to the description of the film scene.

In this excerpt, the successive perturbations affecting Nadine's complex turn are visible not only in the suspension of her talk, but also in the delays of her lateral gesture at the beginning and in the retractions of her gesturing right hand. Only after having (for a moment) secured her turn's trajectory, Nadine (re)positions her right hand in a vertical open palm position and executes a lateral movement.

6. Current speaker suspends his gesture and then abandons it

As we saw in the last excerpt, the perturbation occasioned by the overlap can provoke a suspension of the speaker's gesture, who can repristinate it as he continues to speak. In other cases, the gesture is not only suspended but also abandoned, as here:

(4) (PM_150_024643_extrZ)

```
1   YVE  *et le  *moment
         *and the*moment
    yve  >>gaze VIC----->
         *.......*palm open
2           où*+  t'#+a[s t'as deux +#nanas,#**OU T'AS]     **la *+&
         where*+you'#+v[e you've two+#girls,#**WHERE YOU'VE]**the*+&
3   VIC                [(et)::::::,                ]
                       [(and)::::::,               ]
    yve        *2hands alterning upon the table-------------------*
                                             **Rhand tow VIC**
         gazeVIC+,,,+gaze table------+gaze VIC--------------------+
    fig             #fig19            #fig20   #fig21
```

#fig19

#fig20

#fig21

```
4    YVE      &+*prin[cesse,# (.) et   le:+#::::       ]
              &+*prin[cess, # (.) and the:+#::::       ]
5    VIC             [mais dans d'autres comédies musi]cales c'est comme ça?+
                     [but  in   other     musical  come]dies it's like that? +
     yve      *both hands' gestures frozen on the table--->
              +looks at table------------+looks at VIC---------------------+
     fig                   #fig22          #fig23
```

#fig22

#fig23

```
6             +(0.2) *# (0.4)*
     yve      ------>*,,,,,,,*
              +looks away--->>
     fig             #fig24
7    YVE      *ouais mais là # c'est c- tu vois c'est c'est j'trouve c'est:
              *yeah but there# it's you see it's  it's  i think  it's:
     yve      *2hands resting on the table---->>
     fig                     #fig25
```

#fig24

#fig25

Yves is talking about a musical comedy's dancing scene with two girls. He is overlapped twice by Victor. The first time (l. 2) he continues his turn, but the second time (l. 4) he drops out, letting Victor continue in the clear. Later on (l. 7), when Yves self-selects again, he does not finish his description of the dancing scene, but produces a second pair part responding to Victor's question (l. 7).

At the beginning of the fragment, Yves has the floor (l. 1–2, 4) and gesticulates iconically, moving his two hands on the table within an alternate movement as he talks about the two girls (Figure 19). Victor comes in with a lengthy stretched "(and):::::," (l. 3), projecting more to come. During this first overlap, Yves maintains his gesture. However, he orients to the incipient overlap: shortly before the overlap, and as Victor raises his head and possibly opens his mouth, projecting an imminent self-selection, Yves stops looking at him and gazes at his hands (Figure 19). During the overlap, Yves amplifies his gesture, directing his right hand towards Victor (l. 2, while saying "WHERE YOU'VE" with a louder voice) (Figure 20–21) and looking again in his direction. Through this triple amplification, Yves treats Victor's overlap as competitive.

Victor withdraws but self-selects again (l. 5), so that Yves's turn continues in the clear only for a few beats and is then overlapped again. In this environment, he freezes his gesture on the table, maintaining the hands' shape but not moving them any more during Victor's turn (Figure 22–23). His gestural conduct is thus different during this second overlap: the gesture is not continued but frozen. Likewise, his TCU is abandoned (long stretch of the article "the:::::", l. 4).

After Victor's turn at the TRP that follows during a pause (l. 6, Figure 24), Yves doesn't continue his gesture, but withdraws and abandons it. Although Victor clearly designs an opportunity for Yves to self-select, he also designs a new sequential environment projecting a next action which is not the continuation of Yves's previous one; on the contrary, Victor produces a disagreeing argument, to which Yves now responds. Yves's gestures are sensitive to this sequential reorientation and show that he is abandoning the previous action: both hands rest on the table, not moving any more. He also glances away, neither looking at Victor, nor at his hands (Figure 25).

In this sense, Yves's gestures show him to be sensitive to the unfolding sequential environment, orienting not only to his rights and obligations as a current speaker, but also to the changing sequential implicativeness of talk. He abandons the trajectory of his talk before its completeness and reorients to the next action Victor has projected in overlap.

7. Discussion

The preceding analyses have shown that gestures in overlap can indicate the degree of perturbation of the overlapping talk for the ongoing speaker. When he treats the simultaneous speaker as non-competitive, he continues gesticulating during the overlap (case 1). Cases 2 and 3 imply a higher degree of perturbation: there is still a continuation of the gesture, but the more problematic character of the overlapping talk is visible in the gesture perturbation (case 2) or even its brief suspension (case 3). The speaker's orientation to the overlap as being even more problematic is shown when he first suspends and then abandons the gesture initiated during his turn (case 4).

Table 1 illustrates this continuum from the less problematic to the increasing problematic cases of overlap. In the first case, there is a continuation of gesture across the overlapping talk, whereas in the further cases gestures are more and more disturbed, being finally abandoned:

Table 1. Possible trajectories of the ongoing speaker's gestures in overlap

Continuation	*Cont. with perturbation*	*Suspension with continuation*	*Suspension with abandon*
Case 1	*Case 2*	*Case 3*	*Case 4*

–problematic ――――――――――――――► **+problematic**

Thus, our findings show systematic variations related to speakership as a practical accomplishment in talk, sensitive to the projective potential of both turn organization and sequential organization.

Gesture perturbation is not only related to possible competitive incomings of other speakers, but also displays sensitivity to the sequential implicativeness of the overlapping talk. An overlapping turn projecting either more to come or a second pair part offers the ongoing speaker two possibilities: he can either follow his own turn's relevance (e.g. by continuing it, eventually through skip connecting), or he can address the emerging relevance set up by the overlapping talk. The bodily perturbation manifested by the ongoing speaker reveals in which way he treats these multiple projections. If he abandons the conditional relevance set up by his own turn and orients to the other's turn, this can be visible in the abandonment of his gesture; while the continuation of his projected trajectory will be reflexively achieved within the continuity of his movements, even if they are momentarily suspended and even if the speaker inserts a response to the overlapping speaker. The speaker can also display his physical disalignment with the overlapping co-participant, turning his gaze and his body away from him. The visible perturbation (suspension or abandonment) does not always manifest that the overlapped party is abandoning his status as a speaker: it can also embody a sequential reorientation in his talk and his changing status from a speaker initiating and projecting more to come to a speaker responding to a previous action.

When looking at our cases, we observe that when the current speaker continues talking within a longer overlap, he does not maintain the gaze towards the overlapper, but turns it away during simultaneous talk (ex. 1, 2, 3). In cases where the gaze to the overlapper is held more continuously, the overlapped speaker seems merely to reorient to the sequential implicativeness of the simultaneous talk of his co-participant (ex. 4). This can be observed in cases where the incoming speaker is explicitly addressing his turn to the overlapped (f. ex. designed as a first pair part).

8. Conclusion

The fine-grained multimodal analysis we presented here shows how speakers deploy gestures in overlapping sequences. Gestures exhibit the participants' orientation to the current speaker's rights and obligations and their changes. Video analysis shows how speakers' gestures exhibit their treatment of different kinds of overlap as being more or less problematic, as being collaborative or competitive. These embodied displays orient to the timed details of turn design: Gestures (and gaze) are used to reflexively define the completeness of TCUs or its anticipation.

As we sketched out, the exact position of the gesture within the sequential development of talk-in-interaction and its precise timing with the overlapping talk have to be accounted for in the transcriptions and during the analysis. If we are interested in how speakers resolve overlap not only via verbal resources but also with gestures, changes in gestures have to be finely related to the overlap onset and to the within-overlap talk. Indeed, we demonstrated that for overlap resolution, speakers can make use of a variety of resources – not only increasing/decreasing volume, prosody, repetitions and restarts, sound stretches and rush-throughs (Schegloff 2000), but also various kinds of gestures, changes of bodily postures, movements, object manipulations, and gaze. Even if overlap is a phenomenon primarily defined by the juxtaposition of verbal and vocal resources, participants also deploy visible resources in order to manage it. Therefore, adequate video recordings, together with a fine-grained analysis, can contribute to the understanding of how participants use complex multimodal resources in order to manage their talk-in-interaction and their rights as speakers.

Transcription conventions

Talk has been transcribed according to Jefferson's conventions. For gesture, the following conventions have been used:

+ +	beginning and end of gaze
* *	beginning and end of gesture (body movement etc.)
...	beginning of gesture (of gaze, movement etc)
,,,	end/retraction of gesture
–	continuation of gesture
–>	continuation of gesture in the next line
–> l.5	continuation of gesture until line 5
–>>	continuation of gesture until the end of the extract
ppp	points
h	hand
L R	left, right
2h	both hands

References

Egbert, M. 1997. "Schisming: the collaborative transformation form a single conversation to multiple conversations." *Research on Language and Social Interaction* 30: 1–51.

Goodwin, C. 1981. *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

Goodwin, C. 1986. "Between and within: alternative treatments of continuers and assessments." *Human Studies* 9: 205–217.

Jefferson, G. 1973. "A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences." *Semiotica* 9: 47–96.

Jefferson, G. 1983. "Notes on some orderlinesses of overlap onset." *Tilburg Papers in Language and Literature* 28.

Jefferson, G. 2004. "A sketch of some orderly aspects of overlap in natural conversation (1975)." In *Conversation Analysis: Studies from the First Generation*, G. H. Lerner (ed), 43–59. Amsterdam/Philadelphia: John Benjamins.

Kendon, A. 2005. *Visible Action as Utterance*. Cambridge: Cambridge University Press.

Local, J., Kelly, J. and Wells, B. 1986. "Towards a phonology of conversation: turn-taking in urban Tyneside speech." *Journal of Linguistics* 22 (2): 411–437.

Mondada, L. 2007. "Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers." *Discourse Studies* 9 (2): 195–226.

Oloff, F. 2009. *Contribution à l'étude systématique de l'organisation des tours de parole: Les chevauchements en français et en allemand*. PhD Dissertation, University of Lyon.

Sacks, H. Schegloff, E. A. and Jefferson, G. 1974. "A simplest systematics for the organization of turn-taking for conversation." *Language* 50 (4): 696–735.

Schegloff, E. A. 1982. "Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences." In *Analyzing Discourse: Text and Talk*, D. Tannen (ed), 71–93. Washington: Georgetown University Press.

Schegloff, E. A. 1984. "On some gestures' relation to talk." In *Structures of Social Action* J. M. Atkinson and J. Heritage (eds), 266–296. Cambridge: Cambridge University Press.

Schegloff, E. A. 1987. "Recycled turn beginnings." In *Talk and Social Organization,* G. Button and J. R. E. Lee (eds), 70–85. Clevedon: Multilingual Matters.

Schegloff, E. A. 2000. "Overlapping talk and the organization of turn-taking for conversation." *Language in Society* 29 (1): 1–63.

Schegloff, E. A. 2002 "Accounts of conduct in interaction: Interruption, overlap and turn-taking." In *Handbook of Sociological Theory*, J. H. Turner (ed), 287–321. New York: Plenum.

Schmitt, R. 2005. "Zur multimodalen struktur von turn-taking". *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 6: 17–61 (www.gespraechsforschung-ozs.de).

PART VI

Gestural analysis of music and dance

CHAPTER 25

Music and leadership

The choir conductor's multimodal communication

Isabella Poggi

This chapter views the choir conductor as the leader of a cooperative group, whose role is not simply to provide technical instruction, but also to motivate singers, provide feedback, express the pleasure of music. It assumes that every choir conductor pursues a specific plan of action and that his/her multimodal behavior is aimed at fulfilling the goals in that plan. It proposes an annotation scheme for analyzing a conductor's head, eyebrow, eye, mouth, trunk and hand actions in terms of their physical parameters, their literal and indirect meanings, and their goals within the conductor's plan. The scheme allows for the outlining of the body styles of different conductors, distinguishing them in terms of the goals fulfilled in their conduction.

Keywords: Multimodal communication, annotation scheme, leader, Conductor, music performance

1. The musician's body

Body behavior in music performance has been studied in players and singers (Bresin & Battell 2000, Davidson 1994, Duranti & Burrell 2006, Caterina et al. 2004, Dahl & Friberg 2007, Dahl et al. 2009), and orchestra conductors (Poggi 2002, Boyes Braem & Braem 2004) for theoretical concern on the mechanisms of body movement to enhance musicians' technical skills, and to build devices for home conducting and virtual conductors (Friberg 2005).

Among musicians' movements, some simply support the technical movements (King 2006), while others have a clear expressive or communicative function (Davis & Ashley 2005, Poggi 2006), by displaying cognitive and emotional processes. A pianist for instance may display concentration while playing a difficult passage or bow on the piano to listen to the music produced, or express 'felt' and 'enacted' emotions. An 'enacted emotion' is one the musician must feel or pretend to feel, and express to impress it onto the music (Poggi 2006). It is '*meaning oriented*' if he simulates or induces it in himself to convey the 'meaning' of that emotion through music (e.g., feeling sad to play a sad music more credibly); '*movement oriented*' when the technical movements required by a music passage are contained in, or favored by, the expression of that emotion (e.g., she performs a frown, looking angry when playing a *fortissimo* because the expression of anger mobilizes the energy required for a strong touch). Sometimes, though, a musician expresses really 'felt' emotions or sensations: those linked to the *process* of playing – positive like relaxation or flow, negative like tension or fear of mistaking – and those caused by the *outcome* – shame for a mistake, satisfaction for a beautiful sound.

In players not all body behaviors are *communicative*: some simply help motor processes; others are *expressive* movements, displaying mental states but not deliberately devoted to communicating anything to others.

A conductor's movements, instead, are by definition *communicative*: the orchestra conductor uses gestures and all his body to communicate various types of information to players.

In this work I analyze the choir conductor's behavior in terms of the goals he must pursue in conducting, and based on them, the types of information a conductor conveys to singers. Then I present an observational study on the multimodal behavior of a choir conductor to explore multimodality and social interaction in music performance.

2. Singing together: A cooperative plan of action

What are the body behaviors typical of a choir conductor? What should he do to have singers sing well?

To analyze a choir conductor's behavior I adopt a model of mind and social interaction in terms of goals and beliefs (Conte & Castelfranchi 1995, Poggi 2007a), according to which the life of every system (i.e., individual or collective) consists of pursuing goals, with a goal defined as any state that regulates, triggers and sustains action. This definition of goal subsumes several psychological notions (e.g., drives, instincts, motivations, interests, needs, intentions). All of these notions are types of goals. Often a goal cannot be achieved through a single action but needs to be pursued through a hierarchy of goals, a plan where each action aims at a goal, but this in turn may be aimed to a further goal (supergoal), with all actions aiming at the same end goal. Goals are pursued through internal and external resources. The former include the system's beliefs and action capacities, the latter material resources, world conditions, and 'adoption' (i.e. help from others), that multiplies the systems' power of achieving goals. When system S lacks resources for its goals, it depends on another system C's adoption (i.e., C's putting its resources to the service of S's goals). Several kinds of adoption exist: exchange, altruism, norm compliance, cooperation. Cooperative adoption holds when C adopts S's goal in order to reach another goal that both C and S have.

A choir is a cooperative system, where many individuals aim at the same goal. Its achievement depends on all individuals' action, and each individual adopts the goal of another in order to achieve the shared goal. For example, Choir Conductor C, moving his hand, adopts the goal of singers "knowing when to start"; singer S1 looks at the Conductor to adopt his goal of gaining attention; singer S2 sings softly to let the theme sung by singer Sn be better heard, all actions aiming at making beautiful music, a common goal of Conductor C and Singers S1, S2, Sn.

Since the choir is a group, the conductor can be seen as the leader of the group; the leader and the group components are highly interdependent in that a good leader's actions fulfill the followers' needs. For example, followers need (have the goal) to know what to do, where to go, what goals to pursue, and the leader proposes a mission, a vision, a goal to be pursued. Further, to feel better as persons, followers need to identify themselves with the leader, to absorb his admirable capacities, and the good leader presents himself as a model to imitate. Followers need to feel that someone takes care of them, and the leader is empathic. In other words, the leader's actions and goals perfectly respond and correspond to the followers' needs.

I hypothesize that the conductor's behavior in conducting pursues a plan where each action fulfills the singers' cognitive and affective needs. In Section 3, I illustrate the hierarchy of goals a conductor ideally pursues in conducting, making the hypothesis that this plan gets instantiated into the actions he performs and the types of information he conveys to singers. In Section 4, I present an observational study on the multimodal behavior of a choir conductor, showing how the communicative and non-communicative behaviors hypothesised are all present in real conducting.

3. The conductor's plan

Figure 1 represents the conductor's plan of action in conducting a choir.

In the round, I represent the Conductor's goals, in italics the actions he performs to fulfill them.

The end goal (G1) is for singers to sing at their best, which is pursued through three sub-goals. An obvious sub-goal (G3) is that singers know how to sing (i.e., they have technical information about the sound to produce, *who* is going to sing, when, what is the *conten*t expressed by the words to sing, *what sound* to produce, and *how*). The specific sound to produce implies various goals, corresponding to the parameters of music: the Conductor asks for a particular *melody*, *rhythm*, *tempo*, *timbre*, *intensity*, *expression*, or reminds the singers about aspects of the *musical structure* of the piece (e.g., coming back to the tonic or changing from minor to major). The Conductor's gestures, face or body movements may point at one or another of these parameters.

While the singers' need for technical information is quite obvious, a Conductor has further goals in interacting with the Choir, i.e., motivating singers (G2) and providing feedback (G4). The former (motivating people to do something) is a typical goal of any leader. Being motivated to do something means to attribute a high value to some goal, even at the expense of other goals: if one is more motivated to sing in a

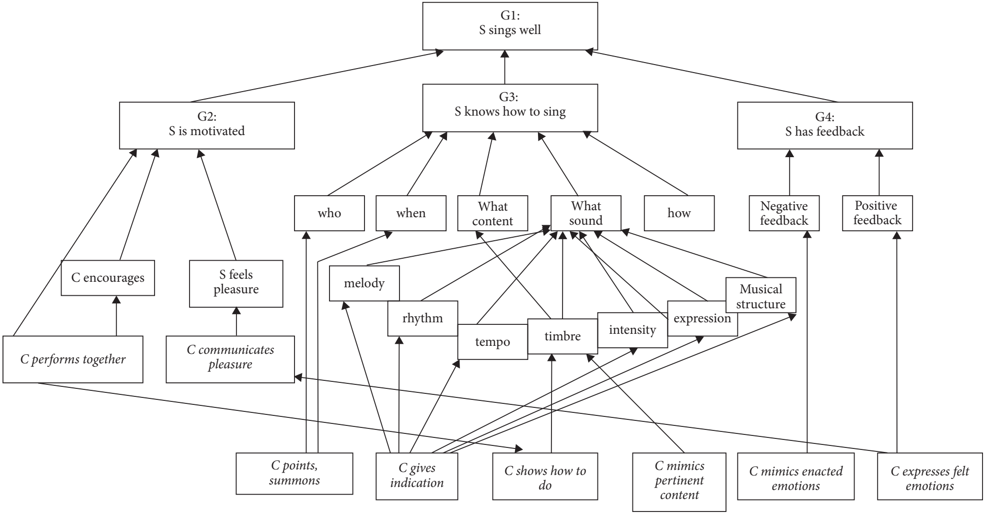


Figure 1. The Choir Conductor's plan.

Choir than to play with friends at home, one may prefer to get out in a rainy night and give up a comfortable evening home. A participant in a group must be motivated to pursue the common goals of the group as opposed to other goals of his own. A good leader enhances motivation: group members should feel that the common goal has a higher value than their individual goals, hence be willing to pursue it with involvement and enthusiasm. Enthusiasm (Poggi 2007b) is a kind of anticipated joy felt when you think that the goal you are pursuing is beautiful, noble, important, and that you can achieve it, as evidenced by your achieving a subgoal of it. Take enthusiasm after a goal during a football game: the partial success: (1) increases the sense of self-efficacy, (2) enhances motivation by further increasing the value of the goal, and (3) increases strength, energy and concentration in goal pursuit.

In the choir job, if singers feel the music is beautiful and worth singing well, they will strive to do their best. This is why it is important for the conductor, as for any good leader, to motivate singers and induce the joy of singing. He can do so by making them notice how beautiful the music being sung is through expressing his pleasure. Another way to induce motivation in the group members is that a leader can be an example for them, which entails making them feel he is like them and doing himself what he is asking from them. So a good leader not only tells the followers how to do something, but performs the same actions together with them.

Once the singers are duly motivated and instructed on how to sing, the conductor's last subgoal for their singing well is to give them feedback (G4) about how they are singing to evaluate their work. A negative evaluation – e.g., communicating they are singing too loud or too fast – is useful to their goal of knowing how to sing; a positive evaluation may both provide relevant information – e.g., "go on this way" – and, by enhancing self-efficacy, encourage (i.e., increase motivation).

4. An annotation scheme of the Choir Conductor's communicative behavior

My hypothesis is that the choir conductor's body behaviors all aim to fulfill the goals and supergoals of the plan above. To test this hypothesis and see how the conductor's plan is brought about in real conducting, I carried out an observational study on the multimodal communication of a choir conductor in concert: an execution of Rossini's *"Petite Messe Solennelle"* by the Choir "Orazio Vecchi", conducted by Roberto Anniballi (University Roma Tre, December 19th, 2006).

To analyze the conductor's multimodal communication, I developed an annotation scheme aimed to capture the Conductor's behaviors, their communicative and non-communicative goals and their function in terms of the conductor's plan of action illustrated above.

The scheme (Table 1) includes 11 lines and 13 columns. The first four lines contain, respectively: (1) number of the bar sung by the choir, (2) words sung, (3) notes sung by singers or played by the pianist, (4) time in the video.

Table 1. An annotation scheme of the Choir Conductor's multimodal behavior

1. BAR	*63*			*63 64*			*64*			*65 66*		
2. WORDS	*Domi*			*ne De*			*Us*			*(Piano)*		
3. NOTES	*B B*			*C# D*			*B*			*A#; B C# A#*		
4. TIME	*11.16*			*11.18*			*11.19*			*11.19-20*		
	Behavior	**Goal/ Meaning**	**Type**	*Behavior*	**Goal/ Meaning**	**Type**	*Behavior*	**Goal/ Meaning**	**Type**	*Behavior*	**Goal/ Meaning**	**Type**
5. Trunk	*Backward right*	I with-draw → I suffer	**Enacted Emotion**	*Leans forward*	I implore you	**Content**						
6. Head	*Down-ward right*	I am suffering	**Enacted Emotion**	*Raised forward upward right*	I address you God	**Content**	*Défault Frontal forward*	I relax now → the musical phrase is over	**Enacted Emotion Musical structure.**	*Shakes head fast*	No, no → highly → I express an intense pleasure	**Felt Emotion**
7. Eyebrows	*Oblique*	I am sad	**Enacted Emotion**	*Oblique*	I am sad	**Enacted Emo-tion**	*Défault.*	I relax now → the musical phrase is over	**Enacted Emotion Musical structure**	*Raised*	I am proud → We are singing well	**Felt Emotion → Positive Feedback**

8. Eyes	*Closed squeezed*	I strive to cope with my suffering	**Enacted Emotion**	*Closed squeezed*	I strive to cope with my suffering	**Enacted Emo-tion**	*Closed relaxed*	I relax now → the musical phrase is over	**Enacted Emotion Musical structure**	*Half-closed*	I am concen-trating on this music	**Metacog-nitive**
9. Mouth	*Smile with lip corners down*	I am suffering	**Enacted Emotion**	*shape of a large a*	Shape your mouth like for an a	**Sing together → Show how to sing**	*Shape of a narrow a*	Shape your mouth like for an a	**Show how to sing**	*Smile*	I feel pleasure	**Commu-nicate pleasure**
10. Left Hand	*Cupped tense*	I am in tension	**Enacted Emotion**	*Cupped hand palm up in grasping shape*	I implore	**Content**				*Cupped hand nearly open palm up, up / down as if sustaining something*	Come on, I sustain you	**Who**
11. Right Hand	*Cupped tense*	I am in tension	**Enacted Emotion**	*Cupped hand palm up in grasping shape*	I implore	**Content**	“					

Lines 5 through 11 concern the modalities whose signals are analyzed. In columns, each group of three columns describes the multimodal behavior – one modality on each line – performed by the conductor while the notes and words of a bar are being sung by singers. In each group, column 1 (Behavior), in lines 5 through 11, respectively, describes the movements of the conductor's trunk, head, eyebrows, eyes, mouth, left and right hand. Each behavior is analyzed in terms of the parameters relevant for its modality. For example, for a gaze item, iris direction, eyelid openness, eyebrow position are reported (Poggi 2007), for a gesture, handshape, location, orientation and movement are described, and in some cases the gesture expressivity (e.g., amplitude, fluidity, velocity, tension, repetition) is also analyzed (Hartmann et al. 2002). In column 2 (Goal/Meaning), each movement is analyzed as to its communicative or non-communicative goal: if considered a non-communicative action, its presumable purpose is written down; if assumed to be a communicative action, a verbal paraphrase of its meaning is provided. Moreover, also the possible indirect meanings are taken into account. In fact, as previously argued (Poggi 2007), a communicative body behavior may also have, beyond its literal meaning, a further meaning that is communicated indirectly through metaphor or other non-literal device. Thus in column 2, besides the apparent meaning of a smile, head movement, or gesture, an arrow sometimes points at its indirect meaning. Finally in column 3 (Type), the goal or meaning written in column 2 is classified in terms of the conductor's plan proposed in Figure 1.

Let us see the analysis of bars 63–65, when the soloist tenor sings *"Domine Deus"*. At time 11.16, bar 63, while he is singing *"Domi-"*, the beginning of the invocation 'Lord God,' the Conductor moves his trunk (line 5) backward rightward and his head (line 6) downward, his eyebrows (line 7) are oblique, an expression typical of sadness, and his eyes are closed and squeezed (line 8) like in the effort of coping with pain. Finally his smile (line 9), with lip corners down, expresses bitterness, and his hands cupped are in tension (lines 10, 11), expressing the enacted emotion of pain. As mentioned, a musician *enacts* an emotion to impress it into the music. In the conductor this device is recursive: he *recites* to be feeling some emotion in order for the singers to feel it and express it through music, and finally transmit it to the audience. At bar 63, everything in the Conductor's body enacts the sadness of someone who is imploring God.

At bar 64 (see the next three columns, time 11.18) he leans forward (line 5) and raises his head (6) as if imploring (*trunk forward*) while addressing God (*raised head*). Thus he is providing information about the content of the words the singer is singing, while again he is enacting sadness with oblique eyebrows (line 7) and closed squeezed eyes (line 8). A performative of imploration – again informing on the content of the words to sing – is also conveyed by his cupped hands with palms up in a grasping shape (10, 11). Simultaneously, he opens his mouth wide and round as if singing an "a" (9): a technical suggestion, reminding the tenor how he should produce the sound; but at the same time a typical feature of this specific Conductor, that characterizes him as opposed to others: his singing together with singers, to let them feel accompanied and tuned with him.

In the last part of bar 64 (next 3 columns, time 11.19), while his mouth (line 9) takes the shape of a narrow "a", aimed at suggesting how to sing and singing together, the Conductor's head and eyebrows come back to the default position, and the eyes are closed in a relaxed way (lines 6, 7, 8). All this facial behavior as its literal meaning conveys an enacted state or relaxation, but this could in turn convey a metadiscursive meaning concerning the *musical structure* of the piece performed: "I relax, so you may relax because the musical phrase is over." In fact, at bars 65 and 66 the piano solo intervenes. During this pause of the tenor, the Conductor conveys various meanings. With his cupped left hand palm up going up and down (line 10), he summons the pianist, communicating *who* is to play, but also sustains and incites him to play in a distinct way. With his half-closed eyes (line 8) he conveys a metacognitive meaning, "I am concentrating on this music."

An orchestra conductor often displays concentration before giving a start (Poggi 2002). That is, he shows concentration to be mirrored by players or singers so that they get prepared to the start too. Here, though, the Conductor seems to show concentration to have singers taste and feel the pleasure of music, like he does. This interpretation of the metacognitive display is made plausible by the context of other simultaneous expressive signals. His smile (line 9) expresses pleasure, and the rapid head-shake (6) can be seen as an intensifier of it (shaking head often conveys meanings like "much," "many," "highly," "a lot": Kendon 2002, Heylen 2005). Meanwhile, the raised eyebrows (line 7) express pride. These displays of pride and pleasure, along with the intensifier, are expressions of *felt emotions*, but might further aim at encouraging the performers through providing *positive feedback*; and the whole set of meanings can be seen as a multimodal discourse (Poggi 2007) conveying, "I am concentrating on listening to the music; please you do so too, I am happy and proud because the music is beautiful, and we all are performing well."

5. Conduction body style

Adopting this annotation scheme, two fragments were analyzed from the concert above (respectively, one fragment of 1'12" and one of 0'26", 1'38" in total), and all body signals with a communicative import were computed (Table 2).

The two meanings most frequently conveyed, either literally or indirectly, are *enacted emotions* – emotions expressed by the conductor to make singers express them in music in their turn (43) and *intensity* – reminding players of playing *piano* or *forte* (36). The next more frequent signals concern *content* and *expression* (18), *rhythm* (17), and *timbre* (14). *Metacognitive* information is conveyed with the same frequency as *timbre* (14), followed by *showing how to sing* (11), *tempo* and *sing together* (7), *melody* and *communicate pleasure* (5), *positive feedback* and *musical structure* (4), and finally *who* is to sing (1). All in all, technical information is conveyed 120 times, non-technical 94 times.

Table 2. Anniballi's multimodal communication

Type of meaning	Literal n.	Indirect n.	Total	%	
Who	1		1	0.46	Technical
Melody	5		5	2.33	
Intensity	18	18	36	16.82	
Rhythm	17		17	7.94	
Tempo	7		7	3.27	
Expression	17	1	18	8.41	
Timbre	12	2	14	6.54	
Musical Structure	1	3	4	1.86	
Content	15	3	18	8.41	120
Metacognitive	14		14	6.54	Non-Technical
Enacted emotion	43		43	20.09	
Felt emotion	10		10	0.67	
Comm.pleasure	1	4	5	2.33	
Positive feedback	1	3	4	1.86	
Show how to sing	7	4	11	5.14	
Sing together	7		7	3.27	94
TOTAL	176	38	214		

Some meanings are communicated only in a direct way. *Who* is to play, *melody, rhythm* and *tempo*, *enacted* and *felt emotions* are never conveyed indirectly.

Out of the meanings conveyed indirectly, the most frequent is *intensity*, that is generally communicated either through *metacognitive* communication (twice) or, more often, through *enacted emotions* (16 times). For example, the Conductor mimics concentration, attention, caution to remind players to play *piano*; and mimics anger or determination to make them play loud. *Enacted emotions*, which are only present at the level of direct communication in 23 cases, have the role of conveying indirect meanings. Specifically, 16 convey *intensity*, 3 *musical structure*, 3 *content* and 1 *expression*. For example, respectively, by mimicking anger the Conductor asks musicians to play loud, by showing inspired to play softly, by showing relaxed he conveys that a musical phrase is over, by expressing sorrow the performative of imploration, by showing determination with gaze fixed downward, he asks for a determined expression.

Such an analysis of a conductor's multimodal behavior gives us a tool to characterise different body styles of conduction.

The particular Conductor observed here shows very characterizing peculiarities. He produces a high number of body signals per time unit and a higher amount of signals of enacted emotion. He also continuously sings together with singers and finally often provides positive feedback, expresses felt emotions and communicates pleasure.

This pattern of open and enthusiastic communication is completely consistent with his behavior in other contexts. His empathic attitude, which leads him to continuously share the singers' job with full participation, is also clear from his behavior before and outside the performance. For example, when people come in to hear the rehearsal, he warmly welcomes any new person, showing a sincere pleasure of meeting together. Furthermore, in his everyday work as a music teacher in Grammar School, he is well known as one who inspires enthusiasm of making music in all children.

Coming back to his behavior as a conductor, our multimodal analysis definitely credits him the look of a charismatic leader, particularly keen to encourage and motivate singers. Moreover, the whole pattern of his multimodal behavior strictly corresponds to the hierarchy of goals of the choir conductor presented above, in that all the behaviors predicted are actually represented in his conducting.

6. Conclusion

There are many things a choir conductor must do during performance for the choir to perform well, including much more than simply beating time, giving the starts or indicating *piano* and *forte*. The Conductor must encourage, motivate, explicate how to sing, provide feedback, accompany the singers, and let them feel and enjoy the music they are singing. These goals of conduction may be fulfilled in different ways, both in a quantitative and qualitative sense, by different conductors. I have analyzed some fragments of a concert using an annotation scheme of the choir conductor's multimodal behavior. This tool allows us to characterize the idiosyncratic body style of each conductor and to see which of those goals are more often or more deeply fulfilled by each conductor. In future works the scheme could be applied to the analysis of other choir conductors to see how different conducting body styles relate to different variables such as conductor's personality, social situation (rehearsal vs. concert vs. music class), and type of music (e.g., jam-session vs. renaissance madrigals).

Acknowledgements

This research is partially supported by SSPNet (European NoE n. 231287). Moreover, I am indebted to M° Alessandro Anniballi and his choir "Orazio Vecchi", as well as to Prof. Raffaele Pozzi who, by organising beautiful Concerts at Roma Tre University, gave me the chance to carry out my research.

References

Boyes Braem, P. and Braem, T. 2004. "Expressive gestures used by classical orchestra conductors." In *Proceedings of the Symposium on the Semantics and Pragmatics of everyday Gestures*, C. Mueller and R. Posner (eds), 127–143. Berlin: Weidler.

Bresin, R. and Battel, G. U. 2000. "Articulation strategies in expressive piano performance. Analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart's sonata in g major (k 545)." *Journal of New Music Research* 29 (3): 211–224.

Caterina, R., Bonfiglioli, L., Baroni, M. and Addessi, A. 2004. "Mimic expression and piano performance." In *Proceedings of the 8th International Conference on Music Perception and Cognition*, M. Baroni, A. R. Addessi, R. Caterina, and M.Costa (eds), 258–261. August 04, 2004, volume cd edition.

Conte, R. and Castelfranchi, C. 1995. *Cognitive and Social Action*. London: University College.

Dahl, S., Bevilacqua, F., Bresin, R., Clayton, M., Leante, L., Poggi, I. and Rasamimanana, N. 2009. "Gesture in performance." In *Musical Gestures. Sound, Movement, and Meaning*, R. I. Godoy and M. Leman (eds), 38–68. London: Routledge.

Dahl, S. and Friberg, A. (2007). "Visual perception of expressiveness in musicians' body movements." *Music Perception* 24 (5): 433–454.

Davidson, J. W. 1994. "What type of information is conveyed in the body movements of solo musician performers?" *Journal of Human Movement Studies* 6: 279 – 301.

Davidson, J. W. 2006. "'She's the one': Multiple functions of body movement in a stage performance by Robbie Williams." In *Music and Gesture*, A. Gritten and E. King (eds), 207–224. Aldershot: Ashgate Publishing Company.

Davis, C. and Ashley, R. 2005. "Jazz as a musical conversation: How do jazz musicians communicate with each other? Performance Matters!" Paper given at the International Conference, Porto, Portugal, 14–17 September 2005.

Duranti, A. and Burrell, K. 2006. "Jazz improvisation: a search for hidden harmony and a unique self." *Ricerche di Psicologia* 3 (27): 71–101.

Friberg, A. 2005. "Home conducting: Control the overall musical expression with gestures." *Proceedings of the 2005 International Computer Music Conference*, 479–482. San Francisco, California: International Computer Music Association.

Hartmann, B., Mancini, M. and Pelachaud, C. 2002. "Formational Parameters and Adaptive prototype instantiation for MPEG-4 compliant gesture synthesis." *Computer Animation* 2002: 111–119.

Heylen, D. K. J. 2005. "Challenges ahead. Head movements and other social acts in conversations." In *Proceedings of Joint Symposium on Virtual Social Agents*, L. Halle, P. Wallis, S. Woods, S. C. Marsella, C. Pelachaud and D. K. J. Heylen (eds), 45–52. The Society for the Study of Intelligence and the Simulation of Behaviour.

Kendon, A. 2002. "Some uses of the head shake." *Gesture* 2 (2): 147–182.

King, E. 2006. "Supporting gestures: Breathing in piano performance." In *Music and Gesture*, A. Gritten and E. King (eds), 142–164. Aldershot: Ashgate Publishing Company.

Poggi, I. 2002. "The lexicon of the conductor's face." In *Language,Vision, and Music. Selected Papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing*, P. McKevitt, S. O' Nuallàin and C. Mulvihill (eds), 271–284. Amsterdam: John Benjamins.

Poggi, I. 2006. "Body and mind in the Pianist's performance." In *Proceedings of the 9th International Conference on Music Perception and Cognition*, M. Baroni, M. R. Addessi, R. Caterina and M. Costa (eds), 1044–1051. Bologna, 22–26 August, 2006.

Poggi, I. 2007a. *Mind, Hands, Face and Body. A Goal and Belief View of Multimodal Communication*. Berlin: Weidler.

Poggi, I. 2007b. "Enthusiasm and its contagion. Nature and function." In *Affective Computing and Intelligent Interaction*, A. Paiva, R. Prada and R. W. Picard (eds), 410–421. Berlin: Springer.

CHAPTER 26

Handjabber

Exploring metaphoric gesture and non-verbal communication via an interactive art installation

Ellen Campana, Jessica Mumford, Cristóbal Martínez, Stjepan Rajko, Todd Ingalls, Lisa Tolentino and Harvey Thornburg

We describe an immersive art installation, *Handjabber*, which is inspired by research in gesture and nonverbal communication. It explores how people use their bodies to communicate and collaborate, specifically via metaphoric gesture, interpersonal space, and orientation. In the piece, participants' individual and collective actions give rise to immediate changes in their perceptual environment. These changes are designed to highlight communicative aspects of experience that often go unnoticed in everyday life, allowing both participants and observers to gain a deeper understanding of how they naturally use their bodies to communicate. We describe artistic motivations, theoretical inspirations and technical details. We also discuss how people have experienced the piece.

Introduction

Since World War II and the onset of the Cognitive Revolution, artists have sought to understand the connection between the scientific understanding of human perception and the arts (Weschler 2009). Much of the Abstract Impressionism movement can be understood as an exploration of this connection. In the 50's and 60's the artists of the Light and Space movement (e.g. Larry Bell, James Turrell, John McCracken, Peter Alexander, Doug Wheeler, Maria Nordman, DeWain Valentine, and Eric Or) were at the cutting edge of art in this vein. Their works highlighted aspects of perceptual experience, prompting reflection and discussion about the underlying principles of how our perceptual systems and what exists together give rise to what we experience. For instance, Larry Bell has produced a body of cubical structures with surfaces with different reflective properties, interior subdivisions, and windows. Some are small, suspended at torso height on a clear pedestal. Others are massive. The goal of these works is to "cause one to make a considerable effort to discern and so to become conscious of the process of seeing." (Compton 1970). In this way, Bell's art explores the connection between perception and experience.

The Enactive Art Movement is a current international art movement that builds on many of the same ideas (for review see Noë, 2002). It focuses not only on human perception but rather on the entire perception-action loop. Often using as a medium computational technologies like motion-capture, physical sensing, real-time music synthesis and dynamic visual media, the works effectively highlight the connections between action, perception, and experience. The philosopher Alva Noë (2002) has argued that there is a role for art in general, and enactive art in particular, in cognitive science. Art, he argues, allows us to explore the role of consciousness in perceptual experience, a topic that has been elusive for scientific disciplines. Understanding consciousness is essential for bridging laboratory research and everyday experience.

Our art installation, *Handjabber*, is an example of Enactive art. The piece focuses on human-human communication and the role of the body in communication, with the goal of raising consciousness and provoking discussion about these issues among participants and audience members. It uses infrared cameras to sense human movement and produces musical compositions based on these movements. In this respect it is similar to David Rokeby's well-known musical compositions using the Very Nervous System (Rokeby 2000), which map the quality of movement to sound. While there is some aesthetic resemblance between *Handjabber* and these compositions, both the artistic goals and the underlying computations that give rise to sound are different. Our piece focuses on the connection between dyads and how action is linked to communication.

Perhaps the most obvious connection between action and communication is hand gesture. There are many types of hand gesture, including beats, emblems, iconics and metaphorics (McNeill 1992). For *Handjabber*, we focused on metaphoric gesture, a major category of gesture that has been studied by many researchers (Webb 1996a, 1996b; Sweetser 1990; Ishino 2007; Cienki & Müller 2008; McCafferty 2008). Metaphoric gestures are ideal for our purposes because they often seem to go unnoticed in everyday interaction even though they are common. Moreover, people generally have difficulty articulating the meaning of metaphoric gestures verbally, yet the gestures do contain meaning. *Handjabber* highlights this apparent paradox to allow people to contemplate and explore it. Thus, we are less concerned with abstract-referential meanings of some co-speech metaphoric gesture and more concerned with the semantic components that underlie them.

Webb (1996a, 1996b) argues that the meaning of metaphoric gesture is componential. Individual features, such as movement direction, size, hand shape, and location, together form the meaning of a particular gesture within a given culture. Calbris has also studied componential meaning in metaphoric gesture (1990). One of the components that Calbris identified is the direction of the gesture. The direction of gesture can metaphorically refer to time. Gestures forward and away from the body represent the future, gestures into the body and downward represent the present, and gestures to behind the body represent the past. Gestures moving forward and backward can also emphasize the concepts of progression and regression. Moving the hands upward may refer to the size or quantity of something, while moving the hands out and away from each other may describe the width or immensity of something. Although other gesture components such as palm orientation, shape of the hand, and continuity of gesture are also important, we were inspired by the potential links between Calbris' analysis and current research in music perception.

In addition to gesture, there are many other ways that actions can communicate, including interpersonal space (e.g. Hall 1963, Watson & Graves 1966), orientation (e.g. Remland, Jones, & Brinkman 1995; Wellens & Goldberg 1978), gaze (e.g. Thompson, Emmorey, & Kluender 2006; Carlisle & Levin 2007), goal-directed action (e.g. Berger 2002; Falck-Ytter, Gredebäck & von Hofsten 2006; see also Wood, Glynn, Phillips, & Hauser 2007), posture (e.g. Bull 1987), facial expression (e.g. Ekman 1975; DePaulo et al. 2003), and even fidgeting (e.g. DePaulo et al. 2003). In *Handjabber* we chose to highlight just two of these: (a) interpersonal space and (b) orientation. Both are fundamental to human interaction and well studied from a variety of disciplinary perspectives.

Interpersonal space is defined most simply as the physical distance between people (Knapp & Hall 2007). It can be broadly subcategorized as intimate, casual, social or public. Both the category boundaries and typical variation within them are heavily influenced by culture. For instance, for Asian cultures the boundary between intimate and casual space is much closer to the body than for Western cultures. Interpersonal space can fluctuate based on noise in the environment, what task people are doing, how familiar they are to each other, their emotional states, and how they are feeling about each other or the interaction. In the piece we highlighted interpersonal space to help people contemplate and discuss how it affects communication. For instance, participants might discover that changes in distance change how they interpret the actions or speech of others. Participants from different cultures might discover that they have different boundaries between intimate and casual space.

Actions can also provide information about how engaged people are in an interaction. When people are deeply engaged in a conversation they tend to nod their heads, backchannel, and mirror each other's posture and gestures. People also tend to make eye contact and follow each other's gaze when they are deeply engaged, and look away when they are not engaged or they are but they want to disengage (Goodwin 1980). This can be intentional, as when people avoid eye contact of salespeople, advocates, and pollsters to avoid initiating unwanted conversation. It can also be unintentional, even unconscious. Sometimes the orientation of the entire body, not just the eyes, can convey information about whether a person is, or wants to be, engaged in an interaction. Handjabber allows people to explore the effects their orientation has on others' feeling of being able to communicate.

Technical details

HandJabber is a participatory installation for two people. The gestures the two participants produce, and aspects of how they interact with each other, have immediate affects on the auditory perceptual environment. The experience takes place in a 10'x10' area of a black box theater that is open on all sides. To participate, two people don t-shirts and pairs of wristbands that have reflective markers attached to them (Figure 1). The motion-capture system, a 10-camera IR array operated via the EvaRT 4.6 software, provides the location in 3D space of each marker at 100 Hz. This data is processed in real-time via software we have developed, which computes a body-centered coordinate system for each participant and uses these coordinate systems to extract the features we describe below. These features are linked to audio in real-time via Max/MSP and the Native Instruments software sampler *KONTAKT.* The sounds are played, in mono, through 6 speakers that are directed toward the participants. The details of the sounds are described below.

Sensing and analysis of features

The automatic recognition of gesture information for *Handjabber* draws on a theoretical framework introduced by Rudolph Laban (Laban 1966, 1974) and extended by Irmgard Bartenieff (Hackney 1998, Newlove & Dalby 2003). Laban Movement Analysis (LMA) includes four categories: body, effort, shape, and space. We focus on analysis of shape qualities, a subcategory of shape. Shape, in general, describes the forms taken by the body and how they change over time. Within the Laban framework, shape is often critical for understanding meaningful movement. Shape qualities, more specifically, describe the dynamics of the body as it moves with intent.

LMA includes six shape qualities: rising, sinking, advancing, retreating, spreading, and enclosing, which describe the general folding and unfolding of the body

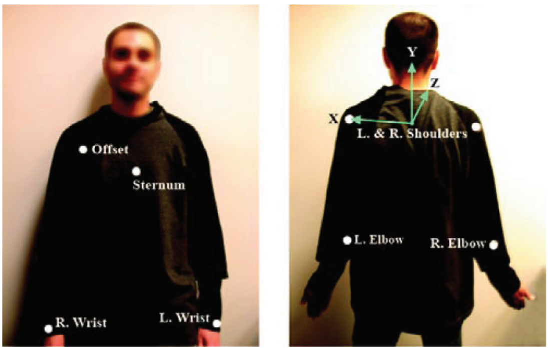


Figure 1. Reflective markers for motion capture.

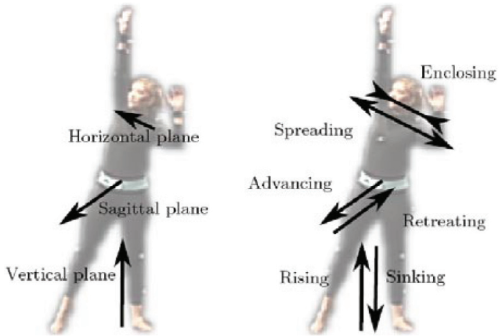


Figure 2. Laban Shape Qualities.

(Figure 2). In prior work, we have developed software for recognizing these shape qualities in real time from motion capture data (Swaminathan et al. 2009), and we have previously used art installation/performance contexts involving full-body movement of a single experienced dancer (Ingalls et al. 2007). For *Handjabber*, we adapted the software to provide features that relate specifically to meaningful components of metaphoric gesture: direction and size. To represent direction we used the rising, sinking, advancing, and retreating shape qualities. To represent size we used the spreading and enclosing shape qualities.

To represent interpersonal space, *Handjabber* uses the distance between the sternum markers of the two participants. We chose the sternum marker rather than the origin to calculate distance so that if participants leaned forward during interaction it would have the same effect as if they stepped closer. Both of these actions are likely to similarly affect participants' sense of interpersonal space, although they might also bear different meanings in everyday interaction.

To represent body orientation *Handjabber* uses one feature for each of the two participants, indicating the degree to which he is facing the other. The calculation of this feature is based on the cosine of the angle between the line between the origin of his coordinate system and the origin of the other's coordinate system and the line that extends forward from his sternum (the Z-axis). It takes the value of 1 if the he is facing the other and –1 if he is facing away (Figure 3).

Auditory perceptual environment

Music communicates abstract meaning that is semantically related to our physical experience (Meyer 1974, 1961; Eitan & Granot 2006). When listening to a performance of Sergei Prokofiev's "Peter and the Wolf" high and rapidly changing notes are immediately understood to represent a bird, while low base notes with a sinister tonal quality represent the wolf. These semantic connections typically do not need to be explained to listeners from a Western musical tradition. Based on the relationships that

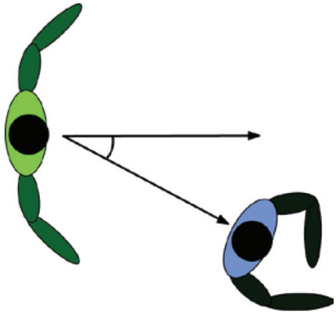


Figure 3. The facing feature.

have been identified in the literature, and connections and some of the authors' prior experience as composers, we created an algorithmic musical score to reflect and highlight the abstract meaning that is communicated through gesture, assuming a Western musical tradition.

In *Handjabber*, hand movements give rise to musical tones and speech samples, organized in a compositional framework, or score. The tones are selected from a library of computer-generated tones, which we pre-classified according to their abstract meaning. Movement in a particular direction triggers sounds that are semantically related to movement in that direction. For instance, rising will call up random notes that are located in the upper register of the C major scale, while sinking will trigger random notes in the lower register of C major. In this way, the two participants each control one synthesized "instrument." Random human voice samples are also interspersed into the score, with a different human voice for each participant. The human speech helps to distinguish the two "voices" and to focus the experience on issues of communication. Speech sounds are treated as notes within the score. The score prevents sounds from overlapping and restricts the rate of sampling, keeping the audio relatively simple.

The gesture size feature affects the timbre of the instruments, controlling a one-pole low-pass filter and a reverberation effect. When the gesture is large, both the cutoff frequency of the low-pass filter and the reverberation room size increase, giving the participant's "voice" a sense of physical volume. Gesture size also affects the ratio of speech samples to tones. When the gesture is large, speech samples have a higher probability of being played, suggesting an increased intent to communicate

Interpersonal space and orientation both relate to the connection between the two participants. In our experience, interpersonal space and distance have the greatest impact on connection when there is also an intention to communicate. In *Handjabber*, both of these features are used to modify aspects of the musical "conversation" that is driven by the two participants' hand gestures.

Interpersonal distance controls the clarity of all of the audio. When the two participants are close together, their music conversation is clear. If they move farther apart, the relationship between their music parts become blurred and hard to discern. This is implemented through convolving all of the gesture-driven audio with an elongated cello note sample and linking the ratio of wet-to-dry signal to the distance feature.

Orientation controls the quality of the audio. Each participant affects the others' "voice" in the conversation. When the two participants face each other, both of their "voices" are equally high in quality. However, if one person tries to communicate while the other is not facing him, or one person turns away while the other is trying to communicate, the voice will be distorted or silent. This is implemented by linking the bit rate and sample rate of each participant's audio to the other's orientation feature.

Participant experience

Over 100 people have participated in *Handjabber* through local exhibitions (joint and solo), workshops, university classes and demonstrations. The participants spanned ages (3–60 years) and levels of movement understanding (trained dancers and those with no movement training). Some pairs of participants knew each other, others did not. The feedback we have received has generally been enthusiastic and positive; participants seem to enjoy the experience and it prompted spontaneous discussion, among participants and audiences, about communication more generally. In these respects, *Handjabber* was a success. It was also successful at deepening our own understanding.

As participants prepared for the interaction, we told them which features affected the sounds they would hear. They were asked to simply play and explore with their partner. Movements were not constrained and participants were allowed to talk to one another. Often an audience would informally gather along the edges of the space. Overall, participants' movements were large and exaggerated, almost theatrical, regardless of whether an audience was present.

Even though they knew which features were sensed and used, participants frequently worked with each other to explore other aspects of nonverbal communication beyond interpersonal space and orientation. Participants often performed the same movements either simultaneously (mirroring) or separated in time (mimicry). Interestingly, we observed partial mirroring and mimicry, in which participants matched specific aspects of their partner's movement (e.g. speed, use of repetition, direction, emotional intent). For instance, two people might take turns making "happy" movements, which had few physical similarities.

Many of the participants explored how changes in height changes and/or verticality convey dominance and territoriality. We often observed a dynamic in which one participant hovered over another with arms raised, while the other crouched close to the ground in fetal position. Another common dynamic was one participant forcing the other to the edges of the space and using nonverbal communication to prevent them from re-entering the center of the space. These explorations of dominance and territoriality are consonant with how people interact in some natural environments (Hall, Coats & LeBeau 2005; Knapp & Hall 2007).

Critique

In general the work accomplished the goal of raising consciousness of and provoking discussion about the role of the body in communication. There was not as much exploration of the metaphoric gestural \aspects of the interaction as we had hoped. One potential reason was that participants found it difficult to distinguish their own "voice" from the voice of their partner. This was verified with more systematic pilot testing in which 10 pairs of participants were asked to identify replayed sections of their own "voices" after the interaction. They could not do this above chance, although they were able to properly identify the effects of both distance and facing. This highlights an important difference between our artistic context and the context of real gesture use. When people use gesture during communication, it forms part of a unified multimodal message (McNeill 1992). To gain a deeper understanding of metaphoric gesture, therefore, participants may need more time to understand the connection between their actions and the auditory perceptual environment in isolation.

When participants did explore the gestural aspects, they made whole body movements (e.g. a confident stride, crumpling into a ball). When the effect was speech, rather than tones, (e.g. a booming "yes", a whimpering "ohnoohnoohno") participants' smiled, and later confirmed that the speech was what they were aiming for. Thus, even when participants did explore gesture, it took on a more iconic character rather than the more subtle metaphoric connection we sought to highlight.

Discussion and future work

Our experience of developing, installing and observing participants interact with *Handjabber* suggests that the enactive art that Noë envisions – an art that is profoundly, and reciprocally, connected to cognitive science – is possible. *Handjabber* brought some implicit aspects of communicative behavior to the realm of conscious reflection. We also observed new patterns of behavior that were not "build into" the interaction, and suggest fruitful areas for future research. For instance, partial mimicry and mirroring are underexplored. Do participants perceive the similarity we observed? What constraints might govern this perception (e.g. semantic similarity, spatial transformations)?

Our experience with *Handjabber* also highlighted the gaps between current enactive arts practice and the vision that Noë champions. Although we were able to make some interesting informal observations, we cannot form and test theoretical predictions about what factors might have given rise to particular dynamics (e.g. communicative intent, learning, audience participation), or compare results to those found in other similar contexts. The videos simply do not contain enough contextual information.

In order to make Noë's vision a reality, there are two critical challenges that must be addressed: (1) recording elements of the computational context in synchrony with video in a way that supports quantitative analysis and generalization and (2) developing methods for tracking aspects of participants' experience (e.g. intent, learning) without interrupting the experience. Our future work will address both of these challenges.

References

Berger, C. R. 2002. "Goals in knowledge structures in social interaction." In *Handbook of Interpersonal Communication*, M. L. Knapp and J. A. Daly (eds), 181–212. Thousand Oaks, CA: Sage Publications.

Bull, P. E. 1987. *Posture and Gesture*. *International Series in Experimental Social Psychology, Vol. 16*. Elmsford NY: Pergamon Press.

Calbris, G. 1990. *The Semiotics of French Gestures.* (O. Doyle, Translation). Bloomington, Indianapolis: Indiana University Press.

Carlisle, N. and Levin, D. 2007. "Predicting human action from gaze." *Journal of Vision*, 7 (9): 659, 659a,

Compton, M. 1970. Catalog essay from *Larry Bell, Robert Irwin, Doug Wheeler* exhibition at Tate Gallery, London.

Cienki, A. J. and Müller, C. (eds.) 2008. *Metaphor and Gesture.* Amsterdam: John Benjamins.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K. and Cooper, H. 2003. "Cues to deception." *Psychological Bulletin* 129: 74–118.

Eitan, Z. G. and Granot, R. Y. 2006. "How music moves: A study of the relation between musical parameters and visuo-kinetic associations." *Music Perception* 23 (3): 221–248.

Ekman, P. 2003. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Cues.* Originally published 1975 by Prentice Hall, New Jersey; reprint edition, Cambridge MA: Malor Books.

Falck-Ytter, T., Gredebäck, G. and von Hofsten, C. 2006. "Infants predict other people's action goals." *Nature Neuroscience* 9: 878–879.

Goodwin, C. 1980 "Restarts, pauses, and the achievement of state of mutual gaze at turn-beginning." *Sociological Inquiry* 50 (3–4): 272–302.

Hackney, P. 1998. *Making Connections: Total Body Integration through Bartenieff Fundamentals*. Routledge, London, UK.

Hall, E. T. 1963. "A system for the notation of proxemic behavior." *American Anthropologist*, New Series, 65 (5): 1003–1026.

Hall, J. A., Coats J. E., and LeBeau, L. S. 2005. "Nonverbal behaviors and the vertical dimensions of social relations: A Meta-Analysis." *Psychological Bulletin*. 131 (6): 898–924.

Ingalls, T., James, J., Qian, G., Thornburg, H., Campana, E., Rajko, S., Swaminathan, D., Mumford, J., Afanador, K. and Narayanan, S. 2007. "Response: An applied example of computational somatics." *4th International Conference on Enactive Interfaces.*

Ishino, M. 2007. *Metaphor and Metonymy in Gesture and Discourse*. Unpublished Ph.D. dissertation, University of Chicago.

Knapp, M. L. and Hall, J. A. 2007. *Nonverbal Communication in Human Interaction (5th ed.)*. Belmont, CA: Wadsworth/Thomas Learning.

Laban, R. 1966. *Choreutics*. London: Macdonald and Evans.

Laban, R. 1974. *The Language of Movement: A Guidebook to Choreutics*. Boston, MA: Plays.

McCafferty, S. G. 2008. "Mimesis and second language acquisition: A sociocultural perspective." *Studies in Second Language Acquisition* 30 (2): 147–167.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Meyer, L. B. 1961. *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

Meyer, L. B. 1974. *Music, the Arts, and Ideas: Patterns and Predictions in Twentieth-Century Culture (4th Edition)*. Chicago, IL: University of Chicago Press.

Newlove, J. and Dalby, J. 2003. *Laban for All*. London: Nick Hern Books Limited.

Noë, A. 2002. "Art as enaction." Paper presented at Art and Cognition virtual conference. Accessed April 11, 2010 at http://www.interdisciplines.org/artcog/papers/8.

Remland, M. S., Jones, T. S. and Brinkman, H. 1995. "Interpersonal distance, body orientation, and touch: effects of culture, gender, and age." *The Journal of Social Psychology*: 135 (3) 281–297.

Rokeby, D. 2000. *Installations: Very Nervous System (1986–1990)*. Accessed April 11, 2010 at http://homepage.mac.com/davidrokeby/vns.html.

Sweetser, E. E. 1990. *From Etymology to Pragmatics: The Mind-as-Body Metaphor in Semantic Structure and Semantic Change* Cambridge: Cambridge University Press.

Swaminathan, D., Thornburg, H., Mumford, J., Rajko, S., James, J., Ingalls T., Campana, E., Qian, G., Sampath, P. and Peng, B. 2009. "A dynamic Bayesian approach to computational Laban shape quality analysis." *Advances in Human-Computer Interaction*, Hindawi Publishing Co. doi:10.1155/2009/362651.

Thompson, R., Emmorey, K. and Kluender, R. 2006. "The relationship between eye gaze and verb agreement in American Sign Language: An eye-tracking study." *Natural Language & Linguistic Theory* 24 (2): 571–604.

Watson, O. M. and Graves, T. D. 1966. "Quantitative research in proxemic behavior." *American Anthropologist*, New Series, 68 (4): 971–985.

Webb, R. 1996a. "Linguistic features of metaphoric gestures." In *Proceedings of the Workshop on the Integration of Gesture in Language and Speech*, Lynn Messing (ed.), 79–93. Newark: University of Delaware.

Webb, R. 1996b. *Linguistic Features of Metaphoric Gestures*. Unpublished Doctoral Dissertation, University of Rochester.

Weschler, L. 2009. *Seeing is Forgetting the Name of the Thing One Sees: Expanded Edition: Over Thirty Years of Conversation with Robert Irwin*. Berkeley, CA: University of California Press.

Wellens, A. R. and Goldberg, M. L. 1978. "The effect of interpersonal distance and orientation upon the perception of social relationships." *The Journal of Psychology* 99: 39–47.

Wood, J. N., Glynn, D. D., Phillips, B. C. and Hauser, M. D. "The perception of rational, goal-directed action in nonhuman primates." *Science* 317 (5843): 1402–1405.

Name index

Subject index

Gestures are ubiquitous and natural in our everyday life. They convey information about culture, discourse, thought, intentionality, emotion, intersubjectivity, cognition, and first and second language acquisition. Additionally, they are used by non-human primates to communicate with their peers and with humans. Consequently, the modern field of gesture studies has attracted researchers from a number of different disciplines such as anthropology, cognitive science, communication, neuroscience, psycholinguistics, primatology, psychology, robotics, sociology and semiotics. This volume presents an overview of the depth and breadth of current research in gesture. Its focus is on the interdisciplinary nature of gesture. The twenty-six chapters included in the volume are divided into six sections or themes: the nature and functions of gesture, first language development and gesture, second language effects on gesture, gesture in the classroom and in problem solving, gesture aspects of discourse and interaction, and gestural analysis of music and dance.

"The present volume gives an impressive survey of the kinds and functions of gestures occurring in humans and other primates and introduces the reader into the leading paradigms of contemporary gesture research. The contributors include prominent gesture researchers as well as promising young professionals with an interdisciplinary background and exemplify the successful international cooperation taking place in this fascinating field. The volume is of particular value for readers interested in first and second language development, social cognition, and problem-solving by means of gestures."

Roland Posner, *Honorary President of the International Association for Semiotic Studies IASS, Berlin Institute of Technology*

"This outstanding volume presents a vast overview of contemporary research on gesture, covering multiple disciplines and different theoretical and methodological perspectives. It demonstrates the breadth and sophistication of studies that examine visible bodily actions and their intricate relationship to communication and cognition. A treasure trove of observations concerning forms and functions of gestures, their role in development, interaction, problem-solving, and even music-making, it's a volume to return to again and again. Essential reading for all interested in the nature and function of gestures!"

Marianne Gullberg, *Lund University*

ISBN 978 90 272 2845 1

9 789027 228451

John Benjamins Publishing Company