

# Children's Gestures from 18 to 30 months

Mats Andréén  
Centre for Languages and Literature  
Centre for Cognitive Semiotics  
Lund University

---

*Dedicated to Evelhin and Alice.*

# Preface

After hours and hours (and hours) of watching video recordings of interaction between children and parents, the signification of bodies still remains mysterious to me in many ways. I continue to be amazed by how acts and movements simultaneously appear deeply familiar and “typical” on the one hand, open-ended in their meaning and slippery when it comes to their specification on the other. The recognition of certain movements as meaningful, and somehow structured, obviously requires the intuitions of a “recognizer”. This holds not only for the participants in the interaction, but also for me, the researcher. I have been thoroughly delighted for having had the opportunity to experience first-hand how the study of visible action also points inward, to my own ways of seeing such things in the mundane activities of daily life — the world otherwise taken for granted — as well as transforming this vision. Writing this thesis has been a wonderful challenge. I loved it!

I also got my first gray hairs while writing it.

## Acknowledgments

Jordan Zlatev, my supervisor, is always busy, but don't be fooled! No matter how busy he is, he has always been available for discussing issues and commenting on drafts — sometimes, when time has been short, even in the middle of the night! The scope of Jordan's academic knowledge is remarkably broad, and there is no aspect of the nature of language and communication that escapes his interest — gesture included. I have learned a tremendous amount of interesting things from him. He has also introduced me to other, more concrete, aspects of the academic reality, such as editing books, organizing conferences, starting organizations (SALC), and much more. I am very grateful for all of these things. Jordan even helped carry heavy furniture when my family and I moved to a new house. I think that speaks for itself regarding the massive support I have received. I look forward to continue working with Jordan!

---

I wish to thank the senior researchers involved in the (now completed) SEDSU project (Stages in the Evolution and Development of Sign Use), as well as the current CCS project (Centre for Cognitive Semiotics) and the PSUII project (Intersubjectivity and Imitation as Precursors of Sign Use). Together you have provided an environment for stimulating research. In particular I want to thank Göran Sonesson, who introduced me to the inspiring world of semiotics with depth and clarity, and who has been very helpful in providing the funding required for my various positions in Lund. Among these senior researchers I thank also Peter Gärdenfors, Ingar Brinck, Sven Strömquist, and of course Jordan Zlatev (again). Recently, Marianne Gullberg returned to Lund University, providing a very welcome and substantial addition to the gesture expertise here.

Joel Parthemore did a fantastic job in proof-reading many parts of the thesis in very short time. I am very grateful for that. Unfortunately time did not permit me to take all of his suggestions for changes into account, even though they were all very good. He should therefore not be blamed for the errors that still remain. When it comes to statistical calculations, Joost van de Weijer's advice has been very valuable. Victoria Johanson and Gilbert Ambrazaitis have been very helpful with all sorts of hints, tips, and information when it comes to the process of thesis writing in general. Frida Mårtensson helped me out during the last hectic hours of writing when there were simply too many different things that needed to be done in too little time. The staff at the Humanities Laboratory has been incredibly helpful in many ways.

When I started working on this thesis, I had immediate access to substantial amounts of video data that were already collected, transcribed, and systematized in various ways. For this, I am greatly indebted to Ulla Richthoff, Sven Strömquist, Åsa Wikström, Tom Sköld, and Jordan Zlatev. I hope this thesis does justice to their commendable idealism regarding the sharing of useful research data.

I would like to direct a personal thank you to Ulrika Nettelbladt, who has been a constant source of enthusiasm and encouragement throughout my thesis work. Three persons from my time at Linköping University have had a strong influence on my thinking: Per Linell, Richard Hirsch, and Oskar Lindwall.

In addition to all my colleagues at the Centre for Languages and Literature — it would simply be too much to list all of you here — I would like to thank a number of other people I have been in touch with during the past five years. Your assistance has included everything from collaborations, discussions, passing conversations, comments on manuscripts, answering questions, socializing at conferences and seminars, and so on:

---

Jens Allwood, Janet Bavelas, Paul Bouissac, Jana Bressem, Ingar Brinck, Mathias Broth, Anders Carlsson, Johan Cedervall, Alan Cienki, Herbert H. Clark, Stephen Cowley, Carla Cristilli, Jakob Cromdal, Nils Dahlbäck, Nevena Dimitrova, Merlin Donald, Anna Ekström, Margaret Fleck, Ellen Fricke, Maria Fusaro, Riccardo Fusaroli, Shaun Gallagher, Marie Gelang, Tove Gerholm, Jennifer Gerwing, Susan Goldin-Meadow, Charles Goodwin, Maria Graziano, Gerlind Große, Marianne Gullberg, Peter Gärdenfors, Simon Harrison, Lars-Åke Henningson, Richard Hirsch, Jessica Hobson, Peter Hobson, Judith Holler, Nils Hansson, Nils Holmberg, Kenneth Holmqvist, Jana Holsanova, Sara Howard, Gisela Håkansson, Jonas Ivarsson, Victoria Johansson, Ingrid Johnsrude, Karin Junefelt, Arne Jönsson, Petr Kaderka, Anna A. Kaal, Sari Karjalainen, Adam Kendon, Sotaro Kita, Timothy Koschmann, Anna Kuhlen, Silva Ladewig, Sara Lenninger, Patrik Lilja, Paulina Lindström, Oskar Lindwall, Jan Lindström, Per Linell, Matt Longo, Brian MacWhinney, Elaine Alenkær Madsen, David McNeill, Irene Mittelberg, Richard Moore, Aliyah Morgenstern, Christiane Moro, Cornelia Müller, Ann-Christin Månsson, Ulrika Nettelbladt, Iris Nomikou, Jesper Olsson, Mathias Osvath, Cajsa Ottesjö, Ezequiel di Paolo, Michael Perkins, Tomas Persson, Simone Pika, Elena Pizzuto, Juana Salas Poblete, Roland Posner, John P. Rae, Anna Cabak Rédei, Cintia Rodríguez, Yvan Rose, Caroline Rossi, Hans Rystedt, Kazuki Sekine, Chris Sinha, Göran Sonesson, Laura Sparaci, Gale Stam, Oliver StJohn, Jürgen Streeck, Sven Strömqvist, Susanne Tag, Michael Tomasello, Mark Tutton, Kristian Tylén, Liesbet Quaeghebeur, Virginia Volterra, Dan Zahavi, Patricia Zukow-Goldring, Şeyda Özçalışkan, and those that I have forgotten.

I would like to make a special mention of Adam Kendon, whose outstanding work on gesture was the first I came across when I first started reading about gesture research a few years ago. It immediately made me helplessly interested in the nature of gesture. Approaching gestures in his particular ways, Kendon has produced an astonishing number of valuable analyses of gesture. I agree with Heath (1992, p. 705) when he writes that “despite Kendon’s generosity in attributing his methodological framework to the work of such scholars as Goffman, Schefflen, Birdwhistell, and others, his research reveals a unique analytic orientation to the study of social interaction”. Without Kendon’s work and publications, the field of gesture studies would certainly not exist in the form it has today. In addition, I would like to take this occasion to thank Kendon for his encouraging comments on the occasions when we have met, for his constructive remarks on several of my writings, and not

---

least for agreeing to be the opponent at the defense of this thesis.

Finally, I express my deep gratitude to my parents Christina and Per for all their love and support, both in recent times and in the past; and to my brother Johan and my two sisters Anna and Karin for always being there for me as well. To my daughter Alice and my wife Evelhin: nothing could be easier than loving the two of you!

# Contents

<b>I</b>	<b>Setting the stage</b>	<b>1</b>
<b>1</b>	<b>The aim of this thesis</b>	<b>3</b>
1.1	Formulation of the overall agenda . . . . .	3
1.2	Why this aim? . . . . .	5
1.3	Roadmap . . . . .	8
<b>2</b>	<b>”What is gesture?” and other conceptual preliminaries</b>	<b>11</b>
2.1	The idea of an upper and lower limit of gesture . . . . .	11
2.2	Unpacking the lower limit . . . . .	15
2.2.1	Intentionality . . . . .	17
	Aboutness and intent . . . . .	17
	First-, second-, and third-person intentionality . . . . .	19
	Public intentionality . . . . .	20
	Intentionality within the natural and the scientific attitude	22
2.2.2	Levels of communicative explicitness . . . . .	23
	Comm#1 — Communication as a side-effect of co-presence	24
	Comm#2 — Action framed by mutual attunement . . . . .	25
	Comm#3 — Visibly other-oriented action . . . . .	26
	Comm#4 — Reciprocated action . . . . .	27
2.2.3	Levels of semiotic complexity . . . . .	30
	Sem#1 — Situation-specific aspects of action . . . . .	30
	Sem#2 — Typified aspects of action ( <i>count-as</i> ) . . . . .	31
	Sem#3 — Differentiated aspects of action ( <i>semiotic signs</i> )	34
2.2.4	Children and the lower limit . . . . .	42
2.2.5	Defining gesture in the context of the lower limit . . . . .	42
2.3	Unpacking the upper limit . . . . .	44
2.3.1	The status of conventionalized gestures . . . . .	46
2.3.2	From imitation to imitative processes . . . . .	51
2.3.3	Levels of conventionalization . . . . .	54

	Conv#1 — Normality . . . . .	54
	Conv#2 — Typification . . . . .	55
	Conv#3 — Normativity . . . . .	57
	Conv#4 — Explicated rules . . . . .	60
	The depersonalized character of conventionality . . . . .	61
2.3.4	Combinations . . . . .	62
2.4	Putting the pieces back together again . . . . .	67
<b>3</b>	<b>Action gestalts, gesture, and intersubjectivity</b>	<b>71</b>
3.1	Structures of action . . . . .	71
3.2	Action, intersubjectivity and the world within reach . . . . .	73
3.3	Manual excursions and Action Gestalts . . . . .	79
3.4	Operation-act-activity interdependence . . . . .	86
3.5	Summary . . . . .	91
<b>4</b>	<b>Data, annotation, and other methodological issues</b>	<b>93</b>
4.1	Data . . . . .	93
4.2	Transcription of speech . . . . .	95
4.3	Gesture annotation . . . . .	96
4.3.1	Annotation as indexing and as coding . . . . .	96
4.3.2	Main annotation categories . . . . .	96
	Deictic . . . . .	99
	Iconic . . . . .	100
	Conventionalized . . . . .	102
	Other annotations . . . . .	103
4.3.3	Comparison with other classification systems . . . . .	103
4.4	Rate measures . . . . .	105
4.4.1	Gestures per Minute (GPM) . . . . .	105
4.4.2	Gestures per multimodal Utterance (GPU) . . . . .	105
4.4.3	Mean Length of Utterance (MLU) . . . . .	106
4.5	Presentation of examples and data . . . . .	107
4.5.1	Visual presentation of examples . . . . .	107
4.5.2	Notation for gesture+speech utterances in text . . . . .	110
4.5.3	Graphs: Averaging of data from several children . . . . .	111



<b>II</b>	<b>Analysis</b>	<b>115</b>
<b>5</b>	<b>Developmental patterns and transitions</b>	<b>117</b>
5.1	Questions asked and contents of the chapter . . . . .	117
5.2	Transition periods and terminology . . . . .	119
5.3	GPM, GPU, and MLU . . . . .	120
5.3.1	GPM, GPU, and MLU for each child . . . . .	120
5.3.2	GPM, GPU, and MLU as a function of age . . . . .	122
5.3.3	Summary . . . . .	125
5.4	Gesture with and without speech . . . . .	126
5.4.1	Gesture only, gesture+speech, and speech only utterances	126
5.4.2	Summary . . . . .	128
5.5	Multi-gesture and multi-word utterances . . . . .	128
5.5.1	Multi-word combinations . . . . .	129
5.5.2	Multi-gesture combinations . . . . .	131
5.5.3	Relations between multi-word and multi-gesture combinations . . . . .	133
5.5.4	Summary . . . . .	135
5.6	Deictic, iconic, and conventionalized aspects . . . . .	136
5.6.1	An overview of the semiotic aspects . . . . .	137
5.6.2	A longitudinal comparison between the semiotic aspects .	138
5.6.3	Sub-types of deictic aspects . . . . .	142
5.6.4	Sub-types of iconic aspects . . . . .	144
5.6.5	Sub-types of conventionalized aspects . . . . .	146
5.6.6	Semiotic aspects with and without speech . . . . .	147
5.6.7	Summary . . . . .	150
5.7	Gestures that involve handling of objects . . . . .	152
5.7.1	Object-gestures versus empty-handed gestures . . . . .	154
5.7.2	Deictic, iconic, and conventionalized aspects in object-gestures . . . . .	154
5.7.3	Summary . . . . .	157
5.8	Summary and conclusions . . . . .	158
5.8.1	Overall findings . . . . .	158
5.8.2	Three transition periods: A developmental trajectory . . .	162
	Transition Period #1 . . . . .	163
	Transition Period #2 . . . . .	167
	Transition Period #3 . . . . .	170
	The nature of the transitions . . . . .	174

5.8.3	Findings relating to the upper and lower limit of gesture . . . . .	175
<b>6</b>	<b>Deictic aspects of the children's gestures</b>	<b>179</b>
6.1	Questions asked . . . . .	179
6.2	What is a "deictic" gesture really? . . . . .	182
6.3	Observations on the use of index finger pointing . . . . .	184
6.4	Pointing gestures that involve TRACING . . . . .	192
6.5	Whole hand pointing . . . . .	198
6.6	Directedness in emblems and iconic gestures . . . . .	202
6.7	Object-gesture variants of pointing . . . . .	205
6.8	Discussion and conclusions . . . . .	208
<b>7</b>	<b>Iconic aspects of the children's gestures</b>	<b>213</b>
7.1	Questions asked . . . . .	213
7.2	What provides for the "transparency" of iconic gestures? . . . . .	214
7.2.1	Different views on the nature of iconicity in gesture . . . . .	214
7.2.2	Mirror-neurons: A natural or convention-based affair? . . . . .	216
7.2.3	Different forms of iconicity . . . . .	219
7.2.4	Children and the transparency of iconic gestures . . . . .	224
7.3	Gestures with ACTION_BASED iconic aspects . . . . .	225
7.3.1	Transparency in ACTION_BASED iconic gestures . . . . .	225
7.3.2	The continuity between typified actions and typified gestures . . . . .	229
7.3.3	ACTION_BASED gestures from an alter-centric perspective . . . . .	232
7.4	Gestures with INDIRECT iconic aspects . . . . .	235
7.4.1	From INDIRECT object-gestures to INDIRECT empty-handed gestures . . . . .	235
7.4.2	INDIRECT gestures and imitation . . . . .	238
7.4.3	ACTION_BASED and INDIRECT aspects in the same gesture . . . . .	239
7.5	Conclusions . . . . .	240
<b>8</b>	<b>Conventionalized aspects of the children's gestures</b>	<b>245</b>
8.1	Questions asked . . . . .	245
8.2	Head-shakes and nodding . . . . .	247
8.2.1	Introduction . . . . .	247
8.2.2	When do the head-gestures emerge? . . . . .	250
8.2.3	Nodding and head-shakes as back-channeling? . . . . .	251
8.2.4	From responses to initiatives . . . . .	252

8.2.5	Head-gestures in combination with other gestures . . . . .	256
8.3	Multimodal constructions . . . . .	258
8.3.1	Introduction . . . . .	258
8.3.2	Item-based multimodal constructions . . . . .	261
	Harry's and Tea's SCOLD_POINT . . . . .	261
	Bella's GONE gesture . . . . .	262
	The HELLO gesture in four children . . . . .	262
	Bella's BYE_BYE gesture . . . . .	264
	Discussion of item-based multimodal constructions . . . . .	264
8.3.3	Flexible multimodal constructions . . . . .	267
	Another possible interpretation of gesture in the transi- tion from one-word speech to two-word speech	268
8.4	Conclusions . . . . .	269
<b>9</b>	<b>Communicative action gestalts in the manipulatory area</b>	<b>271</b>
9.1	Questions asked . . . . .	271
9.2	Reaching towards the object . . . . .	278
	9.2.1 Definition . . . . .	278
	9.2.2 Analysis . . . . .	279
9.3	Touching or grabbing the object . . . . .	282
	9.3.1 Definition . . . . .	282
	9.3.2 Analysis . . . . .	283
9.4	Handling the object in center space . . . . .	290
	9.4.1 Definition . . . . .	290
	9.4.2 Analysis . . . . .	291
9.5	Moving the object towards a target . . . . .	295
	9.5.1 Definition . . . . .	295
	9.5.2 Analysis . . . . .	296
9.6	Putting the object down . . . . .	305
	9.6.1 Definition . . . . .	305
	9.6.2 Analysis . . . . .	305
9.7	Withdrawing from the object . . . . .	311
	9.7.1 Definition . . . . .	311
	9.7.2 Analysis . . . . .	312
9.8	Summary and conclusions . . . . .	315
	9.8.1 Ongoing vs. stopped movement . . . . .	317
	9.8.2 Coordination with speech . . . . .	319
	9.8.3 Generic and specific aspects of action . . . . .	324

<b>III Rounding up</b>	<b>329</b>
<b>10 Summary and conclusions</b>	<b>331</b>
10.1 What sorts of gestural actions do the children perform? . . . . .	331
10.2 What are the changes in the gestural repertoire over time, as the children grow older? . . . . .	335
10.3 The lower limit of gesture: action and gesture . . . . .	337
10.4 The upper limit of gesture: gesture and language . . . . .	338
10.5 Final words . . . . .	340
<b>References</b>	<b>341</b>

# Part I

## Setting the stage



# CHAPTER 1

## The aim of this thesis

*There is a progression and change in the use of these gestures over time so that we cannot speak of gestures as a whole.*

Erting & Volterra (1990, p. 299)

### 1.1 Formulation of the overall agenda

In a nutshell, this thesis aims at providing an account of the character of children's gestures during the period between 18 and 30 months. "Gesture" is here taken to mean, as a first approximation, those instances when the body works as an expressive medium through movement. The analysis is based on video recordings of five hearing-enabled Swedish children as they interact with their parents. Each child is recorded approximately once a month throughout the study period. The approach involves empirical analysis, in the form of detailed descriptions of particular situated occurrences and generalizing quantifications. It involves as well conceptual analysis, in the form of investigations of central concepts such as "gesture", "communicative action", "instrumental action" (i.e., practical action), "signs", "conventionality", and "intentionality", along with the relations between them. Rather than identifying the term "gesture" with a single set of essential properties, a comparative semiotic approach is taken (e.g. Kendon, 2004, 2008), which serves to pinpoint and highlight both differences and similarities between different kinds of gestural performances. This leads to a conception of various forms of gesture as a matter of family resemblance (Wittgenstein, 1953). Such a conception, one could argue, is not a vaguer understanding of gesture than more unitary and essentialist conceptions of the nature of gesture, but to the contrary, a more precise one.

This broad account of the character of children's gestures during the period when language starts to "take off" should be of interest, not only to researchers within the field of gesture studies<sup>1</sup> but also, hopefully, to researchers from other fields for which the character of children's gestures is relevant (e.g. developmental psychology, semiotics, linguistics, conversation/interaction analysis, neuroscience, anthropology, speech and language therapy, ethology, among others).<sup>2</sup> Even though it is impossible to cover all the relevant issues involved in specifying the character of children's gestures, I still believe that it is worthwhile to frame the inquiry in this broad way, in order not to lose sight of the larger picture. Such an approach raises fundamental questions about which actions that are to be included (or not), leading one in turn to scrutinize the central concepts of "gesture", "communicative action", "instrumental action", and so forth. Nevertheless, it is obviously necessary to be selective in the questions addressed and the analyses attempted. Consequently, the approach is an opportunistic one, whereby a number of phenomena of particular relevance are brought up and highlighted. Some of these relate specifically to the nature of children's gesture, and some of them relate to the nature of gesture more generally.

The overarching questions or themes addressed by this thesis can, for the sake of clarity, be divided into those that primarily involve empirical analysis and those that primarily involve conceptual analysis, though the reason for addressing them in the first place is of course because they are relevant, even indispensable, in relation to each other. The main empirical themes addressed are:

- (E.1) *The repertoire*: What sorts of gestural actions do the children perform? What gestures are the children *not* performing?
- (E.2) *Development*: What are the changes in the gestural repertoire over time, as the children grow older?
- (E.3) *Multimodality*: What is the character of the gestures' organization in

---

<sup>1</sup>The field of gesture studies has emerged during the last 20–30 years. Pioneering efforts by researchers such as Adam Kendon, David McNeill, Susan Goldin-Meadow, Virginia Volterra, Jürgen Streeck, Charles Goodwin and others have provided the foundation for a steadily growing interest in gesture. The International Society for Gesture Studies (ISGS) was founded in 2002. Since then a number of ISGS conferences has been organized — Austin 2002, Lyon 2005, Chicago 2007, Frankfurt (an der Oder) 2010. A fifth one is planned for Lund 2012. Since 2001 there has been a journal named *Gesture* (John Benjamins Publishing Company), edited by Adam Kendon.

<sup>2</sup>It is thus a prime target for the emerging field of *Cognitive Semiotics*, which is aimed at "integrating methods and theories developed in the disciplines of cognitive science with methods and theories developed in semiotics and the humanities, with the ultimate aim of providing new insights into the realm of human signification and its manifestation in cultural practices." ([www.cognitivesemiotics.org](http://www.cognitivesemiotics.org))



## 1.2. WHY THIS AIM?

---

concert with other semiotic resources, such as speech and objects in the world-at-hand, as part of a social activity? How do the gestures arrive at their meaning, due to factors such as gestural forms and the contextual embedding of the gesture?

The main conceptual themes addressed all concern the nature, and definitions, of gesture. This is tackled from three main points of view:

- (C.1) *The lower limit of gesture:* This concerns if and how to distinguish between those actions that have features that qualify them as gesture proper, and those actions that lack some or all of these features. This leads naturally on to investigations of the notion of intentionality, different levels of communicative explicitness, and different levels of semiotic complexity.
- (C.2) *The upper limit of gesture:* This concerns if and how to distinguish between those actions that are considered as gesture proper and those actions that are somehow too complex and language-like to be considered gesture. This leads into discussion of different levels of conventionalization, as well as the notions of ritualization, imitation, and “combinations” (i.e., communicative acts with multiple components).
- (C.3) *Factors contributing to the meaning of gestures:* In order to disentangle some of the factors that contribute to gesture meaning, a number of issues are discussed, such as intersubjectivity and interpersonal understanding, gestalt properties of gesture, as well as the respective roles of form and context.

The empirical analysis was carried out with these conceptual questions in mind. At the same time, the formulation of these questions was, to a large extent, the result of grappling with the task of characterizing the gestural repertoires of children empirically.

## 1.2 Why this aim?

As the research questions in the previous section suggest, the analysis is primarily concerned with *what*-type questions. To put this another way, the thesis primarily

concerns explanation in the sense of “make clear”.<sup>3</sup> It is only to a lesser extent concerned with *why*-type questions (reasons and causal mechanisms), although they are discussed where appropriate. One basic motivation for this focus on *what*- rather than *why*-type questions is that I consider descriptive work to be of great value in and of itself. I should also point out that I use “description” in a wide sense here, referring both to empirically grounded specifications of phenomena, on the level of *empirical analysis*, and clarifications and elaborations of the relevant concepts, on the level of *conceptual analysis*. The extent to which understanding may emerge through detailed descriptions of phenomena, along with considerations of the concepts needed to explicate these descriptions, is generally underrated.<sup>4</sup> Descriptive analysis should therefore not be thought of as “mere description”, but rather as a source of understanding and insight in its own right (cf. Wittgenstein, 1953). Notwithstanding this, descriptive analysis is to some extent also a prerequisite to the possibility of asking the right *why*-type questions at a later stage. This need not be taken as the radical position that there must first be a lengthy period of purely descriptive work, that only after this should one be allowed to embark on theory building and hypothesis testing. However, it should be clear that description, in a broad sense, provides the foundation on which any theorizing can be built, and that the value of particular descriptive works may, at least in part, survive “paradigm shifts” at the level of theoretical analysis.

Why study children’s gestures in the age range of 18 to 30 months specifically? Even though research on children’s gestures has gained serious momentum during the last decade or two, most of the work that has been done so far has been concerned with the period *before* the age range studied here. That is, previous research has mainly focused on the period when the first intentional gestures appear, around the end of the first year, when one finds the first pointing gestures and some other social actions such as GIVE and SHOW (e.g. Bates et al., 1979; Volterra & Erting,

---

<sup>3</sup>From the Latin, *explanare* = to ‘make flat’.

<sup>4</sup>It has always seemed to me that achieving consensus on the structure of the *explanandum* is far more pressing than achieving consensus on the vocabulary to use for the *explanans*. For example, there is endless debate in the human sciences whether there is a place for a “mentalistic” vocabulary of explanation or not, somewhat related to the question of whether there is a place for folk-psychological concepts in scientific explanations or not. To me, such debates seem to have surprisingly few consequences for the understanding of the *explanandum* itself, although they are often presented as posing an either/or issue. The question of whether a “mentalistic” or “folk-psychological” vocabulary should be used in descriptions of the explanandum or not is an entirely different issue, and here I would say that both approaches are possible in principle as well as mutually enriching. As long as specifications of the explanandum seems to survive tests of verification, my position is that they should be considered science proper.

## 1.2. WHY THIS AIM?

---

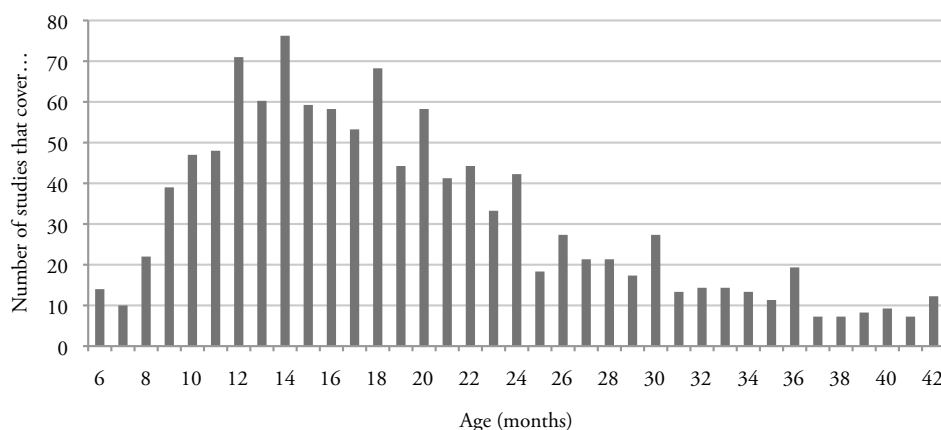


Figure 1.1: Studies of children's gestures ( $n=172$ ) between 6–42 months of age.

1990); up to the point in time when children begin to produce two-word utterances at around 18–20 months (Capirci et al., 1996; Butcher & Goldin-Meadow, 2000; Özçalışkan & Goldin-Meadow, 2005a; Iverson & Goldin-Meadow, 2005). Beyond this point, the number of studies decreases, as illustrated in Figure 1.1. Figure 1.1 shows the age ranges for the 172 studies of children's gestures that I have come across so far involving children between 0.5 and 3.5 years.<sup>5</sup> The majority of research on in this sample of studies concerns the age range of 9 to 24 months (71.6% of the total), even though that is less than half the age range in the figure as a whole. The period studied in this thesis therefore begins at a time which has been studied by a relatively large proportion of researchers — which is useful, for purposes of comparison with other studies — then extends further in time than most research has done so far. Moreover, it should be noted that most of the studies included in the graph have concerned some specific aspect of children's gestures rather than overall characterizations of the repertoire, which means that there are many aspects of the gestures of children between 18 and 30 months that remain largely unexplored (cf.

---

<sup>5</sup>The graph includes studies involving both children's own gestures and gestures directed toward children. Since many studies include children of more than one age, single studies are often represented in several of the bars in the graph. Even though the graph does not represent all the work that has been carried out in this area, it still gives a good idea of the overall situation. Some studies that I have come across are not included in this graph, such as studies on deaf children, children with autism or other special issues. Also excluded are articles that summarize or discuss other research without presenting results from any new study. Regarding biases in the sampling, it may be noted that there are many studies on early pointing in infants around 9–14 months and that I have been slightly less concerned in locating everything that has been written on that age range compared to studies of older children. This means that the true distribution of studies of children's gestures in these ages is likely to be even more uneven, with an even larger proportion of studies focused on younger children, than what is shown in the figure.

Stefanini et al., 2009, p. 169): hence the aim of this thesis, both with respect to the focus on a descriptive approach and the focus on children between 18 and 30 months. One benefit of the data, in addition to covering an age range that has been little studied, is that each child is recorded relatively often. This makes it possible to get a relatively detailed view of various developmental progressions — more so than in many other studies of this age range. Overall, there are also very few studies of gestures with Swedish children. In 1999 Berglund (1999, p. 23) stated that “Systematic research in early play or gestures of Swedish children is, to our knowledge, non-existent” — which was certainly correct at the time. A few more studies have appeared since then, but not many (Allwood & Ahlsén, 1999; Månsson, 2003; Berglund et al., 2005; Gerholm, 2007). Several of these are concerned with children younger than those studied in this thesis.

The period between 18 and 30 months is interesting in many ways. It is during this period that children begin to communicate by means of more complex spoken constructions. Since it is well known that gesture and speech are tightly coordinated in both adults (Kendon, 1980b, 2004; McNeill, 1985, 2005) and children (Volterra & Erting, 1990; Iverson & Goldin-Meadow, 1998), one may wonder what happens with gesture during this period of intense changes in the mode of speaking. Even though there are fewer studies of gesture in older children, as shown in Figure 1.1, that does not mean, of course, that the children’s use of gesture disappears; indeed, there are also a number of aspects of the use of gesture that emerge during this period, as this thesis will demonstrate.

### 1.3 Roadmap

The thesis is divided into three parts. The chapters in Part I (“Setting the stage”) deal first and foremost with conceptual and terminological issues. Even though all of these considerations are motivated by the need for a conceptual apparatus to tackle the task of characterizing children’s gestures, many of the considerations in Part I are relevant to the nature of gesture more generally — not only with respect to children’s gestures. Chapter 2 addresses the question of whether, and how, to distinguish the actions that are to be considered gesture from those that are not. Instead of treating “gestuality” as a single property that may either be present or not in a given action, a range of different properties are identified and elaborated on: all of these relate in some way to gestural qualities of action, including, among other things, different levels of communicative explicitness, semiotic complexity,

and conventionalization, as well as different kinds of intentionality. This relatively lengthy discussion leads to a multi-faceted conceptualization of gesture that opens up for comparisons, both in terms of similarities and differences, between gestures and various sorts of actions that many gesture researchers may not typically consider as gestures. Chapter 3 provides further conceptual investigations of the nature of action in social settings and of gestalt properties of gestures. These considerations lay the foundations for the investigation, in later chapters, of the factors that contribute to gestural meaning. In last chapter of Part I, Chapter 4, various issues related to methodology are discussed and described: the data used for the analyses, transcription conventions, and quantitative measures. A gesture typology employed in different ways throughout the thesis is presented. This typology is used both for purposes of quantification and as a starting point for further analyses of a more qualitative kind.

All of the chapters in Part II (“Analysis”) are concerned with analysis of the data. Chapter 5 presents and discusses a relatively large number of longitudinal developmental patterns, from a quantitative point of view. Then there are three chapters that deal, respectively, with deictic, iconic, and conventionalized aspects of the children’s gestures. These are Chapter 6, 7, and 8. It should be pointed out that deictic, iconic, and conventionalized aspects are not treated as corresponding to separate gesture types. They are instead treated as different kinds of semiotic motivations that may co-exist in one and the same gesture. The three chapters on deictic, iconic, and conventionalized gestures are mainly qualitative in character, except Chapter 8 on conventionalized aspects of gesture, which also includes quantitative descriptions relating to the use of nodding and head-shakes at different ages. The last chapter in Part II, Chapter 9, consists of an in-depth analysis of how manual handling of objects may take on communicative appearances in a range of different ways.

Part III (“Rounding up”) is short compared to the first two parts. It consists of a single concluding chapter, Chapter 10, where the overall empirical findings and conceptual developments are summarized. That chapter may also serve as a more extended roadmap for readers who want to do a more selective reading of the contents of this thesis.



## CHAPTER 2

# ”What is gesture?” and other conceptual preliminaries

*My proposal could be called an elimination of theoretical terms, if you insist; for to define them is to show how to do without them. But it is better called a vindication of theoretical terms; for to define them is to show that there is no good reason to want to do without them. They are no less fully interpreted and no less well understood than the old terms we had beforehand.*

Lewis (1970, p. 427)

### 2.1 The idea of an upper and lower limit of gesture

To idealize matters, one might say that there are two main ways of conceiving of gesture. The first is broad and inclusive: gesture as all sorts of bodily movements and symptoms, including facial expressions, gaze patterns, pointing, postures, just about any action made in a social setting (e.g. Rome-Flanders & Ricard, 1992), proxemics (interpersonal distances in social encounters, Hall, 1963), and in some accounts perhaps even blushing and being moved to tears. This broad view roughly corresponds to what is sometimes also referred to with the less fortunate term “body language”. Such conceptions of gesture have the benefit of not excluding any potentially interesting phenomena from the research agenda, but they also have the drawback of failing to acknowledge quite real differences between various modes of expression. For that reason, this way of speaking of gesture will not be employed here.

The second perspective is narrower and more constrained: gesture as those actions that are (or are perceived to be) performed under volitional control and that have publicly recognizable features marking them as being performed for purposes of expression rather than practical aims (Kendon, 2004, p. 15).<sup>1</sup> These actions are commonly coordinated with speech (Kendon, 1972, 1980b; McNeill, 1985), although the specific ways in which this is done may vary (e.g. Kendon, 1985b, 2004) and the presence of speech is in no way obligatory. This narrower sense of the term gesture corresponds more closely to how it is used by most researchers within the field of gesture studies. It reflects as well how the term is used in this thesis. Whenever gestures are performed as part of social discourse, rather than in "non-social" contexts (Rodríguez & Palacios, 2007; Rodríguez, 2009; Chu & Kita, 2008; Alibali & DiRusso, 1999)<sup>2</sup>, they can be involved in utterance orchestrations in several different ways, most saliently through:

- Establishing a referent (such as by pointing) or saying something *about* a referent by means of a "representational" gesture (Kendon, 2004; McNeill, 2005; Streeck, 2009b).
- Performing pragmatic functions directly, such as the speech-act-like functions of marking a multimodal utterance as a question or a proposal (Kendon, 2004; Streeck, 2009b), or other functions of providing interpretative frames of utterances.
- Regulating the interactive communicative process as such: for example in relation to turn taking (Duncan, 1972; Streeck, 1992; Bavelas et al., 1992), or to initiate repairs (Seo & Koshik, 2010).
- A combination of the above, perhaps also including other kinds of functions.

Since the narrower conception of gesture rests on the idea that not every action counts as a gestural expression, a number of distinctions are commonly employed to demarcate that which is gesture from that which is not. Most of these demarcations can be classified as concerning either what I call the *lower limit* or what I call the *upper limit* of gesture (see Figure 2.1).

---

<sup>1</sup>Practical actions refers to actions such as driving a car, throwing away garbage, or opening a can of beer.

<sup>2</sup>I put "non-social" in scare quotes here, because, of course, even if a human being acts in relative isolation in a particular situation, she still may draw upon skills and abilities that have a social component as part of their genesis on an ontogenetic time-scale.



## 2.1. THE IDEA OF AN UPPER AND LOWER LIMIT OF GESTURE

---

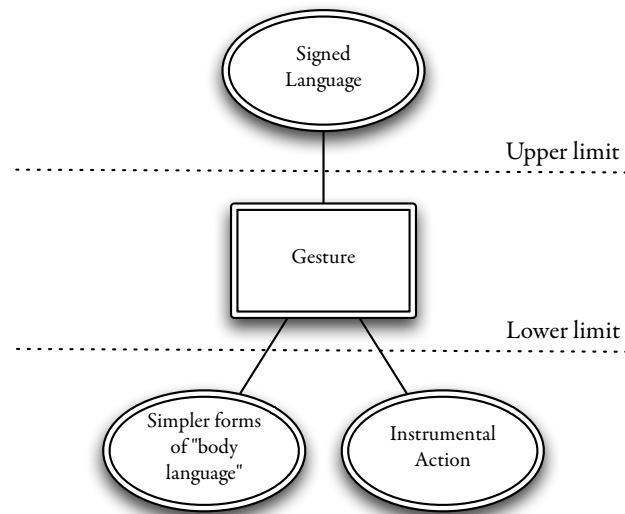


Figure 2.1: The upper and lower limit of the concept of gesture.

The *lower limit* separates those actions and behaviors typically considered to be gesture proper from those that are “too simple” to deserve the label. One class of actions or behaviors that is typically considered to belong *below* the lower limit is that of *communicative behaviors* that lack either some required degree of *volition* (such as blushing) or some required degree or kind of *semiotic complexity* (such as taking an object offered by another person). That is to say, these are acts that do lack both overt and covert differentiation (Piaget, 1962 [1946]; Sonesson, 2007) between expression and referent/content. They are labeled “simpler forms of body language” in the figure, for lack of a better term. Another class of actions that is typically placed below the lower limit is that of *instrumental actions*, which lack either some required form of *communicative explicitness* (such as not being intentionally communicative, either in a manifest and public sense Kendon, 2004, or psychologically Tomasello, 2008), or some required degree or kind of semiotic complexity.<sup>3</sup> In research on child gestures, it is relatively common to talk of gesture in a way that implies the criteria of both explicit communicative status *and* a certain degree of semiotic complexity. That is, it is often said that for something to be a gesture it must be *both* “symbolic” (in one of the many senses of this term) and involve or give some form of indication of a communicative intention, which may, for example, be considered to be present if a child engages in gaze alternation between the target of a pointing gesture and the interlocutor.

---

<sup>3</sup>Other movements that are typically not considered as gesture, but often not classified as “practical action” either, are those involved in locomotion and sustaining posture.

Many researchers also postulate an *upper limit*, which separates those actions or behaviors that are considered “too complex” and *language-like* to be called gesture from those that are not, in one way or another, language-like. The reasons for this are sometimes motivated by empirical findings, and sometimes motivated by theoretical distinctions. The prime example of bodily communicative action considered to belong above this upper limit is signed language, as it has the systematicity and complexity of a language proper (cf. Saussure, 1983 [1916]; Hockett, 1966; Zlatev, 2008b).<sup>4</sup> Even though the signs of signed language fulfill the positive criteria that are required for something to qualify as gesture, and even though both gestures and the signs of signed language are articulated by means of the same bodily medium, many researchers do not consider signed language as gesture due to the presence of *additional* properties and constraints.

The benefits and drawbacks of the more narrower conceptualization of gesture are the reverse of the broad conceptualization. That is, in the case of the narrower conceptualization, distinctions can be made between various types of expressive actions, which is clearly more satisfactory from a theoretical viewpoint than an undifferentiated mass, but as a consequence of the contrast made with other forms of action, movement, or signification, there is sometimes a tendency “to exaggerate differences and obscure areas of overlap”, as Kendon (2008, p. 348) has argued. Kendon’s argument is put forward in a discussion of what is here referred to as the upper limit, but the argument is equally valid with respect to the lower limit. Such exaggerations of differences between categories — simple non-gesture versus gesture versus signed language — without a corresponding focus on similarities, has, as I will argue, resulted in blind spots in gesture research. Gestures having properties typically associated with types of action and movement *other* than “gesture”, such as practical action or language, are much less studied (cf. Andrén, in press a; Kendon, 2008, p. 360). I will therefore try to follow Kendon’s recommendation to complement such an exclusive focus on differences between gestures and other forms of bodily expression with a more detailed “comparative semiotics of the utterance uses of visible bodily action” that “will be better able to articulate the *similarities and differences* between how kinesics is used, according to whether and how it is employed in relation to other communicative modalities such as speech” (*ibid.*, p. 348, my italics).

The question of how to handle the upper and lower limits of gesture is clearly at the same time difficult and central to understanding the nature of gesture. In Figure

---

<sup>4</sup>See the discussion on the upper limit in Section 2.3 for further elaborations on the upper limit.

2.1, the limits are drawn as two idealized lines, but as even the brief discussion so far has revealed, both subsume *several different distinctions*. These different distinctions often constitute separate dimensions that may vary with respect to each other so that they appear in different configurations in different actions. Put another way, the two “lines”, or limits, have internal structure, which means that they are not single lines after all. Both can be unpacked and analyzed in more detail. Various conceptual issues involved in that unpacking occupy the rest of this chapter.

Issues relating to the upper and a lower limit also constitutes a thread that runs through the entire thesis, both on a conceptual level and as part of the empirical investigations, although it certainly not the only question that will be addressed. The aim is not primarily to critique shortcomings of previous research, but rather to contribute positively, by taking a few steps toward a systematic treatment of phenomena that have been less studied due to the aforementioned blind spots in gesture research.

## 2.2 Unpacking the lower limit

As stated in the previous section, the lower limit concerns the “line” between those actions and behaviors that are usually considered to be gesture proper and those that are too simple, in various ways, to qualify. The reasons for rejecting certain behaviors vary between researchers. Sometimes gesture is defined in terms of the degree or kind of *communicative explicitness* involved. Acts with manifest features indicating some sort of communicative intention, may, but need not, be *semiotically complex* in the sense of being semiotic signs, with differentiation between expression and content/referent.<sup>5</sup>

On other occasions gesture is rather defined as bodily movement with a certain degree or kind of *semiotic complexity*, that may, but need not, be used for purposes of communicative interaction with other persons. Indeed, it is well accepted that gestures may appear outside social encounters. Some examples of this are gestures in the context of individual problem solving (Chu & Kita, 2008) and pointing gestures in the context of counting objects, as a way to keep track (Saxe & Kaplan, 1981; Alibali & DiRusso, 1999; Graham, 1999). There is also the notion of private gesture (Rodríguez & Palacios, 2007; Rodríguez, 2009), relating to the idea of private/inner speech (Vygotsky, 1962), which may be considered communicative in the sense of being a kind of “dialogue” with the self, but not in the sense of being a

---

<sup>5</sup>A more encompassing definition of a semiotic signs is provided in Section 2.2.3.

part of a communicative encounter with another human being. In such cases, the issue is rather one of self-regulation through symbolization.

On yet other occasions gesture is talked about as those actions that are both explicitly communicative *and* semiotically complex. This is common in research on children's gestures. Here, the focus is often directed toward actions that are characterized by differentiation between expression and content/referent(s), and, in addition, it is commonly required that the actions be intentionally communicative in some way, as manifested through e.g. gaze alternation between the target of a pointing gesture and the interlocutor (e.g. Bates et al., 1979).

Finally, in some occasions, the reasons for considering something as gesture are simply not explicated, and the reader is left to rely on pre-theoretical intuitions of what will count as a gesture.

There is also research on the relation between gesture and thought — considered either as two different but related phenomena (Goldin-Meadow, 2003a), or as a kind of embodied thought itself (McNeill, 2005; Streeck, 2008b). However, as far as I know, there are no proposed *definitions* of gesture that pick up on issues related to thought specifically, such as, say, "gestures are those motor processes that serve function X in relation to thought". Most authors thus seem to agree that an essential property of gestures is that they are publicly visible and recognizable (but see the notion of *phantom gesture*, Ramachandran & Blakeslee, 1998), and that gesture can be defined on that basis.

What is implied by the phrases "communicative explicitness" and "semiotic complexity"? In Section 2.2.2 and 2.2.3 I elaborate on these phrases by classifying each into levels of increasing explicitness or complexity — beginning with simpler forms that most researchers would agree lie below the lower limit and ending with more complex forms that most researchers would agree lie above the lower limit. The classifications are formulated in ways that are relevant to the purposes of this thesis. They are not intended as one-size-fits-all constructs. Other analytical purposes may require other conceptualizations. The primary motivation behind the classifications is not to force reality into neat boxes, but rather the opposite: to account for more variants with finer nuances than is usually the case whenever dichotomous distinctions are made between communicative and non-communicative, symbolic and non-symbolic, or gestural and non-gestural. More distinctions are always possible, but the granularity chosen here is considered sufficient for the purposes of this thesis.

It is worth pointing out from the outset that advocating clear conceptual distinc-

tions should not be confused with claiming that reality necessarily corresponds to such conceptual distinctions in a clear-cut or non-problematic way. Whether, and to what extent, that is the case will always remain a question that is open for investigation. Still, multiple distinctions may be better than simplistic binary distinctions and certainly better than no distinctions at all, not least because a well-defined conceptual apparatus makes more precise discussions possible, and defining key terms does not require that the analysis itself be locked into these terms.

Before presenting the classification of various levels of communicative explicitness and semiotic complexity, a number of clarifications regarding the concept(s) of intentionality will be required to facilitate the subsequent discussion.

### 2.2.1 Intentionality

#### Aboutness and intent

There are two main senses of the term “intentionality”, as it figures in various theoretical contexts. The first of these is intentionality in the sense of *aboutness* (or *directedness towards*) (e.g. Searle, 1983). This is how Brentano (1995 [1874]) used the term in his famous “intentionality thesis”, as part of his project to create a descriptive psychology. Brentano’s intentionality thesis is the claim that “every mental phenomenon includes something in it as an object [...]” (*ibid.*, p. 88). For Brentano, intentionality is *the* distinctive mark of the mental, which separated mental phenomena from other kinds of entities. Different traditions have picked up different aspects of Brentano’s thesis. Analytic philosophers have sometimes treated it as a justification for a Cartesian mind/body dualism, whereas phenomenologists, starting with Husserl, have rather used it as a starting point for the exploration and description of various modes of givenness of this “object” (Bartok, 2005). At least some of the phenomenological takes on aboutness are in line with various present day theories on embodiment, without any dualistic implications (Embree, 2004; see also Zlatev, 2009, p. 150).

The second main way that the term is used is in the goal-directed sense of *intent* (Condillac, 1971 [1756]; Bentham, 1907 [1780]; Anscombe, 1957; Grice, 1957), also related to the notion of *volition* (Maasen et al., 2003). This is how the term is used in most research on child gesture and children’s communicative development. For example, Tomasello (2008, p. 113) writes: “current theoretical debates about infant pointing and prelinguistic communication center on the question of whether the most accurate interpretation is a cognitively rich or a cognitively lean

one". The differences between the various theoretical positions in the debate that Tomasello refers to are typically phrased in terms of differences in the degree, kind, presence or non-presence of intentionality that is involved in children's actions, in the sense of "having the intent to do X" or "attempting to do X". Tomasello (2008, p. 15), who himself favors a rich interpretation, writes that "when communicators are *attempting to* influence the behavior or psychological states of recipients *intentionally*, we now have the starting point for communication from a psychological point of view" (italics added). The notion of intent may figure in several different ways in the context of communicative actions. One may, for example, distinguish between intending to perform some action as such (*action intent*), and intending "for the other to attend to a referent" as a result of a performed action (*referential intent*, cf. Tomasello, 2008, p. 124), and intending for others to understand that I intend them to understand what I do (*communicative intent*). Communicative intent, understood from this point of view, amount to a kind of "second-order" intentionality, over and above the action intent(s) or referential intent(s) of an action (Grice, 1957; Sperber & Wilson, 1995 [1986]; Tomasello, 2008). Regarding action intent, one may further distinguish between intending the *form* of the action (serving wine in an elegant way), intending the *immediate results* of the action (getting wine into the glass, no matter who serves it or how), and intending more *indirect consequences* of the action (serving a good wine to make my guests feel comfortable, which in turn may make them want to come back on another occasion, which in turn may make my wife happy, and so on).

Being organized around some more or less specific *intent* is only one of several possible ways that something, such as an action, might be said to be *about* something. Intent and aboutness are therefore not necessarily entirely separate issues, but in any case, aboutness is clearly a much broader notion than intent (cf. Searle, 1983, p. 7). Whenever it is important to keep these two meanings of the word "intentionality" apart in the discussions that follow, the term *aboutness* will be used for the first kind of intentionality, and the terms *intent* and *intend* will be used for the second. However, when the word intentionality itself is used, it should be remembered that it need not necessarily imply that an action is produced with a specific goal clearly in mind, only that the action has the general character of being about something.

### First-, second-, and third-person intentionality

Another crucial distinction with respect to intentionality is to distinguish between three different *intentional perspectives*, corresponding to some extent with the grammatical perspectives of *I*, *thou*, and *anyone/everyone* (the *generalized other*, Mead, 1934), respectively. This three-fold distinction is orthogonal with respect to the distinction between aboutness and intent, which is to say that both of them can be regarded from all of the three intentional perspectives. *First-person intentionality* refers to intentionality in the sense of the existential condition of mindful being itself, i.e. the intentionality (intent or aboutness) presupposed by the possibility of having experiences in the first place. *Second-person intentionality* refers to the seeing of another person's conduct as endowed with intentionality (intent or aboutness), and more specifically, perceiving that intentionality as, in some sense, belonging to this specific other unique person as a response to the present situation.<sup>6</sup> What this means may be easier to explain after considering *third-person intentionality*, which refers to the seeing of another person's conduct as endowed with intentionality (intent or aboutness), and more specifically, that this intentionality of the observed conduct is perceived as being an action of the kind that "anyone" (in an idealized sense) would *typically* do in a given type of situation.<sup>7,8</sup> An example of seeing an action from the perspective of third-person intentionality is seeing someone put dough in the oven, and one perceives this as being done in order to bake bread or a cake, which is what "anyone" would *typically* intend by such an action (cf. Schutz, 1943), in appropriate contexts. An example of how the very same event could be seen from the viewpoint of second-person intentionality might be something less typical such as putting the dough in the oven and seeing this, not as being done in order to bake a cake or bread, but in order to hide the dough from the spouse who just came home, because the cake to be baked was intended as a surprise, not to be

---

<sup>6</sup>At the same time, no situation is strictly "unique". Overarching schemata such as reaching towards objects, grabbing them, putting them away, and so forth, span across many or even most situations.

<sup>7</sup>Sometimes the term "third-person perspective" is used to refer to a more objective kind of perspective, as the perspective employed within the natural sciences, but this is not how the term is used in this thesis.

<sup>8</sup>The notion of typicality is also discussed in Section 2.3.3.

seen before it was done.<sup>9</sup>

An important aspect of all of the three intentional perspectives is that they turn back toward the self (cf. Mead, 1934; Schutz, 1953). In the case of first-person intentionality this follows directly from the definition since it is already a property of the self.<sup>10</sup> In the case of second- or third-person intentionality, this is perhaps less obvious. What it means is that when I perform an action, I might not just do it, but I may do it in a way that is sensitive to its status as a *publicly recognizable action that is endowed with intentionality*. That is, I might act in a way that is sensitive to the manner in which I am a "you" or an "anyone" to you. I might do it in a way that is not particularly typical, but is still perceivable as at least minimally rational, given the relevancies of a particular uniquely unfolding social situation (second-person intentionality). I might also perform an action in a more standardized and typified way, such as when I am acting like the typical sender of a letter (Schutz, 1953, pp. 19), doing all the steps that "anyone" would conventionally do when sending a letter (third-person intentionality), such as putting a stamp on the envelope, writing the address according to conventions, putting the letter in a mailbox, and so forth, all in order to ensure that my letter will be recognized by the staff at the post office as a typical letter to be handled like any other. The first-, second-, and third-person perspectives are obviously *not* mutually exclusive, but rather three co-present aspects of most social encounters.

### Public intentionality

While first-person intentionality need not be *publicly recognizable* in all respects, second- and third-person intentionality are publicly recognizable by definition. In fact, even when first-person intentionality is somehow publicly recognizable through the way a person acts, it can only be so indirectly, in the form of second- or third-

---

<sup>9</sup>Describing first- and second-person intentionality by means of ordinary language, the way I do here now, is inherently problematic, because ordinary language is itself a third-person typification that consists of standardized elements — conventionalized words and grammatical patterns with conventionalized meanings — that "anyone" should understand (even more so in its written form, since the communication is then removed from the dynamic potential of the face-to-face encounter, cf. Linell, 1982; Schutz & Luckmann, 1973). That is, when describing the event of putting dough in the oven, or when I describe my feelings to *you*, by means of the conventionalized words of language or by means of conventionalized gestures, it is impossible to do so without invoking generalized third-person frames of reference, since the meaning of the words used are determined not only by the situation of here and now in its uniqueness, but also by their prior history of uses in a community — what "spouses" typically do, what "bakers" typically do, what "ovens" are typically used for, what "surprises" are typically like, and so forth.

<sup>10</sup>This is not to say that there may not be developmental processes of a social kind lurking behind the emergence of some aspects of first-person intentionality.



person intentionality, because I will always be a “you”, or an “anyone”, to you. Second- and third-person intentionality might therefore be referred to collectively as *public intentionality*. There is no direct correspondence between first-person intentionality and public intentionality. This can be seen if one considers that a movement which appeared to be *unintended* may in fact not be unintended, as in the case of deception (or that it was not *about* something from the point of view of the person who performs it). The opposite “dissociation” also holds — when an infant performs a movement of some sort, parents may well ascribe more/other intentionality to this movement than what is warranted (from a scientific point of view) purely due to the publicly visible ways in which the movement is performed.

To be sure, this is not to say that first-person intentionality and public intentionality are so disparate as to be unrelated. In fact, most psychologically oriented research on gesture that treats gesture as a “window to the mind” (e.g. McNeill, 1992, 2005; Goldin-Meadow, 2000) rests on the assumption that gestures are typically more or less direct reflections of the content they seem to express, and, crucially, that this content is the same as the thought of the speaker. If this is taken to mean (A) that there must be some bodily process corresponding to the movements that a person is performing, then it is not controversial at all. I agree that some aspects of gestures may indeed be best understood as a kind of “thinking by hand”, as Streeck (2009b, pp. 151) puts it, in the sense that the situated movements and sensory activities of the hand may be considered an integral part of cognitive processes — that they need not only be seen as indirect reflections of thought that goes on “elsewhere” (*ibid.*, p. 160). However, if “window to the mind” is taken to mean (B) that the first-person intentionality and the public intentionality of (all) gestural performances are not just related, but that they are just two sides of the very same coin, then it is a lot more controversial. I do not want to contest that even studies that adopt B as a vantage point in the study of gesture can yield, and have yielded, interesting and reliable results, and that disregarding the differences between the intentional perspectives might therefore be defensible as a methodological heuristic.

Nonetheless, I do not subscribe to the idea of gesture as a “window to the mind” if it is taken to imply that persons always attend to their own as well as others’ gestures in the same, basically homogeneous and situation/activity-independent, way, and that all gestures therefore “reflect thought” — or rather meaning, as seen from different intentional perspectives — in a single way. The degree to which all of the various forms of intentionality described here are part of gestural performances is quite variable, both from the point of view of the performer and from the point

of view of other participants or bystanders. On some occasions, people hardly attend to their gestures at all, on other occasions, people attend to their gestures and the visibility of the gestures to the Other in very explicit ways (cf. Streeck, 2003; Streeck, 2009b, pp. 151; Gullberg & Holmqvist, 1999). Kendon (1985b, 2004) points out that people often adjust their gestural performances to the (social) situations at hand, not least through the heterogeneous ways in which gesture and speech may be mutually adjusted to form multimodal utterances (see also Andr en, in press c).

The distinction between intentionality as a publicly recognizable affair and as an existential condition of the first-person perspective need *not* correspond to a distinction between processes that take place inside and outside the skin, nor does it correspond to a distinction between subjective and objective, because in both cases the action involved and its lived apprehension may be best understood as at once "mindful" and situated in the world-at-hand. Gilbert Ryle (1999 [1968]) made clear that observable conduct is not necessarily a matter of the behaviorist conception of action in terms of "physical" movements. For these purposes he offered the useful distinction between *thin descriptions* (focusing on "physical" description movement) and *thick descriptions* (intentional, meaningful, and contextually embedded) of observable behavior.

### Intentionality within the natural and the scientific attitude

Finally, one should distinguish between approaching public intentionality in action from within the *natural attitude* of everyday life and from within the *scientific attitude* (Husserl, 1983 [1913]; Schutz, 1932, 1945, 1953).<sup>11</sup> Jones & Zimmerman (2003, pp. 156-157) capture the character of approaching action within the natural attitude of daily life when they write: "we assume that the 'default presupposition' guiding social life is that the actions out of which interaction is constructed are designed to be what they appear to be, and in that sense, are intentional — unless accounted otherwise" (see also Heritage, 1984). Action understanding from within the natural attitude is thus characterized by a *suspension of doubt* "until further notice". That is, participants in interaction normally do not doubt the intentionality (in both senses) of others' actions unless there are some explicit reasons to do so.

Within research on children's communication, often guided by the logic of the scientific attitude, action and gesture is not always understood in this way. In this

<sup>11</sup>While Schutz adopted the term "natural attitude" from Husserl, he also developed this concept in his own direction.

context, the approach to children's actions may sometimes instead be guided by the skeptic's presupposition not to take any of children's abilities on face value, unless a rigorous account can be given to motivate such an interpretation, perhaps on the basis of results from carefully controlled experimental studies (but see the concept of *rich interpretation* for a kind of middle road, Bloom, 1970). Action understanding from within the scientific attitude is guided by *principled doubt*, in direct contrast to the suspension of doubt within the natural attitude. Issues of judging whether a given act is intentional or not, or whether an action should be understood and treated as a communicative act or not, and so on, are by necessity quite different when approached from within the the natural attitude and from within the scientific attitude.

Both modes of action understanding are obviously valid, in their own ways, but it should be noted that the habitat of human action is, after all, within the natural attitude (Schutz, 1953, 1954). Even though scientific practices may legitimately strive to move away from intentionality-as-taken-for-granted in their explanations, they should not strive away from this sort of intentionality as a genuine property of the explanandum, i.e. the phenomenon to be explained.

### 2.2.2 Levels of communicative explicitness

“Communication” is a highly polysemous word (Zlatev, 2009). Even when its meaning is restricted, as it is here, to include only the use of various semiotic resources as part of face-to-face interaction between human beings, a great deal of different conceptions still exist. Instead of making the case for a binary distinction between communicative and non-communicative actions, a classification into four levels of communicative complexity will be presented below. The various levels may be said to form a progression from less explicit to more explicit communication.

The classification is concerned with various kinds of communicative organization, from the point of view of publicly recognizable properties of movements, actions, and activities that give the actions their appearances as this or that sort of action. To clarify: in the case of an action that is performed so as to appear unintended, although it in fact was intended — an example of that is the well-known phenomenon of tackling diving in soccer (Morris & Lewis, 2010) — the aspects of this action that are relevant to the levels of communicative explicitness presented below are rather the publicly recognizable features that give rise to the public appearances of being “unintended” (assuming the deception was successful), on the level of second- and third-person intentionality, rather than the underlying and

non-visible first-person intent to deceive.<sup>12</sup> This is not to say that first-person aspects of human cognition are unimportant, but the point is that what stays covert and therefore goes unnoticed by an interlocutor can hardly be argued to be a *communicative* phenomenon. This focus on public appearances is not necessarily to taking an observer's perspective, since people may also orient to the public appearances of their own actions when they produce them (e.g. Alibali & Don, 2001; Lerner & Zimmerman, 2003). An action may, for example, be produced in such a way as to be publicly recognizable as an action of this or that sort, as discussed in the previous section on intentionality. The example of tackling diving in soccer games is a case in point, but more generally, and perhaps more typically, first-person intentionality and second-/third-person intentionality need not differ from each other the way they do in deception.

As stated above, the levels of communication to be presented below are intended to be focused on public appearances. The public appearances of actions that make up communication have two important properties: (A) participants in interaction have *shared access* to them, and (B) they constitute *sufficient* "evidence" for whatever sense (or non-sense) that participants make out of it, so that in one sense, there is nothing for the scientist to correct, or add to them, apart from uncovering and describing their role in communicative interaction.

### Comm#1 — Communication as a side-effect of co-presence

On a very basic level (Level 1), most or even all of the things that people do when they are in the presence of each other have a kind of implicit communicative value. This is captured in the axiom of *the impossibility of not communicating* when human beings are in the presence of each other (Watzlawick et al., 1967; see also the notion of *unfocused interaction* in Goffman, 1963, and the notion of *indication* in Allwood, 1976, p. 66). Actions (i.e., action intentions) may well be understood in various ways by others even when they are not performed with any sort of visible orientation to an interlocutor. If I am having a glass of lemonade, someone might suddenly enter the room and notice that I am drinking. This may lead them to believe that I am probably thirsty, and if it happens on a hot day, they might think that this is the reason why I am drinking, and so on. They might also be able to see more specific things in this action if they have access to more specific knowledge, for example if they know me well or if they happen to be experts on lemonades or similar.

---

<sup>12</sup>Of course, in the case that the deception is not successful, the act might have precisely the appearance of foul play.

Even though my actions may be seen as comprehensible in various ways, both due to generic knowledge as well as more specific sources of knowledge like personal acquaintance or expertise, they may not be perceived and treated as *communicative* actions per se, neither by myself nor the observer.<sup>13</sup>

Already at this level, people may distinguish movements by other people that seem to be non-voluntary (stumbling on a stone and then falling, or the spasms of an epileptic seizure) from those that appear voluntary (scratching one's back or picking something up in order to use it). While the actions of A may result in some forms of actions on behalf of B as a result of the actions of A, this is still not a matter of communication as an organized activity. Indeed, this type of action understanding is not what most people think of when they think of communication, and the communication at this level is at best an *indirect* kind, emerging as a by-product of mere co-presence of beings capable of intentional interpretation. Still, it is worth pointing out that *any* action, including explicitly communicative actions, may have communicative consequences that are not anticipated or expected — hence not strictly speaking part of the communicative organization of an activity — so, in one sense, aspects of this level are always present whenever people come together.

### **Comm#2 — Action framed by mutual attunement**

A first step in establishing a communicative encounter is typically the achievement of a mutual attunement between at least two persons. Accordingly, at the level of Comm#2 actions, there is other-orientation in the sense that a mutual attunement is established between at least two participants. Then an action occurs during this mutual attunement, but the act itself is not of a kind that makes it publicly recognizable as a specifically communicative action. The participants may be turned toward each other and engaged in coordinated gaze patterns, and so forth. On this level the actions are surrounded by an interpersonal engagement in a negotiated interactive formation (cf. *F-formations*, Kendon, 1990; Ciolek & Kendon, 1980) — a kind of focused interaction (Goffman, 1963) characterized by the presence of a *we-relationship* (Schutz, 1951, 1953, 1955). This *we-relationship* and its manifest establishment in activity is a more basic form of intersubjectivity than the achievement of shared understandings, but it is nevertheless a kind of intersubjectivity (cf. Duranti, 2010). Even though mutual attunement may be established through

---

<sup>13</sup>Obviously there are cases where drinking is indeed done in a stronger communicative sense, such as when demonstrating how Elvis Presley used to drink, or when demonstrating the generic act of drinking. That is not the sort of drinking that is at issue in relation to Level 1.

means recognized as communicative, a given action that is performed once such a mutual attunement has been established does not need to be publicly recognizable as a communicative act. It is precisely such actions, or aspects of action, that belong on this level.

Nevertheless, this does not mean that there will be no other-orientation involved in the action, it only means that this other-orientation will not be visible as such. An example of this may be when I am interacting with someone and then I suddenly look briefly at my watch. I may do this either to indicate that I am in a hurry, even though I may already know very well what time it is, or I may do it because I genuinely wonder what time it is, or it may be something like a habit. Such actions may not have any visible properties to indicate to the Other that a given act was produced as an explicit part of the communicative organization of the encounter, i.e., for the sake of communication. However, this lack of a visible communicative intent notwithstanding, that an action is part of a focused interaction may in itself provide enough motivation for the Other to respond to it: in the case of the watch, to say "oh! I see you're in a hurry!", or to the publicly recognizable fact of the woman's sore throat, to say "hope your throat gets better soon!". Already at this level there can be a kind of indirect reciprocity, such that an action that is part of a focused encounter may be treated as intersubjectively shared and "answerable" despite the lack of manifest properties of communicative intent in the performance of the act itself.

Also belonging to this level are cases where someone acts in order to evoke a stylized presentation of the self "as one really is" (Goffman, 1959, 1971; Habermas, 1991 [1981], p. 86), at least when they take the form of an implicit self-presentation that is not designed to be seen as such, such as taking on a certain facial expression that looks good in the mirror. Such a controlled dissociation between "hidden" and public motives in the organization of an action may in fact be a more cognitively demanding feat than some of the other acts that would belong on this and the *next* level to be presented below (cf. Tomasello, 2008, p. 90) because it requires quite strong awareness of the self as visible to others. However, as pointed out above, this classification is primarily concerned with various forms of public appearances of actions, irrespectively of their cognitive demands.

### **Comm#3 — Visibly other-oriented action**

At the level of Comm#3 actions, there is other-orientation in a stronger sense: not just in the sense that the act *may* be other-oriented in its organization (albeit not vis-

ibly so), but in the sense that the act is now produced in a way that makes it publicly recognizable as an other-oriented act made for the purposes of expression. That is, Comm#3 actions are structured in ways that may be characterized by second-order Gricean *communicative intent* — the communicating person “must intend to induce by x a belief in an audience, and he must also intend his utterance to be recognized as so intended” (Grice, 1957, p. 383). In accordance with Kendon (2004, p. 15), and in accordance with the distinction made in the Section 2.2.1 between intentionality in a psychological sense and intentionality in the sense of publicly recognizable properties of action, it should also be pointed out that such Gricean intentions cannot be magically known to others. From a communicative point of view the crucial issue is how various actions may possess more or fewer of various features of publicly manifest deliberate expressiveness (relative to their context of occurrence), either in the form of a manifest orientation to second or third-person intentionality, or both. When restricting this argument to communicative actions as articulated by visible bodily movements and configurations, this does, in fact, correspond to how Kendon defines *gesture* (*ibid.*, p. 15):

‘Gesture’ we suggest, then, is a label for actions that have the features of manifest deliberate expressiveness. They are those actions or those aspects of another’s actions that, having these features, tend to be directly perceived as being under the guidance of the observed person’s voluntary control and being done for the purposes of expression rather than in the service of some practical aim.

In contrast to Comm#2 actions, the publicly recognizable character of the other-orientedness of Comm#3 actions makes them mutually manifest as communicative (cf. Sperber & Wilson, 1995 [1986]). For this reason, such actions are of a kind that people are normally held responsible for *qua* communicative actions and the act is therefore typically seen as a “‘move’, ‘turn’, or ‘contribution’” in the communicative interaction (Kendon, 2004, p. 7). When the term “communicative action” is used in a non-specified way in this thesis, it refers to actions that qualify as Comm#3 action.

### **Comm#4 — Reciprocated action**

That an action is perceivable as a Comm#3 action typically also makes *responses* to this action conditionally relevant (although not necessarily in all types of activities, cf. Linell, 2009, p. 185; Corrin et al., 2001, p. 204). The conditional relevance projected by Comm#3 action is stronger than in the case of Comm#1 and Comm#2

action, precisely because of the explicitly communicative character of Comm#3 actions. At the level of Comm#4 the status of the act as communicative is not just mutually manifest in the manner of Comm#3 actions, but now the act of A is *responded to* by B in a way that is sensitive to the conditional relevance of the initial Comm#3 action.<sup>14</sup> In the words of Schutz & Luckmann (1989, p. 71):<sup>15</sup>

Social action can be either [unilateral] or [reciprocal], although in concrete cases the boundaries may not be clearly recognizable. At any rate, an action that the other person toward whom the project was directed does not "answer" is unilateral; only when it is answerable can one speak of reciprocity.

At this level, actions are not merely seen as communicative acts, but rather manifestly *treated* as communicative acts by means of responses that orient to the conditional relevance of these acts (including the special case of situations where non-action is the appropriate response). In line with the conceptions of language and human activity of *dialogism* (Linell, 1998, 2009), one may point out that without any sort of response from B to an initial action by A, the action of A is generally incomplete from a social point of view. An example is when someone raises his hand to greet a fellow man, but receives no response whatsoever. Responses serve minimally to acknowledge the very occurrence of a prior action, but they generally also provide some evidence of *how* B has understood A's contribution (Sacks et al., 1974, p. 728). B's subsequent contribution is simultaneously a response to A's initial turn as well as a new initiative in its own right, which means, more generally, that every utterance is at the same time potentially both a response and an initiative (Linell & Gustavsson, 1987). That is to say, B's contribution is both context-shaped and context-shaping (Heritage, 1984), or formulated differently, it is made with respect both to future-oriented "in-order-to" motives and with respect to past-oriented "because-of" motives (Schutz, 1951, p. 163; Linell, 1998, p. 166). It should also be added that "A's and B's actions are sometimes, in a sense, more of analytically distinct aspects of the interaction than sequentially separable units" (Linell, 1998, p. 167). Indeed, a response might be anything from an elaborate line of reasoning, or an entire story, to rather minimal and somewhat implicit

<sup>14</sup>This is not to say that the Comm#3 level is required for various less explicit forms of interaction between persons. It is certainly possible to respond in various ways to less strongly communicative acts of other persons, as already pointed out in the discussion of Comm#1 and Comm#2 actions. Reciprocal processes in a more general sense are therefore not restricted to Comm#3 actions.

<sup>15</sup>This book was finished by Thomas Luckmann long after Schutz' death. It is based on manuscripts written by Schutz from 1958 up to his death in 1959.



acknowledgments embedded in the performance of some act that primarily serves other purposes. B's response may follow A's initial action, or it may be initiated during the very performance of A's initial action, as in the case of one's head nodding in agreement. In some highly routinized activities, responses need not even be overt, but they can sometimes simply be taken for granted unless indications are given to the contrary ("silent confirmation"). It might be argued that, in such cases, the routinized and predictable character of the interaction itself stands in for the functions that would, under other conditions, have resulted from an overt response (Linell, 2009, p. 185; Andrén, Sanne & Linell, 2010).

A yet stronger form of reciprocity occurs when B not only responds in a way that is sensitive to the conditional relevance of A's utterance, but A responds to B's response in a further turn, making manifest that A has recognized B's response and to some extent also how A has understood B's response. A hypothetical example might go like this: First A says something (A: "What kind of bird is that+POINT?"). B responds (B: "I think it is a swan."). Finally A responds to the response from B (A: "I see+NODDING"). It might be argued that it is first at this point of *mutual reciprocity* that the reciprocity of A's initial action has come full circle (Mead, 1934; Linell, 1998, p. 167), not just as a manifestly noticed occurrence (by B) but as a contribution whose meaning has been jointly confirmed in a mutually manifest way. Of course, there are all kinds of particular cases and variations on the relatively idealized example involving "A" and "B", as presented here. For instance, the organization may be quite different in multi-party interactions, and if it is already the case that the responding turn of B can sometimes be omitted, as mentioned above, this is even more true for A's response to B's response to A's initial action.

One could also continue to postulate even more levels of communicative organization at yet higher levels, such as the joint achievement of topical cohesion across many turns of interaction, communicative projects (Linell, 1998, 2009; Schutz, 1951), and so forth, but the distinctions offered here are the ones that are of most direct significance to the analytical scope of this thesis. Whenever more precise elaborations are required, they will be offered.

The way Comm#4 actions are defined moves the focus away from A's initial action, but the processes of reciprocity described here may still legitimately be seen as part of A's initial action in the sense that they take part in the mutual determination of the communicative significance of this act with respect to intent or aboutness (i.e., "negotiation"). Such processes of reciprocity may even serve to establish whether or not an act was communicative at all, such as when A makes a sound of clearing the throat and B responds by asking "Did you just say something to me?",

to which A may respond “no, I was just clearing my throat. I think I had too much candy.” or “Well, I was going to say that...” or “Yes, I didn’t really like what you said before” (before A’s “initial” turn). All of these alternative determinations may serve to establish what the initial action “really was”, after its actual occurrence.

Finally, it should be pointed out that a given action may, and typically will, contain elements from several levels of communication at once and a single multimodal utterance may well be communicative at all the levels proposed here. An action can also be seen as a communicative act at Level Comm#3, without necessarily being interpretable in terms of its action intention on Level Comm#1. That is, B may only be able to tell that “person A clearly just performed a communicative action of some kind, but I do not know what it means”.

### 2.2.3 Levels of semiotic complexity

Gesture researchers, especially in research on children’s gestures, frequently make a distinction between “symbolic” gestures and those that are “non-symbolic”, even though different researchers differ quite radically in what exactly they mean when they make this distinction. In this section, a number of semiotic distinctions will be presented that are related to different kinds of semiotic structures of visible bodily actions, which may help to disentangle various senses of words such as “symbolic”, “signs”, and so forth, as used in this thesis. Just as in the previous section, the distinctions are presented in the form of a classification into levels of increasing complexity. This time, the classification involves three overarching levels.

#### Sem#1 — Situation-specific aspects of action

Level 1 regards actions with primarily *situation-specific* relevance or meaning, which are not organized and recognizable as pre-established *types* of action. At this level, the “code” aspect of action meaning is weakest, and the situated “inference” aspect is strongest. This holds true no matter whether “inference” is interpreted in a mentalistic way (Sperber & Wilson, 1995 [1986]) or in a more praxis-oriented and “direct” way. There is no differentiation between the form of the action and its meaning or purpose, and the action is a direct adaptation to a current local situation in its *uniqueness*. Already at this level actions may invoke the world at hand in various ways, such as making a certain object relevant by means of handling it somehow. Therefore, already at this level, a basic form of *reference* is possible. Arguably, this level is more like a theoretical idealization than an empirical reality, since no action

or situation is ever *entirely* “new” in all kinds of ways — prior experience is always brought into situations, and there are overarching generic schematisms and typifications in most kinds of actions such as reaching towards objects, grabbing them, picking them up, eventually putting them away, and so forth. Nevertheless, even in highly recognizable, typified, and conventionalized actions, there is always a local aspect of the action that relates to how the action is deployed in a particular way in a particular situation. As an action is performed in a local context, it always has the property of pushing the social activity forwards (cf. *communicative dynamism*, Firbas, 1971) in a situation-specific way, since every action — communicative, instrumental, or both — is a move that *does* something. In line with Schegloff (1986) one may also point out that even in the case of routine activities, the performance is always to some extent a local situation-specific achievement. Hence, the motivation for the inclusion of this level in the classification of semiotic complexity, is that it serves to acknowledge the potential of action dynamics to bring about novelty and change as well as the potential of action to orient to the *particular configurations of specific unique situations*, through the (locally) *context-shaped* and (locally) *context-shaping* aspects of action (cf. Heritage, 1984).

### Sem#2 — Typified aspects of action (*count-as*)

Level 2 actions are those that not only have a particular local and situated meaning (Level 1), but *also* are recognizable as pre-established action types, i.e., as tokens of a type. The importance and roles of *typification* in action, signification, and intersubjectivity has perhaps been most rigorously analyzed in the work of Alfred Schutz (1943, 1951, 1953, 1954, 1955, 1959). Typified actions need not strictly speaking be a category of actions with clear-cut boundaries allowing someone to state whether a particular instance is unambiguously an instance of the type in question or not. The essential criterion is rather that of a sense of *familiarity of a kind* — as something that one has seen before, although perhaps not necessarily in the *exact* same form. A prime example is the *canonical use* of certain artifacts, such as using scissors for cutting or a pen for writing or drawing, in contrast to novel creative uses of artifacts, such as using scissors as electrical conductors or placing a pen underneath a flower pot to raise the pot slightly.<sup>16</sup> Typified action schematizations may constitute minimally a “doing it again” of an action performed just moments before, either by oneself or by someone else; they may also be a more situation-

---

<sup>16</sup>Another example: a few minutes ago, I killed a wasp with Wittgenstein (1953), simply because the book constituted material-at-hand that was useful for the purpose.

transcending kind of typification that extends across particular settings, eventually spreading among members of a community or culture (cf. *mimesis*, Donald, 1991; Zlatev, 2007). Even though the exact form of the performance of a typified action may not necessarily be governed by such well-defined *norms* of correctness as those involved in some conventional gestures and signed languages, there is in many cases a conventional (i.e., socially shared) knowledge involved in knowing the form (third-person intentionality), typical purposes and relevancies, and suitable context(s) of such established ways of acting, although these need not necessarily be well-known in the sense that they have a corresponding vernacular name in ordinary language (cf. Schegloff, 2007, p. 7).<sup>17</sup>

Since there is not only reference and local relevance, but also a basic typification of *form* as well as *content* (that is not solely an idiosyncratic creation of the moment) the typified nature of the act makes a first step towards representation. Level 2 actions are meaningful not just in relation to a particular situation in which they occur, but they may also *count-as* (Searle, 1995, p. 28) some particular situation-transcending *type* of action — a type/token relationship — although not in a way that differentiates form from meaning.<sup>18</sup> As pointed out by Sinha (2009a, p. 302), the fact that something may count-as something does not mean that it represents it. Putting this another way, type/token relationships are not enough for something to be a representation in the stronger sense of a *semiotic sign* (Sonesson, 2007), as defined below. An act of drinking coffee surely counts as an act of drinking coffee, at least in appropriate contexts. At the same time, and under most circumstances, it *is* an act of drinking coffee rather than a representation of this act. There is no differentiation between expression and content/referent (Piaget, 1962 [1946]). In later chapters it will be argued that the coordination between count-as actions and speech is in many respects similar to the coordination between more complex kinds of signification (i.e., gesture) and speech, and that this has so far received very little analytic attention despite the general trend toward emphasizing the relation between action and gesture in ontogeny and phylogeny, and on a more theoretical level.

The typified character of actions at this level does not replace local situational relevance (Level 1) as the eventual determinant of meaning, but the typified aspects of actions are rather to be seen as augmenting the concerns of specific situations by virtue of being more or less *situation-transcending* resources for meaning-

<sup>17</sup>A set of distinctions relating to the notion of conventionality are presented in Section 2.3.3.

<sup>18</sup>"X counts as Y in context C" (Searle, 1995, p. 28).

making which are nevertheless employed in specific situations (Linell, 2009, pp. 49). Furthermore, it is not simply the case that such situation-transcending resources and typifications are used as templates for action — they are rather *oriented to* in the sense that strict adherence to them is not necessary (in the end, situational appropriateness has higher priority than, for example, grammatical correctness), and they should rather be thought of as resources that provide frames of interpretation (Coseriu, 1985, pp. xxxiv; Seedhouse, 2007). In a similar vein, instead of speaking of the development of symbolic skills in children only in terms of a one-sided process of *decontextualization* from context-bound skills to generalized abilities and schemas (Werner & Kaplan, 1963), it might be better to speak of *trans-contextualization* (cf. Zlatev, 1997, p. 193), because the latter term captures the aspects of abstracting features of action for use across situation (development as *detachment* and *generalization*) as well as the way any action still needs to be adapted appropriately to the particulars of situations, socio-cultural practices, and general norms of conduct, as evoked in specific situations (development as *attachment* or accumulation of *specification*).<sup>19</sup> No matter how abstract, rule-like, or generalized some aspect of action becomes, it is never applicable to just any moment in any kind of situation. In the words of Schutz (1954, p. 267), “typification depends upon my problem at hand for the definition and solution of which the type has been formed.” Development of action is therefore a two-fold process of increasing degrees of abstractness *and* increasing degrees of specificity and concreteness in skill/knowledge in handling particular types of situations and activities, roughly corresponding to two metaphors of learning: the *acquisition* metaphor and the *participation* metaphor, both of which should be included in an understanding of learning and development (Sfard, 1998). The initial state of learners with respect to some particular not-yet-mastered ability is thus both *non-generalized with respect to general principles* as well as *non-specified with respect to the relevant details*, and the benefit of the concept of typification is that it includes both of these aspects. The emergence of typification in action throughout development is as much a development along the axis of specification in detail as along the axis of generalization and abstraction.

---

<sup>19</sup>Nevertheless, there seems to be a difference between humans and apes when it comes to decontextualization. Petitto (1988, p. 188) reports on her attempts to learn the chimpanzee Nim to make use of a signed language: “Only a few of his 125 signs were ever used regularly (e.g., NIM, MORE, EAT, DRINK, GIVE), and these occurred only in fixed contexts (e.g., eating)”.

**Sem#3 — Differentiated aspects of action (*semiotic signs*)**

At Level 3 the form of the performance of an action is modulated in one of two main ways, or in both of these ways at once. The first kind of modulation is modulation of the act itself: its movement and configurations, with regard to its form as it would have been performed the way one would typically have performed it when performing it for practical purposes. An example is when an act such as writing with a pencil is performed, but where this is done without a pencil in the hand: i.e., as an empty-handed gesture. The second kind of modulation is when a modulation of typicality is a result of factors *outside* the performance of the bodily movements and configurations in themselves, such as various sorts of highlighting, modifications, concretizations or vaguefications, and re-contextualizations of an action, such as may be achieved by means of a concurrent spoken utterance or, perhaps, mere performance of the action in a peculiar context such as in the theater.

The quality of the reduced typicality that results from such modulations can take many forms, such as when only the beginning of an action is performed so as to invoke the action as a whole (*intention movements*, see also page 281), or when an action is performed in its entire duration, but where some aspect is left out that would be required in order to perform it for real, or "seriously" as Piaget (1962 [1946], p. 101) puts it. The modulations need not be subtractive, in the sense of taking something away. The typicality of the action may also be reduced by means of *adding* elements to the performance, such as exaggerations or repetitions of the action that normally do not belong in the action when performed plainly for practical purposes. No matter whether the modulation is achieved by means of additions, subtractions, or both, there is a *reduction* of typicality with respect to some typification, such as how an act would be performed if it were a straightforwardly practical action. This reduction of typicality in Level 3 actions is qualitatively different from the *lack* of typicality in Level 1 actions, since at Level 3 the non-typical aspects of an action stand out as non-typical precisely with reference to the invoked familiarity of some particular typifications. Thus a background of *normal appearances* (Goffman, 1971, pp. 238) of praxis is essential for the possibility of a Level 3 action (cf. *forms of life*, Wittgenstein, 1953), because it is this background that serves to bring about the contrast involved in seeing the modulation aspect of an action that makes it a Level 3 action rather than a Level 1 or Level 2 action.

Precisely due to the achievement of reduced typicality in these actions, the relation between expression and content is now more indirect, and an *as if* quality has entered into the picture — there is now an explicit differentiation between the

expression as such and the content/referent.<sup>20</sup> As argued by Sonesson (2008) in a discussion of iconic semiotic signs, expressions that are *too similar* to what they stand for may, for this very reason, fail to achieve the status of differentiated iconic signs. In such cases, as paradigmatically exemplified by the use of an object to stand for itself, some form of social convention may be required in order for a sign relationship to be recognizable (what he calls secondary iconic signs). “A car, which is not a sign on the street, becomes one at a car exhibition” (Sonesson, 2008, p. 51; see also Sonesson, 1989, p. 139; Sonesson, 1994, p. 279). In cases where there is too much similarity the expression may instead be interpreted in terms of the relationship of identity: i.e., a car on the street may not appear as a representation of a car but it will simply count-as a car — it will present rather than represent.

One common kind of reduction that is of central importance for the whole field of gesture studies is of course the reduction of object-involving actions to *empty-handed gestures*: the primary object of study for this field. However, empty-handedness is only one of several possibilities for reduction of typicality, and it is no way necessary for an act to be empty-handed in order to qualify as Level 3 (Andrén, in press a). The fact that reductions of typicality can also be invoked by means of additions (both of the movements and configurations involved in the action itself and of “external” elements outside the movements of the action such as an utterance or a particular way of coordinating gaze), and not only by subtractions, is crucial because it means that *there is no necessary opposition between instrumental and expressive/communicative aspects of action. These two dimensions may well co-exist in one and the same action* (Andrén, in press a; see also Streeck, 2009b, p. 4, 23, and 82).<sup>21</sup>

Sem#3 actions, in contrast to Sem#1 and Sem#2 actions, are characterized by being *semiotic signs* as a consequence of the indirectness achieved by means of a modulation. The term “semiotic sign” is not used here in the sense of the signs of signed language in a way that contrasts with gesture: hence the use of the label “semiotic sign” rather than just “sign”. It rather corresponds to how some authors use the

---

<sup>20</sup>Ironically, performing an action in such less “typical” ways is in fact one of the main means for invoking an action to signify an *ideal type* itself (Andrén, in press c), to borrow a term from Max Weber (1947 [1922]) who did not refer to ideal in the sense of perfection, but rather to the world of ideas (as opposed to the particulars of tokens). The reason is that the performance may actually be hampered on a token level in the sense that its instrumentality may be reduced, and this effectively reduces the token aspect of the action as well as bringing the type aspect of the action to the foreground.

<sup>21</sup>This is not to say that communicative aspects and practical aspects of action conceptually the same thing. The claim is merely that the presence of one of these two aspects of action in a given action does not require the absence of the other.

terms "symbol" and "symbolic". However, the terms "symbol" and "symbolic" are used in several different ways in the literature on gesture. Sometimes symbols are associated with the differentiation between expression and content/referent (e.g. McNeill, 1992, p. 1). Sometimes they are associated with conventionality (Efron, 1941 [1972], p. 11) or arbitrariness (Bates et al., 1983, p. 77) in contrast to indexicality and iconicity. Sometimes they are associated with the presence of communicative intent (Namy, 2005; DeLoache, 2004, p. 66). A definition is clearly required. The term "sign" will be used here in accordance with the following definition, which is a slight adaptation of the definitions given by Sonesson (2008, p. 49; 2009, p. 138) and Zlatev (2009, pp. 153). This definition furthermore clarifies the use of the term "symbolic" as intended here:

Definition of *semiotic sign*:

A *sign* is present if and only if E (Expression) signifies C (typified Content) and/or R (a Referent in the world) for at least one S (subject) in some context(s) X, if and only if the following three criteria are fulfilled:

Criterion 1: E and C/R are *related* for S in context X: While directly perceiving or enacting E in context X, C and/or R (rather than E) is experienced as thematic. The basis for the establishment of a relation between E and C/R rests on the perceiving of either *indexicality* (contiguity and/or directedness), *iconicity* (similarity), or *symbolicity* (conventionality), or several of these grounds for meaning at once — sometimes multiple times in one and the same sign (cf. Peirce, 1931–35).

*Comment on criterion 1:* This criterion in isolation only amounts to the basic forms of reference invoked by Sem#1 actions or to (symmetrical versions of the) count-as relation (Sem#2), and is not itself sufficient for something to be a semiotic sign (Sem#3). The mere presence of indexicality, iconicity, and/or symbolicity (i.e., conventionality) as such is therefore not sufficient for something to be a semiotic sign.<sup>22</sup> Indexicality is sometimes defined as being constituted by *causality*, but this is not how it is intended here. Following Sonesson (1996), indexicality is rather defined as constituted by *contiguity* and *directedness*, as matters of "the ongoing practice of the ordinary world of our experience, [...] our common Lifeworld". To the extent that "causality" is a

<sup>22</sup>As a consequence, not all meaningful actions are semiotic signs according to the definition provided here.



phenomenon of the Lifeworld, it is not the same thing as causality in the sense of physics. The addition of directedness to the Peircian criterion of contiguity is motivated since “real indicators, such as fingers and arrows, are equally contiguous to a number of objects which they do not indicate” (*ibid.*); and it is often the directedness projected by an index, rather than its contiguity, that accounts for the meaning of this index. In a similar way, Eco (1976) has used the phrase *movement towards* to account for the meaning of pointing gestures. Sometimes *arbitrariness* is used as the criterion for symbolicity, but according to the view adopted here, symbolicity is better characterized by conventionality. Conventionality certainly makes arbitrariness *possible*, but arbitrariness itself is *not required* for something to be symbolic (Calbris, 1990; Zlatev, 2008a; Tomasello, 2008). The meaning of a gesture may be iconically and/or indexically motivated (hence not “arbitrary”, at least in one sense of the term) at the same time as the gesture is established as well-known and conventionalized, with some more or less determined meaning.<sup>23</sup>

Criterion 2: E and C/R are *asymmetrically* related for S in context X: E stands for C and/or R in context X, but C or R does not stand for E (or at least not at the same time).

*Comment on criterion 2:* This criterion in conjunction with criterion 1 amounts to an asymmetrical version of the count-as relationship. Criteria 1 and 2 together are not sufficient for something to be a semiotic sign. Some of the relations, such as contiguity and similarity may intuitively be considered symmetrical, perhaps even per definition. Indeed, it has been argued that iconicity/similarity is a symmetric relation (i.e., that A is as similar to B as B is to A), and thus cannot account for the asymmetric relation between E and C/R in iconic signs (Goodman, 1970; see also Streeck, 2008a, p. 288). This may be the case if we consider this relation in terms of the equivalence relation of formal logic; but from an experiential point of view, E and C/R may well be asymmetrically related for S even though the ground for relatedness is iconicity/similarity (Sonesson, 2008, p. 56).

---

<sup>23</sup>Arbitrariness, in another sense of the term, can be said to be present in any conventionalized entity, in so far as it could have had another form (it may in fact have another form in another culture, to the extent that the meaning of the sign could be said to be “the same” across the cultures), even if some aspects of its form may also be iconically or indexically motivated.

Criterion 3: E and C/R are *differentiated* (Piaget, 1962 [1946]) for S in context X: E is qualitatively different from C and/or R for S in context X. This differentiation may come about through additions of elements that lie outside an action or object conceived as the expression (as discussed above, p. 34), without any modifications of a movement or object itself: thus there may be nothing in the expression itself that tells us that it functions as a semiotic sign on a given occasion (Sonesson, 2007, pp. 13).<sup>24</sup>

*Comment on criterion 3:* This criterion finally marks the distinctive *conceptual* difference between semiotic signs (Sem#3) and the count-as relation (Sem#2), both in its symmetrical and asymmetrical versions. Whether there is such a distinct difference on an *empirical* level between that which is differentiated and this which is not differentiated is another issue. By way of example, the notion of a gradual *distancing* between expression and content/referents in children's development (Werner & Kaplan, 1963) is built on the idea that there may not be a distinct point during this process where a dichotomous switch occurs from strictly "qualitatively similar" to strictly "qualitatively different". I am inclined to agree with Werner and Kaplan on this point. In any case, the conceptual difference should be clear.

All of these criteria are to be interpreted as context dependent and perspectival experiences and expressions on behalf of subjects. In the case where they are part of social encounters, their meaning will to some extent also be a matter of intersubjective negotiation.<sup>25</sup> This stands in contrast to thinking of the meaning of semiotic signs as static entities that belong solely to physical properties of the expressions themselves. All these criteria presuppose subjects with particular *points of view*, engaged in particular activities with their associated relevancies and rationalities, who have come to be what they are due to a previous history of experiences and interactions with the world and other subjects. On the other hand, even though the meaning of semiotic signs is a "subjective" matter, people seem to be relatively consistent in

<sup>24</sup>In non-communicative contexts, the addition may come about purely as a result of thought, without any overtly visible properties. Something may become a sign purely on the basis of considering it as such. The sun may be taken to stand for happiness, even though the sun itself was clearly not created for purposes of having this specific meaning — the meaning it has in this example is purely ascribed to it from the "outside" and is not a part of its constitutive structure.

<sup>25</sup>Even in situations where a subject acts in isolation, the actions performed may derive at least in part from a previous history of social interactions (cf. Schutz & Luckmann, 1989). That is, it is not obvious how to differentiate "social" from "non-social" action.

what they see in certain gestural forms. Aspects of form should not be downplayed just because form does not fully determine meaning.

One seemingly minor but actually quite consequential aspect in this formulation of the definition of semiotic signs, with respect to research on gesture, is that there is an “and/or” inserted between C(ontent) and R(eferent). Instead of requiring that every sign involve *both* a referent and a content, it is useful to detach these two aspects of meaning: hence the “and/or” rather than “and”. For example, even though pointing gestures can sometimes clearly involve typified content, and even though pointing gestures can be very complex in semiotic terms<sup>26</sup>, there are some instances where it is still possible to use pointing to refer to an object, without having any idea whatsoever what the target might be other than where it is located: i.e., without invoking typified content (cf. Wundt, 1973 [1921], p. 75). This kind of pointing is particularly useful for infants and very young children who may use pointing gestures to refer to things still largely unknown to them, to ask their parents what something is, or what it is called, and so on.<sup>27</sup> In such cases, the pointing gesture is effectively relying on the immediate presence of the referent, and it may be less mediated by a typification on the content level. It is rather the pointing gesture itself (i.e., its form) that is typified as well as the schematized relation between the expression and the referent (POINT → Target, where the target may be just about anything). This stands in contrast to what are sometimes called *representational gestures* (Kita, 2000; Nobe, 2000; Capirci et al., 1996; Kendon, 2004) (often defined so as to include so-called *abstract pointing* gestures to non-present entities and qualities, McNeill et al., 1993), *descriptive gestures* (Wundt, 1973 [1921]), or *content-loaded gestures* (Pizzuto & Capobianco, 2005), all of which involve typified content of some sort. It also stands in contrast to considering all pointing gestures as “object-referring terms, akin to nouns or pronouns” (Goldin-Meadow, 2007b, p. 743), since this rests on the assumption that pointing is always used to refer to some-

---

<sup>26</sup>For example, people sometimes point to a target of some sort, which in turn stands for something else: the actual referent rather than the target pointed to (Kendon, 1980a, p. 86). A common practice, at least in Sweden, is pointing to a wristwatch (or where one would be) to indicate that time is running out (rather than just referring to a watch as such). Pointing gestures may also invoke various complex spatial frameworks as frames of reference (Haviland, 1993, 1998). Speakers of Arrernte, a central Australian (Pama-Nyungan) language, make use of a singular/plural contrast through different conventionalized forms of pointing (Wilkins, 2003). Pointing gestures may also be used to indicate the past or future with reference to cultural metaphors that take the future as “ahead” and the past as “behind”, or vice versa in cultures that construe the flow of time differently (Núñez & Sweetser, 2006; McNeill, 2005, p. 46).

<sup>27</sup>The episodes from Superman where people point to “something” (which happens to be Superman) in the air as part of their attempts to identify this “something” also comes to mind.

what known items (especially for nouns): i.e., to point to a target *X as an instance of* category Y, where Y is a known category.

It is possible to interpret the concept of "content" in other ways, so that content would always be considered to be present in a semiotic sign, but the interpretation I have offered is suitable precisely because it allows for discussion of gestures that qualify as semiotic signs according to the definition given here, both in terms of what they share — they can all be characterized as semiotic signs by means of the three criteria above — and in terms of how they differ — they may primarily be invoking a referent or they may invoke both a referent and a content. This distinction is further relevant because there is an overall development from more exclusively referential modes of expression into more content-loaded *and* referential ones in children's development, and the first communicative acts, such as pointing, giving, and showing are all of a more referential kind (e.g. Bates et al., 1979; Caselli, 1990).<sup>28</sup> Long ago, Wundt (1973 [1921], p. 75) argued in a similar way regarding the less demanding nature of (some) pointing gestures for children:

Their greater primitiveness [of pointing gestures compared to content-loaded gestures] is easily explained by the psychological conditions of their origin. Where an object referred to by a given gesture is in the realm of visible things, direct pointing at it with the index finger is the simplest, surest, and most straight-forward way of drawing attention to it.

Content-loaded (or representational) gestures themselves may vary in specificity. That is, they may vary with regard to whether they primarily serve to express a general concept (content), not referring to any specific referent, or whether they primarily serve to express a more particular referent. As an example, an iconic gesture may depict the general shape of an arched vault; not referring to any particular vault, or it may depict a particular vault that actually exists at a certain location (i.e., a referent) (cf. Fricke, 2009). Content-loaded gestures referring to a specific referent thus include content to the extent that these gestures involve a representation of the referent, and not just a reference to the uniquely present referent in the current situation as may the case may be for pointing gestures.

Finally, once expression and content/referent have been "detached" by some kind of differentiation, actions may be changed along a veritable array of further dimensions of variation into ever more complex and abstract forms of semiotic signs.

---

<sup>28</sup>The development from more indicative modes of displays to more representational ones may sometimes also be observed within single situations (Hirsch, 1994, p. 481; LeBaron & Streeck, 2000; Roth, 2002).

Content may become metaphorical (McNeill & Levy, 1982; Cienki & Müller, 2008) or metonymic so that a gesture invokes some content which in turn stands for something else (cf. the concept of a “base”, (Kendon, 1980a; Bergman, 1982; Müller et al., manuscript)). The techniques of realization involved in the articulation of various expressions (Kendon, 2004; Müller, 1998a; Wundt, 1973 [1921]) may depart from action-based logic into much more indirect forms of realization, where the hand is no longer seen as a hand. The space in which the action is performed may change from the concrete *action space* (where the actor is directed to the actual physical surroundings), to a *supporting space* (where non-present entities are invoked with support of present entities: a real table may serve as an ocean on which a boat is sailing), to a purely *abstract space* (where there is no relation between the signified and the actual current physical surroundings in which the expression is articulated).

As expressive actions begin to be detached from the concrete spatio-temporal domain of practical action in the world within reach, i.e., the core of the Lifeworld (Schutz & Luckmann, 1973), there are increasing degrees of freedom in the many ways in which semiosis may be achieved. Therefore it might be said that semiotic complexity explodes into a multidimensional progression along many different dimensions at this point in the classification of levels of semiotic complexity. From this point on (if not before), it is not possible to arrange the increasing semiotic complexity along a single dimension of variation. Thus the classification into levels on an ordinal one-dimensional scale stops here. Another reason is that actions that qualify as Sem#3 fulfill criteria that most researchers would consider sufficient to qualify as gesture, whereas Sem#1 actions clearly lie below the lower limit. The status of Sem#2 actions is more unclear, since the count-as relation involved in Sem#2 is typically not distinguished from Sem#1 and Sem#3 in the accounts of other researchers. Hence, the levels of semiotic complexity presented here provide an analytic vocabulary that covers the (continuous) span of action that lie below, “around”, and above the lower limit of gesture.

It may be worth repeating that although Sem#3 actions rely heavily on a world known in common and taken for granted in various ways (i.e., the Lifeworld), it is not the case that semiotic signs only draw on pre-existing meaning. The act of producing a sign is always in itself also a meaning-creating act — an act of construal which leaves some things out and emphasizes others, and which may have unintended and unexpected consequences (cf. *communicative dynamism*, Firbas, 1971), as pointed out in the discussion of Sem#1 actions. However, to some extent the same goes for count-as actions (Sem#2 actions). An action may itself serve as an establishment of a *type* of action, even though there was no such type in existence

before the action was performed, and so it creates the category of which it will itself count as token. A case in point is when a type of action is demonstrated to another person; "you unlock the door like this+GESTURE".

### 2.2.4 Children and the lower limit

There are numerous studies that confirm that all the levels of *communicative explicitness* are already often achieved when children are 18 months of age or younger. This is evidenced with respect both to children's performances and comprehension of communicative actions and to parents' comprehension of children's actions (e.g. Bates et al., 1975, 1979; Masur, 1983; Masataka, 2003; Behne et al., 2005; Liszkowski et al., 2006; Gräfenhain et al., 2009). Since this thesis takes the age of 18 months as its starting point, the question of when this sort of sensitivity to the intentional structure of communicative actions emerges is not a central concern, although it is clearly one of the main concerns of many studies on gesture in younger children. In a similar way, numerous studies confirm that most children are capable of performing actions up through Sem#3 actions of *semiotic complexity* before 18 months, both when it comes to gesture (e.g. Bates et al., 1979; Volterra & Erting, 1990; Iverson & Goldin-Meadow, 1998) and speech (e.g. Bates et al., 1988).

Abilities to anticipate goals of familiar and typified everyday actions (action intentions) emerge already in infants first year of life, even though what is familiar or not to an infant primarily depends on experience and only indirectly on age as such (e.g. Gredebäck & Melinder, 2010); and infants are obviously not yet familiar with many sorts of actions (e.g. the notion of life-time learning). The various distinctions in terms of levels that have been presented in connection with the issue of the lower limit of gesture are thus not primarily used in this thesis to analyze stages in children's development. Instead, they are used first and foremost for descriptive purposes, to characterize a variable array of actions in the repertoires of children all of which may be present at one and the same point in time.

### 2.2.5 Defining gesture in the context of the lower limit

In consequence of the distinctions made so far, it becomes clearly possible to conceive of gesture as constituted by properties that belong to two separate continua — one continuum of communicative explicitness and one of semiotic complexity. Even though these two continua are closely related in practice, they do not necessarily overlap or follow each other. That is, already at Level 1 of semiotic complexity, an

expressive action may in principle qualify for Level 3 (or higher) of communicative explicitness. Something which is far from being a sign, can still be relatively explicit in communicative terms. The opposite also holds true. An expressive action can be semiotically complex in, without being part of the communicative organization of a social encounter. In order to be clearer about which aspects of gesture are discussed in later chapters, a few terms other than “gesture” will be used when it is necessary to be more specific about a particular expressive action:

**ACTS OF BODILY COMMUNICATION (ABC):** Actions that qualify for Level 3 of communicative explicitness (visibly other-oriented action). Such actions need not be semiotically complex, although they can be.

**ACTS OF BODILY SIGNIFICATION (ABS):** Actions that qualify for Level 3 of semiotic complexity (signs, as defined in Section 2.2.3). Such acts are not necessarily part of the communicative organization of a social encounter, although they can be. Those ABSs that are not correspond to the class of actions sometimes referred to as private gesture (e.g. Rodríguez, 2009).

**ACTS OF BODILY COMMUNICATIVE SIGNIFICATION (ABCS):** Actions that qualify *both* as ABCs and as ABSs. ABCSs are thus instances of *communicative signs*, rather than being communicative only (ABCs) or signs only (ABSs). Communicative signs need not be instances of *bodily* expression; they can also be realized through speech, writing, pictures, and other expressive modalities (semiotic resources), as well as through the co-expressive synergies of multiple modalities.

**GESTURE:** Actions that are either ABCs, ABCs, or both (i.e., ABCSs). This captures the common but somewhat ambiguous usage of the term “gesture”. The term “expressive” will be used to capture the ambiguity between *signification* (even when it is performed for the self) and *communication* (with an orientation to the Other), so that an expressive act is either (at least) a Sem#3 action or a Comm#3 action (or both).

As mentioned before, these conceptual distinctions and definitions should be understood precisely at a conceptual level. To what extent actually occurring actions lend themselves to clear-cut conceptual classification of this kind is an empirical question. Furthermore, it is not only an empirical problem for scientists, from within the scientific attitude; it is also to some extent a practical problem that faces participants in everyday interactions, from within the natural attitude.

## 2.3 Unpacking the upper limit

In Section 2.1 it was stated that many researchers treat gesture as something different from signed language, even though gesture (so conceived) and signed language also have many similarities. This distinction between "gesture" and "language" was termed *the upper limit of gesture*. One essential underlying similarity between gesture and signed language is that both are instantiated by the same bodily expressive medium, meaning that they share a set of similar conditions and potentialities for expression (Kendon, 2004, Chapter 14–15). Furthermore, both the signs of signed languages and at least some gestures (i.e., ABSs) satisfy the positive criteria for being *semiotic signs* (according to the definition provided in Section 2.2.5).<sup>29</sup> The differences between gesture and signed language are primarily formulated in terms of properties that are lacking in gesture, i.e., negative criteria, but that are present in (signed) language. An example of such a characterization of gesture is made by McNeill (2005), who describes a spontaneously created gesture, corresponding to what he calls *gesticulation*.<sup>30</sup> He then uses this gesture as a principal example of a kind of gesture that is different from more language-like forms of bodily expression such as conventionalized gestures and signed language (*ibid.*, p. 7, italics added):

The bends-it-back gesture *lacks all linguistic properties*. It was *non*morphemic, *not* realized through a system of phonological form constraints, and had *no* potential for syntactic combination with other gestures.

This means that the question of differences and similarities between gesture and signed languages is tightly coupled with the question of how to define language (and gesture). Language definitions almost always consist of *several different properties* that collectively make some expressive repertoire a language (Saussure, 1983 [1916]; Hockett, 1966). However, some of these properties, such as *differentiation* between expression and content or the fact that signs may be *conventionalized*, may be present in a gesture without any implications of other language defining properties, such as *systematicity*.<sup>31</sup> Thus, conventionalized acts of bodily signification will have some but not all properties typically associated with language. These ges-

<sup>29</sup>Positive criteria are criteria formulated in terms of the presence of a certain property (what gesture is), as opposed to negative criteria that concern the lack of a certain property (what gesture is not).

<sup>30</sup>Authors who have used the term "gesticulation" before McNeill include Efron (1941 [1972]) and Kendon (1980b).

<sup>31</sup>Properties such as *arbitrariness* (Saussure, 1983 [1916]) and *displacement* (Hockett, 1966; see also Blake, 2000, p. 78) may also occur in gesture without implications of a necessary presence of systematicity. But in terms of this thesis, arbitrariness and displacement are considered as emerg-



tures are often called *emblems* in the gesture literature (Efron, 1941 [1972]; Ekman & Friesen, 1969; McNeill, 1992). A repertoire of conventional gestures without much systematicity, either on the level of form or on the level of meaning, correspond roughly to what is sometimes called “protolanguage” (Corballis, 2002; Bickerton, 2003; Zlatev et al., 2008a, p. 221). As long as *systematicity* is not present in a repertoire of signs, most gesture researchers do not consider it to be a language. What this means that some properties that are commonly part of language definitions actually occur *below* the upper limit gesture.

Indeed, *systematicity* is the property typically invoked as definitive of the upper limit of gesture — the boundary between that which is considered to be gesture and that which is considered to be signed language (e.g. McNeill, 2000, 2005; Zlatev, 2008a,b). Yet, systematicity is a multifaceted and heterogeneous notion that includes a range of different conventionalized aspects of expression such as word-order patterns, morphology/inflection, parts of speech, double articulation, contrasts between forms as well as between meanings in the system, and so forth. Even though it is not always clear where exactly to draw the line between having and not having some or all of these properties in particular cases (Kendon, 2004, Chapter 14–15), it is clear that what is usually thought of as language has a lot of these properties, and that gesture has far fewer of them. The difference is thus at least relatively clear on a conceptual level, though less so on an empirical level due to the many intermediate forms of systematicity that exist (*ibid.*).

In any case, the issue of the upper limit of gesture is relevant to this thesis, but not in the same way as the lower limit. The reason is that the children studied here are all hearing-enabled children who have not been exposed to signed language, which means that they are not, after all, really crossing the “border” of the upper limit in the sense of learning a full-fledged signed language. Therefore, there is no real reason to assume that any of these children’s more complex bodily expressions ought to be classified as anything other than “gesture”. What is of interest here is rather if and how, and to what extent, there are rudimentary language-like properties in the gestures of these children. The two main topics that will be analyzed in this thesis, with respect to the upper limit, are different forms of *conventionality* in the children’s gestures; and second, the nature of different *combinations* of expressive units within, and across, the expressive modalities of gesture and speech.

---

ing from the more basic properties of differentiation and conventionalization. Displacement is the possibility of communicating about things beyond the here and now.

### 2.3.1 The status of conventionalized gestures

In the quote from McNeill (2005) above, a gesture was described that served to exemplify what McNeill calls a gesticulation: gestures that lack "all" linguistic properties. In his further discussion, McNeill adds: "There are no conventions telling the speaker what form bending back is to take" (*ibid.*, p. 10); hence he makes a distinction between the non-conventional gestures that he call gesticulation and the conventionalized gestures he calls emblems (cf. Efron, 1941 [1972]; Ekman & Friesen, 1969; Johnson et al., 1975; Morris et al., 1979). McNeill's research is concerned with gesticulation (*ibid.*, p. 4), but he clearly acknowledges the existence of emblems as existing between gesticulation and language — close to the notion of "protolanguage" (Bickerton, 2003; Zlatev, 2008a, p. 221) as mentioned before. He further acknowledges that conventionalized gestures may be used in concert with speech (what is called *co-speech gesture*). McNeill (2005, p. 12) writes:

Emblems are at an intermediate position on the various dimensions of contrasting gestures. They are partly like gesticulations, partly like signs [of signed language]. For many individuals, emblems are the closest thing to a sign 'language' they possess, although it is crucial to emphasize the non-linguistic character of these gestures: they lack a fully contrastive system [...] and syntactic potential (the impossibility of combining two emblems into a gesture sentence).

The term "emblem" refers specifically to those gestures that are *word-like* in the sense that a specific gestural form is related to a more or less specific content by means of conventionality (Efron, 1941 [1972]). Just as there are dictionaries for words, there are dictionaries of emblems as well as documentation on how their meanings differ between cultures and geographical areas (e.g. Efron, 1941 [1972]; Saitz & Cervenka, 1962; Green, 1968; Morris et al., 1979; Payrató, 1993). Ekman & Friesen (1969, p. 63) also make use of the term emblems, and define it as "those nonverbal acts which have a direct verbal translation, or dictionary definition, usually consisting of a word or two, or perhaps a phrase." Neuropsychological research indicates that deaf people who use American Sign Language "process" emblems similarly to ASL signs and that this differs from the "processing" of iconic gestures (Husain et al., 2009).

There are, however, different degrees of conventionalization in gesture, and gestures may be conventional in different ways (Kendon, 2004; Andrén, in press c). In effect, not all aspects of conventionality in gesture fit nicely into the concept of emblems. Therefore, I will treat "conventionalized gestures" as a broader notion than

“emblems”. Indeed, the notion of conventionalized gestures is broader than the notion of emblems even when the considerations of conventionality are restricted to the conventionality present within the sign relationship, in the form of (A) conventionalized form, (B) conventionalized meaning, (C) conventionalized relations between form and meaning, and (D) conventionalized contexts of use. Emblems are typically conventionalized with respect to all four of these types of conventionality, whereas the notion of conventionalized gestures, as used here, does not require all of these aspects to be present. Apart from the conventionality that may be present within the sign relationship, gestures may also be sensitive to conventions at other levels such as norms of proper conduct (cf. Kita, 2009). In addition, one could distinguish between different levels of conventionality, and such a classification is presented below. It should further be pointed out that when the term “conventional gesture” *it is not implied that the gesture is only conventional — the gesture may also invoke indexicality and iconicity as grounds for meaning*, in accordance with the definition of semiotic signs (Section 2.2.3).

McNeill’s way of conceptualizing gesture has become very influential, especially in psychologically oriented research on gesture. It could be argued that in research that adopts this conceptualization there is a tendency, at least implicitly, to equate the otherwise more general notion of gesture with McNeill’s more narrow notion of gesticulation, amounting to an exclusion of considerations of conventional aspects of gesture. McNeill’s project of bringing the dynamic and creative dimension (the non-conventionalized dimension) of gestural expression into the foreground certainly has its merits. Above all this is so because McNeill’s perspective provides an antidote to one-sided structuralist approaches to gesture only in terms of “pre-established” forms and meanings. However, as a result of this conceptualization the pendulum sometimes swings too far, and there may be a bias towards downplaying the role of conventionality in gesture in general and co-speech gesture in particular. At worst, it may result in somewhat homogenized and sometimes even circular statements about the nature of gesture, because conventional aspects of gestures are often excluded from the analytical scope at the outset. Kendon (1996) makes a similar point with respect to McNeill’s way of conceptualizing gesture:

It is often said that gesticulation is idiosyncratic, each speaker improvising his own forms. So far as I know, no one has ever really tested this claim. My own experience in gesture-watching suggests to me that people are far more consistent in what they do gesturally than this ‘idiosyncrasy’ claim would lead one to imagine. One’s own experience in noticing differences in ‘gesture style’ from one culture to another, the work

of David Efron (1941 [1972]), etc. actually confirms this point. It suggests that [...] there is conventionalization to a degree affecting all kinds of gesturing [...].<sup>32</sup>

Kendon (2008, p. 360) also points out that "this exclusion of so-called 'emblems' has had the unfortunate consequence that they have been little studied". In fact, McNeill (2005, p. 51) seems to agree with this, and suggests that "to settle the form question, and at what level conventions apply, several probes of conventional forms could be attempted. [...] Until very recently, no one has used these probes and the answers to such questions have not been sought".

Still, the blind spot in the research territory suggested in these quotes from McNeill and Kendon is not quite as severe as the blind spot surrounding some of the expressive acts that reside around the lower limit of gesture. Most gesture researchers do, after all, clearly acknowledge the existence of conventionalization in gesture without equating such gestures with language, even though many researchers then take McNeill's lead in restricting their studies so as to exclude conventionalized aspects of gesture from their analytical scope.

Furthermore, McNeill (2005, p. 7; see also Morris et al., 1979, p. xvi and Brewer, 1951, p. 234) characterizes the relation between emblems and speech as "optional": conventionalized gestures may be performed without any co-occurring speech and still be clearly intelligible due to their pre-established and commonly-known character. For this reason emblems have sometimes also been called "autonomous gestures" (Kendon, 1985a, p. 8; Payrató, 1993). For non-conventionalized gestures, the situation is often different. It has been found that many gestures that are produced in concert with speech (*co-speech gesture*) are difficult to interpret without this co-occurring speech. They are, so to speak, designed to be expressive in concert with speech and may therefore not be shaped so as to be self-sufficient autonomously expressive units (cf. McNeill, 1985, p. 351; Singleton et al., 1995; Goldin-Meadow et al., 1996; see also Kendon, 2004, Chapter 14). Nonetheless, there are cases where gestures *without conventionalized properties* may function in an autonomous fashion. One case is when gestures are inserted into a "syntactic slot"

---

<sup>32</sup>The "idiosyncrasy claim" refers to McNeill's (1992; 2005) conception of "gesticulation" as those gestures that are created by speakers rather than being culturally spread conventions. McNeill's use of the term "idiosyncratic" seems slightly ambiguous though. Sometimes it seems to refer to gestures that are created in the moment (McNeill, 2005, p. 8), and sometimes it also seems to include gestures that are not conventional in the sense of being socially shared forms of expression, but that may still be some sort of individual image-schematic habits (McNeill & Sowa, 2007), hence not necessarily creations of the moment although they may at least in theory be restricted to one individual (see also Parrill, 2007).

in an utterance where a word has been, so to speak, left out. For example: when saying “he went [GESTURE] a few moments ago” (Slama-Cazacu, 1976; Ladewig, 2010). It could be argued that this is a kind of “co-speech” gesture too, even though the gesture may not overlap in time with any of the words, because the meaning of the surrounding speech contributes to the interpretation of the gesture. There are also cases observed in the data studied in this thesis of gesture-only utterances without any speech at all, where the gestures performed are not conventionalized. These may provide a better example to make the point that *non-conventionalized gestures may be used as self-sufficient autonomous utterances*.

Let us return to conventionalized gestures and their relation to speech. Characterizing the presence of speech in co-occurrence with conventional gestures as “optional” may give the unfortunate impression that conventionality in co-speech gesture is a marginal phenomenon and that it would somehow be atypical for conventionalized gestures to be “designed” to be usable in concert with speech (cf. Kendon, 2004, p. 106; Seyfeddinipur, 2004). I suggest, therefore, that the use of the term emblem as a general label for conventional gestures is not productive, since it tends to lead to an exaggerated emphasis on the “autonomous” character of (all) gestures with conventionalized properties, and in effect, an exaggerated emphasis on the differences between conventional gestures and the co-speech gestures that McNeill calls gesticulation.<sup>33</sup> In this spirit, Kendon argues that:

Observations of speakers in communities where there is a considerable repertoire of such [conventionalized] forms show that *their use is often fully integrated into the flow of everyday discourse* and that their sharp separation from non-conventionalized forms that many gesture-classification systems imply, cannot be sustained. (Kendon, 2008, p. 360, italics added)

[...] it is far from being the case that gestures that are always associated with speech are less conventionalized than those that are not. (Kendon, 2004, p. 106; see also Kendon, 1984, p. 94)

Arguably, such coordination of conventionalized gestures with speech is found not only in “gesture rich” cultures, of which southern Italy is perhaps the prototypical example (cf. Kendon, 1995, 2004; Iverson et al., 2008). The presence of various forms of conventionality in co-speech gesture may be more frequent and general than is commonly assumed to be the case. One recent description of conventionalized co-speech gestures can be found in the work of Harrison (2009b). Harrison

---

<sup>33</sup>This is not to deny that there is a class of conventional gestures that can be used autonomously.

documents the performance of conventionalized gestures of rejection and negation as used in tight coordination with speech among adult native speakers of English. Harrison shows how various conventionalized gestural expressions of negation may be temporally organized so as to cover the negated parts of syntactic clauses in the spoken utterances (what Harrison calls the *scope* of negation). That is, even though some of the gestural forms that Harrison describes may be potentially understandable, at least to some extent, without speech, it is misleading to characterize the presence of speech in relation to gestures with conventionalized aspects as merely "optional", as this makes the ordered performance of conventional gestures in coordination with speech seem almost superficial.

There is indeed quite a large number of gestures that are conventionalized in some respect, and that are frequently, or even typically, used in coordination with speech. Examples include gestures that perform such "pragmatic functions" as marking an utterance as a question, handing over the speech turn to someone else, holding the turn, or even presenting an abstract idea (e.g. Duncan, 1972; Kendon, 1995, 2002, 2004; Müller, 2004; Neumann, 2004; Seyfeddinipur, 2004; Mondada, 2006; Seo & Koshik, 2010). Other gestures that are routinely coordinated with speech include highly frequent and conventionalized gestures such as *head gestures* including nodding, head-shakes, head-tosses, and so forth (Darwin, 1872; Morris et al., 1979; Junefelt, 1987; Anderson & Reilly, 1997; Kendon, 2002; McClave, 2007); and *pointing gestures* which are at least partly conventional in form (cf. McNeill, 2005, p. 12), as shown by studies of different forms of pointing in different cultures (Sherzer, 1972, 1991; Johnson et al., 1975; Enfield, 2001; Kendon & Versante, 2003; Wilkins, 2003). Blind children are, for example, not as inclined to use the index finger form of pointing as sighted children are (Junefelt, 1987; Iverson, 1998), which may be interpreted as a lack of conventional "domestication" of their pointing gestures due to the absence of visual experience of others' gestures. In a longitudinal study of three Swedish and three Thai children between 18 and 28 months of age (Zlatev & Andrén, 2009), pointing and conventional gestures (in particular nodding and head-shakes) were by far the most frequent gestures expressed in both cultures.<sup>34</sup> Similarly, in a study of five bilingual (English and French) children between 24 and 42 months of age Nicoladis et al. (1999, p. 518) report that "the most frequent type of gesture produced by all the children with English and French utterances was the point, followed by emblem and give gestures"; none of these gestures are of course spontaneous creations of the moment. Findings such as these

---

<sup>34</sup>The three Swedish children that were part of the study carried out by Zlatev and Andrén are also part of the data analyzed in this thesis.

have lead Guidetti & Nicoladis (2008, p. 109) to suggest that:

If our reasoning is correct, then infants may use primarily conventional gestures, as well as gestures that they have learned by acting in the world (such as ‘pick-me-up’). There is a curious lack of the predominant kind of gesture used by adults: spontaneous, non-conventional gestures that seem to be created on the spot to convey meaning [...].

As will be seen, this quote also finds support in the analyses presented in several of the chapters of this thesis, although it should also be added that indexicality (in particular) and iconicity will play important roles too. To be clear: the aim of this thesis, with respect to the upper limit of gesture, is not to provide a one-sided emphasis on conventionalized aspects of gesture, or to say that such conventionalized aspects have no implications on how gestures can be used with or without speech; rather the aim is to put conventionality on a more equal footing with indexicality and iconicity in the understanding of gesture in general and of co-speech gesture in particular. To make further progress in research on conventionality in gesture it may be a good idea to distinguish between different kinds and levels of conventionality. But before presenting a number of such distinctions, it may be useful to offer a few considerations on how preexisting conventions might be learned.

### 2.3.2 From imitation to imitative processes

What are the processes by which conventions that pre-exist the individual may be learned? A first distinction to make, raised in the quote from Guidetti and Nicoladis above, is the one between *imitation* and *ritualization* (cf. Tomasello, 1996). Imitation and ritualization constitute two analytically distinct routes to the emergence of socially schematized and conventionalized actions: the first in a symmetrical way, the other one in an asymmetrical way, as described below.

Imitation (or “copying”) refers to the skill of appropriating actions performed by others into one’s own repertoire. When imitation is understood as an autonomous individual ability, it presupposes the ability to conceive of the action of the Other as a *type*, rather than as a unique token. That type is “distinct from the body of the other in its specificity, so that it can be repeated by the self” (Sonesson, 2007, p. 116). Imitation, in a more general sense, is not always a result of purely individual abilities though, as it may also involve the active assistance of others (Zukow-Goldring, 2006; Zukow-Goldring & Arbib, 2007). That is, *imitative processes* may involve what Vygotsky (1962, 1978) has called *scaffolding*, whereby a child manages

to do things beyond the capabilities of the child as an individual thanks to the support of someone else. In a similar vein, Hay & Murray (1982) performed a study on 12-month-old infants and the results suggested that "the modeling of a social action alone, without explicit prompts for imitation and/or a game-like pacing of modeled events, may not be sufficient to induce infants to perform that action" and that "the impact of the social behaviors infants see modeled is likely to be tempered by the network of interactive events in which they are embedded." Imitative processes are *symmetric*, in the sense that the outcome of the process is the imitator doing something from the repertoire of someone else, not by accident but through some motivated process. One may distinguish between different kinds of imitative processes, such as imitating the *form* of an action versus imitating its *goal* or *result*; and some authors have made even finer-grained distinctions (Whiten et al., 2004). It is not necessary that the imitated action be a conventionalized one; in principle any (not overly acrobatic) action can be imitated. It is, however, clear that imitative processes are of fundamental importance for the possibility of conventionality, since the socially shared (i.e., symmetric) character of conventionality presupposes the possibility of copying others' actions (Zlatev, 2005; Ziemke et al., 2007).<sup>35</sup> Hence, processes of imitation are one way in which socially schematized and typified actions may emerge in the child.

Ritualization refers to social learning that is not based on copying others actions, but that is rather a matter of habitual action emerging from routine activities in which the participants play complementary roles. Returning to the example from the Guidetti & Nicoladis (2008) quote above, after a parent has lifted an infant from a chair many times, the child may start raising her arms in anticipation, and finally start lifting the arms as a communicative act that signals to the adult that the child wants to be picked up. In this example, it is quite clear that the emergence of pick-me-up is not a matter of symmetric copying, but rather of a schematized and *asymmetric* interaction where the child and the adult plays complementary roles. It is likely that most children have never observed an adult lifting his or her hands to be picked up in this way.

As stated above, imitation and ritualization are analytically distinct routes to social learning, but this does not mean that particular skills, especially those that are learned over time, result from *either* ritualization *or* imitation but not both. Both

---

<sup>35</sup>Not all conventional relationships are symmetric in this simple sense. Conventions may also concern asymmetric relationships, such as the way younger people are supposed to address elderly people. In such cases it would be wrong of the child to act like an elderly person in addressing her, i.e., it would be wrong to "copy" the way an elderly person acts.



kinds of processes may be involved. A possible candidate for such a “dual” explanation is the emergence of pointing in infants. This is commonly viewed as an either-or question (e.g. Tomasello, 2008, p. 112) — is pointing a result of ritualization (from reaching) or is it a result of imitating what others do? In a similar way, one could ask the more general question — Are conventions learned through imitation or through ritualization? — and argue that the answer is a bit of both. While it seems fairly evident that imitative processes are involved, children do not simply hear conventionalized words and see conventionalized actions and gestures and learn solely from this input; they also get a lot of feedback from others in response to what they do. In this way, a heterogeneous array of complementary/asymmetric actions (actions other than the target action) figure as components in the process that lead to the eventual copying of the target action.<sup>36</sup> Junefelt (1987) describes how a blind child was explicitly socialized into using index finger pointing and other conventional gestures such as head-shakes and nodding, demonstrating how the imitative “transfer” of various actions may be implemented in alternative ways, that still end up in a learned convention. Hearing- and sight-enabled 16-month-olds have been shown to imitate actions selectively depending on how the act is embedded in an interactive sequence (Carpenter et al., 1998). The children in the study were much more inclined to imitate actions produced together with the exclamation “There!” than the same actions when produced with the exclamation “Whoops!”. That is, factors “outside” the action to be imitated are also part of the process that leads to successful imitation.

In sum, the emergence of conventionalized actions in the repertoires of children should not be understood as resulting from a single process. It may be more useful to characterize the abilities of children for learning conventions as an ability to integrate a heterogeneous array of quite disparate forms of “evidence”, on many levels, into coherent conduct. The notion of *imitative processes*, as opposed to the narrower concept of imitation, will be used here as an umbrella term for this heterogeneous

---

<sup>36</sup>An interesting case in this regard is the deaf children of hearing parents who do not teach their children signed language (called *homesign*). These children tend to create rudimentary signed languages of their own (Heider & Heider, 1941). These children are often said to “lack input” in the sense that they lack input of a conventionalized language (Goldin-Meadow & Feldman, 1975; Goldin-Meadow, 2003b), i.e., symmetrical input of target actions. That is of course true to a considerable degree, especially for the auditory modality. However, there is inevitably a lot of asymmetrical input in the form of all sorts of responses to the children’s actions (Da Cuhna Pereira & De Lemos, 1990) as well as symmetrical input in the form of gestures and actions that may be imitated. These other forms of input are potentially of huge importance for the children’s ability to “create” an expressive and stable repertoire of signs, along with “rules” for combining these signs, even though the repertoires do not correspond in any particular way to the language spoken by their parents.

set of factors that contribute to the transmission of conventions. Strictly speaking, some specific imitative processes do not by necessity include direct imitation as a component. Someone may, for example, teach someone else how to do a target action, not by showing them the action itself, but by telling them how it is done, or helping them in other ways to perform it.

### 2.3.3 Levels of conventionalization

The term *conventionality* is used in many different ways. Some bear of a resemblance to concepts such as sharedness, normativity, regularity, recurrence, formality, routine, habituality, artificiality, and so forth, but these concepts are only partially overlapping in meaning. An essential property running through all the senses of "conventionality" discussed in this section is that they concern *recurrent* patterns that recur *across individuals* that are members of the same group(s), but not across individuals of different groups (other than by co-incidence). To gain a bit of precision, one can distinguish between at least four levels of conventionality that differ in their degree of explicitness. Level 1 is least explicit, and Level 4 is most explicit. On an empirical level, these levels may represent more of a continuum, although on a conceptual level, they are distinct.

#### Conv#1 — Normality

Since members of the same group(s) share many things by virtue of engaging in similar kinds of activities, in similar settings (cf. *forms of life*, Wittgenstein, 1953; *communities of practice*, Wenger, 1998), there is a level of socially shared regularities which is essentially a matter of *normality* (cf. *normal appearances*, Goffman, 1971, pp. 238; *body techniques*, Mauss, 1979 [1936]; *habitus*, Bourdieu, 1977), although they may not appear "normal" or "typical" from the point of view of members of other groups (cf. Schutz, 1944, on the structure of being a stranger). Such assumptions about normality may go entirely unnoticed unless the normal order is somehow disturbed (cf. Garfinkel, 1964, p. 236). Normality, as defined here, is to be understood as the lack of the atypical rather than as some explicitly specified experiential structure. Consider how the train drivers in the London Underground may notice that a station platform they are approaching is particularly busy or empty. Crucially, this is possible only against some sort of background assumptions of what constitutes "normal activity" on a platform (Heath et al., 1999, p. 563), however vague these assumptions may be. Members of a group do not need to be meta-aware

of these things: they need not know that they know them. They are typically, rather, a matter of pre-reflective knowledge.<sup>37</sup> Normality as such does not directly specify what members “ought” to do: if some atypical phenomenon would start to appear on a more regular basis, it would soon become typical, no matter how undesirable or “wrong” it may be judged to be. Deadly violence is an everyday experience for some groups of people (whereas for many others, it is not). Already this level of conventionality presupposes a certain degree of *recurrence* of some phenomenon. Recurrence is a property shared by all the levels of conventionality presented here, despite the other differences between the levels that make them conceptually distinct.

A “group” may be anything from a mother and a child, a family, a couple of friends, colleagues at work, to the citizens of a certain city or country. The only requirement for something to be or have the potential of becoming a group is that there is at least some minimal, and sometimes very indirect and mediated, contact between the members. Any individual person is a member of a large number of groups, making it problematic to conceive of people as members of a single group or culture.

#### Conv#2 — Typification

At this level, acting in certain ways does not just appear as more or less “normal”. Actions on this level are acts of this or that *type* for the members of a group, i.e., a shared (and context-related) *categorization* has taken place. When a Conv#2 action is performed, it *counts as* a token of a certain type of action. The knowledge involved here is therefore of a more structured and typified kind than on the previous level. This is in line with Bach & Harnish (1979, p. 108) who state that “for us, conventions are *counts-as* rules and nothing else” These “rules” specify only that something is an instance of a type, with no implication whether performing a certain typified action in context X is socially obligatory (cf. *ibid.*, p. 121), and no implication that deviations from typified ways of performing the action will necessarily be perceived as wrong. Cases where such restrictions apply belong rather to the next levels (Conv#3 and #4).

Consider when someone presses the power button on a TV. This counts not just

---

<sup>37</sup>Something that may be conventionalized at this level are those phenomena that are studied under the label of *proxemics* (e.g. Hall, 1963; Mehrabian, 1971; Rosenfeld et al., 1976; Shuter, 1976), namely issues related to the appearances and sustainment of certain interpersonal distances in social encounters as well as the cultural differences that exist in these respects.

as a normal way of acting, but more precisely as "turning the TV on".<sup>38</sup> Pressing the power button on a TV is *not* typically seen as some random movement. It is more likely to be taken as an action that relates to the typical uses and rationalities associated with the power buttons on a TV. Many mundane everyday actions are of this recognizable, seen before, familiar, and already established kind: they are tokens of types. People frequently operate on the (often implicit) assumption that others will see such actions in this light. Of course, it may be less clear what is going on from the point of view of someone who is not familiar with the workings of a TV, such as an infant or a person from a culture where TVs are unknown. Conv#2 actions are usually a lot easier to express succinctly than Conv#1 actions, and some may even have names (usually as verbs) in ordinary language. When one member of a group tells another member of the group that some person "turned the TV on", the other will immediately know what this refers to. Likewise, consider the the typified action of "swimming". Although the particular *style* of swimming that is predominant in a culture in a certain historical period *may* be more of a pre-reflective Conv#1 phenomenon (cf. Mauss, 1979 [1936]), that certain ways of acting, in certain contexts, may *count as* instances of swimming (tokens of the "swimming" type) is a Conv#2 phenomenon.

Typified action and the associated count-as relation were also discussed in Section 2.2.3, in the context of Level Sem#2 of semiotic complexity. There it was pointed out that "typified action schematizations may constitute minimally a 'doing it again' of an action performed just moments before, either by oneself or by someone else; they may also be a more situation-transcending kind of typification that extends across particular settings, eventually spreading among members of a community or culture." It may well be undesirable to characterize a single repetition of an action that someone else performed just a moment before as a conventionalized action, since the concept of conventionality is typically associated with (A) at least somewhat larger groups than just two people, (B) more than a single recurrence, (C) recurrence not just within, but also across, situations, and (D) a greater degree of specification. Yet, so far as I am aware, there is no well-motivated criterion available which might provide a clue where exactly to draw the boundary between a convention proper and that which is only approaching the status of a convention. One may conclude that as a matter of empirical reality there is no strict boundary

---

<sup>38</sup>This is not intended to imply any completely pre-determined specification of how to interpret actions. Depending on context, the same action could be interpreted as "testing whether the electricity has come back after a power failure", "doing something against someone else's will", "following an instruction", and so forth.

to be found. Nevertheless, there are, of course, clear cases to be found, i.e., actions that are indisputably conventionalized.<sup>39</sup>

The more explicitly conventionalized something becomes, the less personal it will be. That is, in the process of conventionalization, there is a shift from first- and second-person intentionality to third-person intentionality, as defined in Section 2.2.1. The more conventionalized an act becomes, the more it will have the meaning of what “anyone” would mean by doing the act in question. The THUMBS\_UP gesture means “good”, not just for me, or for me and you, but for anyone that is a member of the same cultural group as us. One prime example of such third-person intentionality is the de-personalized and to some extent situation-transcending aspects of word-meaning sometimes called *linguistic meaning* (Coseriu, 1985) or *meaning potential* (Linell, 2009, Chapter 15), in contrast to the situated meanings of actual (multimodal) utterances. However, many of the more conventionalized aspects of language are not sufficiently described simply as more or less pre-established types. A yet stronger kind of conventionalization is involved, and that takes us to Level Conv#3.

#### Conv#3 — Normativity

Conv#3 actions are not just recognizable, for members of some particular group, as (tokens of) specific types of actions. At this level, actions are deemed “right” or “correct” (or “wrong”). This stronger form of conventionalization is therefore *normative*. Even though all the levels of conventionalization pose a certain resistance to the will of individuals (cf. *social facts*, Durkheim, 1965 [1895]), this is most strongly so at the higher levels of conventionality. Whereas typified conventions (Conv#2) specify ways of doing things, normative conventions (Conv#3) specify *that* things should be done.<sup>40</sup> Norms have a rule-like character. At the same time, this level implies no more than that members of a group are able to judge and treat certain ways of doing things as correct, and other ways of doing things as incorrect. They need not be able to give any explicit account of rules as such.

One sort of normativity has to do with *obligatory performance* (for prescriptive norms) or *obligatory non-performance* (for prohibitive norms). That is to say that some actions are obligatory: whenever context X is the case, some specific type of

---

<sup>39</sup>It may also be pointed out that it is not only the case that an action may be classified as conventional. To some extent, the performance of a conventionalized action also creates and sustains the convention.

<sup>40</sup>It should be noted that (cf. Bach & Harnish, 1979, p. 122) do not distinguish between Conv#3 and Conv#4 normativity, in the way I do in this chapter.



Figure 2.2: Correct form



Figure 2.3: Partially incorrect form

act or class of actions *must* be produced (or *must not* be produced). Many aspects of grammar are of this kind (Söderbergh, 1969; Zlatev, 2007; Sinha, 2009a). “An utterance provides an obligatory context for a particular morpheme if the omission of that morpheme would result in an ungrammatical sentence, from an adult point of view” (Bates et al., 1988, p. 167). In American Sign Language (ASL) negation is signaled by manual signs performed in concert with an obligatory HEADSHAKE (used as a grammatical marker). Children who learn ASL typically learn the manual sign NOT one to eight months before they start combining it with the obligatory HEADSHAKE, even though the children typically learn to use the HEADSHAKE already at around 12 months of age, well before they learn the NOT sign (Anderson & Reilly, 1997). Even though the children know both the manual sign for NOT and the HEADSHAKE, it takes a while before they learn that the NOT sign creates a context X in which the performance of a HEADSHAKE is obligatory. Context X could be just about anything, such as “everyone should rise when the King enters the room” or “bicycles are forbidden on the highway” etc.

Another sort of normativity has to do with *correctness of form*. In some cases, it is not obligatory to perform a certain act A just because a certain context X is present, but *if* act A is performed then it must take a certain form in order to be correct. Figure 2.2 shows the conventionalized HUSH gesture.<sup>41</sup> One particular vocalization is commonly performed together with this gesture. The vocalization is optional, but if it is there it should take a certain form: for example, it should be produced through exhalation and not through inhalation. On the other hand, the gesture is

<sup>41</sup>Figure 2.2 and Figure 2.3 are posed, and not based on video recordings of a spontaneous communicative setting. Nonetheless, the gesture should be familiar to most people, at least in Sweden.

also optional: the vocalization may also be performed alone, with a similar overall meaning. When the gesture is performed, it should have the form shown in Figure 2.2, although some variation is allowed. For example, the thumb could be folded behind the hand rather than sticking out to the side, and it would still count as a correct version of the same gesture. Some variations are not acceptable though. Figure 2.3 shows a partially incorrect version of the same gesture: the hand faces the wrong direction.<sup>42</sup> In fact, the gesture performed in Figure 2.3 is at least similar to *another* conventionalized gesture, commonly used when contemplating something. One of the children included in this study ("Alice") made the HUSH gesture consistently in this way for over a year, starting some time before 18 months of age, before she eventually switched to the correct form.<sup>43</sup> Despite her incorrect performance, the gesture was still readily recognizable as a token of the HUSH gesture type, rather than a token of the contemplating gesture, partly due to the different contexts of use for these two gestures and partly due to the different vocalizations that tend to go together with them: the "shh" vocalization of the hush gesture versus a vocalizations such as "hum", "hmm" and "uhmm" for the contemplating gesture. Last but not least, the similarity of this gestural performance to the correct performance, although only partial, also plays a role in identifying this gesture as an instance of the HUSH gesture. In this way, a given action need not be fully compliant with the idealized norm, for the conventional and normative aspects to provide an *interpretative frame*.

Furthermore, not all deviation from prototypical performances of conventionalized and normatively regulated actions are "incorrect" in the same way. For the signs of signed languages that are not only conventionalized, but also at least partially iconic, variations in performance with respect to the prototypical form of the sign are judged as less bad when the variations are compatible with the iconic dimension of the gesture (Arendsen, 2009, p. 94).<sup>44</sup> Some variations are still considered appropriate due to other grounds for meaning than the conventionalized aspect of the sign. In the end, "pragmatic" appropriateness overrides normative correctness whenever there is a conflict between the two (Coseriu, 1985). Generally, so long as some communicative actions are considered appropriate to a given situation, there will be no objections, no matter how incorrect they are. To conclude, normativ-

---

<sup>42</sup>This partial incorrectness is not entirely unlike how children sometimes write letters in reverse direction (mirrored) before they learn that the orientation is part of the normative constraints on writing.

<sup>43</sup>In a study of Swedish children, the word (or vocalization) "shh" was in the repertoire of at least 20% of the 16 months old children (Eriksson & Berglund, 1998, p. 15).

<sup>44</sup>Arendsen also found that individuals varied in their tolerance.

ity and conventionality provide an interpretative frames, but their primary job may not be so much to validate or invalidate actions *per se* (many errors are tolerated in practice), but rather to facilitate interpersonal processes of interpretation.

#### Conv#4 — Explicated rules<sup>45</sup>

At this level, members of a group not only have intuitions of correctness, they also *formulate* rules and norms explicitly, as objects of communication and thought in their own right. Hence, at this level normative aspects of action become products in their own right. In the most explicit case, these rules may even exist in written form, as *prescriptions* (cf. *formalization*, Andrén, Sanne & Linell, 2010, p. 224; Linell, 2009, p. 210), such as juridical laws, traffic rules, or the Ten Commandments. Other variants include conversational asides about some aspect of proper conduct, and signposts, such as stop signs, that have an explicit rule as their core meaning: for the stop sign, a rule that is known by anyone who drives a car. At Level Conv#4, the rules are not simply part of conduct, but are to some extent independent of the actions they refer to. In cases where the rule is communicated prior to the performance of an action, as a “pre-scription”, the rule resembles an *instruction*, even though instructions are not necessarily normative in character.

The distinction between Level Conv#3 and Conv#4 regarding the explicitness of the convention need not entail the presence of the “ought-ness” of normativity (Conv#3). It is also possible to give explicit accounts of other forms of conventionality (i.e., Conv#1 and Conv#2). The presentation here of four different levels is a mixture of two separate continua: a continuum of *increasing explicitness of the conventions* and another of *increasing “ought-ness”* (normativity) of the conventions. One could even add a third continuum of *increasing sizes of the communities that share a convention* (two persons, a handful, the people of a country, humanity, or some other point along this line). To construe these different aspects of conventionality in terms of a single continuum, as done here, is therefore a compromise, but one that is sufficient for the purposes of this thesis. Even though the ought-ness introduced in Conv#3 and the the explicitness of Conv#4 refer to some extent to different continua, the distinction of this fourth level from the third in this way helps clarifying the fact that “knowing a grammar” or some other normatively regulated phenomenon can be understood in quite different ways. It can be under-

---

<sup>45</sup>I am grateful to Jordan Zlatev for the suggestion to include this fourth level in the conceptual scheme. Its inclusion makes the nature of Conv#3 clearer, as I intended it, by showing how it is distinct from Conv#4.



stood pre-reflectively and implicitly (Conv#3 and lower), or reflectively and explicitly (Conv#4). A stronger form of explicitness provides one of the ways in which ought-ness becomes stronger, as in the case of law books or the Ten Commandments.

An interesting aspect of prescribed actions, where the rules exist separately of the actions and activities that they refer to, is that one might learn these actions (at least partially) through learning the rules that are associated with them, rather than by observing others performing the actions, or by inventing the structure of the action oneself.<sup>46</sup> Tomasello & Barton (1994) have shown that at least at 24 months of age, children comprehend enough language to be able to learn to use some words directly by being told what the words mean, rather than by hearing the word in an ostensive context where the referent is concretely present. The separation of rules from the actions that they refer to also means that the reliance on *recurrence* as an underlying factor in this form of conventionality may be very weak, or even non-existent. That is, one may make up rules for a game that has never been played before, so that there is not yet any recurrent phenomenon in existence to which the rule may refer. In such cases the rules are indeed “pre-scribed”. Accordingly, one might distinguish between stipulated “pre-scribed” rules (a top-down process) and rules that attempt to formulate the inherent normativity and regularities of already existing phenomena (a bottom-up process).<sup>47</sup>

#### **The depersonalized character of conventionality**

As argued above, the more explicitly conventionalized certain phenomena become, in increasingly larger communities, the less personal they get. Normativity (Conv#3) and explicated rules (Conv#4), as defined here, are not primarily matters of first and second-person intentionality, but rather strong forms of third-person intentionality. Therefore, in addition to the ‘I’ and ‘you’ in interpersonal relations and communicative encounters, “we could think of the language, including the meaning potentials of all its lexical resources, as belonging to an ‘invisible third party’ [...], or a

---

<sup>46</sup>Then again, in order to understand an instruction, one must know what the words that make up the instruction refer to, which nevertheless points back to embodied experiences of action. Still, the embodied experience which is required here is not of the same sort, and a rule may very well describe an action of a particular sort that has not been experienced before, and the rule may also add the ought-ness aspect to the actions that the rule is concerned with, which may not have been there before the formulation of the rule.

<sup>47</sup>See also Searle (1995, p. 27) for a similar but not identical distinction between “regulative rules”, which adds some sort of regulation to an already existing activity, and “constitutive rules”, that create new activities that did not exist before the stipulation of the rule.

generalized other", that concerns "how 'we' or 'one' [or generic 'you'] would use the words in the common language" (Linell, 2009, p. 95; see also Salgado & Hermans, 2005 and Mead, 1934). At the same time, of course in the use of language and in other more or less conventionalized expressions, first and second-person intentionality are also present: actions are constituted by personal will and by adaptation to particular unique situations and recipients. For participants in interactions, norms and conventions provide resources that they may orient to, for the sake of order, rationality, and comprehensibility, rather than being strictly governed by them (i.e., the problem of agency versus structure in sociology). The acknowledgment of all three intentional perspectives (see Section 2.2.1) in communicative action provides a framework that accounts for the nature of communication both as an initiating expression of the self, as a responsive adaptation to the Other in a social encounter, as well as a re-articulation of more situation-transcending cultural patterns, including language and conventionalized aspects of gesture.

Finally, one may ask: are norms *something more* than the "sum, average or common denominator of all the individual beliefs" (Sinha, 2009a, p. 301) of the members of a group? As Sinha argues, in a way they are: most conventions, in their various forms, pre-date the individual's lifespan (cf. Schutz, 1932; Heidegger, 1962 [1927]; Linell 2009, p. 57), providing an already structured world that the child will eventually be able to appropriate and exploit. Hence, many if not most conventions exist in a way that does not depend on any specific individual. At the same time, even though conventions (or "social facts", in Sinha's terminology, borrowing it from Durkheim, 1965 [1895]) are "in some way prior to individual cognitions about them [...] it cannot be claimed that social facts are independent of cognitions, in this case of social cognitions, since their normative status is dependent upon agreement in cognition" (Sinha, 2009a, p. 301).

### 2.3.4 Combinations

One characteristic property of language, according to most definitions, is that its expressive units may combine to form more complex meanings, such as propositions. Children's communicative development is often characterized in terms of *preverbal*, *one-word*, and *combinatorial* (or *two-word*) developmental stages (a multi-word utterance stage), or in a similar vein, as development through *presymbolic*, *symbolic*, and *propositional* stages (cf. Lock, 1997). Peters (1983) uses the expressions *one-unit* and *multi-unit* stages rather than one-word and two-word stages, since the

units of production (especially in children) may not always correspond to “words”.<sup>48</sup> The combinatorial, or propositional, stage is then typically considered to be that stage which corresponds to the achievement of language proper, although there may still be a long way to go until truly adult-like speech is achieved in all its concrete detail and abstract generalizations, including structural aspects like discourse level structure. The combinatorial, or propositional, stage emerges some time during the second year of life, although there is considerable individual variation (Bates et al., 1988). Some authors make a further distinction between a developmental stage of *combinatorial* sequences of units and a further developmental stage of *hierarchically* ordered sequences (e.g. McCune, 1995).

It should be stressed that combining two expressive “units” into a larger whole is not synonymous with grammar, in the sense of *conventionalized* and, to a large extent, *normatively regulated* word order patterns. It is perfectly possible to combine expressive units in a creative and non-conventionalized fashion, which would not count as grammar in the strict linguistic sense. This means that the issue of “combinations” is to be found both below the upper limit of gesture, and above the upper limit, as part of the conventionalized grammar of the language. The words “combine” and “combination” are used in a number of other ways in the scientific discourse, more or less related to each other, and the aim of the present section is to clarify some of these different uses.

When gesture and speech are used in a coordinated fashion it is often said that they are “combined” with each other. Several studies distinguish between *unimodal combinations* within a single modality (word+word or gesture+gesture) and *bimodal combinations* across modalities (word+gesture) (e.g. Greenfield & Smith, 1976; Greenfield et al., 1985; Goldin-Meadow & Morford, 1985; Capirci et al., 1996; Butcher & Goldin-Meadow, 2000; Pizzuto et al., 2005; Guidetti, 2005; Iverson et al., 2008; Stefanini et al., 2009). One conclusion that comes out of this line of research is that hearing-enabled children produce many more gesture+word and word+word combinations than gesture+gesture combinations (e.g. Goldin-Meadow & Morford, 1990; Capirci et al., 1996, 2005; Stefanini et al., 2009), and in most cases when gestures are combined, one or both of the gestures are deictic rather than content-loaded (e.g. Volterra, 1981; Masur, 1983; Petitto, 1988; Goldin-Meadow

---

<sup>48</sup>The notion of “developmental stages” can be, and has been, interpreted in many different ways (cf. Zlatev & Andr n, 2009). For example, does a developmental stage begin when the first few instances of some criterial property appears, or does it begin when the average utterance of a child possesses the criterial property? Does the new developmental stage replace an earlier stage, or is the new stage rather building on top of the previous one rather than replacing it? Are the transitions between stages a matter of gradual change, or are they more like sudden reorganizations?

& Morford, 1990; Morford & Goldin-Meadow, 1992; Capirci et al., 1996). Several studies have also shown that gesture+speech combinations reliably predict the onset of word+word combinations at around 17 to 22 months of age, i.e., two-word speech (Capirci et al., 1996; Butcher & Goldin-Meadow, 2000; Iverson et al., 2008). That conclusion does not regard all gesture+speech combinations though, but rather more specifically those combinations in which the content/referent of the gesture and word are not *complementary* (e.g. holding up an empty glass while saying "glass"), but *supplementary* (e.g. holding up an empty glass while saying "juice").<sup>49</sup> Typically the gesture involved in a supplementary combination is deictic rather than content-loaded. Supplementary combinations are effectively multimodal predicates, emerging a few months before the ability to express such meanings purely within the vocal modality.

Another distinction to make is between *serial combinations*: sequential "strings" of words or gestures; and *parallel combinations*: simultaneous expressions that overlap in time. Both types of combinations can take place within and across modalities. An example of serial combination across modalities are gestures that fill a grammatical "slot" in a spoken utterance (cf. *mixed syntax*, Slama-Cazacu, 1976), such as saying "he went..." and directly after the word "went" pointing in a certain direction. In such cases, even non-conventionalized gestures may function as if they were "words", in a relatively autonomous fashion, as pointed out earlier in this chapter.<sup>50</sup> Note that while both gesture+word, and gesture+gesture combinations may be performed in the same instant, i.e., as parallel combinations, word+word combinations can not be produced in parallel (at least not in any existing spoken language).<sup>51</sup>

It is not obvious where exactly to draw the line between what is to be considered a parallel simultaneous combination and what is to be considered a serial one. Different authors draw the line differently, depending in parts on their research in-

<sup>49</sup>It is not completely straightforward how to determine whether a gesture means "the same thing" as some particular word. See Pizzuto et al. (2005, p. 229) for a discussion of some of the differences in how this has been interpreted by various researchers.

<sup>50</sup>McNeill (2007, p. 22) has recently extended his scale from non-conventionalized gesticulation to conventionalized signed language to include such cases of 'mixed syntax' (he calls them *language slotted gestures*), as a slightly more language-like gesture than the class of gestures that he calls 'gesticulation'. Also, Ladewig (2010) is currently working on a PhD thesis on the topic of mixed syntax.

<sup>51</sup>It may be speculated that the onset of gesture+speech combinations before speech+speech combinations might have something to do with the fact that speech+speech combinations require the expression to be laid out in a temporally extended and sequential manner (cf. the *linearization problem*, Levelt, 1980, 1981), while gesture+speech combinations may be expressed, as it were, simultaneously.

terests. Researchers with a psycholinguistic interest in gesture may be interested in very precise measurements of gesture+speech timing, either on a scale of milliseconds (cf. McNeill & Levy, 1982, p. 284; Seyfeddinipur, 2006) or on a syllable level (Gullberg et al., 2008, p. 213). Others use more liberal criteria. For example, Guidetti (2002, p. 273) uses a criterion according to which the gesture and the word (or possibly also non-linguistic vocalization) are labeled as a gesture+word or a gesture+vocalization combination if the gesture and the speech occur within two seconds of each other. Blake (2000, p. 98) uses a criterion with a maximum gap of one second. These more “liberal” criteria may seem rough from a psycholinguistic point of view, but it should be noted that a word and gesture may well appear to “occur together”, or at least “belong together”, even though they are not performed in exact coordination in terms of objective time. Neither of these two approaches should therefore be considered the one and only “correct” one. Different criteria for what should count as a parallel (simultaneous, synchronous) gesture-speech combination capture different aspects of the temporal relations between gesture and speech.

Related to the distinction between an “objective” and “experiential” level of characterizing combinations is the distinction between combination in a *temporal* sense, and in a *semantic* sense. While a gesture may be performed in temporal overlap with some specific word, it may be the case that the semantics of the gesture relates to *another* word, or of several words of which the gesture only co-occurs with a subset. The *lexical affiliate* (Schegloff, 1984), i.e., the “semantic partner” of a gesture, may or may not be the word(s) with which the gesture co-occurs.

There is a debate in research on children’s gestures over whether simultaneous gesture+speech combinations exist from the outset, or whether there is first a developmental period where gestures and speech are performed independently of one another before they eventually “come together”. There is evidence in favor both of the early integration scenario (Masataka, 2003; Iverson & Thelen, 1999; Capirci et al., 2005; Pizzuto et al., 2005), and of the late integration scenario (Butcher & Goldin-Meadow, 2000; McNeill, 2005; Rowe et al., 2008). Some studies suggest an interpretation somewhere in between (Guidetti, 2002; Rodrigo et al., 2004). The answers to this debate depend in part on whether gesture+vocalization (rather than gesture+word), movement+vocalization, and movement+word are to be considered to be “integration”, or only gesture (proper) + word (proper). The answer depends as well on whether the issue concerns deictic or “representational” gestures and deictic or “representational” words (Pizzuto et al., 2005). In any case, the children studied in this thesis all clearly combine words and gestures at the first obser-

vations (18 months), such combinations including supplementary as well as complementary combination. At the same time, given that the beginning of the period studied in this thesis overlaps with the end of some of the studies on gesture+speech integration, some comparisons can be made (see Chapter 5).

One may further distinguish between combinations in an *additive* sense, where two complementary meaning units are added to bring about a larger meaning unit, and combinations where the units acquire *syntagmatic* values: i.e., "values they don't have outside the combination — as 'direct object' is the value of a noun only in opposition to a verb, combined in a verb phrase" (McNeill & Sowa, 2007).

At a more fundamental level, one may also consider what it means to say that different aspects of a bimodal expression are "combined". To say that anything is a "combination" implies that there are first two separate "units", and that these are subsequently "put together". From an analytical point of view it is of course possible to treat any identifiable aspects of an expression as separable, so that it may well be warranted to speak of these as different elements that are "combined". Nevertheless, to use the word "combination" for all coordinated uses of gesture and speech hides some of the variation that exists in the ways that gesture and speech may be said to be combined. So one may distinguish between *combinations* proper which are the result of putting together two or more units that have a well-defined status on their own, and those which may not be best thought of as combinations at all, but rather as *multimodal wholes*. The issue of multimodal wholes is the topic of Section 8.3, where it will be argued that in some cases the relationship between gesture and speech may be one of a multimodal whole, and more specifically, that the relationship is sometimes that of a *conventionalized multimodal whole*. Such instances will be referred as *item-based multimodal constructions*, that phrase being inspired by construction grammar and other usage-based approaches to language (e.g. Tomasello, 2001, 2003). Multimodal constructions are to be understood as holophrastic phenomena that extend across modalities. The existence of such conventionalized multimodal wholes puts into question any theory that posits a necessary opposition between the semiotic modes of co-speech gesture and speech, according to which the mode of expression involved in co-speech gestures is fundamentally different from the mode of expression of the words that it accompanies.

## 2.4 Putting the pieces back together again

So, what *is* gesture? What I hope to have made clear so far is that there is a long list of properties involved behind the scenes, as it were, in what might at first glance appear as relatively simple and intuitive demarcations between gesture and non-gesture, with respect to both the upper and lower limit of gesture. Rather than identifying gesture with a single set of characterizing properties, a *comparative semiotic* approach will be taken here (cf. Kendon, 2008), serving highlight both differences and similarities between different kinds of gestural performances. Arguably, this results in an understanding of the notion of “gesture” as a matter of *family resemblance* (Wittgenstein, 1953). Such a conception of gesture, I believe, is not vaguer than what is offered by more unitary and seemingly well-define conceptions of gesture, but to the contrary, more precise. Gestural performances then need to be specified with a more fine-grained set of properties. The task then becomes to identify interesting dimensions of variation that may be used to characterize and compare different gestural performances (Kendon, 2004, p. 104). Metaphorically speaking, the notion of gesture can be thought of as a cake that can be divided in different ways depending on how it is analyzed; but the cake itself, as it is discussed in the gesture literature, is an aggregate of several different features that may vary with respect to each other. It would be misguided to try to treat all those actions that are referred to as “gesture” as one *category* of action with a singular and essential core. Table 2.1 summarizes the main terms and distinctions that have been made in this chapter.

Although clear distinctions between different modes of expression are beneficial for clarity and descriptive precision, they may sometimes have a tendency to obscure *similarities* that exist, empirically, across the distinguished phenomena in actually occurring actions, so that the conceptual distinctions that one has made result in certain blind spots. Blind spots may emerge from other conceptual dichotomies, apart from the issues that have been discussed in this chapter, such as the distinction between pragmatics and grammar where “pragmatics” tends to serve as a waste-basket for everything that does not fit into narrow accounts of grammar as a self-contained domain. On such a view, little consideration is given to how grammar and pragmatic issues are interrelated.<sup>52</sup> In a similar vein, Haviland (1998, p.

---

<sup>52</sup>For example, one might ask how structural properties of grammar are motivated in terms of their (recurrent) deployment within the relevancies related to the turn-taking of communicative interaction. Fortunately, this issue has received more attention in recent years (Ochs et al., 1996;

<b>Distinctions relating to the Lower limit</b>		
<b>Intentionality</b>		Intent/volition versus aboutness
		First- versus second- versus third-person intentionality
		(Intentionality in) the natural versus the scientific attitude
<b>Communication</b>	<i>Comm#1</i>	Communication as a side-effect of co-presence
	<i>Comm#2</i>	Action framed by mutual attunement
	<i>Comm#3</i>	Visibly other-oriented action
	<i>Comm#4</i>	Reciprocated Level 3 actions
<b>Semiotic Complexity</b>	<i>Sem#1</i>	Situation-specific aspects of action
	<i>Sem#2</i>	Typified aspects of action (count-as)
	<i>Sem#3</i>	Differentiated aspects of action (semiotics signs)
<b>Expressive action</b>	ABC	Acts of bodily communication (Comm#3 or more)
	ABS	Acts of bodily signification (Sem#3 or more)
	ABCS	Acts of bodily communicative signification (ABS+ABC)
	Gesture	ABS or ABC, or both
<b>Distinctions relating to the Upper limit</b>		
<b>Imitation</b>		Imitation versus ritualization
		Imitation versus imitative processes
<b>Conventionality</b>	Conv#1	Normality
	Conv#2	Typification
	Conv#3	Normativity
	Conv#4	Explicated rules
<b>Combinations</b>		Unimodal versus bimodal combinations
		Complementary versus supplementary combinations
		Serial versus parallel combinations
		Temporal versus semantic combinations
		Additive versus syntagmatic combinations
		Multimodal combinations versus multimodal wholes

Table 2.1: Summary of terms and distinctions

163) suggests that :

[...] what I sometimes dub "subtractive" thinking has erected dichotomous walls between allegedly distinct phenomena: words versus gestures, syntax versus mere combination, "real" language versus general

(Auer, 2007; Lindström, 2008), resulting in many interesting findings.



communication, and so on. The standard procedure is to define one term strictly and exclusively, and to relegate to the other everything left over once the strictly defined parts have been subtracted from the phenomena in question.

In order to reconcile the apparent conflict between clear conceptual distinctions and overly dichotomous ways of treating empirical phenomena I have attempted, in this chapter, to go in the direction of a more fine-grained conceptual apparatus, and avoid any simplistic binary distinction between “gesture” and “non-gesture”. Although this maneuver does not get rid of the problem of “subtractive thinking” altogether, it will make the subtractive error smaller. Even though the distinctions that are made will still serve to emphasize differences between the distinguished entities, there will in effect be a richer set of options for capturing nuances. At the same time, this maneuver retains the (necessary, I would argue) benefits of a clearly defined analytic vocabulary. Indeed, without retaining some trust in the value of clear conceptual distinctions, it would be hard to bring any clarity to issues of similarities and differences in the first place since there would never be any determinable shift from “similar” to “different”, but only vague continua. *The crucial issue here is not to think of distinctions on a conceptual level as discrete phenomena on an empirical level.* That is, even if the issues of the lower limit of gesture may not be a matter of a continuous scale on a *conceptual level*, it may certainly be so on an *empirical level*: any particular gestural actions may exhibit more or less of the properties pinpointed by the conceptual distinctions. It is precisely this emphasis on both similarities and differences that a more fine-grained set of conceptual distinctions makes possible, thereby avoiding any undue emphasis on differences alone (with the associated blind spots that emerge from confusing conceptual distinctions with empirical phenomena) or on gradual continua alone (with the consequent problem of providing clear analytical accounts due to the lack of conceptual distinctions), that constitutes the comparative semiotic perspective. Although this chapter has been presented in isolation from any empirical data, all of the conceptual distinctions that have been offered are motivated by observations resulting from the analysis of the data to be described in Part II of this thesis.



## CHAPTER 3

# Action gestalts, gesture, and intersubjectivity

*In activity, subject and object interpenetrate.*

Payne (1968, p. 85)

*Thus the crucial question is: How to construe persons as being social without abandoning their obvious personal autonomy, separateness from any social unit (group, crowd, community), while being members of such units. The conceptual imperative of the commonsense "either-or" thinking holds scientific terminologies in its iron grip.*

Valsiner & van der Veer (2000, p. 7)

### 3.1 Structures of action

Gesture is a kind of action. As such, gestures share a number of properties with other forms of acting. For example, all actions are by definition meaningful in some way, to be distinguished from involuntary movements or "mere movement". Formulated in terms of one of the distinctions introduced in the previous chapter, that is to say that both action in general and gesture in particular cannot be sufficiently accounted for by *thin* descriptions of observable conduct (meaningless form), but rather presuppose *thick* (intentional, meaningful) descriptions (cf. Ryle, 1968). At the same time, the particular ways in which a gesture or an action is performed play

a crucial role in shaping what a gesture or an action may come to mean or be about. Furthermore, what a gesture or an action may mean or be about also depends on the overall activity that they are a part of. Hence the meaning of a gesture or an action is not solely attributable to form properties of the gestural performance itself. Such relations between form and function, and between acts and activities, will be discussed in this chapter in terms of the “formula” of *operation-act-activity interdependence*.

Another similarity between gesture and other ways of acting is that they are typically experienced as coherent wholes that have some sort of beginning and end, even in cases where this happens in a continuous stream of movement. This can be traced both to aspects of the movement dynamics involved in the performance of the act or gesture (cf. Baldwin et al., 2008), as well as to the possibility to attribute meaning to them (cf. Hendriks-Jansen, 1996), often in terms of culturally defined meanings. Kendon (1978, 1980b, 2004) has introduced a descriptive vocabulary of *gesture phases*, which enables a description of gestures relating to issues such as specifying when they begin and when they end, i.e., to specify what part(s) of a stretch of movement(s) that stand out from prior and subsequent movements as the articulation of a gesture. This terminological apparatus is essential to descriptions of gesture and it will also be described and discussed briefly in this chapter.

Since most gestures and many actions take place in social contexts, it is also relevant to consider the relation between action and *intersubjectivity*, i.e., action as a way of being together. Indeed, infants do not encounter the world alone. This means that the development of action cannot be understood merely as a matter of sensorimotor interaction between an organism and the world without regard to the (active) role of the social Other in this process. In research on children’s development, children’s abilities to understand other people are often conceptualized in terms of *Theory of Mind* (ToM) (e.g. Leslie, 1987; Baron-Cohen et al., 1985; Baron-Cohen, 1995). Simply put, this approach posits that children have a particular inferential mechanism fully developed by the age of 3 which enables them to understand the intentionality of other people more or less “across the board”, as if it did not matter how the interaction is carried out and what sort of activity that is involved. The analytic route chosen for the purposes of this thesis does not conform to this conceptualization. This issue will be discussed first, before turning to operation-act-activity nexus, actions gestalts, and gesture phases.

## 3.2 Action, intersubjectivity and the world within reach

Alfred Schutz was concerned with establishing a phenomenology of the natural attitude in the activities of everyday life.<sup>1</sup> As part of this endeavor he provided an analysis of *the world within my reach* (Schutz, 1945), and its intersubjective properties. “The world within reach” refers to the world as experienced through the senses as well as kinaesthetic, locomotive, and operative movements that “gear out into the world”.<sup>2</sup> His analysis of this particular part of the life-world was inspired by George Herbert Mead (1932, 1938) who distinguished between the *manipulatory area* and more distal zones that were perceivable, but not within the range of direct action. Schutz regarded the manipulatory area as being of utmost importance. In fact, he describes it as “the kernel of the reality of the life-world” (Schutz & Luckmann, 1973, p. 42; see also Schutz, 1945, p. 546), and it is in this primordial reality that the newborn child will find itself.

Although terms such as “zone”, “sphere”, and “area” may suggest that this zone is primarily a spatial and topological zone, this is not only so. The manipulatory area is also endowed with the *temporal* experiential character of the present, including various relations to past experiences and future possibilities in terms such as “the world within potential reach”, “the world within restorable reach” (as opposed to “the world within reach” and “the world within actual reach”) and through temporally situated motives such as doing things “in order to” (the action performed in order to achieve a future result) or “because of” (the action as motivated by some previous state of affairs). As such, the concept of the world within reach is framed in a solid spatio-temporal and experiential framework.

The most crucial aspect of Schutz’ (and Mead’s) analyses of the zone of operation, however, is that it does not only concern the relation between the subject and the world. It is also a highly social zone that includes involvement with other people. This essential acknowledgment stands in contrast to a number of other accounts of the relation between the subject and the world, such as, for example, Johnson’s (1987) “body in the mind” (see Linell, 2009 for a relevant critique), the “ecological” framework of Gibson (1979) (see Costall, 1995 for a relevant critique), and the sensorimotor account account of perception of Noë (2004) (see Gallagher, 2008 for a relevant critique). Although all of these accounts emphasize the role of action and sensorimotor interaction with the world in human sense-making and concept

---

<sup>1</sup>The notion of the natural attitude was also discussed in Section 2.2.1.

<sup>2</sup>“Within reach” is a translation from the German word *erreichbar* and it seems to be the standard translation of Schutz’ term.

formation, all of them lack a developed notion of social action. Schutz paid close attention to the partly overlapping, and partly disjoint, character of participants' experiences and perspectives in face-to-face encounters, carefully working out, for example, the consequential nature of the distinction between what is *here* for me is *there* for you, and vice versa.<sup>3</sup> Questions of intersubjectivity constitute the very heart of Schutz' theoretical enterprise.

In the literature that does stress the importance of intersubjectivity for the understanding of human action and interaction one can, ideally speaking, discern two "poles" regarding how the problem of intersubjectivity is construed.<sup>4</sup> The first pole consists of approaches to intersubjectivity that start out with the solipsistic assumption of "no access". According to this position, minds first exist in separation, and some sort of mental bridge is required to close this "gap". Typically, in such accounts, it is "minds" or "brains", rather than situated and acting persons or bodies, that make intersubjectivity possible, and emphasis lies on domain-general *inferential mechanisms* (the "inferential model of communication", Sperber & Wilson, 1995 [1986]; Theory of Mind (ToM), Baron-Cohen et al., 1985; Baron-Cohen, 1995; Leslie, 1987; and to some extent the relatively detached notion "shared intentionality infrastructure" proposed by Tomasello, 2008). This results in a conceptualization of the problem of interpersonal understanding which is more or less "content-free" in the sense that no reference is made to the child's more specific previous experiences of engagement in a multitude of specific activities. That is, the Theory of Mind conceptualization of the problem of interpersonal understanding assumes that interpersonal understanding is a unitary phenomenon. It may be informative to note that Schutz speaks, not of "interpersonal *understanding*" in this generalized way, but of "*knowledge of other minds*" (my italics, see Schutz, 1954), and more specifically, that this knowledge is described as a "*stock of knowledge*" (my italics, Schutz & Luckmann, 1973) that involves familiarity with a *heterogeneous* array of activities. As remarked by Zahavi (2010), the Schutzian position involves the recognition that there is not just one form of interpersonal understanding, but many, and in a similar way, Sharrock & Coulter (2009, p. 87) argue that:

'Understanding other people' is no unitary task, accomplished in any

---

<sup>3</sup>Kendon (1990, p. 211) presents a similar notion in a discussion of spatial organization in communication. When an individual's *transactional segment* (the space into which a participant looks and speaks, into which he reaches to handle objects) overlap with another individual's transactional segment in a joint transactional segment (an *o-space*), we can speak of a jointly regulated *F-formation* (a stable spatial arrangement of a focused encounter).

<sup>4</sup>See Gallagher (2004) and Zlatev et al. (2008b) for similarly structured arguments regarding theorizing on intersubjectivity.

one way, by any single ‘mechanism,’ but is a multifarious one, and it is perhaps the tendency to think of this in an almost wholly abstract way [e.g. “Theory of Mind”] which creates the false impression that it is a single, well-defined affair.

The second pole consists of approaches that start out with an assumption of “direct access” to the Other. According to this position, it is a mistake to think that persons are “separate” from each other in the first place, when they are in fact physically and experientially *available* to each other, including emotional expression and the like. These approaches generally emphasize some form of “direct perception” of the Other and it is mainly within the phenomenological tradition that such accounts are found. For Jean-Paul Sartre (according to Schutz, 1948, p. 191), “Expressive gestures in particular, do not indicate a hidden affect lived through by any psyche. The frowning brows, the clenched fists, etc., do not *indicate* the Other’s wrath; they *are* his wrath.” Max Scheler defends a similar view:

For we certainly believe ourselves to be directly acquainted with another person’s joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands... And with the tenor of his thoughts in the sound of his words. If anyone tells me that this is not “perception”, for it cannot be so, in view of the fact that a perception is simply a “complex of physical sensations” ... I would beg him to turn aside from such questionable theories and address himself to the phenomenological facts. (Scheler, 1954, cited in Gallagher, 2005, p. 228; see also Gallagher, 2004; Zlatev et al., 2008b)

This way of formulating the problem avoids some of the problems associated with the previous approach — the “no access” approach — because it does not disregard the concrete ways in which interpersonal understanding comes about as a result of actually being together. However, this approach also comes with its own problems. First of all, the outcome of such views are sometimes dangerously close to accounting for intersubjectivity merely by saying that it is not a “problem” in the first place. That may be true to some extent, from the viewpoint of participants in routine activities in the natural attitude, but it is not necessarily productive from a scientific point of view to merely take intersubjectivity for granted. It is also dangerously close to a behaviorist position — Schutz (1948) also points this out — in which the mental life of a subject at any given moment is reduced to what is visible from the surface of the person’s body. This problem is sometimes confused for the issue of Cartesianism. Just because we may not want to separate “mind” from body, in

a Cartesian way, we do not want to commit the mistake of reducing the mental to that which is visible on the *surface* of the body in each given moment, neither from a scientific point of view, nor as a description of how participants orient to each other. According to the way Schutz construes the “problem” of intersubjectivity, however, we are neither in a position of “no access” to each other, nor do we have “direct access” to the inner worlds of each other. The way he formulates the problem *is* productive in the sense that it locates a phenomenon which is massively available for analysis, each time human beings engage with each other (which is not to say that other ways of studying intersubjectivity must be excluded).

What we have “between” us, according to Schutz, is the mundane social world of everyday life itself. A central resource for the achievement of order in this everyday life is the various typified means for coping with this life. This includes typified means of interpretation (typical motives) and typified ways of acting.<sup>5</sup> Schutz (1955, p. 173) states that “fully successful communication” is not possible, but that “The common-sense praxis of everyday life, however, solves this problem to such an extent that for nearly all good and useful purposes we can establish communication with our fellow-men and come to terms with them.” Further, he points out that this is only possible if the communicative process is based on a set of typifications, abstractions, and standardizations. Thus, he locates the problem of intersubjectivity right in the middle of the spatio-temporal, personal, and mundane contingencies of everyday life in all its particulars.

When we act, in Schutz’ account, we act in ways that are sensitive to — or “orient to” in ethnomethodological parlance (cf. Seedhouse, 2007) — more or less standardized, tried-out-before, and typified ways of acting (rather than being strictly governed by them). We do this because it works, and because others can understand us when we do so. We also understand, to a large extent, actions of *others* in terms of such standardized motives and typified rationalities (Schutz, 1953), i.e., the third-person intentional perspective that was discussed in Chapter 2. Through the socially shared (i.e., conventionalized) character of such typified understandings of action we are able to achieve and maintain a social order that is sufficient for the “pragmatic” purposes at hand.<sup>6</sup> This is not to say that all sense-making in face-

---

<sup>5</sup>See also De Jaegher (2008) a recent and phenomenologically inspired account that emphasizes the importance of joint action for interpersonal understanding.

<sup>6</sup>This way of thinking was picked up by Harold Garfinkel, the founder of ethnomethodology, and made into a central point of departure of this school. According to ethnomethodological thinking social life is organized with respect to being “visibly-rational-and-reportable-for-all-practical-purposes” (1967, p. vii). Garfinkel’s (1952) PhD thesis was concerned with providing an account of social order that to a large extent based on some of Schutz’ ideas, although he was also partly critical



to-face encounters is a matter of what I have dubbed third-person intentionality. If that was the case, it would be hard to account for how (interpersonal) understanding of action in terms of third-person intentionality could emerge in the first place, unless it rested on more basic forms of social interaction of a not-yet-typified character (e.g. Trevarthen, 1979).<sup>7</sup> Indeed, it is in the more direct we-relationship of the face-to-face encounter and the world within reach that children find themselves from the very outset (Schutz, 1967 [1932], 1966, p. 80). It should be noted, though, that even in the face-to-face encounter Schutz did not uncritically accept the idea of perception of the other as “direct”, as the “direct access” account of intersubjectivity would have it, also from a phenomenological perspective (Schutz & Luckmann, 1973, pp. 63):

Although we speak of the “immediate” experience of a fellow-man, this experience is internally, also in the precise meaning of the word, “mediated.” I grasp my fellow-man’s flow of lived experiences only “mediately,” in that I explicate his movements, his expression, his communications as indications of the subjectively meaningful experiences of an alter ego. But among all my experiences of the other I, what is mediated least is the encounter of the fellow-man in the simultaneity of the we-relation. Thus we will continue to speak, even though it is not completely accurate, of an immediate experience of the fellow-man.

Also, the further away from the world within actual reach that interpersonal interaction gets (telephone calls, conversation through written text, etc), the more indirect it will be, with a stronger dependence on indirect processes based solely on “already established” typification, such as conventionalized language, standardized acts and activities, including the ways in which we typify our own conduct in institutional settings in the sense of “acting like the typical sender of a letter” or “acting like a typical customer in the grocery store” and similar depersonalized ways of acting (Schutz & Luckmann, 1973; see also Barker, 1978).

Before ending this section it should be pointed out that even though at last some of these typifications may be considered “mere constructions” and may have “flaws” such as being counterfactual, this does not make them somehow unreal and irrelevant for explanations of human action. This is captured in the so-called *Thomas theorem* according to which “If men define situations as real, they are real in their

---

towards Schutz (and Mead) (cf. Garfinkel, 2006 [1948]).

<sup>7</sup>Non-typified aspects of face-to-face interaction are also discussed in relation to level #1 of semi-otic complexity (Sem#1) in Section 2.2.3.

consequences” (Thomas, 1928, p. 572; see also Schutz, 1955, p. 194; Bruner, 1990). A variant of the same idea is central to the perspective of *symbolic interactionism* (Blumer, 1969), according to which human beings do not act in direct response to the actions of others, but rather towards the meanings they ascribe to those actions, which means that held beliefs, no matter if they are counterfactual or not, will have very real consequences. In fact, it is precisely the property of conventions (i.e., “social constructions”) of constituting a (shared) choice in a somewhat “arbitrary” and open-ended set of possibilities of cultural beliefs that will eventually allow people to align themselves with each other in more complex forms of mutual understanding.

To conclude, in the Schutzian perspective of intersubjectivity and interpersonal understanding that is adopted here, there is a strong role for social processes of typification in the achievement of social order and interpersonal understanding, and as a consequence, also in scientific explanations of these phenomena. In particular, social processes give rise to third-person intentionality, which is perhaps the single most important factor underlying the possibility of interpersonal understanding of a more general kind since it concerns knowledge of what people typically intend when they do such and such. That is, when we express ourselves, we typically do this, at least in part, via depersonalized typifications. Third-person intentionality is by its very nature not something that can be achieved or understood purely by abstract “inference” on behalf of an “isolated” individual, as the “no access” accounts of intersubjectivity would have it, since it is rather based on what Schutz calls a “the stock of knowledge at hand”, which in turn presupposes prior pragmatic engagement with others in a heterogeneous multitude of different activities. On the other hand, the “direct access” accounts of intersubjectivity do not provide a satisfactory explanation either. For example, they do not explain why children do not “already” understand all aspects of social life more or less from day one, or at least from the moment when they first perceive them, since these accounts assume a non-problematic model of communication.

As argued by Linell (2009, pp. 81), “the term ‘intersubjectivity’ can either be taken in a sense that stresses commonality, sharedness and perhaps consensus, or it can be understood in a more neutral sense of mutual other-orientedness that can also accommodate alterity”. That is, the understanding of intersubjectivity that is adopted here is in line with the latter interpretation of the term, in which it takes active engagement in social interaction to achieve mutual understanding, both in specific situations, but even more so on a longer ontogenetic time-span. De Jaegher (2008, p. 541) comes to the same conclusion and summarizes it like this:

Others are not opaque to us. But they are also not continually trans-

parent to us. We still work at understanding them. But rather than doing that from our strictly individual vantage point, we do it in interaction (and sometimes we collaborate towards this goal, and sometimes we don't).

In this regard it should also be added that “the idea is not that we *simultaneously* come to the *same* understanding of any given situation (although this can happen)” (Duranti, 2010, p. 6). The way intersubjectivity is understood in this thesis is therefore as something which is more basic than “mutual understanding” as it is rather the conditions that make mutual understanding possible. As I would argue then, these conditions are constituted precisely by the possibility of “throwing oneself into” (or indeed already being in) engagement in joint activities, not as detached puzzle solvers, but through actual engagement. Indeed, “since we are caught up in such pragmatic circumstances from the very beginning (consider for example the infant’s dependency on others for nourishment), the original situation cannot be characterized as an isolated subject confronting an alien being” (Gallagher, 2004, p. 31).

This way of viewing intersubjectivity is “premised on the abandonment of the search for any pristine ‘fail-safe’ technique through which [interpersonal understanding] could be said to be assured or guaranteed” (Heritage, 1984, p. 54; see also Wittgenstein, 1953) and instead takes an interest in how it is achieved in, and possible through, semiotic processes of action.

### 3.3 Manual excursions and Action Gestalts

Schutz & Luckmann (1973, : 52) state that it is “as Husserl has shown, a universal principle of consciousness that in my conscious acts I ‘live’ attentive to their intentional Objects, not to the acts themselves.” In the case of gesture, for example, this means that we generally do not attend to the hands or the movements of the hands as such, but we rather see “through them”, to some sort of content/referent (cf. Kendon, 2004, p. 358; Streeck, 2009b, p. 139). This quality of gestures may be called *gestural transparency*. However, not any movement that is part of gesture performances has this quality. Kendon (1978) made gesture researchers thoroughly aware of the fact that only *some parts* of the manual excursions that make up gestural movement are actually perceived as “the gesture”, standing out as an expressive figure against a ground and having the quality of gestural transparency, even though there is generally other movement going on both before and after this part of the move-

ment excursion. Kendon has baptized such expressively articulated parts of bodily movements the *stroke(s)* (Kendon, 1980b), defining it as the part of the manual excursion “in which the movement dynamics of ‘effort’ and ‘shape’ are manifested with greatest clarity” (Kendon, 2004, p. 112). The movement phase coming before the stroke is called the *preparation* phase, and this is the movement of the articulating body part, typically the hand, that is required to arrive at the place where articulation begins. After the stroke, there may either be another preparation phase (in cases where several gestures are performed in sequence), or there may be a *retraction* phase<sup>8</sup>, which consists in the returning of the articulating part of the body to its *rest position* (Kendon, 1980b, p. 212). Although it is not strictly necessary, there is a strong tendency for gesture retractions to end where the initial preparation phase began, and this is the reason for the use of the term “excursions”.

The stroke together with the preparation is called a *gesture phrase* (note that the term “gesture phrase” is distinct from the term “gesture phase”, which refers to phases such as preparation, stroke, or retraction). If there is a retraction after the gesture phrase, the excursion has reached its closure, and the gesture phrase plus the retraction then constitutes a full *gesture unit*. A simple example of a gesture unit with a preparation, a stroke, and a recovery phase is shown in Figure 3.1. There can also be several gesture phrases in a row, without a retraction in between, so that there are several gesture phrases in the same gesture unit. Whenever a retraction is present, the gesture unit is finalized. The only obligatory part of this scheme is the stroke. All other phases are optional. The reason for this is that a gesture may be performed directly where the hand is located, without requiring a preparation phase or a retraction.

At various junctures during the movement excursion of a Gesture Unit there may also be *holds*, where the hand (which is the most common articulator of gestures) momentarily “pauses” in the course of a not yet finished movement excursion. 3.2 shows a more complete rendering of the gesture phase vocabulary that also includes holds. *Post-stroke holds* are holds that follow after the stroke (Kita, 1993; Kita et al., 1998). These are typically seen as a retention, or prolongation, of the gesture. Hence, the stroke together with any existing post-stroke hold constitutes the *nucleus* of the gesture phrase (Kendon, 2004). *Pre-stroke holds* are strokes that come before the stroke (Kita, 1993; Kita et al., 1998). Since the stroke has not yet started to be articulated when the pre-stroke hold occurs, if it occurs, the pre-stroke

---

<sup>8</sup>The term “retraction” seems to be used by most gesture researchers (e.g. McNeill, 2005), although Kendon uses the term *recovery*.

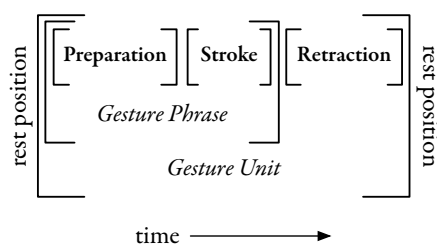


Figure 3.1: A simple example of a Gesture Unit.

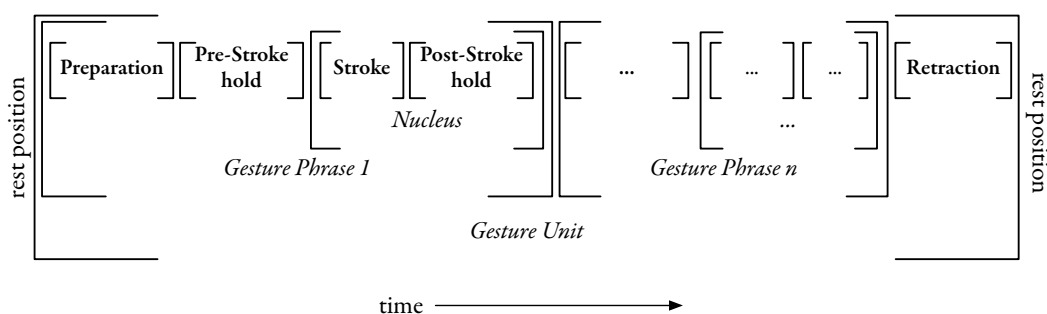


Figure 3.2: Full scheme of the gesture phase vocabulary.

hold typically does not appear to be part of “the gesture” (the stroke) in the same way as post-stroke holds. Hence the pre-stroke holds appear outside the nucleus. Holding the movement for a brief moment often serves to align the gesture with some particular part of a spoken utterance. However, there are also so-called *stroke holds* (McNeill, 2005, p. 32), which are a variant of strokes rather than a phase separate from strokes, the way pre-stroke and post-stroke holds are. Stroke holds refers to static strokes, where the hand is held still as part of the articulation of gesture. An example of this, provided by McNeill (*ibid.*), is when “raising a hand into the upper gesture space (preparation) and then holding it in place with the meaning of the upper floor of a building. There is no motion, but there is content and ‘effort’ in terms of focused energy.” That is, in the case of stroke holds, there is “an information-laden component” (*ibid.*, p. 283). Stroke holds are not shown in 3.2 since they are effectively a stroke variant rather than a separate phase. On the other hand, the part of a movement excursion that appears to an observer as a stroke hold can simultaneously also take on the functions of a post-stroke hold, which means that stroke holds may effectively serve as some sort of blend between strokes and post-stroke holds. In any case, they are part of the nucleus of the gesture phrase.

It is not always easy to identify the boundaries of these phases in a gestural movement excursion. This is partly due to the seamless character of some gestural move-

ment excursions, but also due to more intricate complexities such as the fact that the hand-shape to be used in a gesture may begin to be articulated before a certain movement that is also part of the stroke is actually carried out. For example, in a pointing gesture, the hand-shape may gradually build up during the preparation phase, before the gesture reaches its climax, so that it is hard to tell exactly when the stroke started. Nevertheless, as an analytic tool, the vocabulary of gesture phases is often indispensable, since it allows for a much more precise analysis than what is possible when equipped only with the general term “gesture”. For example, this terminology helps disambiguating between “the onset of the gesture” in the sense of “the onset of the preparation” and “the onset of the stroke” (cf. Kendon, 2004), which may sometimes be crucial to avoid misunderstandings.

Since the analytic vocabulary of gesture phases has mainly been developed for purposes of describing empty-handed gestures one may ask how they apply to object-gestures, i.e., those expressive acts that involve handling of an object. Kendon (2004, p. 14) argues that it is precisely these kinds of “dynamic features” of gestural movements, captured in the analytic vocabulary of gesture phases just described, that make observers see some movements as intentionally expressive. He says that “deliberate expressiveness is *manifest*” and due to this we perceive “the quality of the action as intentional (not the specific intention, necessarily)”, without having to *infer* an intention explicitly. As I have suggested elsewhere (Andrén, in press a), the mere fact that many gestural excursions are performed with empty hands may itself be one such manifest feature of “deliberate expressiveness”. Perhaps even a particularly important one, as it may explain why the term “gesture” is often treated as synonymous with what I call “empty-handed gesture”. Empty-handed excursion-like movements will, in themselves, have a very strong tendency to appear (a) *volitional*, due to their excursion-like character, and (b) *performed for expressive rather than practical purposes*, due to the absence of any handling of a physical object in the performance. This should not be interpreted as grounds for treating the absence of objects as a necessary feature in definitions of gesture, even though this is unfortunately precisely what many gesture researchers do (e.g. Acredolo & Goodwyn, 1990; Morford & Goldin-Meadow, 1992; Goldin-Meadow & Iverson, 1998; Nicoladis et al., 1999) (sometimes for theoretical reasons and sometimes for methodological reasons), with the exception of the particular acts of GIVE and SHOW which are commonly included in the category of “gesture” despite not being empty-handed. However, “defining away” object-gestures obviously does not make them go away. Kendon’s (2004, p. 15) definition of gesture as those actions that have features of manifest deliberate expressiveness does not commit the mistake of excluding com-

municative acts that involve objects from the research agenda.<sup>9</sup> However, in practice he does not analyze this class of actions himself, and therefore it remains unclear from his account of “gesture phases” how and if they might apply to communicative acts that do involve objects.

Although not the only one, Jürgen Streeck is one of the few gesture researchers who have taken an interest in object-gestures (e.g. Streeck, 1996). He points out that “in the context of conversation, it is usually not difficult at all to identify gesture units and their boundaries. In contexts of work, however, gestural communication may consist in nothing more than a repetition or a slight embellishment or exaggeration of an instrumental act” (Streeck, 2009b, p. 23). This is true, but it should also be emphasized that not all object-gestures are necessarily a matter of such “slight embellishments”. In many cases object-gestures are as clearly articulated as empty-handed gesture.

Acts that involve objects (in general, not only object-gestures) also often have the character of excursions and they often have a phase-like structure in terms of preparation, stroke (“the actual act”), and retraction. Sacks & Schegloff (2002) formulated the concept of “home position”, which is more or less equivalent to Kendon’s concept of “rest position”, around the same time as Kendon, in the 1970s (although in their case it was not published until much later). Sacks and Schegloff say that they also noted the phenomenon of manual excursions — i.e., that when performing an action the hands often end up after the action where they began before the action, in a “home position” — when studying gestures, but they also add that the excursion-like character is frequent in many other kinds of actions as well (ibid., p. 141). In short, the overall logic of manual excursion and the vocabulary of gesture phases is applicable to many kinds of actions, not only empty-handed gestures, and not even communicative action only. Nevertheless, as the analysis in Chapter 9 will show, it is not always straightforward to do so, since it may depend on whether the practical aspects that may be present in a object-gestures are considered, or if one considers the expressive aspects of the object-gestures. For example, not all “stroke-like” handling of an object will appear foregrounded as expressive acts, entering into the

---

<sup>9</sup>For example, Kendon (1984, p. 81) points out that:

It is possible for actions to be performed in such a way that one cannot be quite certain whether they were shaped by a communicative intention or not. It is also possible for purely practical actions, such as sniffing a wine to test its bouquet or drawing upon a cigarette, to be informed to a greater or lesser degree with “flourish” or “style.” To the extent that this is done, to this extent such actions may come to be seen as mainly gestural.

*main attentional track* of the joint activity (cf. Kendon, 1978; Goffman, 1974), and sometimes the reaching towards an object (i.e. the “preparation” of the practical act) may serve as the “stroke” on an expressive level, if, for example, the reaching movement is frozen in the middle of the performance, which may serve as a pointing-like gesture (i.e., REACH).

One key feature that emerges from Kendon’s analysis of empty-handed gestures is that they have *gestalt properties*, even though he does not use this specific term. Some stretch of movement, the stroke, appear as some sort of unitary and foregrounded expressive “whole”, as an *action gestalt*.<sup>10</sup> That is, even though the stroke may involve several movements and changing hand-shapes that unfold over time, and even though there may have been movement before (preparation) and after (retraction) the movements that make up the stroke, strokes are somehow experienced as unitary wholes. For this reason, it could be argued that gesture (and action) is perceived at a certain “level”. This is not to say that gestures are interpreted at a *static* level (i.e., always interpreted at a fixed level of abstraction or granularity), or that gestures and actions cannot be interpreted in several different ways at different occasions. It is also worth keeping in mind that in actual situations, these acts are recognized as they unfold in time, rather than as units that are immediately given as wholes (the process/product distinction). The meaning of a communicative action is different when it is ongoing compared to when there has been a retraction and the action has thus transformed into an after-the-fact entity (Andrén, in press b,i). The point that I wish to make is just that in each particular case, at a particular moment, acts are experienced in *some* particular way, at a certain level of granularity. When someone picks up a phone and puts it to the ear, it is not generally perceived as a highly granular and subdivided sequence of “reaching”, “grabbing”, “lifting”, but rather as a conveniently unitized and unitary act of “phoning”, or “bringing the phone to the ear”, or possibly some moderately subdivided sequence such as “picking up the phone, and putting it to the ear”. It *could* be perceived in a more zoomed in and fine-grained fashion, but it is generally not, and even when it is (for example as part of gesture analysis), it just means that one attends to a different aspect of the action, but now this aspect of the action becomes an action gestalt in its own right, so the same basic experiential “action gestalt” structure remains.

Gestures and multimodal utterances may be characterized in terms of *weak* and *strong* gestalts (Köhler, 1938 [1920]). Strong gestalts are holistic entities that could not, strictly speaking, be said to have “parts” at all. Strong gestalts may in fact be

---

<sup>10</sup>Hirsch (1993) has also discussed gesture in terms of gestalt structures, although from a slightly different point of view than discussed in this section.



hard to find in the context of human action and experience, but there are for example electromagnetic phenomena that may be argued to qualify as strong gestalts (cf. Ash, 1995). More relevant to the study of gesture then are weak gestalts. Weak gestalts are rather the result of a weaker dialectic between parts and wholes, such that parts contribute to the interpretation of the act as a whole (bottom-up), but the overall act also contributes to the interpretation of the parts (top-down). In the case of weak gestalts, the parts are possible to perceive as unitary phenomena in themselves (often even in multiple ways) when attention is shifted, as described in the previous paragraph, even though they are also part of a whole on a higher level. The holistic or gestalt-like properties of gestures are thus generally towards the weak side of the spectrum in that we are not completely blind for the “parts”, i.e., the move(s) and hand-shape(s) and so forth that make up gestures as acts, even though we tend to see “through” them, to the overall acts as intentional wholes, and primarily to the stroke phase of these movements. Multimodal utterances, which consist in both bodily movement and speech are even more clear examples of weak gestalts, since the different modalities (speech, bodily movement etc.) of the parts that make up the multimodal utterance almost per definition make them perceivable as somewhat separate entities.

It should however be noted that Köhler discussed “physical Gestalten”, and since we are dealing here with (inter)subjective and meaningful phenomena such as gesture we should also say that gestures are *dependent gestalts*, in that they depend on “subjective articulation” to receive their gestalt like properties. It should also be noted that this *subjective* aspect of action/gesture is not necessarily equivalent to perception of a gesture or an action. One could also follow Koffka (1938 [1915], p. 377) and argue that actions (which is what he discusses rather than gestures in particular) are not only gestalt-like in their perceived appearances, but also in the constitution of their production:

We may in fact place the experiencing of Gestalten squarely beside that of creating Gestalten; to sing or play a melody, dash off a sketch, write, and so forth, are not cases where one sings or plays *tones*, or draws or writes *strokes*. The motor act is an *organized whole-process*; the many individual movements can be understood only as *parts* of the process which embraces them, and it is indeed only thus that they attain their particularity.

This brings us to the question of parts and wholes in gestural action and social activity, and that is the theme of the next section. There I will be argued that gestures

are best understood both as (weak) *wholes* (action gestalts), constituted by underlying “operations” (its parts), but also that the gestures themselves serve as *parts* in an overarching activity, which also has the character of a (weak) whole.

### 3.4 Operation-act-activity interdependence

Leontiev (1981) made a threefold distinction between *operations*, *actions*, and *activity* for the purposes of describing the organization of activities. I suggest that it is also useful for the description of gestures and multimodal utterances. These terms will be used here, although in somewhat unorthodox ways, not strictly compliant with Leontiev’s own use of them. Through some convenient bending of the meanings of these concepts, they can be made to fit well with the idea of gestures and multimodal utterances as weak dependent gestalts. Thus, we may distinguish a level of *actions*, the level of action gestalts, which is the “level” in which we, as I intend to use these concepts, tend to perceive action both in the instrumental and the communicative sense (see Figure 3.3).

Further, there is the level “below” the actions, which Leontiev calls *operations*.<sup>11</sup> This will be referred to as the operational level. Operations in the case of gesture would be the actual movements that realize them, its parts, corresponding to the level of Kendon’s (2004) vocabulary of gesture phases (preparations, strokes, holds, etc). Whereas the action level is concerned with goals (or in the case of gesture, expression of meaning) operations are concerned with conditions. In the words of Leontiev (1981, p. 63), “If we imagine a case in which the goal remains the same and the conditions under which it is given change, then only the operational composition of the action changes.” In the case of single acts that consist in several operational steps, each individual step would generally be perceived as “insufficient”, and perhaps even an incomprehensible or incomplete “movement”, on its own, unless re-interpreted in terms of an action gestalt (if possible).

There is also the level “above” the level of actions, which is termed *activity*. Activities are, naturally, made up of a number of actions that form a whole on a higher level, sometimes with a more or less well-defined (and jointly established) overall motive. These acts need not involve gestures, but could be of any modality. It should be noted that the term “activity” refers both to the actual situated unfolding activity and to more situation-transcending kinds of “frames” (Goffman, 1974) relating to

---

<sup>11</sup>Habermas (1991 [1981], p. 97) makes a similar distinction, when he states that “a bodily movement is an element of an action but not an action.”

well-known and recurrent kinds of activities.

One could go on and introduce more distinctions other than the one between the single act and the overarching activity, as well as the one between the operation level and the action level. For example, if such a commonplace event such as a spoken utterance is considered, which could be seen as “one action”, depending on how it is viewed, it is often structured on multiple levels, such as phonemic organization, words, clauses, utterances, narratives, and so forth.<sup>12</sup> However, the threefold distinction just presented is enough for the purposes of this thesis. The point is not so much that a particular activity or action may not be more complex than what this threefold distinction is able to capture, or that the levels are always easily identifiable, but rather that a given act or multimodal utterance is always situated in the tension between the parts that make it up (operations) and overarching wholes in which it itself is a part of (activities, and other forms of cultural contexts, genres, traditions, etc.)

As already mentioned, Leontiev’s distinction between operations, actions, and activities, is used in an unorthodox way here, and while Leontiev emphasized a principled functional dissociation between the levels<sup>13</sup>, I rather prefer to use these terms in a way that does not downplay the possibility of dependencies between the levels that may exist in particular cases. That is, just because actions can be carried out in different ways, by means of different operations, that does not mean that the significance of an action is entirely independent of how it is performed, and the same thing goes for the actions that make up an activity.<sup>14</sup>

The relations between the levels of action and activity can be understood through the concept of *act-activity interdependence* (Linell, 1998, 2009; see also Duranti, 1991; Markova & Linell, 2007). An activity is partly constituted by its “constituent acts” (its parts), but the meaning of the “constituent acts” is also partly depending on their place in an overarching activity. For example, the utterance involved in pleading guilty in a court is part of what makes this activity a trial in the first place,

---

<sup>12</sup>That which is at the operation level for an adult, such a phoneme, may be at the action level for an infant; see Zinchenko (1981) for a discussion of automatization in relation to Leontiev’s terminology.

<sup>13</sup>According to the way Leontiev views these distinctions, “the same” action can be carried out in several ways, by means of different operations, and “the same” activity can be carried out in several ways, by means of different actions (see Wertsch, 1981).

<sup>14</sup>As a somewhat extreme example, one may consider how ritual is a type of activity that generally does not permit permutations, or exchanges of constituent acts for other alternatives, in the performative sequence (Sørensen, 2010), even though there is always some elbow-room for slight variation even in such formal contexts (Andrén, Sanne & Linell, 2010).

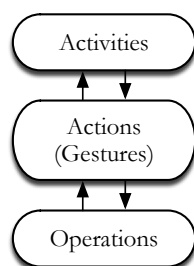


Figure 3.3: Operation-act-activity interdependence

and the act of pleading guilty becomes a true act of pleading guilty since it is part of a trial. Drawing on Leontiev’s distinction, I would argue that an analogous notion of *operation-act interdependence* could also be employed, or better yet, that the concept of act-activity interdependence could be extended into a more full-blown concept of *operation-act-activity interdependence*, placing action in the middle of this field of interdependence.

It is clearly not desirable to treat any of these interdependence relations as unidirectional arrows, since doing so results in invalid forms of reductionism. For example, downward reductionism that traces the meaning of action only to the action itself and its underlying constituents fails to acknowledge how the meaning of an action, such as a gesture, is not only a result of the properties of the movement that make the gesture up, but also how it figures as part of an activity. One example of a line of theorizing that has committed this mistake is several of the mainstream pragmatic theories, such as those of Searle, Grice, and Sperber & Wilson, that tend to ignore the holistic aspects of (communicative) action, as they try to derive the meanings exclusively from constituent acts (Linell, 2009, p. 187). Upward reductionism that traces the meaning of action only to their place in an overarching whole, such as activities, or even “societies” or “class” and so forth, is equally misguided. As pointed out earlier, the term “activity” is ambiguous in that it refers both to (A) the situated sense-making of a particular situation as well as to (B) more situation-transcending and recurrent kinds of activities that may serve as a framing of individual actions. With respect to A, upwards reductionism amounts to interactional reductionism (cf. Levinson’s 2005, p. 451 critique of Schegloff, 1987), where any action on behalf of an individual is solely treated in terms of its interactional nature. With respect to B, upwards reductionism amounts to rigid forms of structuralism, where the acts of an individual is only considered in terms of its third-person intentionality qualities (what Garfinkel, 1967 termed a “cultural dope” or “judgemental dope”), such as “acting as a middle-class man”, which effectively denies the actor of

agency (cf. Alfred Schutz' critique of Talcott Parson's structuralism, Grathoff, 1978 [1940-41]; Garfinkel, 1952; Heritage, 1984) and uniqueness.<sup>15</sup>

Linell's position in this matter, which I find productive for the reasons discussed above, is one of *moderate holism* (Linell, 2009, p. 18). In my interpretation, adhering to a principle of moderate holism means that both the (weak) holistic gestalt character of different levels (operations, actions, activities) as well as the dual-directional interdependence between all the three levels (operation-action-activity interdependence) are acknowledged.<sup>16</sup> It is still possible, of course, to conduct analysis of various phenomena in ways that pay more attention to the relationship between the operation-act interdependence than the act-activity interdependence, or vice versa, as a matter of delimiting the research questions in particular studies, but on a meta-theoretical level the whole triad and its interdependencies should be acknowledged. It may be pointed out that many, or perhaps even most, researchers would probably agree that an understanding of human action involves both matters of how the action is carried out by an individual as well as a matter of various forms of context-dependence. Therefore the principles argued for here may perhaps seem almost trivial and little more than commonsensical, but since so many analyses of human action have missed out on one or the other of these aspects as a matter of actual practice, it is still important to spell these things out as an underlying principle for both empirical analysis and theorizing. Not least because they are so fundamental to any characterization of human action, including gesture.

The operation-act-activity interdependence nexus also captures the fact that individuals are active as authors of their own actions (i.e., not only matters of third-person intentionality), at the same time as their actions are also constrained as well as supported by the social conditions and material contingencies of situations. That is, the view of agency in action endorsed here emphasizes both the *initiating/creative* and *responsive/sensitive* aspects of action and agency (cf. Andrén, in press b).

Before ending this section it will be useful to return briefly to the issues of gestural manual excursions, as discussed in the previous section, and to do so in light of the operation-act-activity scheme. As noted in the previous section, it is not always easy to identify the boundaries of these phases in a gesture. Nevertheless, even though the analytic vocabulary of gesture phases is often viewed as an analytic vocabulary that deals with "form" (e.g. Seyfeddinipur, 2006), it is here treated

---

<sup>15</sup>See also the discussion of situation-specific aspects of action (Sem#1) in Section 2.2.3.

<sup>16</sup>However, since the concept of operation-act interdependence is not directly taken from Linell (2009), but is rather my own extension of some of his (and Leontiev's) concepts, Linell should not be blamed for any shortcomings that might be associated with this extension of the conceptualization.

as a matter of “experienced form”, corresponding to the operation level, i.e., form as a perceived phenomenon rather than “objective form”. From this point of view, the gesture phase terminology mainly serves to capture the fact that the quality of the movements involved in the movement excursion changes over time, and hence needs to be distinguished as part of the analysis, and the fact that it may not always be possible to define exact boundaries where one phase shifts to another may not be so much of a “methodological problem” as an empirical fact about gesture.<sup>17</sup>

Also, as Kendon (2007, p. 5) points out, “it appears that these [gestural] movements can be recognised as such on the basis of their appearance alone”. That is, there are certain aspects of form that are sufficient for a movement to be seen by an observer as a gesture (see also Kendon, 1978). However, this does not necessarily account for all aspects of gestural meaning. In an elaboration on this issue, Kendon (2004, p. 15, my italics) also writes that what is perceived is “the quality of the action as intentional (*not the specific intention, necessarily*)” (my italics), which is to say that not all aspects of meaning derive directly from the operational moves that make up the gesture.<sup>18</sup> To be sure, in order to recognize certain gestures for what they are, such as recurrent gestures, recognition of the “stroke” of gestures on the action-level must also be taken into account because an observer must also be able to recognize the (familiar) identity of the action performed in the stroke, which requires previous experience of seeing precisely this type of action, and for this reason local aspects of form are not enough. Indeed, the exact form of a recurrent gesture may even vary somewhat across instances, but still be recognized as an instance of a particular type, on the action-level (this was also discussed in Section 2.3.3). Recog-

---

<sup>17</sup>Another consequence of treating the gesture phases in terms of experienced form (i.e., appearances) rather than objective form is that it is not assumed that a set of rules could easily serve to produce automatic gesture phase coding by a computer. The problems involved are similar to those that faces speech-recognition systems or computerized vision, where a certain degree of semantics always have to be included, sooner or later, as well as contextual factors. With that being said, I do not claim that criteria for coding gesture phases should be avoided as such. Developing coding criteria may certainly help to provide methodological consistency as well as an improved understanding of the nature of gesture phases.

<sup>18</sup>Even though it is clear that most gestures are indeed recognizable as gestures by virtue of form alone, one may still dispute that *all* expressive bodily actions are recognizable as gestures only by virtue of their form. One class of actions that may sometimes be less clear in this regard are those acts that have both instrumental and communicative properties, since it may sometimes be contextual aspects that makes it clear that an action is not just an instrumental action. An example of this may be a demonstration of how to perform a certain instrumental action, for which the communicative status may be more due to contextual factors. It also depends on whether “surrounding” aspects of the performance is included as being “part of the act” or not, such as the way gaze is coordinated during the performance of the action. However, Kendon’s discussion is mainly concerned with empty-handed gestures, and for empty-handed gestures the form of a movement may indeed be enough for an observer to be able to see it as a gesture.

nition of a gesture as a particular type of action on the action-level is also relevant to many gestures with ACTION\_BASED iconic aspects, even if the gestural enaction of some sort of action may itself not necessarily be a conventionalized gesture, since many ACTION\_BASED gestures are enactments of recurrent and well-known (i.e., typified, Conv#2) sorts of action (i.e., on the content level) that needs to be known to an observer in order for the observer to see it as an “instance” of a particular type of action. In fact, most gestures require the observer to somehow “recognize” something in the gesture, which means, per definition, that it is not only the local form of the gesture that underlies the possibility to see a certain kind of meaning in the gesture, although the recognition of an action as a more specific type of action is of course mostly relevant to gestures that are either recognizable as a certain kind of gesture (i.e., recurrent gestures) or as ACTION\_BASED iconic enactments of a recognizable recurrent type of action. Furthermore, as Kendon (2004, p. 16) also points out, “How actions that vary in terms of the features of manifest deliberate expressiveness are interpreted, however, will depend upon context”, which is effectively the recognition that the activity level must also be added to the scheme in order to arrive at a proper understanding of how it is that gestures come to have a certain meaning in the situation in which they occur.<sup>19</sup>

## 3.5 Summary

In this chapter, I have argued for a view of action and gesture which can be characterized as a *moderate holism* (cf. Linell, 2009). More specifically, this means that actions are viewed as a kind of gestalts (*action gestalts*). However, they do not exist in a vacuum, but emerge as a result of several factors. This was captured in terms of a principle of *operation-act-activity interdependence*, where each level to some extent co-determines the other levels. This has direct consequences for the interpretation of gestural actions, because it effectively serves as an imperative to acknowledge the interplay between all these three levels of meaning in any analysis of what a given gesture comes to mean (at least ideally). This view stands in contrast to only describing the meaning of a gesture from the point of view of how it figures in, and is relevant in, the sequential flow of actions in a social activity. It also stands in contrast to an account of the meaning of the gesture purely in terms of hand-shape and bodily movement.

---

<sup>19</sup>Kendon’s analyses are indeed systematically carried out within his framework of contexts-of-use analysis (cf. Kendon, 2004, p. 226).

Furthermore, gesture performance and gesture interpretation is in most cases also an interpersonal affair. Therefore I have offered an account of how to think about notions such as intersubjectivity and interpersonal understanding, by the theorizing of Alfred Schutz. This account provides a middle row between present day theories on Theory of Mind, that views interpersonal understanding as a matter of mental inferences, and accounts of intersubjectivity that views interpersonal understanding as a matter of a direct and unproblematic perception of the Other. Simply put, Schutz' emphasizes that it takes active work to achieve interpersonal understanding, and in relation to this, the central importance of the so-called *manipulatory area*. Hence, interpersonal understanding cannot be sufficiently understood as a purely mental affair, without taking into account the actual activity that people engage in when they are in the presence of each other. Edwards (1997, p. 118) eloquently characterizes this artificial way of thinking about interpersonal understanding as a "problematics of the frozen moment" where no action is possible. At the same time, interpersonal understanding cannot be sufficiently understood as an unproblematic direct perception of others' thoughts and intentions. Once again interpersonal understanding must rather be understood as something which also involves active work on behalf of the participants. Hence, the emphasis on social activity serves as an antidote to the one-sided character of both of the views on intersubjectivity that were criticized.

Schutz theorizing on intersubjectivity is also relevant to this thesis because it emphasizes the key role of processes of typification in social life. The notion of typification has direct relevance for the understanding of gesture, and to some extent it has also already been discussed in relation to Level #2 of semiotic complexity in Chapter 2. The relevance of the concept of typification to gesture and action will be further explicated all of the empirically oriented chapters in Part II of this thesis, although to a lesser extent in the quantitatively oriented Chapter 5.



## CHAPTER 4

# Data, annotation, and other methodological issues

### 4.1 Data

The data analyzed for this thesis consists of longitudinal video recordings of Swedish parent-child interaction at home. Five children are included, and each child was recorded at least once a month from 18 to 30 months of age. Common activities include playing with toys, book reading, eating, and conversation on various topics. Some activities recur across sessions in similar form. Most of the data concerns interactions where the participants sit side by side by a table (either along a one side of a table, or in an L-shape around a corner of a table, cf. *interactive formations*, Kendon, 1990, p. 213) and in many cases the interaction is oriented to, and organized around, various objects. However, some other forms of interactive alignments occasionally appear in the data too, for example when playing with toys on the floor, where movement tends to be less restricted. The activities are spontaneously organized, although the children have the attention of their parents secured throughout, since the parents know that they are recorded and make sure that there is some kind of interaction going on all the time. Hence, the vast majority of the data consists of “focused interaction” (cf. Goffman, 1963; Kendon, 1979).

The five children are here called Alice, Bella, Hanna, Harry, and Tea (this is not their real names).<sup>1</sup> Information on the recordings of these children is provided in Table 4.1. Recordings of three of the children (Bella, Harry, and Tea) comes from

---

<sup>1</sup>In other publications (Andrén, in press a, b, c), I have used other more English-sounding names for these children (Alice=Amy, Bella=Betty, Hanna=Helen, Harry=Howard, Tea=Tess).

Name	Sex	Age range	No. of sessions	Total time annotated
Alice	F	17;03–30;14	9	1h 30min
Bella	F	18;09–29;24	14	2h 20min
Hanna	F	18;18–30;21	13	2h 10min
Harry	M	17;26–30;10	13	2h 10min
Tea	F	18;10–30;23	12	2h 0min
Total:			61	10h 10min

Table 4.1: Subjects

the Strömquist-Richthoff corpus, collected by Ulla Richthoff at the Department of Linguistics at Gothenburg University (Richthoff, 2000). Recordings of a fourth child (Hanna) were collected by Åsa Wikström at the Centre for Languages and Literature at Lund University (Wikström, 2008). Recordings of a fifth child (Alice) were collected by myself. Although the data has been collected by different persons and for different analytical purposes, they are still very similar. They have been filmed in similar ways, they include similar kinds of activities (no particular instructions has been given to the participants about what they should do in any of the recordings), and so forth. All the five children come from middle class families in the southern (Alice and Hanna) or western parts of Sweden (Bella, Harry, and Tea). Four of the five children are first-born. The exception is Tea, who is the second child in her family. All of the children have grown up with both their parents living in the same household.

Instead of transcribing and annotating long stretches of interaction in few recorded sessions, the strategy here has been to annotate less time for many recorded sessions (10 minutes for each session). In doing so, there is both a richer variety of activities included than what would otherwise be the case, which is good for qualitative analyses that are focused on the use of gestures within the organization of different activities. The richer variety of activities also means that single kinds of activities will have less influence on the overall results in quantitative analyses that are primarily focused on the effect of age on various rate measures than the influence of specific activities. Perhaps more importantly, including more sessions means that there is a finer temporal resolution in the longitudinal analyses presented in Chapter 5 than in most other studies of gesture in children of these ages. In other studies, the ages of children are typically either observed or quantized in 3 (or more) month intervals (e.g. Acredolo & Goodwyn, 1990; Hannan, 1992; Nicoladis et al., 1999; Guidetti, 2002; Rodrigo et al., 2004; Rowe et al., 2008; Özçalışkan & Goldin-Meadow, 2009), and some of these studies also do not involve following the same individuals over

time, but many different children at different ages and compare gesture characteristics at various ages across individuals, i.e., cross-sectional rather than longitudinal, as in the present study. There are some studies with more dense observations, but the vast majority of these begin earlier than the period studied here and extend up to around 20–24 months (e.g. Zinober & Martlew, 1985a,b; Caselli, 1990; Blake et al., 1992; Butcher & Goldin-Meadow, 2000; Camaioni et al., 2003; Capirci et al., 2005; Pizzuto et al., 2005; Pizzuto & Capobianco, 2008; Iverson & Goldin-Meadow, 2005; Iverson et al., 2008), thus ending 10–6 months before the period studied here. In sum, the temporal denseness of the observations, in particular for the age range between 24 and 30 months, is the most unique contribution of this thesis when it comes to the kind of data that is analyzed, in addition to the fact that the age range as a whole is less studied than earlier periods, as argued before (see Section 1.2).

## 4.2 Transcription of speech

The transcriptions were all in CHAT format, which is a notation format that can be processed by the CLAN software (MacWhinney, 2000).<sup>2</sup> Bella, Harry, and Tea were transcribed by Ulla Richthoff, Hanna was transcribed by Åsa Wikström, and Alice was transcribed by myself. In addition, I have modified all of the transcriptions for consistence, when necessary, and I have also made all the gesture annotations for all the CHAT files. For all children except for Alice, between 20–40 minutes of talk was transcribed for each session, which is more than the 10 minutes of each session that had annotations for gestures. This means that some measures, such as mean length of spoken utterances (MLU) could benefit from more data. Following Özçalışkan & Goldin-Meadow (2009, p. 194), the transcriptions are divided into utterances defined as word(s) or gesture(s) “alone or in combination, that was preceded and followed by a pause, a change in conversational turn, or a change in intonational pattern.” One difference is that there is no distinction made here between vocalizations and words — a notoriously tricky issue in research on research on children’s communicative development (e.g. Vihman & McCune, 1994). Nevertheless, the children are clearly way beyond the babbling stage, and for the most part, the non-word vocalizations still have a somewhat word-like character, although they are not readily recognizable as a particular word.

It should also be added that in the case of utterances that only include gesture

---

<sup>2</sup>CLAN can be retrieved free of charge from: <http://childes.psy.cmu.edu>

but no speech, two or more gestures have been considered to belong to the same utterance when they are performed in direct sequence without a retracting movement in between (i.e., *recovery*, Kendon, 1980b, 2004). Furthermore, in cases where a single gesture extends across several spoken utterances, this has been counted as several utterances, corresponding to the number of spoken utterances involved. An example of the latter is when a pointing gesture remains directed to its target across more than one spoken utterance.

## 4.3 Gesture annotation

### 4.3.1 Annotation as indexing and as coding

As mentioned above, the recorded sessions are transcribed in the CHAT notation format and these transcriptions also contain gesture annotations. The annotations are used in two different ways in this thesis, which may be called *coding* and *indexing* respectively.

**CODING** In Chapter 5, which has a quantitative focus, the annotations are used directly to compute research results in the form of various graphs that show broad changes over time in the children's use of gesture. That is, when a particular gesture has been "coded", this particular segment of the video recording is not used anymore for analysis, but any further analysis rather consists in performing calculations on the *codes*.

**INDEXING** In the more qualitatively oriented analytic chapters, the annotations are rather used as a starting point for further analysis of the videos (and for more elaborated accounts of the semiotic properties of various gestures). In these chapters, the annotations serve as an "indexation" that facilitates the navigation of the data, so that several instances of similar or contrasting phenomena in the videos can be identified, revisited, and analyzed further in more detail.

### 4.3.2 Main annotation categories

The first step in the annotation process was to specify every instance of a gesture in terms of various deictic, iconic, and conventional features. The list below specifies the main categories and sub-categories that were applied to all gestures performed by the children that were encountered in the data. *These categories are not mutually exclusive* (cf. Zlatev & Andrén, 2009). They are not to be understood as a typology

of the kind where a given gesture must be assigned to one and only one of these categories. A pointing gesture may have iconic and conventional elements, a conventionalized gesture may have iconic elements, and so forth.<sup>3</sup> Yet it is often the case that one of the aspects are perceived as dominant in a gesture.

The categories were not predetermined before the annotation procedure begun, but emerged in an iterative and “dialectic” fashion during the annotation process itself. This means that all categories are essentially motivated by what was found in the data. It also means that there was no action that was judged to be a gesture (in the sense defined in Chapter 2) but which was not covered by at least one of the annotation categories. The categories are summarized in Table 4.2. Arguably, the annotation categories are relatively coarse, in the sense that they hide a relatively large number of more specific distinctions that may nonetheless play a significant role in the nature and meaning of various gestural movement. The distinctions are therefore mostly to be considered heuristics that may be used to capture some large scale quantitative patterns in the developmental data. It should also be kept in mind that these annotations are used in this thesis both for purposes of *coding* and *indexing*, as explained in the previous section. It is only in the case of using the annotations as coding that the analysis is restricted to these terms, and furthermore, keeping the number of categories relatively low facilitates quantitative treatment (as in Chapter 5). When the annotations are used as indexing, they pose little restrictions on the possibility of further treatment in terms of more detailed situated analyses of particular gestures and groups of gestures.

In all essential respects the annotation categories presented in the remainder of this section correspond to the categories used in an earlier study (Zlatev & Andr n, 2009), which involved three of the children from the present study (Bella, Harry,

---

<sup>3</sup>Despite insights from research on adults’ gestures according to which one and the same gesture may simultaneously be meaningful in many ways (Kendon, 2004; McNeill, 2005), very few studies of children’s gestures have a design that take into account the fact that a single gesture may well have more than one of the properties: deictic, iconic, and/or conventionalized. In some cases this may have methodological motivations, because typologies with mutually exclusive categories may sometimes be required for certain statistical calculations, but it is clear that it also creates some relatively artificial problems. For example, some researchers consider REQUEST (extension of a palm up open hand to request an object) to be a conventional gesture (Goldin-Meadow & Morford, 1990;  zcaliřkan & Goldin-Meadow, 2009, , although they call this gesture GIVE — note that these researchers name this gesture “GIVE” instead of “REQUEST”, as most other researchers do), and others consider REQUEST to be a deictic gesture (Volterra et al., 2005). Such problems can only be expected as long as research on children’s gestures operates on the assumption that each gesture should be coded as being of *one* type.

Category	Sub-category	Cohen's Kappa
DEICTIC	<i>INDEX FINGER POINTING</i>	$K=0.94$
	<i>OTHER POINTING</i>	$K=0.68$
	<i>OTHER (OFFER, SHOW, PLACE etc.)</i>	$K=0.87$
ICONIC	<i>ACTION-BASED</i>	$K=0.89$
	<i>INDIRECT</i>	$K=1.00$
	<i>TRACING</i>	$K=0.50$
CONVENTIONALIZED $K=0.82$	<i>NODDING</i>	
	<i>HEAD-SHAKE</i>	
	<i>HELLO</i>	
	<i>BYE-BYE</i>	
	<i>etc...</i>	

Table 4.2: Summary of main annotation categories

and Tea) as well as three Thai children of the same ages.<sup>4</sup> In this earlier study there was also a calculation of the inter-rater agreement between the two coders for a subset of the data that included 190 gestures. The measure used was Cohen's Kappa, which is a measure of the extent to which the correspondence between the annotations of two different raters correspond *above chance*. It was calculated on the level of each sub-category, except in the case of conventionalized gestures, where there were many sub-categories. For the conventionalized gestures the Kappa score was instead computed at the top category level. The Kappa scores are shown in Table 4.2.

The Kappa scores were overall *substantial* (between 0.61–0.80) or *almost perfect* (between 0.81–1.00), apart from the TRACING sub-category where the agreement was *moderate* (between 0.41–0.60) according to the criteria guidelines suggested by Landis & Koch (1977). According to the guidelines suggested by Fleiss et al. (1969, p. 281), the same scores would be interpreted as ranging between *excellent* (between 0.75–1.00) and *fair to good* (between 0.40–0.75).<sup>5</sup> The somewhat low Kappa score

<sup>4</sup>In this earlier study, there was also a category of action-based gestures where the signified was not the action performed by the body itself, but rather some object (real or imaginary) to which the action was directed, and more specifically, those cases where the acting towards the object evoked some sort of characteristic of this object, such as shape or size. This category is merged with the action-based category here, partly because there are very few instances of this category.

<sup>5</sup>It should be noted that both of these guidelines are somewhat arbitrary (Bakeman & Gottman,

for TRACING may be explained by the fact that it is sometimes hard to tell whether a movement with a tracing character is (mostly) a matter of tactile/visual exploration (“impression”) or whether it is actually a matter of expression. Sometimes it is also unclear whether a movement that may seem to be tracing is just a matter of a less clearly articulated gesture or if it is actually part of what is expressed. This is so especially in the early half of the period studied here, where gestures are overall less clearly articulated than in the latter half of the period.

Note that whenever the term *articulating medium* is used below, it refers primarily to the body and in particular to the hands, but it may sometimes also include objects that are involved in the articulation.

#### Deictic

If a gesture actualized a *target* referent in the physical surroundings, such as an object, location, direction, it was marked as deictic. Included in the deictic category are both actions where attention is directed towards a target, as in the prototypical case of pointing, and actions where the object is brought to attention, as in the prototypical cases of SHOW or OFFER (cf. Clark, 2003).<sup>6</sup> For every instance of a gesture with deictic aspects, at least one of the following sub-categories were specified:

*INDEX FINGER POINTING* If a pointing gesture was performed with the index finger extended, and the other fingers more or less curled, the gesture was marked as an instance of INDEX FINGER POINTING.

---

1997), since there are no generally agreed upon procedures for interpreting the acceptability of inter-rater agreement measures, partly because even for a Kappa value that is just slightly above chance the Kappa value may, on the one hand, be statistically significantly above chance, but on the other hand, it might still be considered an unsatisfactory low degree of agreement. It should also be noted that Cohen’s Kappa is a slightly too conservative measure of inter-rater agreement. The reason for this is that it assumes that when a rater is not sure of how to rate a gesture, it is rated purely by chance, i.e., that the probability of choosing to apply an annotation category or not would be 50%/50%. Annotation is typically not a matter of pure guessing in that way, neither in general (Uebersax, 1987) nor more specifically for annotation of gestures. To be sure, a rater may, for example, doubt whether or not to call a specific movement a pointing gesture so that the eventual judgment for the pointing category in this case may be something like a 50%/50% probability, but at the same time, the rater may be quite sure that it is not an iconic gesture so that the probability for not judging this gesture as a conventionalized gesture is much higher than 50%. In the end, no really satisfactory measure of inter-rater agreement exist for the kind of annotations that gesture researchers typically use, although there is some work in progress aimed at developing such a measure (Rein & Holle, 2010).

<sup>6</sup>Most other researchers on child gesture follow Bates et al. (1979) in calling acts of offering someone something GIVE, rather than OFFER. This label is not used here, since a “give” event typically requires collaboration from a recipient, and I prefer a label which is neutral as to whether the recipient actually takes the object that is offered or not.

*OTHER POINTING* If a pointing gesture was performed with some other hand-shape than the index finger pointing hand-shape, it was marked as an instance of *OTHER POINTING*. Included here are also cases where the articulating medium includes an object, i.e., pointing with an object. This subcategory turned out to be somewhat heterogeneous, since it involves both pointing gestures with a lack of articulatory specificity as well as pointing gestures that are clearly articulated, but where the hand-shape is different from that of index finger pointing.

*OTHER* Other forms of actions that may serve to ground attention to referents in various ways, such as *OFFER* and *SHOW*, were assigned to this category. Following Clark (2003), some cases of *PLACING* objects were also included, in cases where this had interactional significance, not only being a matter of disengaging from objects. Even though most of the acts in this category involves a physical object in some way or another, it is not necessarily so. One may, for example, extend an empty palm toward the interlocutor to *SHOW* that the hand is empty.

### Iconic

If a gesture evoked a certain referent/content by means of some aspect of *similarity* between expression and content, it was marked as iconic. For every instance of a gesture with iconic aspects, at least one of the following sub-categories were specified:

*ACTION-BASED* The articulating medium is used for expression in a way that is *congruent* with the way the body is used as a practically acting body.<sup>7</sup> That is, the body represents a body, although not necessarily the body of the actor.<sup>8</sup> In such actions there is an explicit or implicit mapping of the whole body onto the signified, even if only a part of the body is thematic in the articulation. The extent to which these gestures actualize a kind of action or actualize an

---

<sup>7</sup>This corresponds, more or less, to what other authors have called *enactment* (Kendon, 2004, p. 160), *character viewpoint iconics* (McNeill, 1992, p. 67), or *the acting mode* (Müller et al., manuscript).

<sup>8</sup>The person who performs an action-based gesture may take the perspective of another being, as in the case of “quoting” (by means of a bodily expression) how Elvis behaved on stage. Another example is when someone pretends to be an animal, such as a dog, and acts in a “doggish” way, but still so that the whole body of the performer is congruently mapped onto the body of a dog, i.e., the mouth of the articulating person corresponds to the mouth of the dog, and the eyes of the articulating persons corresponds to the eyes of the dog, and so forth.



object (real or imaginary) that is involved in the action may vary. Since the spatio/temporal domains of the signifier and the signified are highly congruent these gestures sometimes have very little differentiation between expression and content/referent, and they may therefore sometimes be on the verge to “actual” action (i.e., close to the lower limit), especially in some of the cases where an actual object is involved in the articulation.

*INDIRECT* The articulating medium signifies something which is *not congruent* with the body in its ordinary role as a practically acting body.<sup>9</sup> That is, in indirect iconic acts of bodily signification the hand may signify something different than a hand. In cases where the articulator signifies an object, the object can either be some object typically held in the hand, so that the hands takes on a shape that signifies a pair of scissors or a gun as if they were held in the hand, or it can be something completely different, such as a mountain in Africa. In the first case, only the part of the body that is involved in the articulation becomes something different from a body, and these gestures may be relatively similar to action-based gestures. In the latter case, there may be very little congruence with the spatio/temporal domains of the signifier and the signified. Indirect iconic gestures may vary in specificity, so that the hand-shape can either be something abstract and non-specific, such as an “entity”, or it may be something quite specific, such as a frog. Another kind of gestural use of the body that is included in this category is when an object, such as a doll, is moved by the hands so that the doll is “walking”. In this case too, the movement of the articulating body are nothing like the movements involved in actually walking.

*TRACING* This category refers to gestures that involve tracing a shape or a path of movement in an expressive manner (cf. Wundt, 1973 [1921]). In the children studied here, this is almost always done as part of some sort of pointing gesture to a target that is present in the world, as a way of performing pointing which also exploits the potential of moving the pointing hand in a way that “leave traces” (cf. Müller et al., manuscript), so that the pointing gestures also come to posit iconic qualities (resemblance of the shape/path of the movement and the shape of an object or a path of movement) (cf. Goodwin, 2003).<sup>10</sup> For

---

<sup>9</sup>This corresponds, more or less, to what other authors have called *modeling* and *depiction* (Kendon, 2004), observer viewpoint iconics, (McNeill, 1992), and *the representing mode* (Müller et al., manuscript).

<sup>10</sup>Accordingly, such instances are annotated as both INDEX FINGER POINTING and TRACING.

this reason, they are slightly different from ACTION-BASED and INDIRECT gestures, at least as they are defined here, and they are therefore assigned to their own category in the annotation system. More generally, however, it is not all that clear that tracing gestures of all kinds are always clearly different from both ACTION-BASED and INDIRECT iconic gestures. In some cases, the marking up of the world that is involved in TRACING can be very similar to ACTION-BASED gestures — for example, tracing may look very much like drawing, including a somewhat drawing-like hand-shape and perhaps performed towards an empty sheet of paper lying on a table. At the same time, drawing is itself an expressive/significational kind of action, and the meaning that is being expressed is often entirely distinct/distant from the spatio-temporal domain of the articulating body, such as when tracing the shape of mount Kilimanjaro in the air. The latter variant of TRACING is more similar to INDIRECT iconic gestures, and of course, tracing can be performed in many ways, which means that the more specific properties that are involved in TRACING gestures may vary from time to time. All in all, it may not be possible to defend that TRACING should be considered an altogether distinct kind of gesture, sharing no properties at all with other modes of gesture creation. As argued before, descriptions of gestures which are more in line with the family-resemblance model of “categorization” (e.g. Wittgenstein, 1953) may ultimately be better suited for describing the range of acts of bodily signification and acts of bodily communication than a model of categorization that treats each gesture as belonging to one and only one “type”, whereas a more detailed specification of the gesture may show both similarities and differences with respect to other broad “types” (cf. Kendon, 2004; Zlatev & Andrén, 2009). To repeat, then, the motivation for treating TRACING as a separate annotation category here was primarily motivated by the empirical fact that the tracing gestures of the children did not overlap particularly much with other kinds of gestures, rather forming an own little family of uses growing out of pointing.

### **Conventionalized**

If a gesture had a recognizable conventionalized form-meaning mapping, in the sense of a at least partially normatively regulated form, the gesture was marked as

conventionalized.<sup>11</sup> As described in the discussion of “Levels of Conventionalization” (Section 2.3.3), to say that something is normatively regulated means that it is not only conventionalized in the sense that it is normal or typical to do something in a certain way in a specific cultural community, but that the convention has a strong enough specification (in the community) so that it is also *correct* to do it in a certain way according to culturally established norms, and in particular, *incorrect* to do in some other too deviating way. Each conventionalized gesture was also sub-categorized in a lexicon-like manner, with labels that are specific to each type of gesture such as NODDING, HEAD-SHAKE, HELLO, BYE-BYE, and so forth. That is, the way conventionality was coded here corresponded quite well to the notion of emblems (the analysis of conventionality in gesture is more nuanced in the qualitatively oriented analytic chapters).

### Other annotations

Other annotations that was made included annotations of whether a each particular gesture involved active handling of an object or not. This distinction is obviously of a different kind of distinction than the semiotic properties of a gesture, i.e., its indexical, iconic, and/or conventionalized properties. Another annotation that was made specified whether a particular gesture extended across several spoken utterances, so the same gesture would not be counted twice in cases where this would not be desired.

### 4.3.3 Comparison with other classification systems

Many other studies of children’s gestures make use of only two main categories, namely deictic gestures and “representational gestures”, instead of the three categories used here (deictic, iconic, and conventionalized). Not all researchers use the term “representational gesture” specifically, and variants such as “characterizing gestures” and “content-loaded” gestures are sometimes used instead. These terms are not always used in synonymous ways, but one common interpretation of this “category” is that it involves a lumping together of the categories of “conventional-

---

<sup>11</sup>Index finger pointing gestures were not marked as instances of conventionalized gestures, although they are common in children and although it is likely that the specific form is at least in part conventionalized. The issue of conventionality in pointing is further discussed in Chapter 6. Nevertheless, index finger pointing constitutes a annotation category of its own, which means that it was still possible to add these gestures to calculations of the rate of conventional gestures in cases where this was desired.

ized” and “iconic” into a single category (e.g. Iverson et al., 1994).<sup>12</sup> This is good in the sense that it (implicitly) serves to acknowledge (at least implicitly) the fact that a given gestures need not be either conventionalized or iconic, but that it can be both.<sup>13</sup> However, on the whole, there is no real analytic reason to lump iconic and conventionalized aspects together as a rule, rather than as an exception, as long as the annotation system allow for the same gesture to be coded with several different categories at once, as the annotation system here does.<sup>14</sup> The use of a category of “representational gesture” has sometimes been motivated in order to construe gesture in similar terms as spoken language. For example, Volterra et al. (2005, p. 12), who make use of the term “representational gesture”, argue that:

in order to pursue appropriate comparisons it is necessary to use the same criteria, and a uniform terminology, for identifying and distinguishing gestures, signs and words in the communicative productions of both deaf and hearing children.

That is, children’s spoken language, just like gesture, can be divided into less content-loaded, and typically deictic, words which can be used to refer to many different referents and words with a more representational and content-loaded character such as verbs and nouns, that refer to a more specific referent (or content). Therefore, a twofold distinction between deictic and representational/content-loaded units makes it possible to use the same annotation categories both for gesture and for speech. Yet, it is well-known that, on the whole, gesture (among hearing people who use speech as their primary communicative medium) is typically, but not necessarily, more toward the iconic side of the spectrum when compared to spoken language, whereas words of spoken language are more or less by definition conventionalized in character, and aspects of iconicity are for the most part less central (although not non-existing) in spoken language (e.g. McNeill, 1992, 2005). Lumping iconic aspects and conventionalized aspects into one category may be motivated for the analytic purposes of particular studies, but on the whole, it cannot be consid-

---

<sup>12</sup>Sometimes “representational” gesture is rather used as a synonym for iconic gestures, and conventionalized gestures are assigned to a separate category (e.g. Rowe et al., 2008, p. 187).

<sup>13</sup>Although, at the same time, this dichotomous categorization tends to exaggerate the distinction between deictic and “representational gestures”, and overlaps between these categories, such as pointing gestures with iconic properties (cf. Goodwin, 2003), are rarely discussed in research on children’s gestures.

<sup>14</sup>As a side note, Peirce (1931–35) also treated his threefold distinction between indexicality, iconicity, and symbolicity as aspects that may be present in one and the same semiotic sign, rather than mutually exclusive categories that would only allow one of these aspects in a single semiotic sign.

ered a satisfactory way of characterizing gestures. Some other researchers make use of three categories, corresponding to the three categories used here (e.g. Guidetti, 2002), but then typically in a way which treats the three categories as mutually exclusive, at least on a methodological level.

## 4.4 Rate measures

This section explains the three main rate measures that are used in Chapter 5, namely Gestures per Minute (GPM), Gestures Per multimodal Utterance (GPU), and Mean Length of Utterance (MLU).

### 4.4.1 Gestures per Minute (GPM)

Calculating the number of gesture per minute provides a measure of the “absolute” frequency of gesture use at various ages, only corrected for the length of the observations. It is preferable over a raw count of the absolute number of gestures during an observational session, since it makes it easier to compare values across studies which involve observational sessions of different length. However, GPM may vary quite a lot between sessions, since the degree of children’s involvement may vary between sessions, due to differences in activities (such as “book reading” or “playing with toy building blocks”), differences in how these activities are implemented (one can implement activities such as “book reading” and “playing with toy building blocks” in many ways), mood, and other factors.

### 4.4.2 Gestures per multimodal Utterance (GPU)

Calculating the number of gestures per multimodal utterance provides a measure of the frequency of gestures which is less sensitive than the GPM measure to differences between sessions regarding how active the child is — the degree of involvement. As mentioned above, GPM may vary quite strongly between sessions due to different activities, interactional frameworks employed within types of activities, mood, and so forth. Therefore, the GPU measure is sometimes better, since it measures the extent to which gesture is used *once the child communicates*. On the other hand, GPU may sometimes be undesirable precisely because it hides this variation in degree of involvement.

It should be noted that there is a slight difference between calculating GPU by

using multimodal utterances (speech only, gesture-only, and gesture+speech utterances) as the baseline, or by using spoken utterances (speech only and gesture+speech utterances) as the baseline. That is, sometimes gesture-only utterances are excluded (e.g. McNeill, 1992) and sometimes they are not (e.g. Butcher & Goldin-Meadow, 2000). The version of GPU used in this thesis uses multimodal utterances as the baseline, including also gesture only utterances.

### 4.4.3 Mean Length of Utterance (MLU)

Mean Length of Utterance (MLU) is a widely used measure which consists in calculating the average number of *morphemes* in a child's spoken utterances (Brown, 1973). Although MLU does not cover but one aspect of the linguistic skills of a child, it is nevertheless a useful measure (cf. Bates et al., 1988). The way MLU is calculated here, the "units" of an utterance may be both *words* or (non-word) *vocalizations*. One reason for this is simply that the transcriptions of the speech in the data does not (yet) include any coding that distinguishes word forms from vocalizations. In addition, there is no coding here of morpheme boundaries within multi-morphemic words. That is, a multi-morphemic word such as "cook-ing" is counted as two units in the original MLU measure, whereas it only counts as one unit the way MLU is calculated here. However, for the most part of the period studied, there are evidence available to believe that also multi-morphemic words are mostly learned as "whole chunks" for the children, and that it is only at around 28 months (or later) that grammatical features like verb inflection typically become truly productive more generally (Brown, 1973; Bates et al., 1988; Tomasello, 2003). Therefore, it could even be argued that it is misleading to count the number of morphemes in words before this sort of productivity has appeared.

The lack of a distinction between words and vocalizations might be desirable in future work, but at the same time it should also be observed that notion of a "word" and its distinction from "vocalization" is a difficult chapter in child language acquisition theorizing (Vihman & McCune, 1994). This is so for at least two reasons:

First, it is hard to draw a clear limit between vocalizations and words. Calculating MLU without going into the difficult issue of separating vocalizations from words may therefore be a more reliable measure (although it will be less directly related to grammatical complexity), because it is relatively easy to judge whether there is some sort of spoken vocalization/word or whether there was just silence, whereas the judgment involved in distinguishing words and vocalizations from each other is difficult and even somewhat arbitrary, and it often involves a fair amount of guess-

ing. In particular in the beginning of the period studied here, many spoken sounds may have a clear syllable structure, but at the same time, the degree to which these vocalizations resemble recognizable conventional expressions may be highly variable. An example of this is when Hanna (18 months) says “look doj” (*titta doj*) while holding up a toy figure which is supposed to be a troll, and the mother asks “a troll?” (*ett troll?*), and Hanna replies “yes” (*ja*). In such cases it is often gestures and other forms of “contextual” support that makes the word recognizable and interpretable for the parent and for the researcher. Also, it is not uncommon for children to use their own personal variants of some conventional words. Such variants may or may not be identified by a parent or a researcher, and even if they are in fact recognized, the question remains as to whether they should be considered “words” or not.

Second, the scope of the “units” of children’s speech is often hard to specify due to the existence of formulaic expressions, or rote-learned constructions, that include several “words” (from an adult point of view) (e.g. Peters, 1983; Tomasello, 2003; MacWhinney, 2005), but may rather work as whole “lexicalized” units for children. Such formulaic constructions are by definition “lexicalized” in ways that differ from what might be specified by adult norms. These constructions may not necessarily be entirely rote-learned and static, but there can also be some limited degree of variation in the “arguments” taken by the construction, so that it is somewhere in between truly lexicalized wholes and flexible constructions. That is, what appears as a two word utterance to an adult may be a single “action” (“holophrase”, “item-based construction”) for the child. MLU, as measured here, is therefore best thought of as a rough measure of children’s actual state of spoken production.

In general, non-word vocalizations are still occurring with some frequency in the beginning of the period studied here, but at the end of the period, they are rare.

## 4.5 Presentation of examples and data

### 4.5.1 Visual presentation of examples

While inter-rater agreement measures are always valuable, a more pressing issue is how to achieve “inter-researcher agreement”, across groups of researchers. After all, good scores on inter-rater agreement measures only specifies that two raters agree on how to annotate some data, but it does little to communicate the way that gestures are judged and annotated across separate groups of researchers. Indeed, several researchers have complained about the lack of consistency and shared understand-

ing between researchers regarding the analytic vocabularies and categories used in research on children's gestures (Nicoladis et al., 1999, p. 515; Volterra et al., 2005, p. 9; Guidetti & Nicoladis, 2008, p. 109). Presenting actual examples in somewhat detailed form may be a good way to achieve a better shared understanding of how gestures are analyzed by different researchers, in a way that is less sensitive to the confusion of the current terminological jungle, since the reader of a publication will be better able to judge for himself/herself whether the interpretation of a given gesture seems reasonable than when faced only with quantitative data or with brief descriptions of a prototypical example of a certain analytic category in a sentence or two.

The next question then becomes: *How to present gestures on paper in a somewhat "readable" and intuitively comprehensible form?* The answer to this question will of course depend on the more specific aims of specific studies, and even though some methods may be better than others more generally there is always a trade-off between the level of technical detail that is presented and the readability of the presentation. Nevertheless, instead of following the common practice of putting a transcription of speech at the center of the representation of an example, and then linking pictures to this transcription in various ways, the method which is tried out in this thesis is to do it the other way around. That is, the method for representing examples of gesture performances used here puts images of the interaction at the center, and then links speech to these images or image sequences in speech balloons, which means that it is possible to read both the visually and textually represented aspects of the interaction without switching back and forth between a transcription and various images.

This way of representing social interaction has been inspired by the work of Oskar Lindwall and Jonas Ivarsson (e.g. Lindwall, 2008; Ivarsson, 2010), at the Department of Education at the University of Gothenburg, and it is effectively an appropriation of the medium of comic strips into a scientific context.<sup>15</sup> The visual "language" of comics has been around for more than 100 years by now, and it has been developed precisely to be an easily readable way of representing complex multimodal sequences of action. It is telling that comics have sometimes been called "sequential art" (Eisner, 1985), as also pointed out by Lindwall (2008, p. 68). In addition, most people are already familiar with this "language", in the form of certain conventionalized ways of representing paths of motion and motion dynamics as well as temporal sequences of events that may be hard to capture in transcription

---

<sup>15</sup>Lindwall (2008) also provides a methodological discussion of the use of these kinds of representations of multimodal interactions for the purpose of presenting analyses.



symbols, and this familiarity with the conventions itself contributes to the readability.

On the other hand, certain other characteristics of comics may need to be left out, such as the exaggerated character of the drawings which is typical for many (but not all) forms of comics. Also, some things may need to be added in order to capture the rigor and detail required in the scientific context. For the purposes of the present thesis, the main addition required is to underline the part of speech that goes together with the stroke of gestures and gesture-like movements.<sup>16</sup> The way that comics are used here, most examples involve only a single frame and the language shown in the speech balloons is English, so that they will be readable for researchers who do not know Swedish (when the original Swedish utterances are provided, they are given as part of the text). In almost all cases, the utterances have been translated on a word for word basis, preserving the word order of the original utterances. This is judged more important than creating “naturalized” English versions of the utterances, since it provides a better image of the temporal coordination between gesture and speech in the original utterances. Parent’s utterances are translated less literally in some cases, to provide a more natural English sentence, because the parent’s gestures are for the most part not analyzed in this thesis. In some cases, it is necessary to translate one Swedish word into two English words, such as when translating a noun in definite form, which is marked by a suffix in Swedish (*-en/-et* in the case of singular). For example, *bok-en* is translated into “the book” and if there is a gesture stroke performed in coordination with this word, both “the” and “book” will be underlined in the presentation of the example.<sup>17</sup>

One interesting aspect of this way of representing gestures is that when the drawings are produced, the researcher is effectively forced to “reproduce” the gestures, which tends to bring all sorts of aspects of the performances to attention — aspects that may otherwise often go unnoticed, even if the video data is watched a repeated number of times. As pointed out by Lindwall (2008, p. 68):

---

<sup>16</sup>The concept of a “stroke” was explained in Chapter 3.

<sup>17</sup>A brief presentation of the characteristics of Swedish is given in Josefsson & Platzack (2003, p. 1):

Swedish is a Scandinavian language, most closely related to Norwegian and Danish. In short it could be characterized as having a fairly restricted word order, a relatively simple verbal inflection system, and a more complex nominal inflection system. Compounding is a highly productive word formation process. Of the most prominent syntactic features we might mention the verb second quality — not more than one constituent may precede the tensed verb —, obligatory subjects, and the absence of subject-verb agreement. Like English, Swedish is a VO language.

Like transcripts, the comics are analytic renderings filled with decisions on what is important. The use of comics also necessitates additional considerations with regard to the organization of time and space: What should be put in a panel? How do the actions fit on a page? How is time represented by means of space? What must be stated explicitly in textual comments and what can be shown with pictures?

However, in order for the comics to have some sort of validity, it is of course important that they are directly based on frames that are taken from the video files. The steps that were involved in producing these images goes as follows:

1. Grab a particular frame from a video file.
2. Open it in Adobe Illustrator and create an empty layer on top of it.
3. Draw the desired outlines in this empty layer, on top of the shapes shown in the frame-grab. Take the framegrab away, and only save/export the layer which contains the drawing. This results in vector graphics images of fine resolution.
4. Add speech balloons and text. In some cases this was made in Illustrator too. In other cases this step was performed by means of a program called Comic Life, which allows the user to arrange image files into comic strips, including the handling of speech balloons.

Other positive by-products of these representations, or of the use of drawings instead of frame-grabs more generally, is that the persons involved become anonymized. Drawings may enhance the visibility substantially, compared to frame-grabs from video files that may often turn out relatively blurry in print. Of course, in addition to these comics, there is typically a verbal description of the event, just like transcripts+illustrations are typically supplemented with a written account of the analytical points that are to be brought home from a particular example.

In sum, presentation of interaction is not just a matter of aesthetics. It is also part of the establishment of an intersubjective understanding between researchers. That is to say, it is part of the objectivity of the research itself, seen in a bigger perspective. Chapter 5, which is more quantitatively oriented, does not contain visualized examples, but all of the other analytic chapters do.

#### 4.5.2 Notation for gesture+speech utterances in text

In cases where gesture+speech utterances are presented as part of the written text, the notation may look like this: “what is that thing+POINT?”. Text which is underlined

shows the part of the spoken utterance which goes together with the gesture, and the gesture itself is specified in SMALL CAPS. Throughout the thesis, the spoken utterances has been translated from Swedish to English, to the benefit of readers who do not know Swedish. Still, in most places the original Swedish utterances are provided too.

### 4.5.3 Graphs: Averaging of data from several children

Many of the graphs that involve all the five children analyzed in this thesis are produced by means of a special method which was invented during the work with this thesis (called *true age average* here). Most studies produce curves of average values by putting all observations within a particular age range into the same “box”, i.e., *quantization* of age, so that all observations within a certain age range are treated as if they were collected at the same point in time although they are in fact not. The result of this common procedure is that information about the more exact age at which the data for a particular session was obtained is thrown away. This can sometimes be justified in order for the researcher to be able to apply certain statistical formulas to the data that require quantized age groups, but in cases where such restrictions do not apply, age quantization is simply an unnecessary distortion of data that is in fact available. One example of a context where age quantization is not warranted is when the goal is simply to visualize the averaged data “as is”.

The method used here to produce average curves does not quantize age values, and that makes it possible to include sessions obtained at different times as well as different numbers of sessions for different children. The author is not aware of other developmental studies using this type of method of averaging, although it would not be too surprising if it has been used elsewhere before due to its relatively intuitive nature. In the end, all it does it to average the graph curves of the individual children as they are, without doing the age quantization step. However, since this procedure seems to be non-standard, a more detailed explanation of how it was calculated may be warranted.

Figure 4.1 shows a hypothetical example of how two curves from two different children would be averaged together using this method, based on the likewise hypothetical data shown in Table 4.3 and Table 4.4. As mentioned above, this method also makes it possible to include children with different number of sessions in the same graph — there are 3 sessions for Child 1 and 4 sessions for Child 2 in the example. The steps involved in the production of these graphs goes as follows:

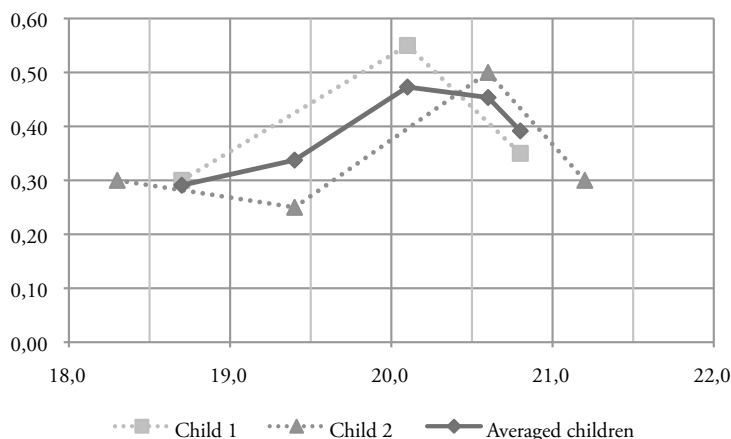


Figure 4.1: Hypothetical example showing how the curves of individual children are averaged.

- Step 1. Calculate age as a decimal number, rather than as month;day. If a session was recorded on day 18 of month 20 in a child's life (i.e., 20;18), the day value instead be converted to a fraction of a month, calculated as:  $(\frac{\text{day}}{\text{days in a month}})$ . Hence, the age of this session would be:  $20 + \frac{18}{30} = 20.6$
- Step 2. Calculate an expanded table, with linear interpolations between the actually observed values (corresponding to the values covered by the interconnecting dotted lines between the marks in Figure 4.1), as shown in column 2 and 3 of Table 4.6. A linear interpolation is calculated by finding out the constant value that need to be added (or subtracted) at each intermediate point between actually observed values in order to end up precisely at the next observed value. The actually observed values are marked with bold font face in this table, and those values are the same ones as are shown in Table 4.3 and Table 4.4.
- Step 3. Calculate average values for all the children at those rows in the table where there is (A) interpolated data or actually observed data available for *all* the children (i.e., excluding the range between 18.3–18.6 and 29.9–21.2) and where there is (B) at least one actually observed value (bold face) is available. The results of this calculation is shown in italics in column 4 of Table 4.6, and it is also these values that are summarized in Table 4.5 and visualized in Figure 4.1. Note that the averaged line in the figure lies in between (y-wise) of the two dotted lines of the two children for each x-position. Also note that the correct kind of graph

#### 4.5. PRESENTATION OF EXAMPLES AND DATA

		Averages			
		Age (months)	Child 1	Child 2	Averages
Child 1		18,3		<b>0.3000</b>	
Age (months)	Measured value	18,4		0.2955	
18.7	0.30	18,5		0.2909	
20.1	0.55	18,6		0.2864	
20.8	0.35	<b>18,7</b>	<b>0.3000</b>	0.2818	<i>0.2909</i>
Table 4.3: Data for Child 1.		18,8	0.3179	0.2773	
		18,9	0.3357	0.2727	
		19,0	0.3536	0.2682	
		19,1	0.3714	0.2636	
		19,2	0.3893	0.2591	
		19,3	0.4071	0.2545	
		19,4	0.4250	<b>0.2500</b>	<i>0.3375</i>
Child 2		19,5	0.4429	0.2708	
Age (months)	Measured value	19,6	0.4607	0.2917	
18.3	0.30	19,7	0.4786	0.3125	
19.4	0.25	19,8	0.4964	0.3333	
20.6	0.50	19,9	0.5143	0.3542	
21.2	0.30	20,0	0.5321	0.3750	
Table 4.4: Data for Child 2.		<b>20,1</b>	<b>0.5500</b>	0.3958	<i>0.4729</i>
		20,2	0.5214	0.4167	
		20,3	0.4929	0.4375	
		20,4	0.4643	0.4583	
		20,5	0.4357	0.4792	
		20,6	0.4071	<b>0.5000</b>	<i>0.4536</i>
		20,7	0.3786	0.4667	
Child 1 & 2 averaged		<b>20,8</b>	<b>0.3500</b>	0.4333	<i>0.3917</i>
Age (months)	Measured value	20,9		0.4000	
18.7	<i>0.2909</i>	21,0		0.3667	
19.4	<i>0.3375</i>	21,1		0.3333	
20.1	<i>0.4729</i>	<b>21,2</b>		<b>0.3000</b>	
20.6	<i>0.4536</i>				
20.8	<i>0.3917</i>				
Table 4.5: Final averaged data.					

Table 4.6: Calculation of intermediate values

for this purpose in Excel is “X Y Scatter”.

To summarize, the benefits of this kind of “true age average” graph is that it is possible to use any existing data for a given child, even though there may be different amounts of sessions available for different children. Furthermore, there is no distortion of age. The only drawback, to my knowledge, is that this kind of graph is not compatible with certain procedures and formulas for calculating various statistical

measures, and also, the calculation of this graph is in itself slightly more complicated than averages where ages are quantized into age groups. Yet, the procedure can be automatized with a script and this is how it was done in the case of this thesis. By means of a self-written Perl script, all the CHAT transcription files could be parsed, and the output of the script was a table of values like the one shown in Table 4.5. This table could then be cut/pasted directly into software like Excel or SPSS for the creation of a graph.

Someone may object that the calculation of interpolated values (see Step 2 above) between the actually observed values are simply assumed by the researcher and that they lack empirical support.<sup>18</sup> To explicate, the assumption involved is the assumption that there is continuity in the developmental “curves” of children, while it is in fact at least theoretically possible that the intermediate values between actually observed values would have been quite different if they would have been based on actual observations. That is true, but then it should be remembered that the objection applies equally much to graphs with quantized ages which also connect the dots in the graph with lines, since this line corresponds precisely to these “assumed” continuous values in between. Hence, a graph showing the true age average makes no more assumptions about these intermediate values than graphs showing data that is quantized for age. In fact, the interconnecting lines are more likely to be valid in the case of true age average, because the distortion of age involved in age quantization will affect the slopes of these interconnecting lines slightly in a way that has nothing to do with empirical data. Therefore, true age average is to be recommended when dealing with the kind of data of data that is analyzed in this thesis.

---

<sup>18</sup>In fact, two persons have raised this objection, so it may be worthwhile to explain.

**Part II**  
**Analysis**





## CHAPTER 5

# Developmental patterns and transitions

### 5.1 Questions asked and contents of the chapter

Whereas the other analytic chapters primarily have a qualitative focus, this chapter provides a quantitative overview of some general characteristics of the data. The overarching question asked here is: *What are the broad changes in the children's use of gesture throughout the period between 18 and 30 months?* Part of this question is also the question of: *How might these changes be understood?* More specific questions include:

- How frequently do the children perform gestures in general, and gestures of different kinds, at different ages?
- Do changes in the use of speech co-occur with changes in the use of gesture?
- Are the children's gestures typically performed in coordination with speech, or are they used autonomously? Are there differences in this regard between different types of gestures?
- To what extent do children combine several "units" into multi-gesture and multi-word utterances?
- Do the gestures typically involve handling of objects or are they typically performed empty-handed?

The unit of analysis in this chapter is "gesture", as defined in Chapter 2 — actions that either qualify as *explicitly communicative* (Comm#3) or as *semiotic signs* (Sem#3),

or both of these at once. Since all the gestures studied here take place within the context of a social encounter, almost all of them are communicative at least in the basic sense of being actions that are framed by mutual attunement (Comm#2).<sup>1</sup>

The quantifications are performed directly on the basis of the annotations that were described in Section 4.3. Most of the analyses that are presented in this chapter are focused on the five children as a group. Treating the children as a group makes the relation between age and various frequency measures clearer than when each child is considered in isolation, and influences of particular socio-cultural activities and individual differences become less salient. This is not to say that the situated character of gestures or individual differences between children are not important or interesting — it is just a matter of the focus of the present chapter, which is changes in the frequencies of various measures as *a function of age*. In cases where one or two children stand out from the rest, comments are provided about this.

There are a handful of other quantitatively oriented studies that cover at least some part of the period between 18 and 30 months. The analysis in this chapter confirms some of the findings from these studies (references are given throughout the chapter, as they are discussed) and it also adds some new ones as well. One benefit of the analysis presented here, even in relation to findings that have already been suggested in other studies, is that the present analysis has a more fine-grained temporal resolution than most of the other studies that have been carried out on children's gesture for these ages. At the same time, since the data used here is based on five children, rather than dozens, firm generalizations cannot be made from this data only. *In cases where explanations are proposed for the patterns that are visible in the data, this should be considered as suggestions for possible interpretations rather than proven facts, and as starting points for further analysis rather than the last word.* Yet, in those cases where previous research exists, it seems that the findings presented here confirm the findings of other research well in most cases, indicating that the sample studied here is representative. All in all, 10 hours and 20 minutes of data has been analyzed, amounting to 4406 gesture performances.

The main aim of the chapter is to provide a broad overview of a relatively large number of different phenomena, rather than going into great depth into specific questions. The chapter is divided into sections that concern different questions. Each section and sub-section ends with a brief summary of the main findings and claims of this (sub-)section. An integrated discussion of all the questions that are

---

<sup>1</sup>The definition of the term “gesture”, as it is used in this thesis, is provided in Section 2.2.5. The underlying distinctions that are used in this definition, i.e., various levels of communicative explicitness and various levels of semiotic complexity, are explicated throughout Section 2.2.

addressed is provided in the last section of the chapter (Section 5.8). In the next section, some terminology is introduced which is used throughout the whole chapter. It is therefore advisable to read the next section. Apart from that, readers who are only interested in some specific aspect of the children's gestures may proceed directly to the sections that best reflect their interests.

## 5.2 Transition periods and terminology

In the analysis, reference will be made to what I call *Transition Period 1, 2, and 3*, or in short, *TP#1, TP#2, and TP#3*. Even though these periods are specified in rather exact ways (see below), this is not meant to imply that all children follow the exact same developmental "schedule" to the day and minute. There are indeed quite huge differences between individual children (cf. Bates et al., 1988). Furthermore, the exactness of the specification is not meant to give the impression of exact boundaries between the transition periods. *The exactness of the numbers simply reflect what is found in the data studied here*, namely periods with especially intense developmental changes, where several things happen more or less at the same time. To what extent these "transition periods" would also show up in studies with more subjects or in comparisons with children in other cultures is a question for future research. The same thing goes for the terms themselves. The choice of the term "transition period" seems to be well motivated as a description of what goes on in the data studied here, but if other studies would find kinds of developmental patterns, then it would of course be inappropriate to insist on the use of these specific terms. Hence, both the terms and the periods that they refer to are first and foremost descriptive. Theoretical considerations on terms such as "stages" and "transitions" in development are provided in the concluding section.

To repeat then: In the analysis, reference will be made to three so-called transition periods, namely TP#1, TP#2, and TP#3. Each of the three transition periods

	start-(core)-end
<i>TP#1</i>	19.8-(20.8)-21.5 months
<i>TP#2</i>	23.5-(24.3)-25.3 months
<i>TP#3</i>	27.3-(28.8)-29.6 months

Table 5.1: Three main transition periods (start-core-end).

are divided into two sub-parts. The boundary between these two sub-parts will be called the *core* of the transition period. Table 5.1 summarizes the onset and end of the three periods, and the value within the parenthesis specifies the core. The periods and their sub parts are also shown in the graphs by dotted vertical lines, so that each transition period is shown as a group of three vertical lines, corresponding to the onset, core, and end of the transition periods, respectively.

## 5.3 GPM, GPU, and MLU

*How frequently do the children perform gestures, of all kinds, at different ages?* This question is addressed by means of measures such as *gestures per minute* (GPM), *gestures per multimodal utterance* (GPU), and *mean length of utterance* (MLU).<sup>2</sup>

### 5.3.1 GPM, GPU, and MLU for each child

Even though this chapter does not focus on individual variation, a brief comparison of the characteristics of the children may be useful as a start, in order to give the reader an idea of the homogeneity and heterogeneity of the data that has been analyzed.

Table 5.2 shows the average, minimum, and maximum GPM, GPU, and MLU obtained for each child throughout the period. The average GPU for the whole data set is 0.51. Bella and Tea produce almost twice as many gestures per minute compared to Alice and Hanna, and Harry (the only male subject) is in between. When it comes to gestures per utterance, the children are much more consistent in comparison to each other, with a span between 0.46 and 0.58. This indicates that

<sup>2</sup>Explanations of these measures were provided in Section 4.4.

	GPM	GPU	MLU
Alice	5.3 (2.4–11.9)	0.46 (0.25–0.76)	2.1 (1.2–2.9)
Bella	9.0 (4.8–14.3)	0.58 (0.36–0.76)	1.8 (1.3–2.5)
Hanna	5.3 (3.3–8.6)	0.46 (0.26–0.63)	2.9 (1.8–4.1)
Harry	6.0 (2.1–10.9)	0.47 (0.26–0.73)	1.5 (1.1–2.5)
Tea	9.2 (6.0–12.1)	0.56 (0.40–0.84)	1.6 (1.1–2.7)
Average child:	7.0 (3.7–11.6)	0.51 (0.31–0.74)	2.0 (1.3–3.0)

Table 5.2: GPM, GPU, and MLU for each child: Average (min–max)

the GPU measure is less sensitive to individual variation than the GPM measure.<sup>3</sup> To the data in the table it might be added that all five children perform less gestures (GPM and GPU) in the latter half of the period studied here than in the first half.

The minimum MLU in any sessions for each child is larger than one. This means that all children produced at least some two-word utterances in all of their sessions, including the earliest ones. Also, Alice and Hanna have the highest MLU of the five children, and also the lowest GPU and GPM. This implies that the ability to produce longer spoken utterance does not necessarily imply more frequent use of gestures, neither in terms of gestures per minute nor gestures per utterance. Caution is warranted here though. GPM and GPU do not directly measure the tendency for gesture to occur in longer or shorter utterances. It is still possible that the tendency to perform gestures with multi-word utterances increases over time, but that the tendency to perform gestures with one-word utterances decreases more, so that the overall average is nevertheless a decrease in GPU although there is in fact possibly a tendency for gestures to be associated with longer utterances. This issue is further probed in Section 5.5.

What rates have been found in other studies of children's gestures? The GPU of 0.51 that was found here is very close to the GPU of 0.48 found in Goldin-Meadow & Morford (1990, p. 253) (three children: 10–30, 14–29, and 17–23 months old).<sup>4</sup> Other studies report other ratios, using slightly different measures. According to McNeill (1992, p. 299), children between 24–48 months produce relatively few gestures compared to older children and adults. More precisely, the estimates reported by McNeill state that 28% of the spoken clauses (i.e., not all kinds of multimodal utterances) are performed in coordination with (one or more) gesture in 24–48 months old children. This is considerably lower than the rates found in the present study, especially given that the divisor he uses (gestures per *spoken utterance*) is bound to produce slightly higher values than the divisor used here (gestures per *multimodal utterance*), since the latter divisor also includes some gesture only utterances. However, there is only a partial correspondence between the age span of 18–30 months that is studied here and the age span of 24–48 months that were investigated by McNeill. The results are therefore not directly comparable, al-

---

<sup>3</sup>This can be considered both a strength and a weakness of the GPU measure, depending on the analytic purposes to which it is applied. More specifically, it implies that the choice between GPM and GPU ought to be an informed choice in studies of children's gestures. At present, most studies on children's gestures do not provide explicit reasons for choosing GPU or GPM (or some slight variation of these).

<sup>4</sup>Goldin-Meadow and Morford do not provide this overall GPU value explicitly, but it can be computed from the data that they present.

though one may suspect that there are more factors involved behind the differences in gesture rates found here than only a mismatch in the age span included, such as differences in the activities that were studied — McNeill primarily studies gesture in the context of retelling episodes from animated cartoons, which may be a hard task for young children<sup>5</sup> — and methodological differences — McNeill does not include gestures that involve manipulation of physical objects, which are frequent in young children in contexts where objects are available (which may not be the case in the context of retelling episodes from animated cartoons).

### 5.3.2 GPM, GPU, and MLU as a function of age

The graph that shows *gestures per minute* (GPM, Figure 5.1) and the graph that shows *gestures per multimodal utterance* (GPU, Figure 5.2) are both similar and different — each of them showing a related, but different, kind of measure of the frequency of children’s gesture performances at different ages. The overall shapes of the two curves are somewhat similar, but between 18 and 21 months they look strikingly different. That is, from the start of the observed period and up to the core of TP#2, the GPU curve is more or less flat, hovering between 0.54 and 0.60, whereas the GPM curve shows a marked increase during the first part of TP#1 (19.8–20.8 months). This means that the children gesture distinctively more often per minute after the core of TP#1 than what they did before, but the average number of gestures in each multimodal utterance still remains fairly stable between the start of the period and up to the core of TP#2. The reason for the discrepancy between the two curves is, by implication, that there must be a corresponding increase in the number of *spoken* utterances in the first part of TP#1 (i.e., between 20 and 21 months). Hence, there is (roughly) an equal amount of gestures performed whenever an utterance is actually “delivered” (GPU) between the start of the observed period and the core of TP#2, even though there is a sharp increase in the number of overall “communications” per minute (GPM) in the first part of TP#1.

These findings are in line with Özçalışkan & Goldin-Meadow (2009, p. 195), who found a significant increase in both gestures and speech between 18 and 22 months. Similarly, Iverson et al. (1994, p. 35) found a significant increase between 16 and 20 months in the overall number of gestural or verbal communicative acts.

Starting from the core of TP#2 (at 24.5 months), there is a general decrease in

---

<sup>5</sup>McNeill (1992, p. 302) notes that “The youngest child we have observed narrating the cartoon story was two and a half years old.”

### 5.3. GPM, GPU, AND MLU

---

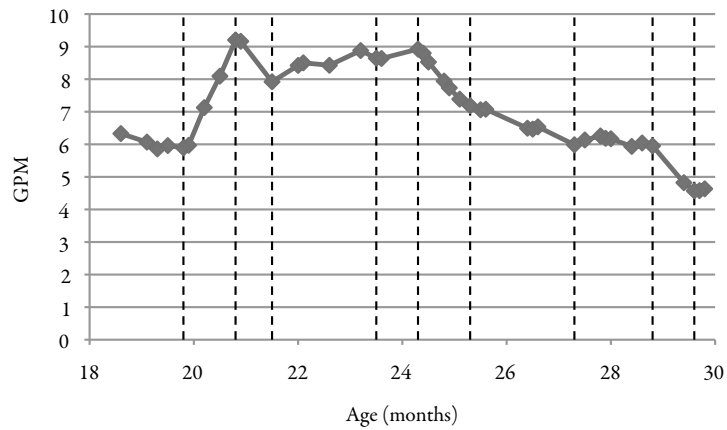


Figure 5.1: Gestures Per Minute (GPM) for all five children.

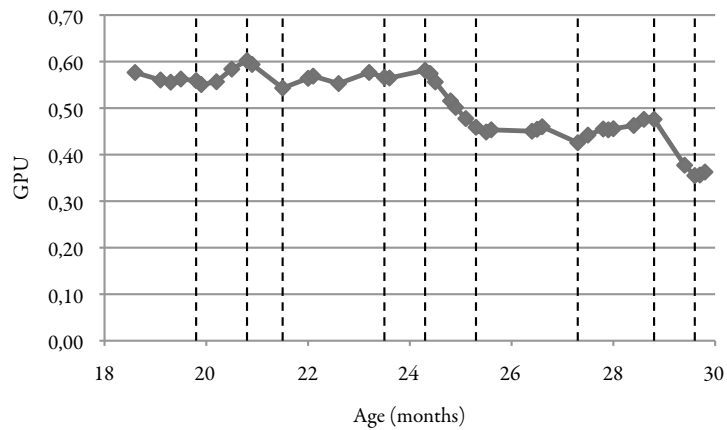


Figure 5.2: Gestures Per Utterance (GPU) for all five children.

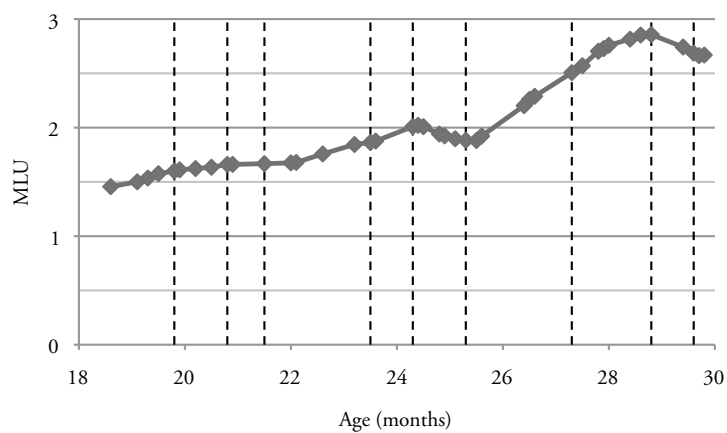


Figure 5.3: Mean length of utterance (MLU) for all five children.

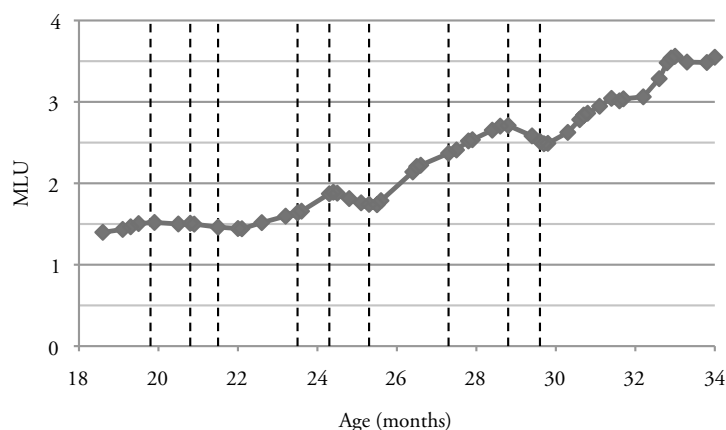


Figure 5.4: MLU between 18 and 34 months for four of the children (excluding Alice)

both GPM and GPU. This decay lasts until the end of the studied period. The decrease is approximately 50% for GPM (from 9 to 4.5) and 40% for GPU (from 0.58 to 0.35). The decline is especially marked in the second part of TP#2 and in the second part of TP#3 and less marked in the period between these two transition periods.

These two more marked decreases appear quite precisely when mean length of utterance (MLU) is just about to reach 2 and 3 respectively (see Figure 5.3). In fact, the otherwise steady increase in the MLU curve temporarily flattens out, and even *decreases* somewhat, precisely during those two more marked decreases in GPM and GPU. After each of the two decreases, MLU turns upward again. The upward turn in MLU that comes after TP#3 cannot be seen in Figure 5.3, since this figure stops at 30 months. However, for four of the five children (Bella, Hanna, Harry, Tess) there are transcriptions of speech available beyond 30 months, and therefore MLU could be calculated up to 34 months for these four children. This extended MLU graph is shown in Figure 5.4 and in this graph it can be seen that the MLU curve does indeed turn upward again after 30 months, after the temporary decline at the second part of TP#3.<sup>6</sup> The simultaneous changes in the MLU curve and in the GPU/GPM curves at the second part of both TP#2 and TP#3 implies that changes in the use of gesture may somehow relate to changes in the organization of speech, and vice versa. That is an issue that will be further discussed on several occasions in this chapter.

Another observation that can be made when considering the MLU curve to-

<sup>6</sup>It should be remembered that the original MLU measure counts the number of morphemes in an utterance, and that the MLU measures presented here are rather on the word/vocalization level, since there was no coding for morpheme boundaries within words.



gether with the GPU curve is that even though the children's spoken utterances become longer and longer during this period, utterances do not, *on average*, tend to contain more gestures. As noted in the previous sub-section, a measure "on average" is not conclusive evidence that gestures may not be more common in longer utterances, since this may only appear to be the case if one-word utterances are more common than multi-word utterances and the tendency to produce gestures with one-word utterances decreases more than the tendency to produce gestures with multi-word utterances increases. As stated before, this question is further probed in Section 5.5.

#### 5.3.3 Summary

Throughout the data corpus, there is an average *gestures per multimodal utterance* (GPU) of 0.51, and an average *gestures per minute* (GPM) of 7.0. The children who produce the longest spoken utterances overall are also those that produce the fewest gestures (both in terms of GPM and GPU), and vice versa. The GPU measure seems to be a more stable measure overall than GPM, which seems more sensitive to variation such as individual variation. All children produced at least some two-word utterances in each of the sessions, including the earliest ones.

At the first part of TP#1, there is a sharp increase in both gestures and spoken utterances per minute, but the GPU rate lies fairly stable from the start of the observed period, up to the core of TP#2. Apart from showing an aspect of the children's development, this also shows that the GPM and GPU measures do not necessarily show the same thing, and that it ought to be an informed choice whenever researchers chose to apply one or the other of these two measures.

However, after TP#2, the GPU and GPM measures show a more similar overall pattern. The frequency of gestures decreases from the core of TP#2 and onwards, with two more marked periods of decay at the second parts of TP#2 and TP#3. These two more marked decays co-occur precisely with two temporary decreases in MLU, and also with the fact that MLU becomes close to 2 at TP#2 and close to 3 at TP#3.

On average, children do not tend to produce more and more gestures per utterance as they grow older, even though their spoken utterances become longer and longer. In this respect there is an asymmetric relationship between gesture and speech.

## 5.4 Gesture with and without speech

*Are the children's gestures typically performed in coordination with speech, or are they performed autonomously?* This is an interesting question in itself. Indeed, understanding the nature of gesture and speech coordination is a central concern in the field of gesture studies (Kendon, 1980b, 2004; McNeill, 1985, 2005).

The question also relates to a debate in research on children's gestures whether there is coordination between vocal activity and bodily movement from early on (from early infancy or more or less from birth) or whether these two modalities "come together" later in development. These two positions will be referred to as the *early integration* scenario and the *late integration* scenario.

Butcher & Goldin-Meadow (2000) are proponents of the late integration scenario, and they report an increase in gesture and speech coordination some time between 16 and 23 months in American children. Prior to this increase they found that gestures were rarely coordinated with meaningful words. On basis of these findings, they suggest the idea of a "convergence point" in children's development some time between 16 and 23 months where "hand and mouth come together" (*ibid.*, p. 235), i.e., a late integration scenario. There are also other proponents of this view (e.g. Murphy, 1978; Lock, 1980; Masur, 1983; Rowe et al., 2008; Zinober & Martlew, 1985b). Other researchers hold the alternative view, that the gestural and vocal modalities are coordinated more or less already when gesture first emerges around 8–12 months of age, or even earlier (Fogel & Hannan, 1985; Capirci et al., 1996; Iverson & Thelen, 1999; Blake, 2000; Masataka, 2003; Iverson & Fagan, 2004; Pizzuto et al., 2005). Since the age range studied in this thesis is 18–20 months, it is not possible to attempt to support any of these positions in a strong sense, but since there is at least some overlap with the period with the period where gesture and speech may "converge" according to the late integration scenario, it is still interesting to consider some of the findings in this section in relation to the early/late integration debate.

### 5.4.1 Gesture only, gesture+speech, and speech only utterances

Figure 5.5 and Figure 5.6 show the distribution of unimodal gesture utterances, unimodal speech utterances, and bimodal gesture+speech utterances, measured "per minute" (GPM) and "per multimodal utterance" (GPU) respectively.

These graphs show that the initial rise in GPM gesture frequency at TP#1 dis-

## 5.4. GESTURE WITH AND WITHOUT SPEECH

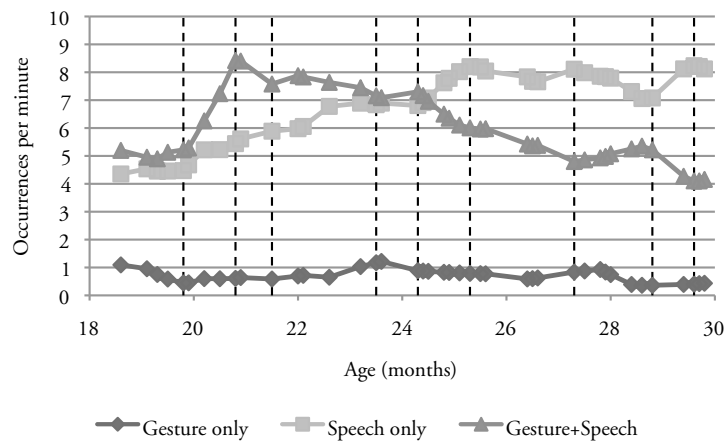


Figure 5.5: Utterances in different modalities (per minute).



Figure 5.6: Utterances in different modalities (per utterance).

cussed in the previous section is mainly a matter of *increasing use of gestures that are coordinated with speech* (per minute). This suggests that gesture and speech may “propel” each other in a synergistic fashion so that each modality creates a relevant context for the other. It is also reminiscent of the late integration scenario, because the increase of gesture+speech utterances per minute may be interpreted as a “convergence point” between gesture and speech around this age. However, the GPU measure for gesture+speech utterances is, once again, much more flat in comparison to the GPM measure, and for this measure there are no hints about a major transition going on, although there is a little peak in gesture+speech utterances at the core of TP#1. Hence, the GPU measure seems more compatible with the early integration scenario in the sense that there is no evidence here for a lack of semantic integration between gesture and speech at the start of the studied period. In fact,

both the GPM graph (Figure 5.5) and the GPU graph (Figure 5.6) show that gesture+speech utterances dominate, although with a relatively small margin, from the very beginning of the period studied here up to the core of TP#2.

After that, speech only utterances dominate to the end of the studied period, and there are small “bursts” in speech-only utterances precisely at those two points in time (the second part of TP#2 and TP#3) when MLU makes a temporary decrease, which was also the two points in time when MLU reaches 2 and 3 respectively. The rate of gesture+speech utterances also decreases more rapidly at precisely these two points in time. It is remarkable that TP#2 and TP#3 show such a similar structure across many different measures, even though this structure is quite complex. This indicates that a similar kind of process is taking place at both of these points in time.

### 5.4.2 Summary

The initial rise in GPM gesture frequency in the first part of TP#1 that has been discussed in previous sections was here shown to be mainly a matter of increasing use of gestures (per minute) that are coordinated with speech, and there was no corresponding increase in gesture-only utterances. Gesture-only utterances are rare throughout the whole period. Up to the core of TP#2, there is a dominance, although with a small margin, of gesture+speech utterances over speech-only utterances. After TP#2, there is a shift in dominance of speech-only utterances over gesture+speech utterances which becomes more and more amplified as the children grow older.

It was noted that the GPU measure produced a much more flat curve than the GPM measure, and that the GPU measure gives less of an impression of a major increase in gesture+speech utterances at the core of TP#1 than GPM does. Therefore, the GPM measure is more in line with late integration scenarios, while the GPU measure is possibly more in line with early integration scenarios. This is discussed in more detail in the final discussion of this chapter.

## 5.5 Multi-gesture and multi-word utterances

*To what extent do children of different ages perform utterances that involve more than one gesture and/or more than one word? More particular questions include: To what extent are multi-gesture utterances performed with and without speech? To what extent are gestures performed in one-word utterances or multi-word utterances? These ques-*

tions all share the general theme of probing the use and emergence of *combinations* in the gesture and in speech, as well as how gesture and speech relate to each other in this respect. Do gesture and speech “follow” each other, so that it is more likely to find multi-gesture combinations in multi-word utterances (a symmetric relation between the way the modalities are used together)? Or is it the other way around, so that more complex speech stands in opposition to more complex gestural combinations, and vice versa (a asymmetric relation between the way the modalities are used together)?

Regarding asymmetries, several researchers have found that it is some time during the latter part of the second year of life that speech starts to dominate over gesture in a range of different respects in hearing children without signed language input. For example, it is around the end of the second year that speech is eventually more frequently used than gesture to refer to objects (Iverson et al., 1994; Capirci et al., 2005). Also, the GPM and GPU measures in the previous section indicated that there may be no overall tendency for more gestures to occur in longer utterances, but it was also noted that the findings in that section were not conclusive in that respect. This section will probe this issue further.

When reading this section, it should be kept in mind that no distinction was made in the speech transcriptions between words and (non-word) vocalizations. That means that “multi-word utterances” are rather something like “multi-unit speech” (see Section 4.2 for more details). In lack of a better term these spoken utterances will nevertheless be referred to as “multi-word utterances”.<sup>7</sup>

### 5.5.1 Multi-word combinations

Figure 5.7 shows the number of one-word and multi-word utterances at different ages. Here it can be seen that the increase in spoken utterances per minute in the first part of TP#1, which was implicated by the findings in Section 5.3, mainly consists of an increase in one-word utterances.<sup>8</sup> Overall, one-word utterances are more common than all other utterance lengths together (i.e., multi-word utterances) from the start of the period, up to TP#3 where multi-word utterances eventually become more common. Looking at the curve for multi-word utterances, there is a small and

---

<sup>7</sup>Variants such as “multi-speech” or “multi-vocalization” seem ambiguous or unintuitive. A more generic term “multi-unit” does not seem satisfactory either, since it is too general and does not tell the reader that it concerns speech rather than gesture.

<sup>8</sup>There is a little increase in multi-word utterances during this period too, although of smaller magnitude.

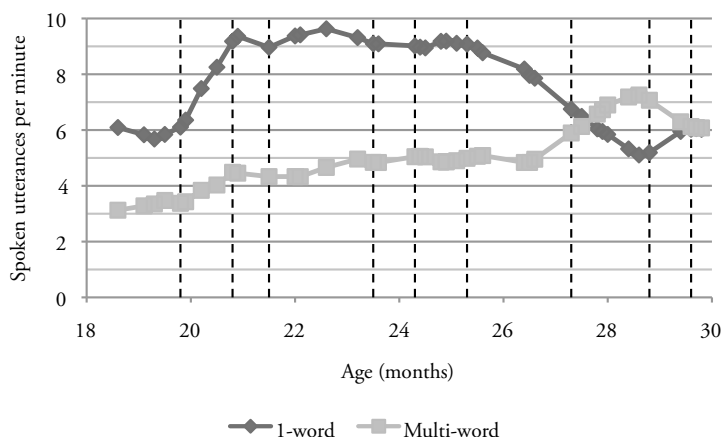


Figure 5.7: One-word and multi-word utterances (per minute).

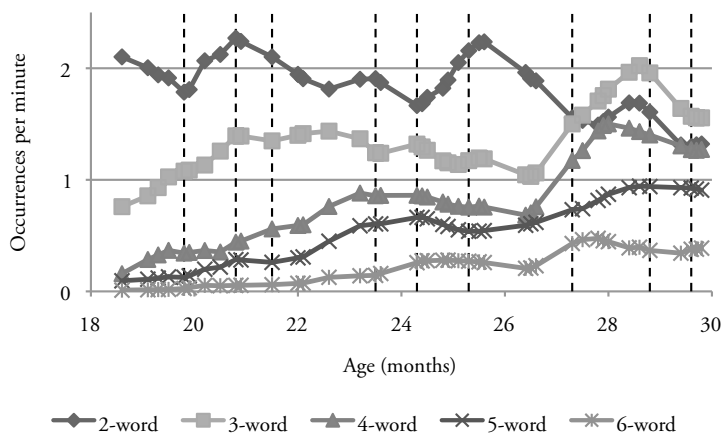


Figure 5.8: Multi-word utterances of different length (per minute).

linear increase from the beginning of the observed period up to 26.4 months, where the number of multi-word utterances suddenly increases more steeply until the core of TP#3. That is, this sudden increase in multi-word utterances is not aligned with the transitional periods, but rather happen in between TP#2 and TP#3. Then, in the last part of TP#3, there is a small (temporary) decline in the number of multi-word utterances again.

Figure 5.8 shows in more detail the amount of the multi-word utterances that are two-word, three-word, and so forth up to six-word utterances. In this graph it can be seen that there is an overall decrease of two-word utterances after a peak at the end of TP#2. Soon after this, at 26.4 months, three- to six-word utterances (especially three- and four-word utterances) show a sudden increase and that the increase ends with a peak at the core of TP#3 (reflecting the overall pattern shown in Figure

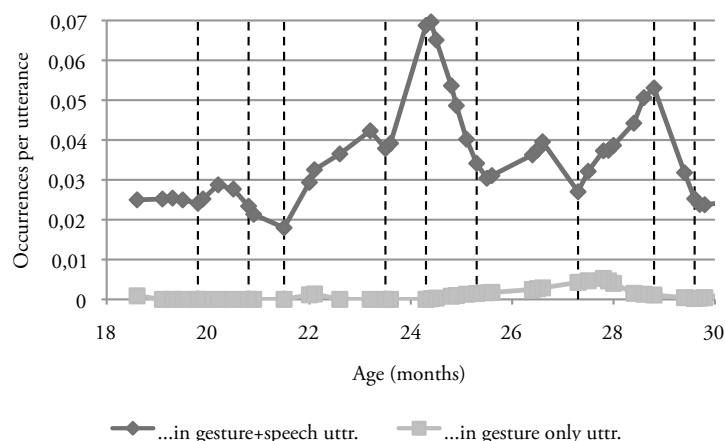


Figure 5.9: Multi-gesture utterances (per utterance).

5.7). It seems that around 26.4 months, the grammatical complexity of the spoken utterances suddenly start to increase. Indeed, at TP#3, two-word utterances are not the single most common utterance length in multi-word utterances anymore. Three- and four-word utterances are more common at that time. This suggests that there are some relatively fundamental changes in the grammatical organization of speech around TP#3.

One-word utterances are still the single most common length of utterances though, as shown in Figure 5.7. That is not surprising given the number of conversational moves that can be handled with a single word, especially in coordination with gesture. It should also be remembered that there are many conversational moves that may consist in a single word also by adult standards, such as a response morpheme like 'yes' or 'no'. That is, far from all one-word utterances are a symptom of a lack of more complex linguistic skills.

### 5.5.2 Multi-gesture combinations

Figure 5.9 shows the number of multi-gesture utterances that are performed in coordination with speech versus the number of multi-gesture utterances that are performed autonomously, without speech. The term “multi-gesture” includes both *simultaneous combinations* where two gesture strokes performed at the same time, such as pointing to something and simultaneously shaking the head, and *sequential combinations* where two gestures appear in the same utterance, without strokes that overlap in time, such as pointing to two targets that are related to each other within

Multi-gesture utterances			
Child	All	With speech	Without speech
<i>Alice</i>	32	32	0
<i>Bella</i>	84	74	10
<i>Hanna</i>	53	53	0
<i>Harry</i>	51	50	1
<i>Tea</i>	85	85	0
<i>Sum:</i>	305	294	11

Table 5.3: The absolute number of multi-gesture utterances in each child.

the same utterance. The first thing to note is that almost all multi-gesture utterances (96.4%) are performed in coordination with speech. Only 3.6% of the multi-gesture utterances are performed autonomously, without speech, and all but one of these instances come from one child (Bella, see Table 5.3), which means that 100% of the multi-gesture utterances contained speech for four of the five children.<sup>9,10</sup> Hence, if multi-gesture utterances are comparatively rare to begin with (as shown in Section 5.4), multi-gesture utterances without speech is an even more rare phenomenon in hearing children.

Looking at the curve showing multi-gesture utterances that are coordinated with speech, there are two marked peaks that are well aligned with TP#2 and TP#3 and both of them occur at the core of these two transition periods. That is, not for the first time in this chapter there is striking similarities in the structure of the developmental graphs at TP#2 and TP#3. Since the magnitude of the numbers is relatively small here — multi-gesture utterances are comparably rare as shown in Table 5.3 — it may be worth pointing out that the curve that describes multi-gesture utterances that are coordinated with speech is not the product of one or a few children only, but all of them contribute substantially to the shape of this curve in the graph. This was not the case for the other curve in the graph — the one showing multi-gesture

<sup>9</sup>Yet, it is interesting to note that most of these autonomous gesture+gesture utterances in Bella come within the range of TP#3 — the transition period associated with an increasing productivity in grammatical combinations. This can be seen in the small “bump” in the curve that shows multi-gesture utterances that are performed without speech in Figure 5.9.

<sup>10</sup>This is in line with findings from other studies, primarily on slightly younger children, showing that gesture+gesture utterances are much more rare than gesture+speech or speech+speech (multi-word) utterances (e.g. Goldin-Meadow & Morford, 1990; Capirci et al., 1996, 2005; Stefanini et al., 2009).



utterances — where the little “bump” in the curve at the beginning of TP#3 is almost wholly due to one child (Bella), as mentioned above.

### 5.5.3 Relations between multi-word and multi-gesture combinations

Figure 5.10 shows the frequency (per minute) of gestures in one-word and multi-word utterances. At TP#1 it is much more common to perform gestures with one-word utterances than with multi-word utterances, at TP#2 it is about as common to perform gestures with one-word utterances as multi-word utterances, and at TP#3 it is clearly more common to perform gesture in coordination with multi-word utterances. This pattern is to be expected given that MLU rises from roughly 1 at TP#1, 2 at TP#2, and 3 at TP#3, as it is to a large extent a reflection of the increase of MLU over time.

Another way to show the distribution of gestures in one-word and multi-word utterances that takes away the effect of the increase in MLU over time can be seen in Figure 5.11. This graph shows two curves. The first curve shows the proportion of one-word utterances that contain gesture. The other curve shows the proportion of multi-word utterances that contain gesture. These two ratios are independent of each other, and furthermore, they do not depend on how common one-word and multi-word utterances are at a certain age (i.e., MLU is disregarded). Since the two curves represent independent ratios, they do not add up to 100%, but they are still presented in the same graph to make comparison easier. The graph shows that the ratio of one-word and multi-word utterances that contain gesture are in fact relatively stable, and similar, from the beginning of the period up to the end of TP#2.<sup>11</sup> After TP#2 the tendency to integrate gestures with multi-word utterances suddenly increases markedly, eventually peaking at the core of TP#3 — *suggesting that gestures are now finding their ways into the more generalized and productive grammatical ways of organizing speech that starts to emerge around TP#3*. At the core of TP#3, as many as 85% of the multi-word utterances are coordinated with gesture, which is a considerably higher proportion than found anywhere else in the studied period, both for one-word and multi-word utterances.

---

<sup>11</sup>At a more detailed level, the following can be observed: There is a small increase of gestures that are coordinated with multi-word utterances at the beginning of TP#2, so that the tendency to perform gestures in multi-word utterances is slightly stronger at TP#2, whereas there is a slightly stronger tendency to perform gestures in one-word utterances at TP#1.

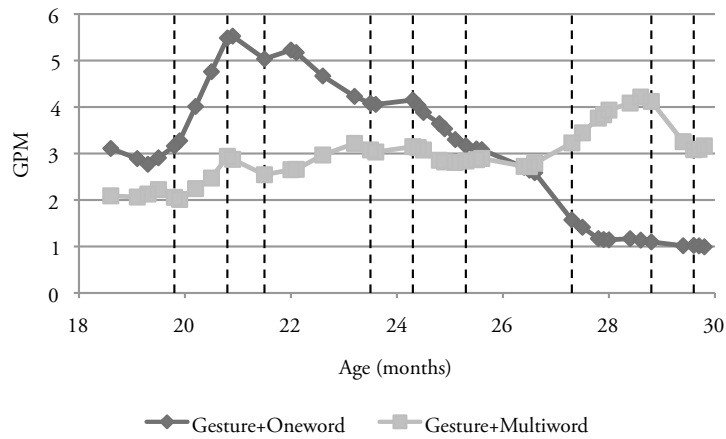


Figure 5.10: Gestures in one-word and multi-word utterances (per minute).



Figure 5.11: Proportion of one-word and multi-word utterances that contain gesture.

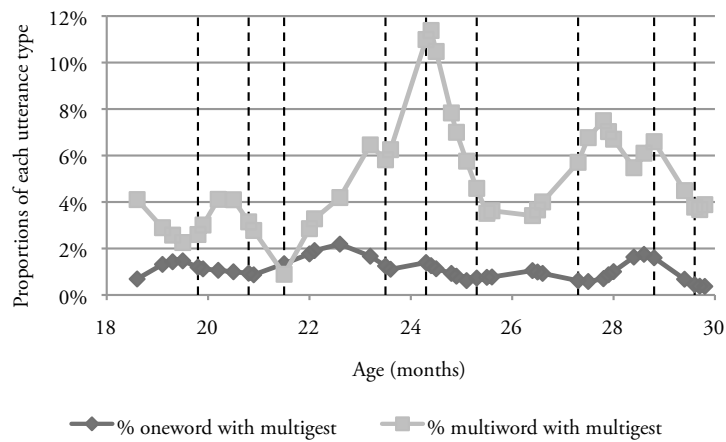


Figure 5.12: Proportion of one-word and multi-word utterances with multi-gesture.

Figure 5.12 shows the proportions of one-word and multi-word utterances that are coordinated with more than one gesture (either simultaneously or in sequence, within the scope of a single utterance). Here it can be seen that the proportion of multi-gesture combinations occurring with multi-word speech is larger than the proportion of multi-gesture combinations that occur with one-word speech. This holds true throughout the whole period, except at the end of TP#1. There is a peak in multi-gesture combinations that occur with multi-word speech at each of the three transition periods, and in particular at TP#2. It may be recalled that TP#2 is also the point in time when MLU reaches 2, i.e., the point in time where multi-word speech is a rule rather than an exception.

Taken together, Figure 5.11 and Figure 5.12 imply that *speech seems to provide a frame within which to organize more complex gesture performances, and that more complex speech combinations do not stand in opposition to more complex gesture combinations*. It does not seem to be the case that multi-word utterances are less likely to contain gesture than one-word utterances. Indeed, the proportion of one-word utterances that contain gesture decreases over the analyzed period (from around 50% to 20%). Gesture rather seem to be particularly useful in the propositional multi-word utterances that emerge most clearly around TP#3, and it is also in multi-word utterances, rather than one-word utterances, that most gesture+gesture combinations are found. In this context it may also be useful to recall that the sudden increase of gestures in the first part of TP#1 (see Section 5.3) was a matter of an increase in gesture+speech utterances, and that there was no increase in the number of autonomous gesture-only utterances (see Section 5.4). Gesture-only utterances are also relatively infrequent overall.

### 5.5.4 Summary

There is an overall increase in utterances (multimodally conceived) per minute in the first part of TP#1, as shown in Section 5.3, and these were mainly a matter of gesture+speech utterances, as shown in Section 5.4. The present sub-section added the further qualification that the vast majority of these utterances are also one-word utterances. There were only a slight increase in multi-word utterances by comparison.

At the end of TP#2 there is a peak in two-word utterances, and soon after this, between TP#2 and TP#3, the number of longer multi-word utterances (primarily three- and four-word speech) starts to increase, eventually becoming more common than one-word-utterances at TP#3. Hence, TP#3 seems to be connected to a more

grammatically complex way of organizing speech. This claim finds ample support in research on the development of spoken language in children, and it will be discussed in more detail in the concluding section of this chapter.

Multi-gesture utterances are rare, and whenever they appear, it is almost always in coordination with speech. During TP#2 and TP#3, there are temporary peaks in the number of multi-gesture utterances.

Throughout the studied period, there is a shift from mostly gesture+speech utterances of the one-word kind at TP#1, to a roughly equal amount of one-word and multi-word gesture+speech utterances at TP#2, and a dominance of multi-word gesture+speech utterances at TP#3.

When considering the ratio of one-word utterances and multi-word utterances (treated independently), it shows that the ratio is quite similar, up to the end of TP#2, where the ratio of multi-word utterances that contain gesture suddenly diverges to become more and more common up to the core of TP#3 whereas the ratio of one-word utterances that contain gesture decreases. The ratio of one-word utterances that contain gesture is rather decreasing during this period. After the core of TP#3, the proportion multi-word utterances that contain gesture decreases again.

When considering the proportion of one-word and multi-word utterances that contain multi-gesture combinations (sequential or simultaneous combinations), it shows that there is a peak of multi-word utterances with multi-gesture combinations at TP#2 and TP#3. Multi-gesture sequences in combination with one-word utterances are rare.

## 5.6 Deictic, iconic, and conventionalized aspects

*How do frequencies of gestures with deictic, iconic, and conventionalized aspects respectively change over time?* In this thesis, deictic, iconic, and conventionalized aspects of gesture are not treated as separate types of gestures, but rather as different aspects of gestures (see also Zlatev & Andrén, 2009). This means that a single gesture may be coded as both deictic and iconic, or iconic and conventionalized, or even all three.<sup>12</sup> This stands in contrast to almost all other literature on child gesture, despite recommendations in the literature on adults' gestures not to treat deictic, iconic, and conventionalized *aspects* of gesture as if only one of these aspects could be present in a given gesture (Kendon, 2004; McNeill et al., 2004). However, even

---

<sup>12</sup>The coding of conventionalized aspects is however perhaps too restrictive here, as discussed further below in this section.

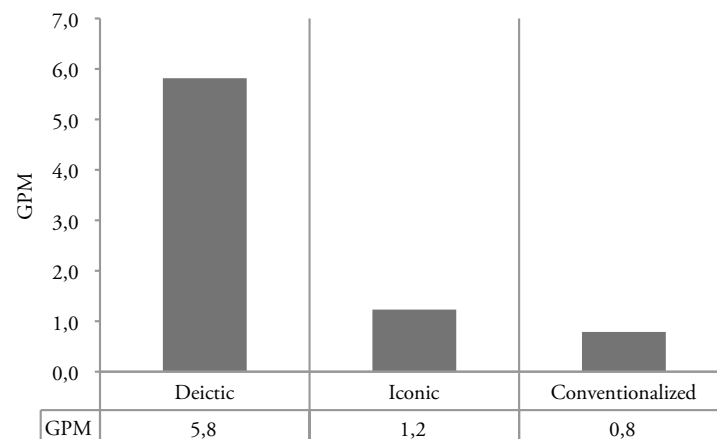


Figure 5.13: Gestures with deictic, iconic, and conventionalized aspects (per minute).

though one and the same gesture can involve one or more of the semiotic properties (deictic, iconic, and/or conventionalized), it is still possible to analyze the properties separately, and that is what is done in this section.

### 5.6.1 An overview of the semiotic aspects

First a general comment on all the six graphs. The graphs show that there are quite different patterns in the three semiotic aspects during this period. Nevertheless, in each of the graphs there are things happening that align well with TP#1, TP#2, and TP#3. Once again, this lends support to the idea that there does indeed seem to be some quite complex changes in the organization and use of gesture at these three “transition periods”.

Deictic aspects of gestures are by far the most common (see Figure 5.13 and Figure 5.17), with an average of 5.8 gestures that include deictic aspects per minute. Iconic aspects and conventionalized aspects appear on average 1.2 times and 0.8 times per minute. That is, gestures with deictic aspects are 4.83 times more common than gestures with iconic aspects and 7.25 times more common than gestures with conventionalized aspects. However, one should be wary of drawing too quick conclusions from this data. First of all, it should be recalled that the coding of conventionalized gestures is quite restricted here (see Section 4.3). For example, many deictic gestures such as index finger pointing, GIVE, and SHOW were not coded as conventional gestures here, even though they might very well be considered to be conventionalized gestures, as some researchers have indeed done (e.g. Johnson

et al., 1975; Greenfield et al., 1985; Guidetti, 2002).<sup>13</sup> Furthermore, only gestures with clearly normatively constrained form-meaning mappings (according to adult standards, i.e., emblematic gestures) were coded as conventionalized gestures here, which excluded aspects of gestures that are conventionalized on the broader level of typified, but not normative, conventions. ACTION\_BASED iconic aspects that consist in gestural performances of conventionalized actions (e.g. pretending to use a remote control) are therefore not coded as “conventionalized” gestures, since there are no strong constraints of form that are comparable with emblematic gestures. Hence, many of the gestures with deictic and iconic aspects are also conventionalized to some degree — at least to the level of typified conventions (Conv#2). Therefore, the data presented here could equally well be interpreted as being in support of the claims of Guidetti & Nicoladis (2008, p. 109), which go as follows:

If our reasoning is correct, then infants may use primarily conventional gestures, as well as gestures that they have learned by acting in the world (such as ‘pick-me-up’). There is a curious lack of the predominant kind of gesture used by adults: spontaneous, non-conventional gestures that seem to be created on the spot to convey meaning [...]

Other researchers have also found very low rates of gestures with iconic aspects for children in these ages (e.g. Nicoladis et al., 1999).

### 5.6.2 A longitudinal comparison between the semiotic aspects

Figures 5.14–5.19 show separate graphs of the frequencies of gestures with deictic, iconic, and conventionalized aspects at different ages, both in terms of GPM and GPU. The first thing to note is that gestures with deictic aspects dominate throughout the whole period, but that they decrease throughout the period, after an initial peak at the core of TP#1. Section 5.4 showed that the increase of gestures per minute at the first part of TP#1 was mainly a matter of an increasing use of gestures that are coordinated with speech, and Section 5.5 showed that this was mainly an increase of gestures that are coordinated with one-word utterances. Figure 5.14 shows that there is a corresponding increase specifically in gestures with deictic aspects precisely at the first part of TP#1, indicating that the gestures that

---

<sup>13</sup>However, it may be problematic to characterize all sorts of pointing gestures as equally conventionalized, or non-conventionalized, since pointing gestures may vary in the extent to which they exploit conventionalized aspects.

## 5.6. DEICTIC, ICONIC, AND CONVENTIONALIZED ASPECTS

---

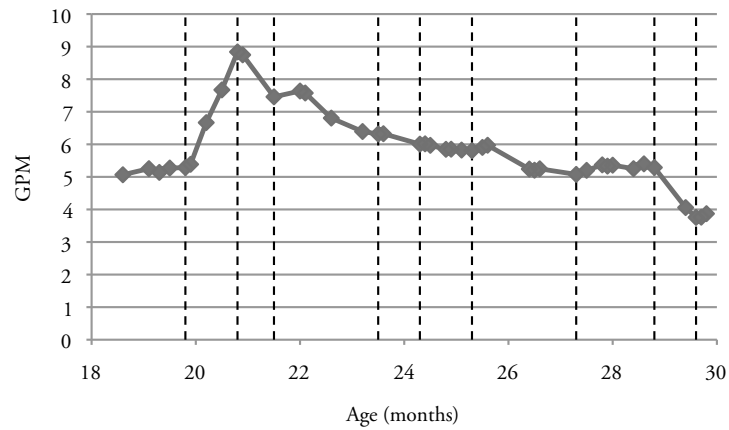


Figure 5.14: Gestures with deictic aspects (per minute).

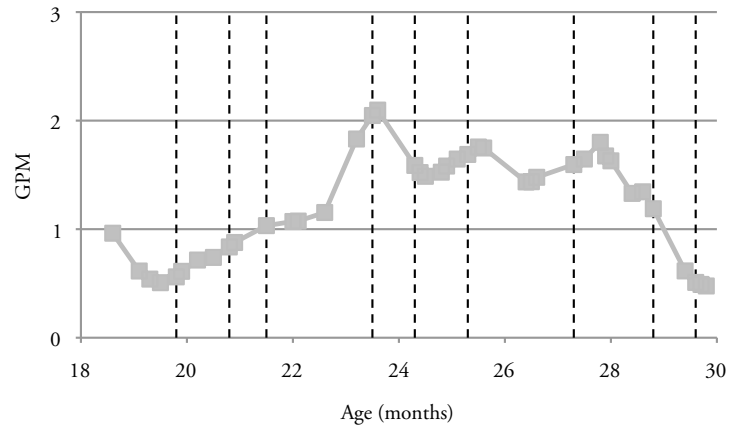


Figure 5.15: Gestures with iconic aspects (per minute).

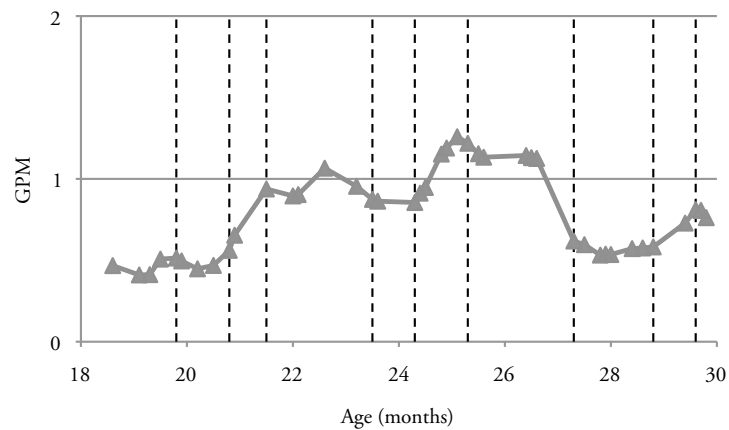


Figure 5.16: Gestures with conventionalized aspects (per minute).

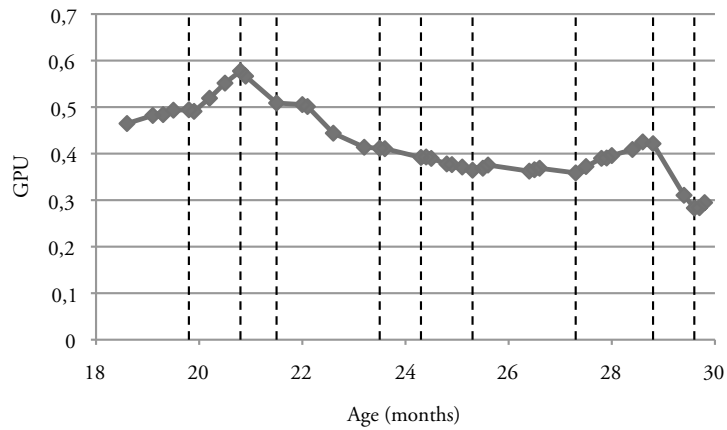


Figure 5.17: Gestures with deictic aspects (per utterance).

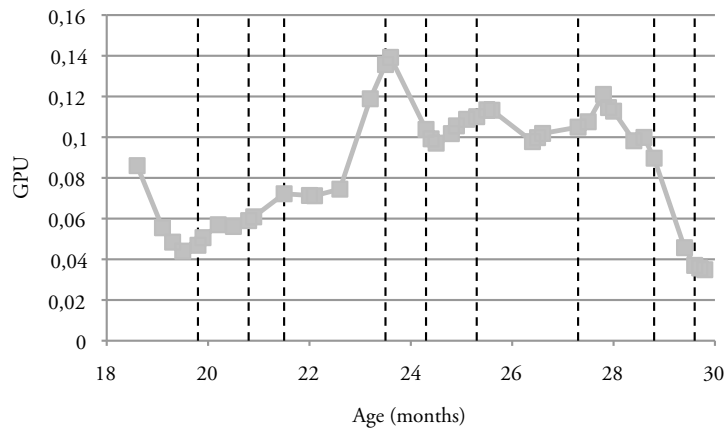


Figure 5.18: Gestures with iconic aspects (per utterance).

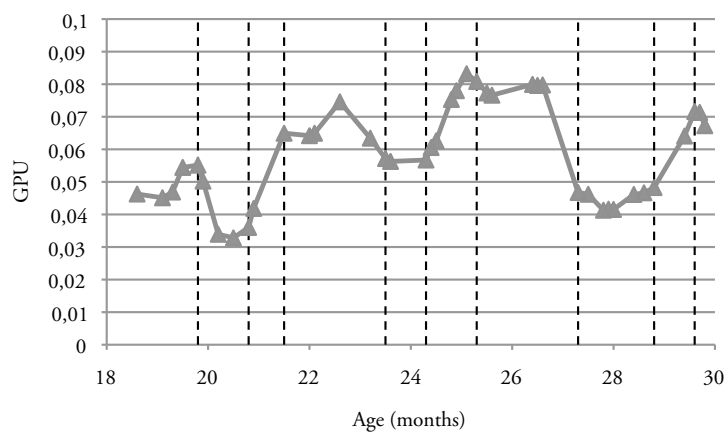


Figure 5.19: Gestures with conventionalized aspects (per utterance).



are used increasingly more often in coordinated with one-word are primarily deictic gestures. After a peak at the core of TP#1, deictic gestures decrease throughout the period. Gestures with conventionalized aspects do not increase at all during the first part of TP#1 and the magnitude of the increase in gestures with iconic aspects at this point is very small compared to the increase in gestures with deictic aspects.

Regarding gestures with iconic aspects (Figure 5.15 and Figure 5.18), it must be pointed out that the first value (at 18.3 months) in both the GPU graph and the GPM graph is almost wholly due to one child (Tea), who repeats a single play-scheme over and over again for quite some time, in a single session.<sup>14</sup> To this, there is the further problematic issue that gestures that are seemingly of the INDIRECT iconic kind from the point of view of the adult, may simply be a imitated action on behalf of the child, so that the child would be able to perform the action by means of the capabilities underlying ACTION\_BASED iconic gesture. The non-creativity that characterizes a high degree of repetition of a single scheme indicates that this may be precisely what is going on in this case. This session is therefore considered to be something like an outlier here, although it was not removed from the graphs.

Apart from this somewhat exceptional session, there is an overall increase of gestures with iconic aspects from the beginning of the period, with a temporary peak at the onset of TP#2, and then there is a quite sharp decrease again at TP#3. The fact that iconic aspects actually increase throughout the period seems to run counter to findings by other authors, who argue that iconic gestures stay relatively low, at least throughout the second year of life (e.g. McNeill, 1992; Iverson et al., 1994; Nicoladis et al., 1999; Volterra et al., 2005). As it will be argued in Section 5.6.4, this may be due to the fact that object-gestures are included in the definition of gesture here.

It would have been informative to know what happens after TP#3 — if the curve stays low, or if the dip is just temporary — before suggesting interpretations of what this dip might mean. The temporary peak in iconic gestures at the onset of TP#2 has not been reported before in the literature (but see Zlatev & Andrén, 2009, p. 322). However, it may be noted that Capirci et al. (2005) found that various typified (iconic) object-gestures tended to precede their empty-handed counterparts, which suggests that the lack of such reports may be due to the fact that many stud-

---

<sup>14</sup>In future work, a type/token distinction will be introduced in the annotation scheme for iconic gestures, so that it is possible to distinguish between increases in the variety of iconic gestures that are used and increases in the number of instances of each recurrent iconic gesture. It is quite clear from the data that I have analyzed that many of the children's iconic gestures are recurrent, both within and across sessions. Still, the number of repetitions in the "outlier" session with Tea is considerably more repetitive than other sessions.

ies focus exclusively on empty-handed gestures. Further details on the nature of this peak are provided in Section 5.6.4, that deals with sub-types of iconic aspects.

Regarding gestures with conventionalized aspects (Figure 5.16 and Figure 5.19), there is an interesting dip in the first part of all three transition periods (TP#1, TP#2, and TP#3), most pronounced in the GPU version of the graph. Why there would be a dip in association with each transition period is very much an open question, but at least it could be said that there seems to be something about these transition periods that does *not* promote the use of gestures with Conv#3 conventionalized aspects. In the upcoming discussion of sub-types of each semiotic category, some more data of gestures with conventionalized aspects is presented, and this may provide slightly more clarity in what is going with these gestures here.

### 5.6.3 Sub-types of deictic aspects

Figure 5.20 shows the proportions of the different sub-types of gestures with deictic aspects that were included in the annotation scheme. The `INDEX_FINGER_POINTING` category and the `OTHER` category were clearly most common at 44% and 49% respectively. The `OTHER_POINTING` category was infrequent in comparison, at only 7%. In Section 4.3, where the annotation scheme was presented, it was suggested that the `OTHER_POINTING` category is somewhat ambiguous. On the one hand it includes less articulated gestures, and on the other hand it includes pointing gestures that are performed with well articulated hand-forms other than the index finger pointing hand-shape (i.e., in the latter case, the hand-shape aspect of the gesture is more clearly on the level of Comm#3 communicative complexity). Cochet & Vauclair (in press) found a decline in less articulated pointing gestures after 21 months, and if it is hypothesized that less articulated hand-shapes are more common during the early half of the period (which seems to be the case), than during the latter half, and vice versa for well-articulated kinds of pointing other than index finger pointing, then the findings of Cochet and Vauclair are in line with what is found here, in Figure 5.21 and Figure 5.22. However, the number of instances that are involved here are quite low and no strong conclusions can be drawn in this regard.

An interesting aspect of the curve that shows `INDEX_FINGER_POINTING` is that it does not simply decrease after its initial peak close to TP#1, the way gestures with deictic aspects do more generally. There is a peak in `INDEX_FINGER_POINTING` again at the core of TP#3, when MLU is about to reach 3 and there is a peak in multi-word utterances more generally. This suggests that pointing has a role to play

## 5.6. DEICTIC, ICONIC, AND CONVENTIONALIZED ASPECTS

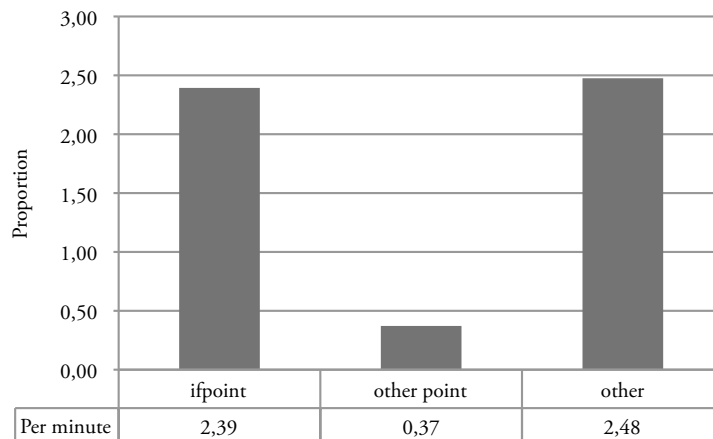


Figure 5.20: Proportions of deictic sub-types.

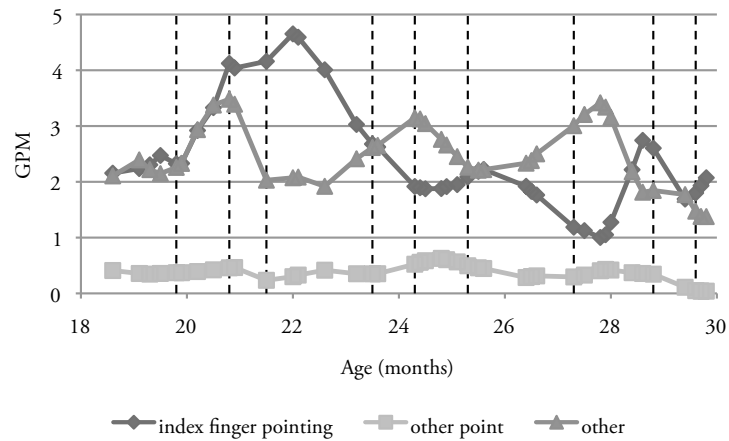


Figure 5.21: Deictic sub-types (per minute).

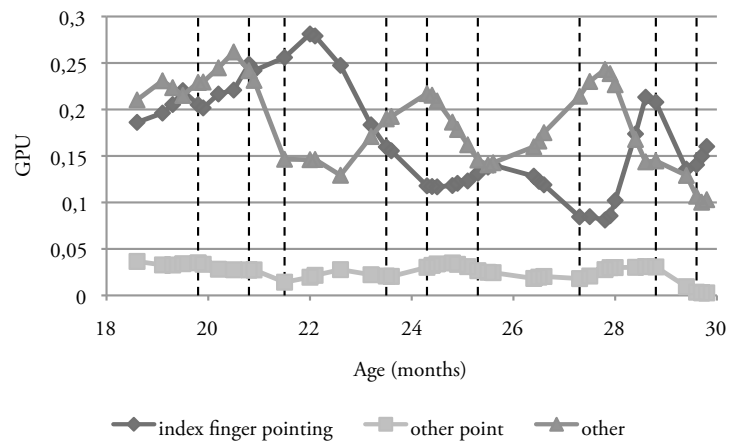


Figure 5.22: Deictic sub-types (per utterance).

once again (after the “word learning” period of pointing that precedes TP#2) at a time in development when the child is beginning to master a new level of complexity in the organization of speech (cf. Özçalışkan & Goldin-Meadow, 2009).<sup>15</sup>

The patterns of the INDEX\_FINGER\_POINTING category and the OTHER category show an interesting inverse relationship from the core of TP#1 and onwards. When one goes up, the other goes down, implying that there is something like a fairly stable “need” for deixis in social interaction, although deictic reference and attention regulation can be achieved by means of different kinds of deictic gestures. As argued by Clark (2003), pointing gestures are only one type of deictic gestures among others, such as PLACING objects to the benefits of others, and of course also “classic” child gestures such as GIVE and SHOW (e.g. Bates et al., 1979; Nicoladis et al., 1999). Note that the combination of these deictic sub-types results in a “curve” which is close to a straight line from the core of TP#1 up to the core of TP#3, as shown in Figure 5.14.

#### 5.6.4 Sub-types of iconic aspects

Figure 5.23 shows the proportions of the different sub-types of gestures with iconic aspects that were included in the annotation scheme. The majority of these are ACTION\_BASED iconic aspects of gesture, at 58%. INDIRECT iconic aspects constitute 30%, and TRACING aspects constitute 12%. Similarly, in a study of Italian children at 27 and 30 months, and older, Stefanini et al. (2009, p. 171) found that ACTION\_BASED gestures were more common than what they call “size-shape gestures”.

Figure 5.24 and Figure 5.25 begin with the temporary high at 18.6 months that was due to the “outlier” session with Tea. Gestures with ACTION\_BASED iconic aspects become more common throughout the period, with a peak at TP#3 and then a final decrease at the core of TP#3. Gestures with INDIRECT iconic aspects start increasing after TP#1 and reach “full height” at the onset of TP#2, but become much less common again at TP#3. The fact that gestures with INDIRECT iconic aspects seem to emerge after gestures with ACTION\_BASED iconic aspects is to be expected if one turns to the commonly invoked notion of an increasing *symbolic distancing* in

---

<sup>15</sup>It has been shown that gestures serve as a precursor to later achievements in language development, such as allowing the child to create two-unit utterances across modalities before the child is able to do so within the spoken modality alone (Capirci et al., 1996; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005b, 2009).

## 5.6. DEICTIC, ICONIC, AND CONVENTIONALIZED ASPECTS

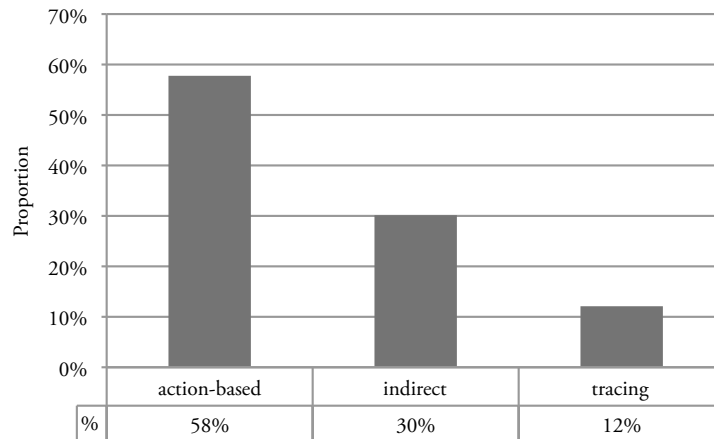


Figure 5.23: Proportions of iconic sub-types.

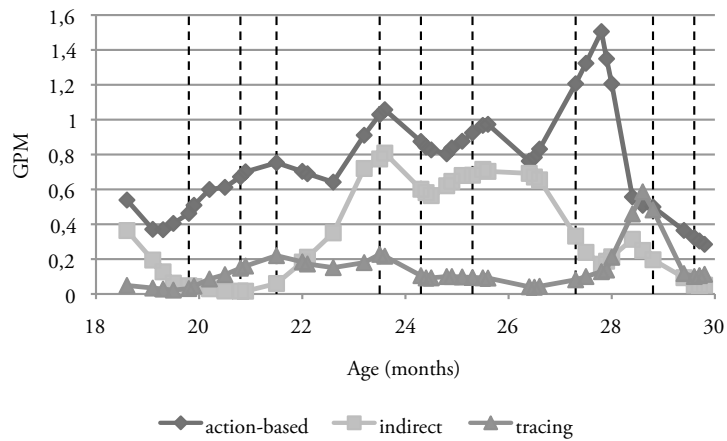


Figure 5.24: Iconic sub-types (per minute).

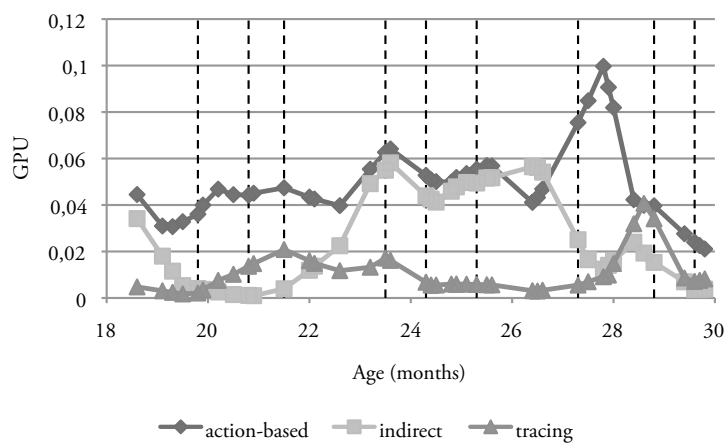


Figure 5.25: Iconic sub-types (per utterance).

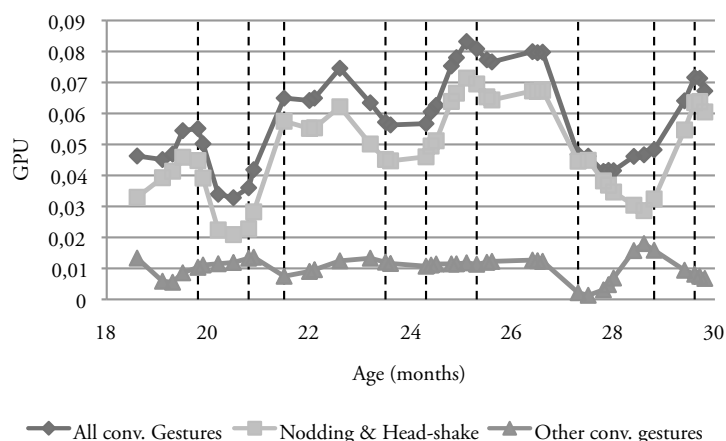


Figure 5.26: The dominance of NODDING and HEAD\_SHAKE (per utterance).

children’s semiotic development in general (Werner & Kaplan, 1963). INDIRECT iconic gestures are indeed more abstract than ACTION\_BASED iconic gestures, since the “distance” between the signifier and the signified is more pronounced.

Regarding gestures that involve aspects of TRACING, the situation is similar to that of the ambiguity in the deictic sub-category OTHER\_POINTING. Many of the early performances were quite ambiguous as to whether the tracing was actually more of a kind of exploration for the self than an expression at level Comm#2 of communicative complexity, whereas the instances of TRACING were more clearly articulated and more clearly communicative in the sense of Comm#3 in the latter half of the period studied here. Hence, this category seems to hide both more “primitive” kinds of action and some more “advanced” ones. Indeed, there is a period between the onset of TP#1 and the core of TP#2 where the curve rises a bit, only to get close to zero again from the core of TP#2 up to the onset of TP#3, where there is suddenly a peak. These two separate periods of activity in the TRACING category may indicate precisely such a difference between a more “primitive” kind of TRACING, and a more “advanced” one, in terms of communicative explicitness.

### 5.6.5 Sub-types of conventionalized aspects

As mentioned before, the coding of the conventionalized aspects of gestures is quite restricted. It only includes gestures that are conventionalized to the degree of Conv#3 normative conventions, and deictic gestures such as INDEX\_FINGER\_POINTING, GIVE, and SHOW were not coded as conventionalized gestures, although they are clearly conventionalized at least at the level of Conv#2 typified conventions.

The conventionalized category has a lot more sub-categories than the deictic and the iconic category. For the conventionalized category, the sub-types is a list of specific conventionalized gestures. As it turns out, NODDING and HEAD\_SHAKE alone amounts to 45.8% and 36.7%, respectively, of all the conventionalized gestures. In Figure 5.26 NODDING and HEAD\_SHAKE are joined into one curve. This curve lies very close to the curve showing all gestures that were coded as conventionalized, which effectively means that that the overall shape of the curve showing “all conventionalized gestures” is in fact more or less constituted by NODDING and HEAD\_SHAKE. The curve that shows the frequency of other conventionalized gestures has a different shape. Hence, the dips at the cores of TP#1, TP#2, and TP#3 in conventionalized gestures is rather a dip in the use of NODDING and HEAD\_SHAKE at these transition periods. Even though this does still not explain why the dips may be there, it does at least narrow the question down to a matter of the organization of NODDING and HEAD\_SHAKE. More generally, Nicoladis et al. (1999) found that the rate of use of conventionalized gestures remains fairly stable throughout the preschool years.

In a way, each conventionalized gesture has a specific nature, almost by definition, since they are particular more or less determined/conventionalized kinds of action which may have particular uses. Gestures with conventionalized aspects therefore constitute a quite heterogeneous category. For example, NODDING and HEAD\_SHAKE are frequently used in all kinds of discourse. In contrast, a gesture like BYE\_BYE is naturally much less frequently used (it amounts to 0.53% of the gestures with conventionalized aspects), since it is used in a much more narrow context. Already these two kinds of conventionalized gestures have quite different relations to speech. For example, NODDING and HEAD\_SHAKE may be used even while *another* person is speaking as a way of agreeing without what the other person says without claiming the turn. Also, except for the most common conventionalized gestures, the children differ quite a lot in which conventionalized gestures they use, and in many cases only one child uses a specific conventionalized gesture, more or less recurrently. Analyses of more specific conventionalized gestures are provided in Chapter 8.

### 5.6.6 Semiotic aspects with and without speech

Even though one cannot assume that the deictic, iconic and conventionalized semiotic aspects are internally homogeneous in how they relate to speech — for example,

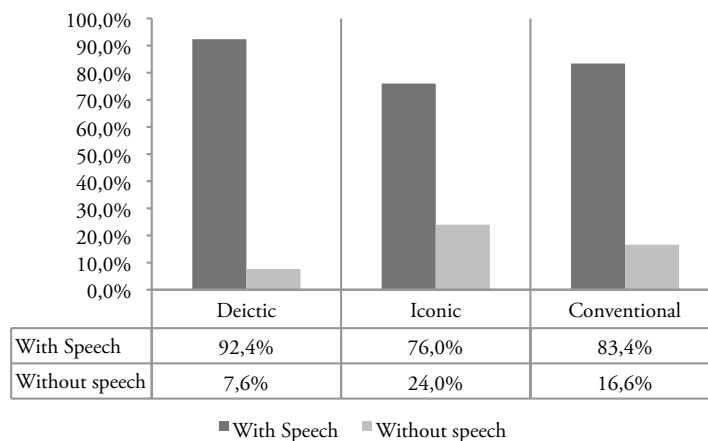


Figure 5.27: Proportion of gestures with and without speech.

far from all gestures with conventionalized aspects relate to speech in the same way, as discussed in the previous section — it is nevertheless interesting to ask the following question: *To what extent do gestures with deictic, iconic, and/or conventionalized aspects tend to occur with or without speech?* Or formulated otherwise: *Which gestures tend to be performed autonomously?* The reason for this is that there are frequently claims about the nature of different types of gestures precisely on this general level. Conventionalized gestures are often claimed to be of a more autonomous kind than iconic gestures, in the sense that conventionalized gestures would be more independent of speech, while iconic gestures would be more dependent on speech (e.g. McNeill, 2005; Singleton et al., 1995). This was discussed in relation to the issue of the upper limit in Section 2.3, where it was also argued that a more promising position is to acknowledge that conventionalized gestures are also very frequently used as co-speech gestures, rather than autonomously, as suggested by (Kendon, 1996, 2008).

As it turns out (see Figure 5.27), it is actually *more* common to perform gestures with iconic aspects autonomously (24.0%), i.e., without speech, than performing gestures with conventionalized aspects autonomously (16.6%). The most common semiotic aspect to be coordinated with speech is the deictic aspect though (92.4%). If more of the deictic gestures would have been coded as conventionalized gestures, as discussed above, the tendency for conventionalized gestures to be coordinated with speech would be even stronger. Figures 5.28–5.30 show how the proportion changes at different ages. In these graphs it can also be seen that autonomous conventionalized gestures even drop to 0% during the most part of TP#1, whereas gestures with iconic aspects never reaches lower than 14%.



## 5.6. DEICTIC, ICONIC, AND CONVENTIONALIZED ASPECTS

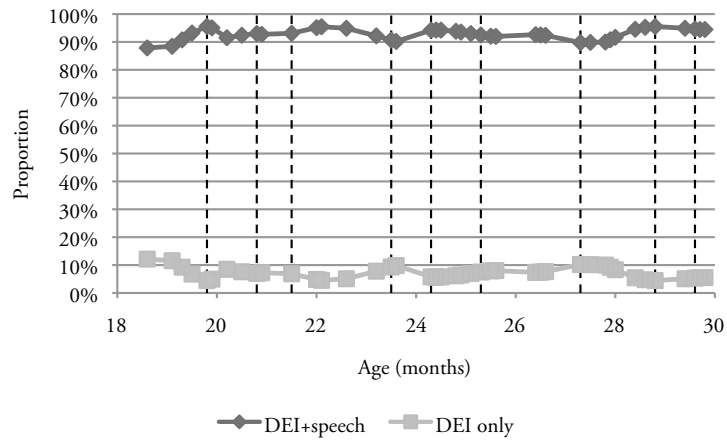


Figure 5.28: The proportion of deictic aspects with/without speech over time.

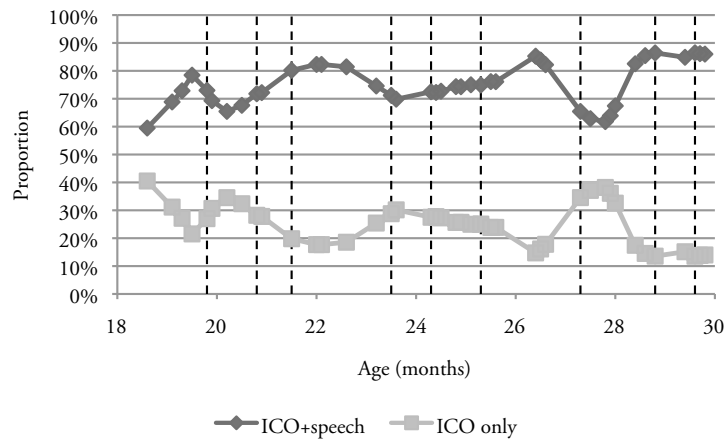


Figure 5.29: The proportion of iconic aspects with/without speech over time.

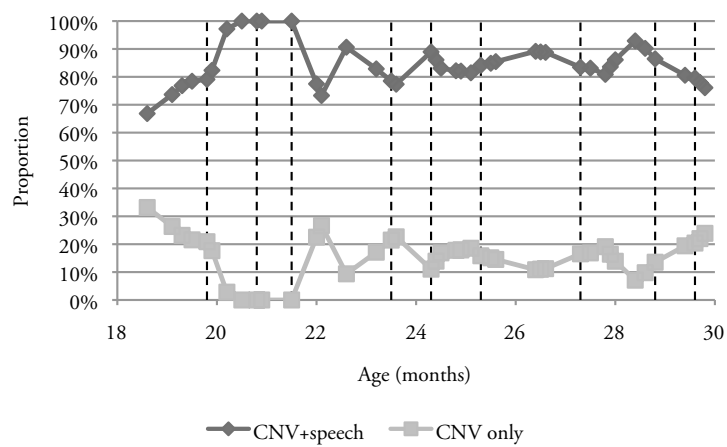


Figure 5.30: The proportion of conventionalized aspects with/without speech over time.

All in all this casts doubt on a generalized and homogenized characterization of iconic gestures according to which iconic gestures would be more dependent on speech, and conventionalized gestures would be less typically used as co-speech gestures. Ladewig (2010) is currently studying the use of gesture in cases where there is a “gap” in speech of various sorts, i.e., during hesitations or when gestures are used “replacing” a word with a gesture in a syntactic slot (e.g. Slama-Cazacu, 1976), and so forth, and she also finds many instances where iconic gestures are performed autonomously. At the very least, the findings here indicate that there is nothing about the fact that a gesture is conventionalized that makes it less *likely* to go together with speech. It still seems reasonable to say that (some) conventionalized gestures may have a different kind of *potential* for autonomy, due to the somewhat less context-dependent kind of meaning that they may convey, but that exploiting this potential is the exception rather than the rule (at least in Swedish children).

Figures 5.28–5.30 also show that the tendency for deictic gestures to be combined with speech is remarkably stable, lying on a consistently high level throughout the period. This supports the idea that deictic gestures constitute a particularly speech-coordinated kind of gesture, that primarily serves its functions in the context of speech. However, for all three semiotic aspects, the majority of the gestures are coordinated with speech.

### 5.6.7 Summary

Deictic aspects are by far the most common in children’s gestures. Iconic and conventionalized aspects are much less common. However, it was argued that the presence or non-presence of conventional aspects is to a large extent a matter of which *level of conventionalization* (as defined in Section 2.3.3) that is required in a definition of gesture. If not only *normatively* regulated gestures are included as “conventional” (i.e., emblematic gestures), but also *typified* (but less constrained) gestures such as SHOW and OFFER/GIVE as it was in the coding system used as a basis for this chapter, then the distribution changes considerably, so that the vast majority of all gestures contain a conventionalized element. Indeed, as Kendon (1996) argues, then “there is conventionalization to a degree affecting all kinds of gesturing”.<sup>16</sup>

Nevertheless, maintaining more strict criteria for the conventionalized status of gestures makes the results more comparable to previous research on gesture, and

---

<sup>16</sup>The electronic version of this publication does not have page numbers, and therefore no page numbers are given for this quote.

that was the reason for coding conventionalized gestures in this more restricted way in this chapter.

Deictic gestures dominate throughout the period, but decrease throughout the period after an initial peak at the core of TP#1. For the GPU measure, there was also a small peak in the use of deictic gestures at the core of TP#3. The increase in use of gesture at TP#1 that has been discussed at several places already were now further qualified to consist mostly in deictic gestures. To recapitulate, this means that the increase in gesture+speech utterances per minute at the first part of TP#1 is a matter of one-word utterances that are combined with deictic gestures.

The INDEX\_FINGER\_POINTING sub-type and the OTHER sub-type were most common, and OTHER\_POINTING rare in comparison. After an initial increase in the first part of TP#1 of both INDEX\_FINGER\_POINTING and OTHER, these two types seems to stand in a somewhat inverse relationship, so that whenever one is more frequent, the other is less frequent. The OTHER\_POINTING category may be ambiguous in the sense that there seems to be one primitive version of it, and one more advanced version. The primitive version consists in pointing gestures with little articulation of hand-shape. The more advanced version consists in well articulated pointing gestures, but with a different hand-shape than the index finger pointing hand-shape. Some forms of deictic gestures are further analyzed in Chapter 6.

Gestures with ACTION\_BASED iconic aspects were clearly most common overall. These gestures increase throughout the period, with a final peak at the beginning of TP#3, and after this there is a sharp decline. INDIRECT iconic aspects emerge around at TP#2. TRACING iconic aspects were argued to consist in two distinct phenomena — one “primitive” form with less clear expressive properties, appearing mostly before TP#2, and one more explicitly expressive form, which mainly appeared at TP#3. This claim was not based on the quantifications presented in this section, but rather on observation of the recordings. Hence, further research will be required to justify this claim, but it was still pointed out since it is relevant to the interpretation of the curve showing TRACING gestures. TRACING gestures of the “advanced” kind are further analyzed in Chapter 6.

Conventionalized aspects of gestures, in the sense of emblematic gestures, turned out to be vastly dominated by NODDING and HEAD\_SHAKE. Together they amount to 85.2% of all instances of emblematic gestures, which indicates that they have a special status in communication. It was argued that each type of conventionalized gesture is in a way unique, in the sense that each type of conventionalized gesture has its own distinct contexts of use (cf. Kendon, 2004). A special treatment of NOD-

DING and HEAD\_SHAKE is provided in Chapter 8. That chapter also includes further analysis of a number of other conventionalized gestures — both emblematic ones, and less strongly conventionalized forms.

It was found that gestures with deictic aspects are most inclined to be coordinated with speech. Gestures with conventionalized aspects are the second most common kind of gestures to be coordinated with speech. The gestures that are most frequently performed autonomously are iconic gestures. These findings runs counter to theories about gesture which postulate an opposition between the presence of conventionalized factors in gesture and the tendency of these gestures to be coordinated with speech, i.e., that conventionalized gestures tend to be more autonomous than iconic gestures (e.g. McNeill, 2005). It seems that a distinction is required between different kinds of autonomy. Even though the results in this section show that gestures with iconic aspects may often be autonomous (see also Ladewig, 2010), it is still indisputable that conventionalized gestures may be autonomous in a less context-dependent way than iconic gestures that are performed autonomously.

Nevertheless, it is still the case that a clear majority of all kinds of gestures are coordinated with speech, and this holds true for each of the semiotic aspects too.

## 5.7 Gestures that involve handling of objects

*Do the children's gestures typically involve handling of objects or are they typically performed empty-handed? Are object-gestures typically used with speech or without? Are there differences between gestures with deictic, iconic, and conventionalized aspects when it comes to object involvement?* In this thesis, “gesture” is not defined in a way that excludes communicative and/or semiotically complex acts that involve handling of objects.<sup>17</sup> Hence there are some gestures that involve active handling and some that do not, and it is this distinction that is the topic of this section. Note that the distinction is not equivalent to a distinction between gestures that relate in some way to the immediate environment versus those that are more “detached” from the present surroundings. There are many empty-handed gestures that still relate to the material surroundings. Index finger pointing is one example of this, and there are also many gestures with iconic aspects that involve some form of directed-

---

<sup>17</sup>The definition of gesture was given in Section 2.2.5.

## 5.7. GESTURES THAT INVOLVE HANDLING OF OBJECTS



Figure 5.31: Gestures with and without objects (per minute).

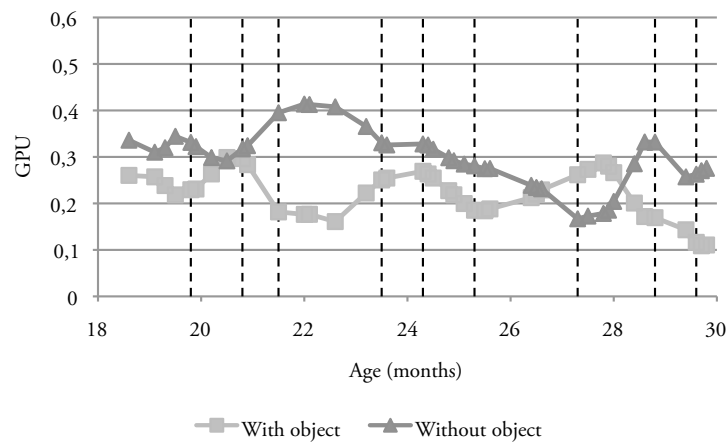


Figure 5.32: Gestures with and without objects (per utterance).

ness toward elements in the present environment.<sup>18</sup>

Gestures that involve handling of objects are referred to as *object-gestures* in this thesis. It should also be noted that gestures performed with other articulators, such as NODDING and HEAD\_SHAKE, are also counted as “empty-handed” gestures since they do not involve handling of a physical object, although they are not articulated by the hands specifically.

### 5.7.1 Object-gestures versus empty-handed gestures

57% of the children's gestures are empty-handed and 43% of the gestures are object-gestures. This means that object-gestures constitute a substantial part of the children's gesture performances, and by implication, that an account of children's repertoires of expressive bodily actions cannot ignore these.

Figure 5.31 (GPM) and Figure 5.32 (GPU) show the rates of gestures with and without objects over time. The rate of *object-gestures* lies fairly stable throughout the period, although with small peaks at the core of TP#1 and TP#2, and in the first part of TP#3. After the last peak, there is a decline at the rest of TP#3 until the end to the lowest level of the whole period.

*Empty-handed* gestures occur most frequently in the period between TP#1 and TP#2. This is also the period where (empty-handed) pointing gestures are most frequent and other (object-involving) deictic gestures such as GIVE, SHOW, and PLACE were less often performed (as shown in Section 5.6), which may explain a lot of the way these two graphs turn out (recall that gestures with deictic aspects are by far the most common). Indeed, the two curves in the graphs have a similar shape to the curves in Figure 5.21 and Figure 5.22, which describe sub-types of deictic aspects. Yet, alluding to deictic gestures does not explain all aspects of Figure 5.31 and Figure 5.32.

### 5.7.2 Deictic, iconic, and conventionalized aspects in object-gestures

Figure 5.33–5.35 show gestures with deictic aspects, ACTION-BASED iconic aspects, and INDIRECT iconic aspects. Each graph shows the number of gestures with different semiotic aspects per utterance that are performed with an object involved versus without an object involved. There are no instances of object-gestures with conventionalized aspects (i.e., Conv#3 normative aspects) in the data, and therefore no such graph is shown here. Once again it should be noted that the first three data points in the graph that show INDIRECT iconic gestures is wholly due to the “outlier” session from Tea, where a single play-scheme is repeated many times, even though no other children perform this kind of gestures at this time, and not Tea either until a few months later.

The first observation to make from these graphs is that the tendency to be com-

---

<sup>18</sup>One example of such an iconic gesture is a gesture such as pretending to hold a spoon towards someone's mouth to feed them.

## 5.7. GESTURES THAT INVOLVE HANDLING OF OBJECTS

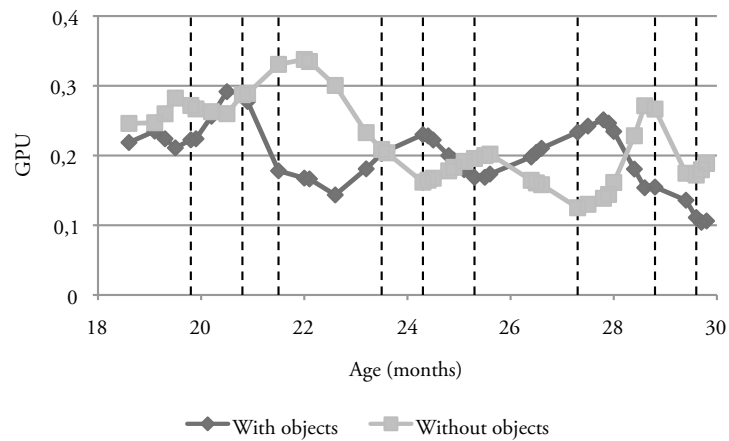


Figure 5.33: Deictic gestures with and without objects (per utterance).

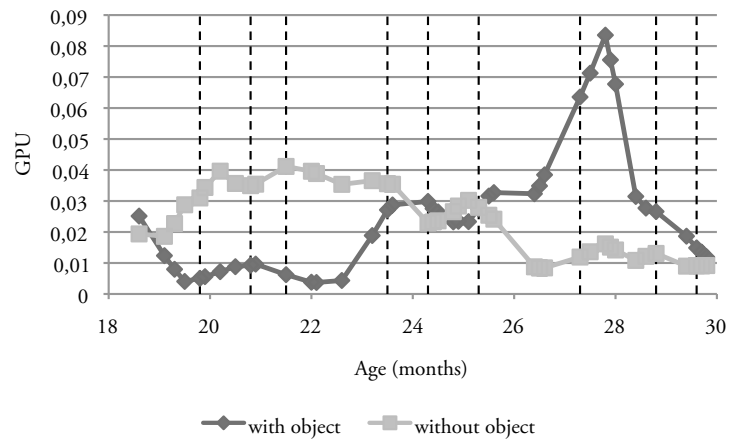


Figure 5.34: ACTION-BASED iconic gestures with and without objects (per utterance).

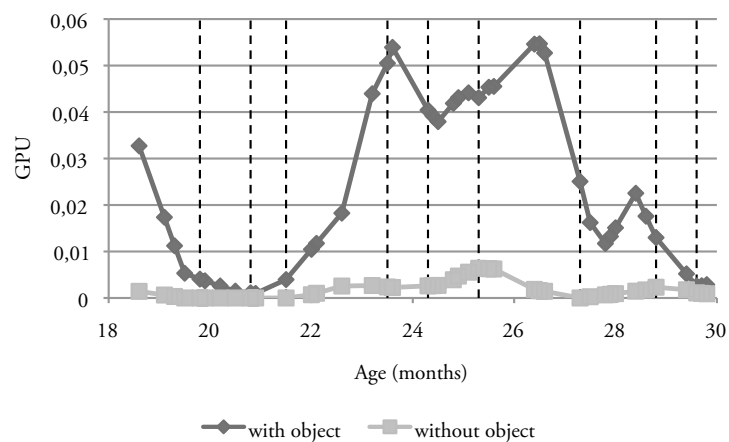


Figure 5.35: INDIRECT iconic gestures with and without objects (per utterance).

bined with speech show rather different patterns for different types of gestures. Object-gestures with deictic aspects tend to be performed both with and without objects throughout the period. Objects-gestures with ACTION-BASED iconic aspects typically appear without objects prior to TP#2 and typically with objects after TP#2, with a marked peak in the first part of TP#3. Finally, object-gestures with INDIRECT iconic aspects almost always appear with objects.<sup>19</sup> In the rare cases where indirect iconic gestures are performed without actively handling an object, they are still performed in a way which is directed to the physical surroundings of the child. That is, all of the instances of indirect iconic gestures that appear in this date are somehow situated in the physical environment in which the child is acting.

One might ask why action-based gestures with objects would tend to appear later in development than action-based gestures without objects, which seems counter to suggestions along the line of the scheme of “from action to gesture to language” (e.g. Clark, 1978; Capirci et al., 2005; Arbib, 2006) and an increasing symbolic distance (i.e., more and more indirect relations between signifier and signified) in development (Werner & Kaplan, 1963). The resolution to the “problem” that is offered here is that the reason for this pattern is due to the fact that there is in fact a new kind of ACTION-BASED gestures appearing after TP#2. The suggestion here is that also ACTION-BASED gestures can be performed from the perspective of *third-person intentionality* (according to the terminology put forward in section 2.2.1). This runs counter to McNeill’s use of the terms Character Viewpoint (C-VPT) and Observer Viewpoint (O-VPT) (e.g. McNeill, 1992, 2005; McNeill et al., 2004). McNeill uses the two terms C-VPT and O-VPT in a way that blurs two separate distinctions into one. First, it refers to the gestural technique which is involved in the articulation of the gesture, i.e., the distinction between ACTION-BASED and INDIRECT iconic gestures. Second, it also refers to a distinction between the psychological perspective that is involved, i.e., acting from a ego-centric first-person perspective or from an alter-centric third-person perspective.<sup>20</sup> I would argue that these two distinctions cannot unproblematically be collapsed into a single distinction and treated as if they were the same distinction. As a consequence of separating these two distinctions, it becomes clear that ACTION-BASED iconic gestures can indeed be performed (in part) from the perspective of third-person intentionality. This will be further substantiated in the discussion in the end of this chapter, as well as

---

<sup>19</sup>Gestures with INDIRECT iconic aspects are mostly constituted by play-acts where the body is used in a way that is not congruent with the “image” that is produced by the gesture, such as moving the legs of a doll to make it appear as if the doll is walking.

<sup>20</sup>See Müller et al. (manuscript) for similar considerations regarding the ambiguity inherent in McNeill’s use of the terms Character Viewpoint and Observer Viewpoint.



in Chapter 7, on the topic of iconicity in the children's gestures. Nonetheless, I agree with McNeill's analysis in so far as it is hard to conceive of INDIRECT iconic gestures as being performed from a first-person perspective, since they by definition involve a kind of dissociation from the way the body ordinarily acts in the world.

### 5.7.3 Summary

Object-gestures are not a marginal phenomenon. 43% of the gestures that were studied here were object-gestures and 57% were empty-handed gestures. The frequency of object-gestures is fairly stable throughout the period, although with peaks at TP#1, TP#2, and TP#3. Empty-handed gestures, show a different pattern. Since INDEX\_FINGER\_POINTING is such a common gesture, the curve for empty-handed gestures has a shape which resembles the curve for INDEX\_FINGER\_POINTING.

Similar to distinctions made previously in this chapter, between more "primitive" and advanced forms of TRACING gestures and the deictic OTHER category, gestures with ACTION\_BASED iconic aspects were also argued to consist in a more "primitive" first-person form, where the child is mainly re-enacting some element of a learned repertoire of actions, and a more advanced third-person form, where the child takes on a more generalized and conventionalized stance in the enactment. The more advanced form of ACTION\_BASED iconic aspects was argued to be the main factor underlying a peak in ACTION\_BASED iconic aspects at the beginning of TP#3. This distinction cannot strictly speaking be deduced from the quantifications shown in the graphs, but was rather used as a possible way to make sense of it. The distinction itself is further discussed in Chapter 7 on iconicity.

INDIRECT iconic aspects appeared in association with TP#2. These gestures were close to exclusively a matter of object-gestures rather than empty-handed gestures. The object-involving variant of INDIRECT iconic gestures may be a precursor to empty-handed INDIRECT iconic gestures, and if that is the case, then it means that the general developmental pattern of "from action to gesture" is not a past issue for children studied here. It could possibly also indicate that the "gesture explosion" in empty-handed gestures somewhere between age 3 and 4, postulated by McNeill (2005, p. 184), may not be as dramatic as it may first seem. The "explosion" effect was obtained in a story-retelling task, presumably with no object manipulation involved. This may make the gap between achievements before and after the "explosion" seem more amplified.

There were no conventionalized gestures (in the sense of emblematic gestures)

that involve objects performed by the children.<sup>21</sup>

## 5.8 Summary and conclusions

The unit of analysis in this chapter has been “gesture”, as defined in Chapter 2 — actions that are either *explicitly communicative* (Comm#3) or *semiotic signs* (Sem#3), or both of these at once (see Section 2.2.5 for more details). The analytic focus has been on changes over time in various aspects of gesture from 18 to 30. The children have been analyzed as a group.

This summary will first summarize some of the general findings. Then there will be a discussion on the three “transition periods” that were identified in the data. Finally, there is a discussion of some implications for the issue of the lower and upper limit of gesture (cf. 2).

### 5.8.1 Overall findings

Some broad and mostly non-longitudinal characteristics in the children’s use of gesture are summarized here. The average gesture rates in all of the sessions were 7.0 gestures per minute (GPM) and 0.5 gestures per multimodal utterance (GPU).<sup>22</sup> The latter measure can be interpreted roughly as meaning one gesture per every second multimodal utterance (on average). The term “multimodal utterances” includes any configuration of gesture and speech: gesture+speech utterances, speech only utterances, and gesture only utterances. The GPU measure provided rates that were much more similar across individual children than the GPM measure. This indicates that the GPU measure measures a more invariant phenomenon than GPM. This is perhaps not surprising, given that GPU normalizes for different degrees of involvement in a communicative activity — which may be due to both individual differences and differences between activities — since the GPU measure is not sensitive to *how often* the child communicates. The GPU measure only measures what the child does once he or she does communicate. This indicates that the GPU measure is a better choice, at least for some applications, than the GPM measure when it comes to measuring the rate of gesture in children, and it also means that child gesture researchers ought to make an informed choice between these measures, rather than simply choosing one of them.

---

<sup>21</sup>Such gestures are rare, but they do exist (Andrén, in press a).

<sup>22</sup>These two measures were described in more detail in Section 4.4.

Gesture-only utterances were rare compared to gesture+speech utterances and speech-only utterances: Only 10% of all gestures were performed without accompanying speech, and only 3.6% of the multi-gesture utterances were performed without speech.<sup>23</sup> In fact, in four of the five children, 0% of the multi-gesture utterances were performed without speech. Hence, *in the overwhelming majority of cases, gesture was used in coordination with speech*. Gesture+speech utterances outnumbered speech-only utterances during the first part of the period up to 24.3 months (the core of TP#2), although with a small margin, and after this point speech-only utterances dominated for the rest of the period.<sup>24</sup> Other researchers have also found gesture-only utterances to be comparatively rare, reporting rates between 2% in Italian children at 27 and 30 months (Stefanini et al., 2009, p. 178) and 19% in Canadian children at 24 months (Nicoladis et al., 1999, p. 518).<sup>25</sup>

Deictic aspects of gestures are much more common than iconic and conventionalized aspects of gestures throughout the whole period. The deictic aspects were around 5 times more common than iconic aspects, and around 7 times more common than conventionalized aspects. After an initial increase in deictic aspects of gesture at TP#1, deictic aspects then decrease throughout the rest of the studied period. Iconic and conventionalized aspects of gestures are rather increasing over time, apart from an eventual drop around TP#3. Although the connection between utterance length and occurrence of different semiotic aspects in gesture were not tested directly here, this overall pattern is in line with results from Nicoladis (2002), who found that preschoolers' utterances with iconic gestures are longer, on average, than those with pointing or no gestures at all.

The overall proportions of the three different semiotic aspects are similar to what has been found in other studies of more or less overlapping age ranges (e.g. Nicoladis et al., 1999, p. 524; Goldin-Meadow & Morford, 1990, p. 255). Yet there seems to be slightly more gestures with iconic aspects here than in the study by Nicoladis et al. (1999) of five Canadian children between 20 and 42 months, where it was found that around 5% of the gestures produced by the children were iconics and beats.<sup>26</sup> This slight difference may be due to the fact that gestures that involved han-

---

<sup>23</sup>Multi-gesture utterances refers to utterances that contain two (or more) gestures performed together either simultaneously or in sequence.

<sup>24</sup>See Section 5.2 for a definition of the three Transition Periods discussed here (TP#1, TP#2, and TP#3).

<sup>25</sup>As a comparison, it may be noted that Özçalışkan & Goldin-Meadow (2005a) found that 90% of the gestures that American *parents* produced when communicating with their children co-occurred with speech, which is the same rate that has also been reported for American adults McNeill (1992, p. 23) (in story-telling contexts).

<sup>26</sup>Beat gestures are typically defined as rhythmical movements that serve to emphasize or mark

dling of objects are not excluded from the scope of gesture studies in this thesis, and the analysis in this chapter showed that a substantial amount of the gestures with iconic aspects did include handling of objects. This was especially true for gestures with INDIRECT iconic aspects, where almost all instances involved handling of objects. On the other hand, Goldin-Meadow & Morford (1990) found quite large differences between three individual children in the proportion of iconic gestures, ranging from 3% to 19% (average 10%), which means that the exact ratios found in the various studies should probably not be interpreted too literally. Especially since none of these studies included more than 5 children. The overall pattern seems well confirmed by now though.

Among the gestures with deictic aspects, INDEX\_FINGER\_POINTING gestures were about as common as the OTHER category consisting of deictic gestures such as SHOW, OFFER, and PLACE. OTHER\_POINTING was much more rare, and INDEX\_FINGER\_POINTING was 6.5 times more common than OTHER\_POINTING. Hence, the vast majority of children's pointing gestures are performed with the index finger pointing hand-shape. Among the gestures with iconic aspects, the majority were ACTION\_BASED. ACTION\_BASED aspects were about twice as common as gestures with INDIRECT iconic aspects, and INDIRECT iconic aspects were about three times as common as gestures with TRACING aspects. This dominance of ACTION\_BASED iconic gestures also seems to continue after the period studied here. In a study of children between 27–90 months, Stefanini et al. (2009, p. 184) found that “action gestures” (i.e., ACTION\_BASED) were much more frequent than other kinds of iconic gestures. In the analysis of conventionalized aspects, not all emblematic gestures were presented, but the analysis was instead focused on the fact that 75% of the instances of emblematic gestures were constituted by HEAD\_SHAKE and NODDING. Most other specific emblematic gestures occur much less often, and in many cases a given emblematic gesture appears only in one of the children, which makes quantifications less informative in this case. HEAD\_SHAKE and NODDING, however, occur in all children with substantial frequency.

It should be remembered that only emblem-like gestures with a *normative* specification of form were treated as conventionalized gestures in this chapter and if the criterion for a gesture to be coded as “conventionalized” would instead also have included *typified* forms of conventionality, a much larger share of gestures would have been coded as having a conventionalized aspect.<sup>27</sup> Hence, the distinctions be-

---

up some particular syllable, word, or utterance. Beat gestures are discussed in more detail in further below.

<sup>27</sup>It could be debated whether it was correct not to treat index finger pointing gestures as conven-

tween different levels of conventionality, as made in Chapter 2, may help to resolve the apparent opposition between different accounts on the role of conventionality in different researchers on child gesture, where the issue of conventionalization is often treated as a dichotomous either-or issue. Some researchers claim that the majority of children's gestures are conventionalized in the sense that they involve learning from people in the surroundings (e.g. Caselli, 1990; Guidetti, 2002; Guidetti & Nicoladis, 2008), whereas most other child gesture researchers tend to emphasize conventionality only to a lesser degree.<sup>28</sup> However, the distinctions between different levels of conventionality were not part of the coding of the gestures used for the quantifications in this chapter, and these distinctions will play a larger role in later chapters in this thesis.

There was a comparison between the three semiotic aspects (deictic, iconic, and conventionalized) regarding the extent to which they tended to be combined with speech. This comparison showed that contrary to the popular idea that iconic gestures are more integrated with speech (i.e., "co-speech gesture") than conventionalized gestures (McNeill, 2000, 2005), it was more common for iconic gestures to be used autonomously from speech than it was for conventionalized gestures to be used without speech, at least during the period studied here.<sup>29</sup> 24.0% of the gestures with iconic aspects were used without speech, and 16.6% of the gestures with conventionalized aspects were used without speech. Gestures with deictic aspects were used least often without speech (7.6%), which indicates that the deictic function of these gestures are especially strongly integrated with spoken language. Indeed, even though most uses of gesture in non-signers are not language-like "enough" to form a

---

tionalized on the level of normativity. Adults may certainly use relatively specific forms of pointing to point out targets in ways that seem sensitive to normative constraints, at least in some cultures (Kendon & Versante, 2003; Wilkins, 2003). Furthermore, blind children do not seem to make spontaneous use of the index-finger pointing hand-shape in their pointing gestures (Junefelt, 1987; Iverson, 1998). Hence, to usage of the index-finger pointing hand-shape seems to require visual access to other people's gestures in order to emerge. Yet there is quite some variation in the extent that the pointing gestures in the Swedish data studied here include a clearly articulated index finger pointing hand-shape, and at least this hand-shape does not seem to be normatively constrained in the sense of being obligatory even though it has a standardized (i.e., typified) form (cf. McNeill, 2005, p. 12). That is to say, in the rare case were children point in a less typical way, such as pointing with the little finger, this appears slightly unusual rather than plainly wrong. The issue of specific hand-shapes employed in pointing in the children studied here is treated in Chapter 6.

<sup>28</sup>Guidetti also include asymmetric social learning such as *ritualization* in her use of the term "conventionality". Hence, in her work, conventionalization is equated with social learning in a wide sense and she does not require conventionalization in the sense of normativity (as in emblems) in order to consider a gesture conventional.

<sup>29</sup>McNeill's claims in this regard seem to be mostly concerned with the gestures of adults though, and it is not clear whether he would or would not apply this conceptual scheme also to the gestures of children.

language when considered on their own, one could argue, as Kendon (2008) does, that gesture is (sometimes) part of language, multimodally conceived, in the specific sense that there are aspects of spoken language that are designed to go together with gesture. In some cases, such as in the use of some of the conventionalized gestures, this integration may be even more stabilized and conventionalized. Gestures such as INDEX\_FINGER\_POINTING, HEAD\_SHAKE and NODDING are indeed used in quite specific ways in a wide range of activity types, and it is also striking that these gestures are articulated with the head, which is especially visible to people who look towards each other when they *speak*. These three gestures were the three most common specific (co-speech) gestures to be used in three Swedish children (Bella, Harry, and Tea, also part of the data used in this thesis) and three Thai children between 18–28 months in a study by Zlatev & Andrén (2009).

Another issue that was investigated was the tendency for gestures to be performed with and without handling of physical objects. It was found that 57% of the gestures were performed in an empty-handed manner and 43% of the gestures were object-gestures, involving handling of an object. Object-gestures included deictic gestures such as SHOW, OFFER, and PLACE, but also iconic performances with objects involved. Among gestures with ACTION\_BASED iconic aspects, the proportion of them that did, and did not, involve objects were about 50/50. INDIRECT iconic aspects were almost always performed with objects, as mentioned above. There were no emblematic object-gestures performed by the children, even though such gestures are both possible and existing in adults (Andrén, in press a), although with low frequency and perhaps most often in activities which are particularly ritualized.

When it comes to spoken language, the studied period covers the development in mean length of utterance from around 1.3 at the start of the period, up to 3.0 at the end of the period. The average MLU for the whole period was 2.0. However, even though the spoken utterances become longer and longer throughout the period studied here, there was no corresponding increase in the number of gestures per utterance, on average. Still it was also found that during the latter half of the studied period there was a much higher ratio of gestures in multi-word utterances than in one-word utterances.

### 5.8.2 Three transition periods: A developmental trajectory

In the analysis presented in this chapter three periods were identified, on an inductive basis. During these three periods there were a remarkable number of changes

in many different measures going on at the same time. These three periods were referred to as Transition Period #1, #2, and #3. This present section is devoted to a discussion of each of these in the order they appear in development. This chapter differs from previous research on gestures in children between 18–30 months in that it is based on more frequently recorded sessions, i.e., a higher temporal resolution. This means that the curves in the graphs that have been shown are more detailed, and also more smooth (since the gap between each data point is smaller). The use of the “true age average” graphs, as explained in Section 4.5.3, also helps preserving the temporal detail of the recorded sessions. Hence, the level of detail in the longitudinal curves that have been presented here is itself one of the main contributions of this chapter.

Also, as shown in the first chapter of this thesis (see Figure 1.1), the major bulk of studies that have been carried out on children’s gestures have been concerned with younger children, and after 20–24 months there are substantially fewer studies in existence. When it comes to Swedish children specifically, I am only aware of one study that deals with children’s gestures in the age span that is covered here. That study is an investigation of children’s emotional expressions specifically (Gerholm, 2007), and more general work on Swedish children’s gestures is therefore non-existent when it comes to the period between 18–30 months. Therefore, another major contribution of the present chapter is to tell the “next chapter” of the developmental story.

As argued in the beginning of this chapter, the exactness of this specification should not be mistaken for a claim that the developmental “schedule” looks the same for all children. The exactness of the specification is simply a reflection of what was found in the data when the children are treated as an averaged group.

### **Transition Period #1**

Transition Period #1 lasts from 19.8 to 21.5 months, and the “core” occurs at 20.8 months. During this period, a number of things happen: In the first part of the period, i.e., between 19.8 and 20.8 months, there is a sudden increase in the number of *gestures per minute* (GPM). This consists in an increase in co-speech gestures rather than autonomous gestures, and it is primarily deictic aspects of gestures that underlie this increase. The increase is around 100% increase for one-word utterances that are coordinated with gesture, and around 50% for multi-word utterances that are coordinated with gesture.

The pattern for *gestures per multimodal utterance* (GPU) at TP#1 is quite dif-

ferent. When looking at the rate of children's gestures from this perspective, the "curve" is instead strikingly flat all the way from before TP#1 up to the core of TP#2, hovering between a GPU of 0.54 and 0.60.<sup>30</sup>

It is interesting that although GPM and GPU are both measures of gesture frequency, each of them provide a quite different image of what goes on at TP#1. On a methodological level, this is yet another reason why child gesture researchers should make an informed choice when they select one of these measures, or variants of them (the first reason was stated in the previous section). These two measures do indeed show different things. The different images that arises from GPM and GPU are also interesting since the increase in GPM at TP#1 is reminiscent of the "late integration scenario" whereas the flat character of the GPU curve is rather reminiscent of, or at least more in line with, the "early integration scenario".<sup>31</sup>

To repeat then, the increase in GPM at the first part of TP#1, which mostly consisted in an increase of gesture+speech utterances, rather than speech-only or gesture-only utterances, could be interpreted as an indication of a kind of "convergence" between gesture and speech. When considering GPU, however, there are no real indications of such a convergence. Yet, it is of course possible that the "convergence point" suggested by Butcher & Goldin-Meadow (2000) has already taken place in the children studied here, before 18 months. Another study that has argued for the late integration scenario is a study of American children by Rowe et al. (2008, p. 190), who report a decrease from an almost complete dominance of gesture-only utterances at 14 months (around 95%), to around 75% at 18 months, to around 40% at 22 months. Between 26–34 months, gesture only utterances were around 20%. No such initial high amount of gesture-only utterances were found in the five children studied in this thesis (gesture-only utterances are never above 10% throughout the whole period), and by implication, no such sharp decrease in gesture-only utterances was found either.

As it turns out, most of the studies that argue in favor of an "early integration" scenario include (non-word) vocalizations in addition to words in the definition of "speech". Also, a few of the studies rather concern motor activity more generally and not just gesture specifically. Most of the studies that argue in favor of a "late integration" scenario, such as the study by Butcher and Goldin-Meadow, focus on gesture+speech in the sense of gesture+*words*, rather than gesture+*words or vocal-*

---

<sup>30</sup>There is a slight increase in gestures per utterance with deictic and iconic aspects at TP#1, but nowhere near as marked as when the rate is measured in terms of GPM.

<sup>31</sup>The two "scenarios" are described in the beginning of Section 5.4.



*izations*.<sup>32</sup>

The debate may therefore be somewhat misguided, and instead of choosing either one *or* the other position, a more reasonable position in this debate may be to accept that there seems to be some truth in both scenarios. More specifically, there is indeed cross-modal integration between non-word vocalization and manual movement early in development, *but* this does not stand in stark contrast to findings that show an increase in gesture+*word* integration during the latter part of children's second year of life, i.e., more complex forms of semantic integration — especially if the frequency is measured as gestures per minute. The finding by Rowe et al. (2008) of an initial high level of gesture-only utterances finds no support by the data studied in this thesis though. Each of the two positions therefore provide a somewhat skewed account of the integration between the modalities when they are stated as opposed positions, since they do not in fact seem to dispute the exact same claim.

Indeed, in an older study Goldin-Meadow & Morford (1990, p. 254) report that 82%, 35%, and 80% of the “gesture-only” utterances of the three studied children's gestures were in fact accompanied by non-word vocalizations (e.g., point at bubbles + ”uh”). It should also be noted that the way utterances were transcribed in the data analyzed in this thesis there was no distinction made between vocalizations and words (see Section 4.2). Therefore, the big difference between the results obtained here (12% gesture-only utterances at 18 months) and the results of Rowe et al. (2008) (75% gesture-only utterances at 18 months) may only be apparent, because Rowe et al. only included gesture+word utterances in their operationalized definition of a gesture+speech combination.<sup>33</sup>

The discussion now turns to the use of different semiotic aspects around TP#1. Deictic aspects of gesture reaches its highest rate at the core of TP#1 compared to

---

<sup>32</sup>Butcher and Goldin-Meadow do in fact include vocalizations in that study, and investigate whether gestures are combined with vocalizations, but the criteria are perhaps overly strict; gesture and speech must occur within one video frame (1/30 seconds in the NTSC videos used in the US) to count as a temporally coordinated gesture+speech utterance, which means that gestures+vocalization utterances with just a slightly less temporal coordination are counted as “non-combinations”. This is a very strict criterion for a measure of the motor capabilities of children at these ages. One may wonder how many gestures and words that occurred 2/30 seconds from each other, but were not considered to be co-occurring. More generally, it is indeed a complication in this debate that the criteria for regarding the two modalities as “co-occurring” vary a lot between different studies.

<sup>33</sup>To be clear, their definition of a “word” did include onomatopoeic “words” such as “woof-woof” as well as word-like expressions like “uh-oh”. These are both essentially word-like in the sense that they are conventionalized (even on a normative level). However, vocalizations that were not recognizable as an instance of a particular conventionalized type were not included.

other points in time throughout the period from 18–30 months. Iconic aspects of gesture are more or less restricted to ACTION\_BASED gestures during TP#1. Almost 100% of gestures with conventionalized aspects (on the level of normativity) are coordinated with speech at TP#1, but not at other times of the studied period. TP#1 also marks the onset of a period which is dominated by empty-handed gesture, which lasts to TP#2. This is largely due to the abundance of pointing gestures during the period between TP#1 and TP#2.

Overall, TP#1 forms the climax of a period where *communication has a referential focus*, as suggested by the peak in the number of gestures with deictic aspects (especially pointing gestures). Although the children do produce a fair amount of multi-word utterances, they still produce more than twice as many one-word utterances, which tend to be less propositional in nature.

The increase in deictic gestures at TP#1 may be viewed in light of the findings of Bates et al. (1988, p. 265) according to which there is an “expansion of the open class as a whole at 20 months”, i.e., an increase primarily in nouns and verbs at TP#1. Since there is evidence that interactive sequences of joint attention are strongly related to vocabulary growth (Tomasello & Todd, 1983), and since deictic gestures almost per definition play a central role in the regulation of joint attention, it is not surprising to find an increase of pointing gestures precisely at a period of intense expansion of the open class. Parents often respond to children’s pointing by naming what they interpret the child to be pointing to (Bruner, 1978a; Hannan, 1992; Marcos, 1991; Marcos et al., 2003; Kishimoto et al., 2007), and decreasingly so after 20 months (Hannan, 1992).

Capirci et al. (2005) investigated changes in the number of word types and word tokens at different ages in three Italian children, and found a marked increase in word *types* around TP#1, i.e., an expansion of spoken language vocabulary, but no comparable increase in gesture types. Children are also known to acquire words more and more rapidly over time, typically towards the end of the second year which is precisely when TP#1 takes place, and this gives rise to an accelerating curve of vocabulary growth (Bates et al., 1994; Tomasello, 2003, p. 50; Ganger & Brent, 2004).<sup>34</sup> This more intense word learning period may well be associated with the increase of pointing gestures around TP#1. In the same vein, Langacker (2004, p.

---

<sup>34</sup>There is a debate as to whether there is a sudden increase in the rate of word learning, called “the vocabulary spurt”, or whether it is rather a matter of a gradual acceleration of the word learning rate with no clear “onset” at a specific point in time. The conclusion at present seems to be that some children show a sudden increase in the word learning rate (Anisfeld et al., 1998), i.e., a “vocabulary spurt”, whereas most children do not (Ganger & Brent, 2004). In either case, it is undisputed that the rate of word learning accelerates over time.

96) speaks of pointing as the archetype of nominal grounding:

The archetype I propose for nominal grounding is a physical pointing gesture. The schema resides in the interlocutors directing their attention to the same conceived entity [...]. Since pointing is an action intended to achieve precisely this, the schema is clearly immanent in the archetype. Among the nominal grounding elements, therefore, the most typical — in the sense of lying closest to the archetypal origin (though probably not in terms of frequency) — are demonstratives accompanied by physical pointing.

Iverson et al. (1994, p. 36) also found an increase in deictic gestures around 20 months (TP#1) in a study of American children between 16 and 20 months, but an overall decrease in the use of gestures of other kinds during the same period. Blake et al. (1992) also found an increase in the use of deictic gestures between 9 and 22 months in a study of Canadian children. Guidetti (2002) studied French children at 16, 24, and 36 months, and found a particularly strong increase in the rate of pointing from 16 months to 24 months.

### Transition Period #2

Transition Period #1 lasts from 23.5 to 25.3 months, and the “core” occurs at 24.3 months. In this transition period, the GPM and GPU measures are more similar than they were during TP#1. There is therefore no separate treatment of these rate measures, and both of them will be referred to simply as “frequency” or “rate”.

At the second part of TP#2, i.e., from the core of TP#2 at 24.3 months to the end of TP#2 at 25.3, there is a marked decrease in the frequency of gestures and also a temporary decrease in the otherwise steadily increasing MLU. There is also an increase in speech only utterances and a decrease in gesture+speech utterances. At the core of TP#2, MLU reaches 2, and there is also a peak in multi-gesture utterances, and related to this, also a peak in multi-word utterances that are coordinated with multi-gestures. Curiously, all these things also hold true for TP#3, except that MLU reaches 3 (almost) instead of 2. This issue will be discussed further in relation to the discussion of TP#3.

There are also a number of other things happening at TP#2. At the core of TP#2, speech only utterances become more common than gesture+speech utterances for the first time, and this remains true for the rest of the studied period. At the beginning of TP#2, the proportion of multi-word utterances that are coordinated with

gesture becomes larger, for the first time, than the proportion of one-word utterances that are coordinated with gesture. At the core of TP#2, there is also a major peak in the proportion of multi-word utterances that are coordinated with multi-gesture performances. The end of TP#2 marks the onset of a period where the proportion of multi-word utterances that are coordinated with gesture suddenly starts increasing intensely (up to the core of TP#3).

All of these things taken together imply that TP#2 is indeed a period of many transitions. There are substantial changes in how gesture and speech tend to be used together, and TP#2 seems to form the starting point for a more clearly speech-dominated kind of communication. To be sure, before TP#1, speech is already dominant in the sense that gesture-only utterances are comparatively rare, but after TP#1, the ability to use speech in ways that are more independent from gesture increases. It is striking that this happens precisely at the time when an MLU of 2 is the rule rather than the exception. Plunkett & Strömqvist (1992) finds an onset of syntactic negation, as opposed to discourse negation, at TP#2 in two Swedish and two Danish children. There is also an increased use of *den* and *det* (which are used both as deictic/anaphoric pronouns and determiners of nouns in noun phrases) in combination with other words at TP#2. Interestingly, *den/det* are extremely frequent in coordination with pointing, which is the dominant gesture used in the period between TP#1 and TP#2. In fact, Plunkett and Strömqvist also finds a little peak in the use of *den/det* as a one-word utterance around TP#1, where the analysis in the present chapter found an overall peak in pointing gestures. These findings adds some more flesh to the idea that the onset of gesture+speech combinations where the gesture and the speech express different meanings (like a “two-word” utterance) precede the onset of two-word utterances in speech (e.g. Capirci et al., 1996; Butcher & Goldin-Meadow, 2000). One may speculate that the reason that cross-modal combinations precede unimodal combinations may be because it also implies a transition from simultaneous expression to sequential expression, and sequential expression may be more difficult for young children (cf. the linearization problem, Levelt, 1980, 1981).

When considering the use of different semiotic aspects, there are also a few things going on precisely at TP#2. For example, there is a peak in OTHER aspects of deictic gestures, such as SHOW, OFFER, and PLACE. Most interesting is perhaps the changes in iconic gestures at TP#2. At the onset of TP#2 there is a sudden emergence of gestures with INDIRECT iconic aspects, and with few exceptions, these are object-gestures. There is also a sudden increase in ACTION\_BASED gestures that involve objects. These findings imply that there is a change in the role of objects

in the children's expressive actions at TP#2. In the analysis, it was suggested that ACTION\_BASED gestures are not necessarily always a matter of first-person expression (as McNeill, 1992, 2005 seems to assume), but that they can also be performed from a third-person intentional perspective.<sup>35</sup> This issue will also be discussed in Chapter 7.

Support for this interpretation can be found in research on *symbolic play*. Research along those lines have shown that the ability for *role-reversal* in play appears around TP#2 (McCune-Nicolich, 1981). At the age of TP#2, children start playing games which involve taking a specific generalized role, such as “playing doctor” or “serving food”, and so forth, in a manner that “anyone” (or at least more than one person) would do in a generalized, typified, and generic way. This involves access to generalized and typified knowledge of ways of using objects which are not invented in the situation, but are rather a matter of more stable socio-cultural practices. Indeed, the many of the ACTION-BASED gestures in the period that follows after TP#2 are ACTION-BASED gestures where the child takes a more generic third-person intentional stance on their own actions and bodies, performing generalized and conventionalized actions, as if taking the role of other actors, or generalized actors in this manner. This development also works the other way around, so that the same generalized knowledge about typified ways of acting can be projected onto others as well. At TP#2 this is not only restricted to interpreting others' actions in this way, but children are also able to make use of this knowledge in a more productive and active way. As argued by McCune-Nicolich (1981, p. 789):<sup>36</sup>

Following the initial development of symbolic games involving dolls, several investigators have observed that by about 24 months [i.e., TP#2], children have begun to give the doll a more “active” role in the game. That is, the doll appears to have its own potential for action rather than being merely the passive recipient of the child's pretend schemes.

It is also at TP#2 that children are able to imitate sequences of action in correct order (O'Connell & Gerard, 1985), and this aptness at imitation also implies an ability to conceive of the action of the Other as a *type* (rather than a unique token) which is “distinct from the body of the other in its specificity, so that it can be repeated by the self” (Sonesson, 2007, p. 116). At TP#2 children are not only able to act in ways that are more detached from the ego, but they also learn to detach them-

---

<sup>35</sup>In fact, in Zlatev & Andr en (2009) we defined ACTION\_BASED gestures along McNeilleian lines too.

<sup>36</sup>Bretherton et al. (1984, p. 285) argue in a similar way.

selves from the now. For example, Christensen (2003) has shown that (Swedish) children master the use of past tense in verbal inflection around TP#2, which is another indication of the emergence of an ability to communicate about things that are more detached from the immediate situation in which the child is situated. As argued by Schutz (1945), the farther away from the here and now that one gets in communication and thought, the greater is the reliance on typification. For example when someone writes a letter to a friend, there is no direct feedback from the friend, and as a result, the person who writes the letter will treat his friend in a more typified way, based on the somewhat stereotypical memory of his friend that will have to take the place of the immediate feedback from the friend. The finding in this chapter that speech becomes more dominant in the communication at the core of TP#2 and onwards fits well with this scenario, since spoken language is the “typifying medium *par excellence*” (Schutz, 1953, p. 10). Furthermore, from this perspective, the finding that gestures with INDIRECT iconic aspects emerge around the time of TP#2 seems logical, since these gestures involve precisely the kind of more abstract and/or generalized relation to one’s own actions and body that is required for a child to be able to act from the point of view of a generalized third-person intentionality. It should be noted that INDIRECT gestures are at best indirectly based on the experience of acting practically in the world, since per definition they do not correspond to such ways of acting in the world. Instead, INDIRECT gestures (and ACTION\_BASED gestures from a third-person perspective) require an ability to map concepts onto bodily movements in a more abstract and/or generalized way.

### Transition Period #3

The proposed Transition Period #3 lasts from 27.3 to 29.6 months, and the “core” occurs at 28.8 months. In the discussion of TP#2, it was noted that there were a number of similar events taking place in TP#2 and TP#3, all related to changes in the relation between gesture and speech. These will not be repeated here, and the discussion will proceed from there.

At TP#3 multi-word utterances become more common than one-word utterances for the first time. In particular, there is a marked increase in three- and four-word utterances that peak at the core of TP#3. At same time, there is also a huge peak in the proportion of multi-word utterances that contain gesture (85%), which is the “end result” of a increase in this proportion during the period between TP#2 and TP#3. By comparison, the proportion of one-word utterances with gesture at the core of TP#3 is only 20%. At the second part of TP#3, speech only utterances

reaches the highest level during the period studied here, and gesture+speech utterances reaches the lowest level.

At the second part of TP#3 there was also a temporary decline in MLU. This should not necessarily be interpreted as an indication that the complexity of the children's speech decreases. The reason for this is that around the time of TP#3 (or later), the grammatical aspects of children's speech become more truly *productive* and rule-like (Brown, 1973; Peters, 1983; Bates et al., 1988; Shore et al., 1990; Tomasello, 2003), as opposed to being (mostly) a matter of rote-learned constructions with a quite restricted degree of flexibility and generality. For example, around 28 months or later, children's words generally become multi-morphemic in a more true sense, so that various inflection suffixes are used in an over-generalized way also with words that should not have this suffix according to the norms of the language. An example of this is when children use regular past-tense forms for verbs that have irregular past-tense forms. Indeed, it is rare for children to commit such grammatical overgeneralization errors before 30 months, or even up to 36 months in some cases (Tomasello, 2001, p. 69; see also Pinker, 1989). All in all, this suggests that a more true form of grammatical productivity has typically not yet entered into the picture before TP#3. More evidence along these lines comes from Christensen (2003) who found that the onset of what she calls "the elaborated stage" of the use of tense begins at TP#3, characterized by the onset of a range of more complex forms of tense specification. Also, Strömquist (1997, p. 67) found a peak in the frequency of inflection morphemes precisely at TP#3 in a study of one Swedish child. Relevant in this regard is that Plunkett & Strömquist (1992, p. 542) found that the acquisition pattern for inflectional morphology in Scandinavian children is similar to what is found in the acquisition of other languages with a similar level of complexity in the system of grammatical markers, which implies that these results may be generalizable to other languages with similar levels of complexity in this regard.<sup>37</sup>

On basis of findings such as those that have been listed here, one may suspect that the temporary decline of multi-word utterances and MLU at the second part of TP#3 that was found in this chapter may very well be a reflection of the emergence of a more truly grammatically organized way of using word roots and inflection, with the complexity that this implies. Nevertheless, it should also be noted that Plunkett & Strömquist (1992) find that different grammatical constructions appear at different times, and it would be a huge simplification to say that "gram-

---

<sup>37</sup>Plunkett & Strömquist (1992) classify Scandinavian languages as being toward the non-complex end of the spectrum when it comes to the system of grammatical markers that is used, when compared to other languages around the world.

mar” emerges at one specific point in time. Some aspects of adult-like speech does not occur until much later (e.g. Josefsson et al., 2003). There may also be sporadic indications of productive use of certain grammatical elements quite some time before TP#3, and the claim is not that it happens all of a sudden. Nonetheless, the proposal is rather that around TP#3 there is, on average, something like a climax that marks the onset of a more substantial mastery of inflection and other syntactical processes.

Children between 20 (TP#1) and 28 (TP#3) months may also use word+word combinations, but then typically of a more “telegraphic” kind, where (mostly) content words such as nouns, verbs, and adjectives are combined. These combinations typically lack most of the normatively required grammatical morphemes that would have been required for a clause to be correct according to the norms of the language. In particular, what seems to be lacking before TP#3 is proper use of closed-class morphology and generalized use of closed-class words, as shown by Bates et al. (1988). Functional closed-class words serve to bind the constituents of more complex utterances in systematic ways. Furthermore, the “telegraphic” utterances are typically produced with a relatively low degree of flexibility. The grammatical (or rather proto-grammatical) competence before TP#3 may be “best characterized as simply an inventory of independent verb island constructions that pair a scene of experience and an item-based construction, with no structural relationships among these constructional islands” (Tomasello, 2003, p. 121; see also Tomasello, 1992).

Hence, the intense increase in the proportion of multi-word utterances that contain gesture in the period between TP#2 (around 40%) and the core of TP#3 (around 85%) indicates that gesture plays an important role in the process that leads up to the more productive use of speech that takes off for real around TP#3 (cf. Özçalışkan & Goldin-Meadow, 2009). Indeed, in the period leading up to TP#3, there is an increase in, and an eventual peak in all kinds of gestures (except NODDING and HEAD\_SHAKE) in association with, or just before, the core of TP#3. Then, directly after the core of TP#3, there is a marked decrease in the rate of all kinds of gestures (except in NODDING and HEAD\_SHAKE), whereas the complexity of speech now sets off in a more productive mode of operation. Another piece of evidence in line with these claims comes from Rodrigo et al. (2004, p. 81) who found that combinations of pointing+content words become dominant at TP#3 (30 months) whereas at TP#2 (24 months) pointing+deictic words were the dominant kind of pointing+speech combination. That is, at TP#2 pointing gestures primarily occur together with a much more restricted set of words, such as “there”, “this” and so forth, in a more recurrent and “construction”-like manner (see also Chapter 8 on



multimodal lexemes). At TP#3, however, pointing gestures are produced together with a much more varied range of word types, such as for example, and increasing use of pointing gestures together with verbs, which is largely absent in the period between TP#1 and TP#2 (cf. Andrén & Zlatev, 2007).

Overall, the changes surrounding TP#3 seems to be organized around the shift to a more productive mode of speaking. Therefore, it is interesting to ask how the development of multi-gesture utterances and multi-word utterances may compare to children's ability to perform and imitate conventionalized sequences of action more generally. In an experimental setting, O'Connell & Gerard (1985) found that at TP#3 (28 months) children were not only able to imitate sequences of action, but they also tended to correct the order in cases where the sequence to be imitated violated conventional orders of action. Violations of conventional orders could be to perform a familiar sequence of action in the reversed order ("dry the bear, soap him, put him in the tub") or combining unrelated actions into a "scrambled" arbitrary sequence ("cover the bear with a blanket, the bear pays money, wipe his mouth").<sup>38</sup> At TP#2 (24 months) children were able to imitate sequences of action that did not violate typical sequences of action, but rarely corrected the sequences if they violated conventionalized orders, implying that the *normative* aspect of sequences of action was typically absent before TP#3 months.<sup>39</sup> At TP#1 (20 months) children appeared to be able to discriminate between meaningful versus reversed or scrambled sequences, but could not reproduce the actions in this order. Recall that Figure 5.9 showed two peaks in multi-gesture utterances around TP#2 and TP#3, but only a weak peak at TP#1. There seems to be an overall improvement in mastering combining familiar units of action into coherent action sequences that starts around TP#2, and continues to improve up to TP#3, where there is a generalized understanding of sequences of actions that also enables the children to *correct* the sequences whenever they violate normative conventions. This is indeed similar to what is required to master an accountable use of the conventionalized norms of grammar and spoken language.

---

<sup>38</sup>O'Connell & Gerard (1985, p. 673) are not always clear about whether the correctness involved is a matter of "things that would work" (similar to what Garfinkel, 2002 calls natural accountability) or a matter of "things that should be in a certain way" according to (partly arbitrary) convention and customary practices. From the examples that O'Connell and Gerard provide, it seems to be a bit of both.

<sup>39</sup>The children may well be able to recognize normative correctness versus incorrectness in single actions before this. The normativity discussed here only concerns sequences of action, i.e., normativity on the activity level, rather than the action level.

### The nature of the transitions

What may be the reason that so many things happen at three quite delimited periods? The question is perhaps especially relevant to TP#2 and TP#3 since they share many similarities? The first possibility that must be considered is that the measures themselves might somehow be logically/mathematically dependent on each other. If that would be the case, it would not only be expected that they should vary in a determined manner with respect to each other simply due to the way they are defined, but they actually could not do otherwise. This is hardly feasible though, because many of the measures reported here are in fact *not* logically dependent on each other. For example, an increase in the proportion of gesture+speech utterances that involve more than one gesture does not implicate logically that there must be some particular corresponding change (or non-change) in the frequency of gesture+speech utterances (either per minute or per utterance). In a similar vein, an increase in speech only utterances does not implicate logically that the MLU measure should be expected to increase, decrease, or remain the same. Yet there are changes in different measures like this around the same transitional period.

A second possibility is that it is only a co-incidence that a lot of things happen to occur around the same time. Although this possibility cannot be completely ruled out, this interpretation is not feasible either, since the likelihood for simultaneous changes decreases exponentially for every new (logically) unrelated measure.

The third main possibility is that the three transitional periods found here do in fact reflect empirically valid phenomena, at least with respect to the five children involved in this study. Nevertheless, it is worth repeating that in order to arrive at more generalizable findings, further research is required with larger numbers of children as well as, for example, studies from children from other cultures. For the same reasons, the relatively “exact” ages that have been specified as starting points, cores, and endpoints of the transition periods should not be interpreted too literally due to the considerable individual differences that are known to obtain between children (Bates et al., 1988). The interpretations that have been offered here are merely suggestions on basis of what was found in the five children studied here. The focus of the analysis has been to provide a broad overview of many different measures at once, rather than going in great detail into specific issues. Further research will have to clarify to which extent the findings and the interpretations of these findings that have been presented in this chapter may be generalizable.

On a higher level, one may briefly consider the nature of the developmental changes involved here, that have been discussed in terms of transition periods Within

developmental psychology and theories on language development there is a long standing controversy between those that argue for stage based models of development (e.g. Piaget, 1962 [1946]), with its associated discontinuity between the various stages, and those that argue for gradual development (e.g. Werner & Kaplan, 1963). From the findings in this chapter it should be relatively clear that the children's communicative development is neither completely linear, without qualitative transformations or convergences, nor is it a matter of completely discrete stages with no continuity between different periods of development. Stating that development is *either* continuous *or* discontinuous is a too much simplified way of stating the problem (cf. Lock, 1997; Zlatev & Andr n, 2009). As argued by Aitchison (1998, p. 27), such dichotomies tend to “naturally fade away as researchers unravel the complexity of the issues involved”.<sup>40</sup> The analysis has shown that there are periods with especially intense changes, as well as periods in between with more gradual changes. Perhaps a better metaphor for the pace and structure of development is one of partially overlapping “waves” (Siegler, 1996; see also Zlatev, 2003).

### 5.8.3 Findings relating to the upper and lower limit of gesture

This section concerns the findings in this chapter that are of relevance to the issue(s) of the upper and lower limit of gesture. With regard to the issue of the lower limit, it is mainly the question of object-gestures that have been treated in this chapter. Object-gestures are clearly not a “past issue” for children of these ages, since roughly half of the gestures were object-gestures. Furthermore, the involvement of objects in expressive actions were not primarily found in the beginning of the studied period. It also appeared in relation to new abilities, such as gestures with INDIRECT iconic aspects that emerged around TP#2. It should also be noted that the dominance of empty-handed ACTION\_BASED gestures over ACTION\_BASED object-gestures during the first half of the studied period should not be interpreted as a lack of such object-gestures. The “problem” is rather that “action based” actions with objects are in many cases simply actions, with a semiotic complexity corresponding to “typified action” (Sem#2) and often a communicative complexity close to “action framed by mutual attunement” (Comm#2). It is not that ACTION\_BASED actions with objects do not exist, but rather that they are less inclined to stand out as semiotically complex (Sem#3) or explicitly communicative (Comm#3), even though they sometimes do. As argued before, the seeming emergence of ACTION\_BASED gestures

---

<sup>40</sup>Aitchison is rather discussing development in the evolutionary time-scale, but the arguments involved could be interpreted more generally.

that involve handling of objects at TP#2, is primarily a matter of ACTION\_BASED gestures that are performed from the perspective of third-person intentionality, and this makes these actions appear more semiotically complex and often more communicative.

When it comes to the upper limit, the issue that has been treated in this chapter which is of most relevance is the issue of combinations. As pointed out in the discussion of combinations (Section 2.3.4), the mere fact that two expressive units are combined in a meaningful fashion is not the same as “grammar”, since grammar is a particular normatively conventionalized system for combination of meaningful units rather than just any combination. It was clear that there are no tendencies, in terms of frequency, towards grammar in this sense when it comes to multi-gesture expressions. A more difficult question is whether the ways in which some conventionalized gestures are combined with speech could be characterized as “grammatical”, since there are clearly norms and cultural differences when it comes to the placement relative to spoken utterances of gestures like NODDING and HEAD\_SHAKE (e.g. Kita & Ide, 2007). This question is postponed to the chapter on conventionalized gestures (Chapter 8).

The question asked here with respect to the upper limit is rather if there is some sort of opposition between complexity in speech and (simultaneous) complexity in gesture, so that one modality will consist in simpler kinds of expressions if the other is more complex, or if they tend to be complex together, so to speak. The term “complexity” is used here to refer to language-like properties, such as normatively constrained combinations.

When it comes to conventionality, it was found that conventionalized gestures were *more* likely to be coordinated with speech than iconic gestures were, and that conventionalized gestures were much more likely to be used together with speech than without. This runs counter to the popular idea of the conventionalization of speech and gesture as opposed poles (McNeill, 2005), according to which co-speech gesture tends to be more spontaneous and non-conventionalized than autonomously used gestures. The finding that conventionalized gestures are typically used with speech (and even more so than the iconic gestures) is interesting in light of the suggestions by Kendon that were discussed in Section 2.3.1. They are repeated here:

Observations of speakers in communities where there is a considerable repertoire of such [conventionalized] forms show that *their use is often fully integrated into the flow of everyday discourse* and that their sharp sep-

aration from non-conventionalized forms that many gesture-classification systems imply, cannot be sustained. (Kendon, 2008, p. 360, italics added)

[...] it is far from being the case that gestures that are always associated with speech are less conventionalized than those that are not. (Kendon, 2004, p. 106; see also Kendon, 1984, p. 94)

When it comes to complexity in terms of combinations it was found that more multi-gesture combinations were almost always coordinated with speech and that the ratio of multi-word utterances that were coordinated with gesture increased substantially after TP#2, along with the increase in three-word and four-word utterances. It was also found that there were peaks in the tendency to produce multi-gesture utterances at the points in time when MLU was about to reach 2 and 3 respectively. To borrow an expression from Iverson & Goldin-Meadow (2005), it does indeed seem as if “Gesture paves the way for language development”. Although combinations and grammar (in a strict linguistic sense) are not the same thing, as stated above, this does not mean that they are entirely unrelated either. In sum, this means that the complexity in speech were symmetrically related to combinatorial complexity in gesture. The difference between the modalities were rather one of magnitude — multi-gesture combinations are far more rare than multi-word combinations, and multi-gesture combinations in the form of gesture-only utterances are even more rare.

Other studies that have argued for an asymmetric relation in the complexity of gesture and speech have mostly focused on the fact that when it is, for some reason, not possible to use speech, iconic gestures tend to change in character so that there is a greater tendency to perform multi-gesture utterances when speech is unfeasible or suppressed (Singleton et al., 1995; Goldin-Meadow et al., 1996). When gestures are used *with* speech, they are commonly performed in ways that make them difficult to interpret without also hearing the co-occurring speech (see also Chapter 7). This sort of “asymmetry” (or perhaps rather complementarity) is a different kind of consideration about the relation between gesture and speech that was discussed in the previous paragraph, since in the contexts studied in this thesis, there are no special restrictions on the use of speech. The considerations in the previous paragraph rather concerned differences between more or less complex speech, than the absence of presence of speech.

To conclude then, the “opposition” between gesture and speech rather seems to be one of magnitude — when speech is the primary medium used, there are fewer gestures, and vice versa — but it is not an opposition in the sense that the most com-

plex utterances would have the fewest gestures and the simplest utterances would tend to have the most complex uses of gestures associated with them.

## CHAPTER 6

# Deictic aspects of the children's gestures

*With children, pointing at objects occurs earliest and most independently in the effort to communicate*

Wundt (1973 [1921], p. 74)

### 6.1 Questions asked

The main question asked in this chapter is *what kinds of deictic gestures do the children perform?* A more specific question that is asked is whether there are any *conventionalized* forms of pointing, other than index finger pointing, used by the children, similar to what has been reported in adults in various cultures (e.g. Kendon & Versante, 2003; Wilkins, 2003). The issue of *iconic* aspects in deictic gestures is also discussed. On the whole, the present chapter is concerned with the description of a number of different gestures that involve directedness to the physical surroundings in various ways, and to provide an overview of some forms of directedness, referent grounding, and achievements of joint attention that appear in the children's gesture repertoires. Deictic gestures that involve handling of objects will be mentioned, but gestural actions such as GIVE, SHOW, and PLACE are instead treated in Chapter 9 and will there be analyzed in terms of a more general framework for studying communicative acts that involve handling of objects. The main focus in this chapter is on pointing and pointing-like gestures.

Pointing gestures and other deictic gestures such as GIVE and SHOW have been

shown to be the first clearly communicative gestures to appear in children, typically appearing around 9–13 months of age, often slightly before the emergence of the first spoken words (Bates et al., 1975; Murphy, 1978; Bates et al., 1979; Leung & Rheingold, 1981; Masur, 1983; Zinober & Martlew, 1985a,b; Volterra & Erting, 1990; Liszkowski, 2005).<sup>1</sup> When the children studied here started pointing is not known, apart from Alice who started pointing at 8 months of age. From this period in infancy and up to the start of the period studied here, at 18 months, where the children are not infants anymore, the children have used deictic gestures for at least 6 months and one may suspect that by this time they will have refined their skills in performing these gestures so that there is a broader variety of uses. In Chapter 5 it was found that deictic aspects of gestures are still by far the most common in the ages that are studied here, and other researchers have found that this seems to be the case at least for some time after this period too (Stefanini et al., 2009). Grassmann & Tomasello (2010) found that both 2 and 4 year old children rely more on a pointing gesture than speech in interpreting acts of reference in cases where gesture and speech contradict each other, which indicates that pointing gestures are somehow treated as more trustworthy than speech, perhaps due to their more “direct” nature of referring to entities “in the realm of visible things” as Wundt (1973 [1921], p. 75) puts it.

Whenever gestures with deictic properties are discussed in research on children's gestures, they are often discussed in a relatively homogenized way, with only a few rare exceptions such as in the detailed ethnographic work of Haviland (1998). One reason for this lack of interest in more detailed accounts of children's gestures may be that most research on children's gestures is quantitatively oriented, and this promotes the use of categorical treatments of gesture (cf. Kendon, 2004, p. 103), with its inherent emphasis on differences *between* categories, and vice versa, its tendency not to bring out internal variation *within* the categories. Children do indeed perform a huge variety of gestures with aspects of directedness to the physical surroundings, but a common treatment of the deictic properties of children's gestures is to make use of only a few broad categories such as POINTING, GIVE and SHOW (e.g. Bates et al., 1979; Goldin-Meadow & Iverson, 1998). In effect, such treatments hide the variation and plurality in the gesture performances. Sometimes there are some further distinctions made, but still on a quite broad level. One of these distinctions is the distinction between *imperative* and *declarative* pointing (Bates et al., 1975; Franco & Butterworth, 1996), where the former category refers to cases when

---

<sup>1</sup>This refers to “communicative” in the sense of level #3 in the levels of communicative explicitness that were presented in Chapter 2.



pointing is used “instrumentally” by the child to get certain objects (typically out of reach), and the latter refers to more explicitly communicative uses of pointing where the pointing gesture is used to “say” something to the Other.<sup>2</sup> Another distinction, of a different kind, that is sometimes made is between index finger pointing specifically and “other” forms of pointing such as “whole hand pointing” (e.g. Cochet & Vauclair, in press). Nevertheless, in general, the variety of uses of pointing in children is rarely studied and is typically treated at a somewhat homogenized level.

This is not to deny that there is, to a certain extent, categoriality involved in children’s gestures in the sense that a great deal of children’s gestures are performed with an orientation to typified ways of acting, such as to POINT, to GIVE, and to SHOW something. As testified by the previous sentence, there are even names in language for a number of these actions and children also learn these names quite early, as manifested by their ability to comply to verbal requests such as “can you give me X?” or “show me the X!”. This sort of categoriality is a tendency towards categoriality on the level of the actions themselves, from the point of view of the participants (in the natural attitude, cf. Chapter 2). Specific actually occurring actions may correspond more or less to these “ideal types”, and the existence of these types does not imply that all gestures correspond in a 1:1 fashion to a small set of types such as POINT, GIVE and SHOW, as assumed by analysts. Such a treatment of children’s gestures may sometimes be insufficient, for several reasons:

First, there are also other typified actions that could be argued to have a similar semiotic status as POINT, GIVE, and SHOW. One example of this is PLACING (cf. Clark, 2003). Strictly speaking, it may not be possible to create an exhaustive list of such typified actions, since they are to some extent an *open-ended set of practices* (cf. Streeck, 2009b), which may, in addition, be more or less *generic* (occurring across a wide range of contexts or perhaps even cultures) or *specific* (occurring in more specific contexts, with more specific meaning).

Second, the homogenized treatment of deictic gestures as a category on its own has downplayed aspects of directedness in gestures that also have iconic and/or conventionalized properties — especially so in the literature on children’s gestures. There is a cline when it comes to the extent to which deictic, iconic, and/or conventionalized properties are foregrounded in a gesture. Some gestures may essentially appear to a spectator as a pointing gesture with minor iconic elements. Other gestures may be experienced as primarily iconic gestures, or primarily conventionalized

---

<sup>2</sup>A further distinction of “*informative*” pointing has also been suggested (see Liszkowski, 2005), but will not be discussed here since the present chapter is not mainly concerned with these distinctions anyway.

gestures such as “emblems”, with only some minor deictic elements. Yet other gestures may be experienced just as much deictic as iconic and/or conventionalized. This means that the extent to which deictic properties are foregrounded in a gesture may vary.

Third, even when the discussion is constrained to a small set of typified actions such as POINT, GIVE, and SHOW, each of these “categories” can be performed in several ways. Furthermore, even though some gestures are crystal clear instances of pointing, for example, that does not necessarily mean that their only function is to single out referents of various sorts, which is how pointing gestures are typically characterized. The performance of a pointing gesture can have additional pragmatic/interactive functions over and above the function to indicate.

These remarks all motivate the theme of the present chapter, which is to provide a number of examples of the variety of child gestures that involve an orientation to the material environment in which the actors are situated, through somehow being “directed towards” this environment. Rather than making an attempt at covering all the varieties varieties of deictic gestures that exist in the data, the chosen examples are rather selected with an aim to make a few principal points. Before the examples are presented, a short discussion of terms such as deixis, indexicality, directedness, and joint attention is provided.

## 6.2 What is a “deictic” gesture really?

What counts as a deictic gesture? Everyone agrees that the prototypical index finger pointing gesture is a deictic gesture, but what else? What about the teacher’s clapping of the hands to draw attention to himself or herself? What about slowly reaching towards an object that you suspect that another person might potentially not want you to take, performed with mutual gaze established throughout the action to ensure that the other person will see what you are doing, to monitor the reactions of this other person? What about simply grabbing an object while saying “this+GRAB”? What about indicating items on a computer screen by using the mouse arrow on the screen? What about performing an “iconic” gesture such as hammering (as if) directed towards a (real) nail as a way to show where to hammer (for real)? There is a huge variety of actions that are directed towards targets, single out referents in speech, or that regulate joint attention towards various aspects of action and the world in various ways. Do they have anything in common? There is certainly a number of interrelated, but different, aspects involved in what

researchers usually speak of as “deictic” gestures.

First, there is the Peircian notion of *indexicality* (Peirce, 1931–35), which is defined as *contiguity in time and/or space*, and if we follow Sonesson (1996) an index is not to be understood in purely physical terms, but rather as a matter of how certain events are experienced and interpreted. For example, Sonesson argues (with reference to Goudge, 1965, p. 55) that:

not all examples of indexical signs given by Peirce are susceptible of receiving a causal explanation: the Pole Star, for instance, may be an index of the north celestial pole, but it is in no way caused by that astronomical location.

Second, not all aspects of deictic gestures can be explained by a principle of contiguity, but needs to be supplemented with the notion of *directedness*. In fact, “real indicators, such as fingers and arrows, are equally contiguous to a number of objects which they do not indicate, for instance to the things which are at the opposite side of the arrowhead, in the direction to which it does not point” (Sonesson, 1996, p. 132). Other researchers have also defined pointing in terms of their directed character (cf. Eco, 1976; Rolfe, 1996; Kendon, 2004; Kita & Lausberg, 2008). Bühler (1982 [1934], pp. 20) has noted that the directed character of deictic gestures may in fact sometimes be used to “create” an imaginary target in cases where no concrete target is present, and he refers to this as *deixis am phantasma* (cf. *abstract deixis*, McNeill et al., 1993).

Third, deictic gestures are often discussed from the point of view of how they may serve to establish *joint attention* and *grounding of referents* in social interaction, including the important special case of grounding referents of spoken utterances. Even though contiguity and directedness are commonly involved in such achievements, the social and communicative aspect of such actions are not reducible to principles of contiguity and directedness alone, but are rather something more. This aspect of deictic gestures is rather a matter of manifest properties of *other-orientedness* (relative to their context of occurrence), i.e., a matter of a certain degree of communicative explicitness (see Chapter 2). Then again, there are also deictic gestures which are not other-oriented in an explicit sense, such as when the index finger pointing is used to keep track of a series of items when counting them (cf. Saxe & Kaplan, 1981; Alibali & DiRusso, 1999).

Further distinctions can be made, but the three features that have just been listed are among the most central ones. The crux of the matter is that not all gestures that are typically called “deictic” gestures share all of these three features. Instead

of going down the road of arguing that one or the other of these aspects constitute the “true essence” of deictic gestures, deictic gestures will be treated here as a family of gestural actions that are characterized to different extents and in different ways by the properties of contiguity, directedness, and/or coordination of attention in social interaction. That is, just like in the discussion of the nature of “gesture” in Chapter 2, a *family resemblance* (cf. Wittgenstein, 1953) model of “gestuality” (cf. Kendon, 2007, p. 6) will be adopted, and in the context of the present chapter more specifically for deictic gestures.

### 6.3 Observations on the use of index finger pointing

In Chapter 5 it was shown that index finger pointing is the single most common type of gesture performed by the children. Yet not all index finger pointing gestures are created alike. The analysis in this section will bring out a number of quite heterogeneous aspects of (index finger) pointing gestures that may vary between specific instances.

Figure 6.1 shows Hanna (19;16, to be read as age in MONTH;DAYS) who produces an index finger pointing gesture which is typical in many ways. For example, just like most pointing gestures of children which tend to be coordinated with concrete nouns and deictic words, Hanna is here producing her pointing gesture together with a noun which names the target of the pointing gesture (“*eyes+POINT*”, *ögon*). In this case the pointing gesture is used to show what eyes are, as part of a general activity which involves a book which is concerned with various body parts. In other cases, the relation can be the opposite, i.e., the pointing gesture can be used to refer to something which is not yet known or labeled for the purposes of a present activity, such as when children point to something and ask what it is and the parent responds by giving an account of what it is. In this way pointing gestures can both be used to display an understanding and to attain an understanding. Even though the pointing gesture itself may not look distinctively different in those two cases, there is still a distinct difference on a functional level between two rather different forms of “reference”.

Hanna's pointing gesture is also typical for children's pointing gestures in the sense that the pointing index finger touches the target. In 91% of the cases when pointing gestures are used to refer to targets that are within actual reach the children studied here touch the target with the pointing hand. This means that the typical



Figure 6.1: Index finger pointing (Hanna 19;16)

pointing gesture of children is both *contiguous to* and *directed to* its target.<sup>3</sup> Still, the fact that Hanna is actually touching her mother's (closed) eye — to the discomfort of her mother — shows that Hanna is not yet orienting to some more adult-like norms of how to touch, or not to touch, different parts of other persons (cf. Morris, 1977, pp. 92). This changes later on, and there are examples from later on in the studied period where the children (such as Tea at 26;12) even avoid touching the eyes of a doll as part of pretense play where the doll is treated as a living being. In another situation which involves pretense play, where a toy frying pan is treated as if it was hot, Tea 24;25 visibly “avoids” touching the frying pan when she points to it. She makes this visible by saying “ouch ouch” (*ajaj*) and each time she says “ouch” she makes a slight movement of the pointing hand *away* from the frying pan, rather than towards it (although the overall gesture is of course directed towards the target), as if to avoid touching the “hot” pan. In this way, the performance of pointing gestures can be visibly oriented to, or “fail” (from an adult's perspective) to orient to, more particular knowledge about whether and how to touch and handle various things and persons.

In Frame #2 of Figure 6.2 Harry points to an empty location in the puzzle that he is doing together with his mother. When he points to this location he says “there!+POINT” (*dä!*). The interesting thing here is how this multimodal utterance

<sup>3</sup>One may speculate that contiguity is the primary factor behind children's early comprehension of pointing gestures, because children do not, for example, generally learn to follow distal pointing accurately until around 12 months (Lempers, 1979; Butterworth & Grover, 1989; Morissette et al., 1995; Thoermer & Sodian, 2001). In distal pointing the gesture is directed towards the target, but not as contiguous to it.

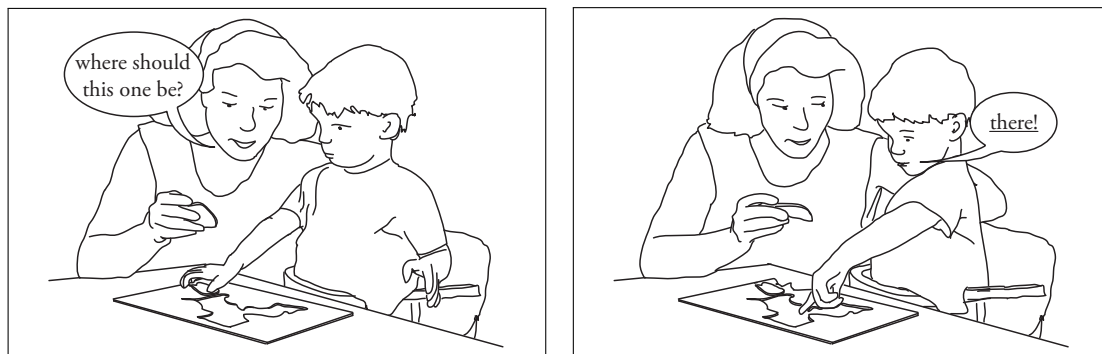


Figure 6.2: Pointing to a location (Harry 24;16)

relates to the mother's previous utterance, because it is to a large extent through this relation to the previous utterance that Harry's utterance gains its meaning, as a matter of *act-activity interdependence* (see Chapter 3). In Frame #1 the mother holds up a piece of the puzzle in front of Harry and asks him "where should this one be?" (*var ska denna sitta nånstans?*). So, when Harry says "there!" and points to the empty location, he is not just referring to a location on the table, but he is also, as if it were, incorporating the mother's utterance in his own communicative action by responding to it. It is as he was effectively saying that "that piece should be there" and it is indeed the piece that the mother SHOWS in Frame #1 that should be "there". This is reminiscent of how Goodwin (1995, 2003) has shown that a man with aphasia who can only produce three words ("yes", "no", and "and", as well as non-word vocalizations) and various gestures is still able to incorporate the meaning of other people's utterances into his own turn to construct more complex meanings than the semiotic resources of his own utterance would be able to achieve on their own. Indeed, with a single deictic word and a single pointing gesture, Harry is able to express a relatively complex "message" through the way the word and the pointing gesture responds to his mother's previous turn. This phenomenon is a highly frequent and it could be argued that even most of the children's gestures derive their meaning in part through their relation to previous turns in this way, i.e., through the way that actions reciprocate previous actions.<sup>4</sup> Children's use of gesture (and speech) to reciprocate others' previous actions is highly under-studied in general, presumably because research on children's gestures is dominated by psychologists who tend to be most interested in the initiating aspects of children's actions, rather than the responding aspects. Research in psychology tends to separate the study of "comprehension" and "production", so when production (i.e., action) is studied, it

<sup>4</sup>Caselli (1990, p. 65) notes similar things, such as "mother asks: 'Cosa vuoi' (What do you want?). The child makes the sucking gesture".

is usually not studied from the point of view of how action displays “comprehension”, that is, how actions relate to previous turns in interactions. Another way to state the same thing is to say that when comprehension is studied, it is usually conceptualized as a matter of sensory “intake” of information, rather than as a property of action. This is to some extent a simplification of the situation, especially since far from all psychologists conceptualize action in the same way, but it is clear that there is a strong tendency in the developmental psychological literature on gesture to miss out on aspects of act-activity interdependence, i.e. the responding and contextualized aspects of gestures, which are nevertheless there in a real and consequential way, and which are just as much an issue for psychology as any other aspect of action.

The reader might recall from Chapter 3 that the term “activity” in “act-activity interdependence” is slightly ambiguous as to whether it refers to the preceding (and potentially subsequent) turns and actions in the present activity in the given situation or if it refers to more situation-transcending kinds of activities (practices). It was argued there that the term should be understood as referring to both of these. In the example here, there is also act-activity interdependence in the sense of a more situation-transcending sort of activity, since the interpretability of Harry’s pointing gesture does not only rest on his mother’s previous turn, but also on the participants’ shared and conventionalized knowledge (both “know-how” and “know-that”) of how to do a puzzle. This knowledge is a matter of third-person intentionality, since it relates to how “anyone” would do a puzzle of this kind.

Another aspect of Harry’s pointing gesture in this example that may be noticed is how pointing to an “empty” location in this way is something like a distant relative to abstract pointing gestures. As mentioned before, abstract pointing gestures are those gestures which project a non-present target (cf. Bühler, 1982 [1934]; McNeill et al., 1993). Of course, Harry’s pointing gesture does point to a concretely present location on the table, but it also invokes the the currently non-present (at that location) piece of the puzzle as well as the potential for this piece of the puzzle to be located there. Again, this projection is possible precisely due to the shared knowledge of the third-person intentionality involved in “doing a puzzle”.

In the next example, shown in Figure 6.3, Bella is playing with plastic toy food together with her mother. Bella has the leading role in this activity and before the episode shown in the figure, Bella gave the toy food to the mother while saying “an’ have that” (*å ha den*). The mother did not quite seem to know what she was expected to do with the food, and she said “aha, so I got that one then?” (*jaså fick jag den då?*) with a questioning/hesitating intonation. Bella responded to the mother

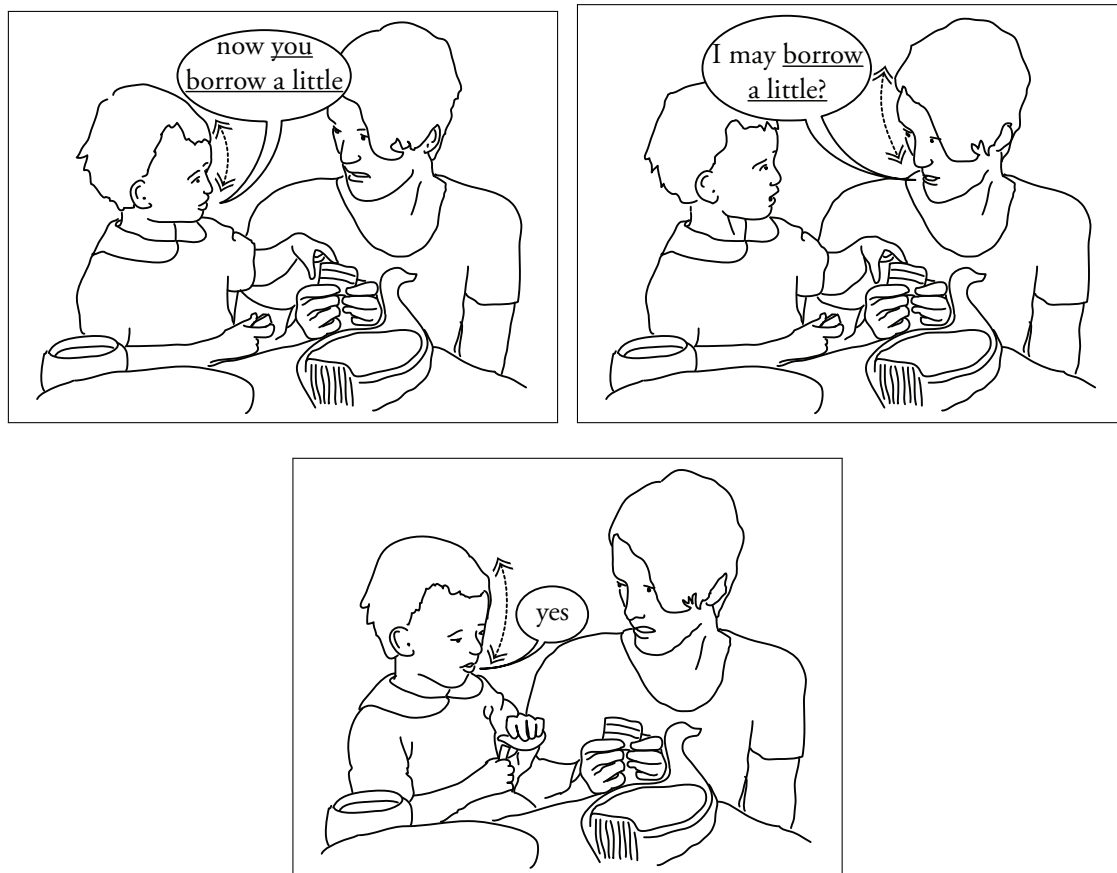


Figure 6.3: POINT continues after the end of the spoken turn (Bella 27;23)

by saying “can borrow a little”, but the mother did not do anything but just looks to Bella, as she seemed to remain confused about what she was supposed to do with the toy food. Then, as shown in Frame #1, Bella POINTS to the toy food that she has given to her mother, and also performs a NODDING gesture while saying “now you borrow a little+POINT+NODDING” (*nu låna du lite*), with a determined voice. The NODDING ends at the same time as the spoken utterance. The pointing gesture, however, is not retracted when the spoken utterance ends, but Bella rather keeps holding the pointing gesture and while she does this she also maintain her gaze fixated at her mother, as if waiting for something. In Frame #2, Bella’s mother responds by more or less repeating Bella’s utterance (including the NODDING gesture), and says “I may borrow a little?+NODDING” (*får ja låna lite?*), apparently still not quite sure what she is supposed to do. Rather than interpreting the mother’s utterance as request for clarification from Bella, Bella seems to treat this as a request for confirmation, as if to ask for a confirmation that she understood what Bella said correctly. Only at the completion of the mother’s utterance does Bella retract her pointing hand while once again nodding and saying “yes+NODDING” (*aa*), seem-



ingly satisfied with the mother's response in the sense that she treats the sequence as finished through the way she then goes on to do other things (not shown in the figure). That is, even though the mother does not seem much more enlightened after this exchange regarding what to do, Bella seems content that the sequence has come to a closure and that understanding has been reciprocated.

This way of maintaining a pointing gesture after one's own utterance has ended is by no means an exceptional way of using pointing gestures. Andrén (in press b) describes a study that was based on a subset of the data used in this thesis, consisting in 393 pointing gestures from Bella, Harry and Tea. It was found that 35% of the pointing gestures lasted past the end of the child's own utterance and continued at least until the first *transition relevance place* (TRP, Sacks et al., 1974) in the parent's responding utterance.<sup>5</sup> 63% of the pointing gestures ended in direct association with some part of the child's own utterance. Only 2% of the pointing gestures continued past the end of the child's own spoken utterance, but did not last until the first TRP in the parent's following response. This indicates that there were two almost categorically distinct ways of organizing the endpoint of the pointing stroke — one *short* variant where the end of the stroke ends when the spoken utterance ends (if not earlier within the own utterance) and one *sustained* variant where the end of the stroke comes after a response has been received from the parent. More crucially, it was also found that the parents gave significantly more elaborated responses in cases where the children performed sustained index finger pointing gestures in comparison to when the children performed short index finger pointing.<sup>6</sup> Furthermore, the children were shown to orient to these responses in two main ways: First, in most cases the children immediately withdrew their sustained pointing gestures when a parental response was given, which indicates that receiving a response was a satisfactory condition for ending the pointing gesture. Second, in cases where there were no such immediate withdrawal after a parent's response, there was a significant inverse relationship between the elaborateness of the response from the parent and the communicative effort invested by the child in the child's subsequent utterance(s).<sup>7</sup> That is, the less response a parent provided 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, such as tapping the target of the pointing gesture again, and/or repeating

---

<sup>5</sup>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. The turn-so-far can be a full grammatical clause, but also, for example, a response morpheme.

<sup>6</sup>This was confirmed by a chi-square test: ( $df=2, n=296$ ) = 22.34 ( $p<0.01$ ).

<sup>7</sup>Spearman Rank=-0.70,  $p<0.01$ .

or elaborating on the utterance. All in all, these findings suggest that children's use of sustained pointing gestures functioned as a request for an (appropriate) response from the parent.

Similar thoughts about the interactive functions of sustained pointing gestures have been provided before, based on observations of adults (Sidnell, 2005; Clark, 2005) and children with Down's syndrome (Wootton, 1990). Bavelas (1994, p. 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."<sup>8</sup> This suggestion was indeed confirmed in the study reported here, focusing specifically on index finger pointing gestures (Andrén, in press b).<sup>9</sup> This is interesting, because it suggests that at least some pointing gestures are not only used to refer to things or just to establish joint attention to a certain target. In that case a simple minimal response such as "mm" or "yes" from the parent would have been enough as a confirmation that the target had been noticed, but as mentioned above, the children were less "satisfied" with minimal responses (such as single response morphemes) than with more elaborated responses. The sustained pointing gestures seem to be doing something more, namely to elicit a (more elaborated) response to the target that is currently being pointed to, as relevant in the context of a current ongoing activity, i.e., an active form of *social referencing*. It is readily acknowledged that this analysis is simplified in some respects. For example, for some communicative acts such as yes/no questions, a response morpheme such as "yes" may indeed serve as a fully acceptable response. However, in most cases, an acceptable response seems to coincide with more elaborated responses (than no response or minimal responses). More generally, however, it may be better to say that the kind of response that will serve as a stopping condition of the sustained pointing gesture is a response that is *relevant and compliant* to the child's communicative act, irrespectively of the elaborateness of the response.

Another observation that may be made regarding the use of sustained pointing gestures in the study (Andrén, in press b) relates to the fact that they are being performed in a way that is sensitive to what someone else is saying during the very performance of the gesture. That is, the endpoints of these pointing gestures must be

---

<sup>8</sup>Interesting in this regard is also the finding by Gullberg & Kita (2009) that holding the stroke of a gesture is one common factor in drawing the recipient's attention directly to the gesture, rather than elsewhere, such as towards the face of the speaker.

<sup>9</sup>That is not to suggest that index finger pointing gestures are the only gestures which may be used in this way, but merely that the study was focused on this phenomenon in the context of index finger pointing.

understood as a result of an ongoing adaptation to what the Other is doing. They are not delivered as “ready-made” wholes, pre-determined by some certain brain process, but are rather (also) resulting from the conditions of a situation that is determined by the way the child and the parent acts together, as a matter of *inter-action*. In the words of Bavelas (2007, p. 127):

[...] what individuals say and do in face-to-face dialogue is intimately affected by what the other person is saying and doing at that moment and by the immediate effect that their own actions will have on the other person.

To conclude, in this section I have argued that not all index finger pointing gestures are created alike. For example, it was argued that pointing gestures can be used both to refer to things as an instance of a token of a certain known kind of a more or less categorized type (i.e., referent+content), and to refer to (at least partially) unknown entities (i.e., referent only). With respect to the issue of whether pointing gestures should be considered to function like content-loaded words (such as nouns, verbs, or adjectives) (Goldin-Meadow, 2007b) or whether they should be considered to be more like deictic words (Pizzuto et al., 2005; Pizzuto & Capobianco, 2005), one may have to accept that this may vary from situation to situation, depending on how the gesture is used in relation to co-occurring speech and in relation to previous turns in the interaction. This may not make much of a difference with respect to the form of the pointing gesture itself, but it is certainly a functional difference in the overall use of the gesture as well as a difference in what factors determine the meaning of the reference achieved by the pointing gesture.

However, it was also argued that referring to various targets is not all there is to pointing gestures. In addition to the function of referring to something, they may also be shaped in ways that are demonstrably dependent on *past-oriented* issues (what has happened before), as in the case of “incorporating” what has been said by someone else in a previous turn as part of the performance, and *now-oriented* and *future-oriented* issues, as in the case of using sustained pointing gestures to elicit a response from the parent. It was also demonstrated how the interpretability of a pointing gesture may rest on shared knowledge of a certain kind of practice — the example that was shown involved “doing puzzles” — as a matter of third-person intentionality. These issues fall under the general heading of act-activity interdependence since they all concern the status of the act of pointing in relation to the overall activity that the pointing gesture is part of.

Pointing gestures were also shown to orient to (or sometimes fail to orient to)

various kinds of knowledge relating to manual interaction with various things, such as touching or not other persons in certain ways, or touching or not touching objects which are known to be “hot” or not, which can sometimes make a difference to the meaning of the gesture. This is of course because pointing gestures are situated real world phenomena — skillful performances — which do not merely serve “abstract” or “pure” semiotic functions, but which are partly subject to practical contingencies as well as more general knowledge of how to handle things and persons in the world.

## 6.4 Pointing gestures that involve TRACING

In the annotation system used for the quantifications in Chapter 5, there was a category called TRACING among the annotated iconic aspects of gestures. In Chapter 4, where this category was defined, it was mentioned that in the children studied here TRACING almost always appears as part of some sort of pointing gesture to a target in the environment within reach. The children very rarely trace a shape “in the air”, in the kind of gesture space I call *abstract space*, where the gesture performance is entirely detached from the physical surroundings. This section takes a closer look at some of those instances of TRACING that occur as part of pointing.

In Chapter 5 a distinction was introduced between “primitive” and “advanced” TRACING. Primitive TRACING are gestures which involves some sort of movement around the target which rather seems to be a matter of exploring the target with the finger than of signifying something. When the tracing is clearly a proper part of to expression, then it is a matter of advanced TRACING. All of the examples in this section are of the “advanced” kind.

Figure 6.4 and Figure 6.5 show two separate moments from the same situation. Hanna and her mother are sitting in front of a sheet of paper, drawing together. In Figure 6.4, Hanna draws a balloon on the paper with a crayon, and while she does this she says “I so draw air balloon+DRAW” (*ja så rita luftballong*).<sup>10</sup> (The outline she draws is shown as a dotted line in the figure although it is not dotted in reality, and note that she is actually drawing here, not just acting as if drawing.) This is one example of how various object-involving actions are often coordinated with speech, much like empty-handed gestures. It is also notable that Hanna is not just doing

---

<sup>10</sup>Another translation of this utterance could be “I like-that draw air balloon”. Also, in Swedish, the word *luftballong* is normally used to refer to a hot air balloon vehicle, but judging from what is said before in their conversation Hanna does not seem to refer to such a hot air balloon, but rather to the more common kind of party balloon that is tied in a string.



Figure 6.4: DRAW (Hanna 23;14)

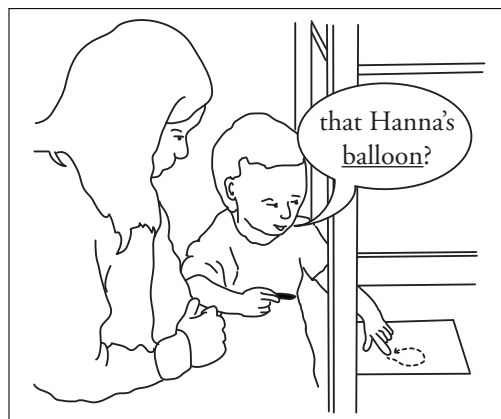


Figure 6.5: TRACE (Hanna 23;14)

something that may be described by an observer as “drawing”, but that she is also categorizing her own act as an instance of the culturally determined *type* of activity of “drawing”, as testified by her labeling of her own action in the spoken part of her utterance. This is similar to how iconic empty-handed gestures are often used to express the contents of conventionalized verbs in ACTION\_BASED gestures, with the only difference that she is actually (also) carrying out the action here. That is, her knowledge of her own action is not just based on previous “private” sensorimotor experience from drawing activities, but it is also in part a socially categorized and conventionalized *kind* of action, and of course, she is not just drawing an abstract shape, but the shape is also a matter of a socially established and conventionalized kind of artifact, as testified by her reference to this shape as a “balloon”. It may be added that although the action is “instrumental” in so far as she uses the crayon to produce (real) marks on the paper, the act of drawing is of course in itself an action

which produces the balloon as an iconic semiotic sign.

After this Hanna's mother draws a girl on the sheet and says that the balloon belongs to the girl (not shown in the figures). Hanna then seems to wonder whether it is Hanna's balloon — possibly wondering if the girl that the mother has drawn on the paper is supposed to be Hanna. Accordingly, in Figure 6.5 Hanna performs an index finger pointing gesture towards the balloon, tracing the outlined shape of the previously drawn balloon with her finger. While she performs this gesture she says “that Hanna's balloon?+POINT/TRACING” (*de Hannas ballong?*) with a questioning intonation.

TRACING gestures have been discussed previously (Wundt, 1973 [1921]; Müller, 1998a,b; Kendon, 2004; Streeck, 2008a), but it remains unclear whether these gestures are somehow ontogenetically related to acts of drawing or not. Therefore, it is interesting to note that the TRACING gesture shown here appears precisely in the context of drawing. In fact, almost all TRACING gestures performed by the children in the data studied in this thesis are performed towards pictures of various kinds, in contexts of creating and looking at pictures. Then again, the parents also produce TRACING gestures in these contexts (see also Gelman et al., 1998), so it is not obviously the case that the children invent the gestural technique (cf. Müller, 1998a; Müller et al., manuscript) of TRACING, based on “private” experience of drawing — as a kind of ritualization that may involve a gradually increasing “distance” (cf. Werner & Kaplan, 1963) from the act of actually drawing to the more abstract and detached empty-handed TRACING gestures. Since the parents also frequently perform gestures like this, ranging from the more concrete object-gesture variants to the more detached empty-handed TRACING gestures, the children's use of TRACING may very well originate from imitative processes of the parent's TRACING gestures in the same contexts. The most likely scenario is perhaps that the ontogenetic process involves both sensorimotor schematization/ritualization (i.e., an ego-world process) and imitative processes (i.e., an ego-alter-world process).

In the next example, shown in Figure 6.6, Alice and her mother are also involved in drawing. Alice's mother has previously drawn a liquorice pipe on the paper and stated that she likes them and would like one. Then the mother asks what Alice likes and Alice suggests that the mother should draw mushroom cheese (*champinjonost*) and the next turn Alice also adds that she likes that. The mother then asks if she should draw a mushroom (not mushroom cheese), and Alice says “yes”. Then, before the mother starts drawing the mushroom, Alice makes a tracing movement along the liquorice pipe while she says “this here I don't like+TRACING” (*de här tyckern'te om*), as shown in the figure. The paint-brush that Alice is holding is dry, and not

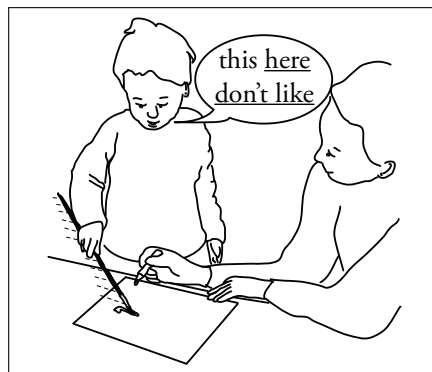


Figure 6.6: TRACE with object (Alice 23;06)

used for actually painting anything. Thus, this object-gesture is an interesting “in between” case between actually drawing and the slightly more detached index finger TRACING performed by Hanna in the previous example.

The two examples of tracing shown so far took place at a relatively early age (both at 24 months). The next example, shown in Figure 6.7, takes place at an older age (28;19), and here the tracing is even more detached from actual drawing, but still appearing in the context of interacting around pictures. In Frame #1, Tea performs a TRACING index finger pointing gesture along the throat of a duck which is depicted in the book in front of Tea. As she does this, she says “loooong+POINT/TRACING” (*lääääng*), with an iconic prolongation of her pronunciation of the word “long”. This iconic modification of the word is likely not a creative invention of herself, as there are a number of examples in the data where her parents also use the word “long” in this specific prolonged way.<sup>11</sup>

When she reaches the upper end of the duck’s throat, she immediately reconfigures her hand into the iconic gesture shown in Frame #2. This gesture starts as a “precision grip” between the the index finger and the thumb, and then the grip widens, until it reaches the state shown in the picture, apparently also showing the extent of the duck’s throat. During her performance of this gesture, which is performed without speech, her father responds by saying “it has a long throat” (*den har lång hals*) in a confirming manner. In Frame #3 then, Tea turns to her father and tilts her head backwards, saying “yes such long arm+POINT/TRACE” (*ja sån lång arm*) in concert with yet another index finger pointing TRACING gesture along her

<sup>11</sup>There are a number of other antonymic adjectives which are sometimes pronounced in iconically motivated ways by the parents such as saying “BIIIG” with a loud and deep voice, and saying “smaaall” with a bright and tiny voice.



Figure 6.7: TRACE (Tea 28;19)

own throat. She clearly uses the wrong word — “arm” instead of “throat” as one would have expected — but the way she points to her throat shows that she is actually talking about the throat and not about an arm. The TRACING gesture performed in Frame #3 is thus not performed in direct relation to a picture, but rather steps “out of” the picture, to an object in the world (her own throat), seemingly motivated by the category membership shared by the two throats. Still, all three gestures performed here are directed towards targets that are present within actual reach in the world. They are thus not just expressing the concept of “long” in an abstract and detached gesture space, but rather “X is long”, where X is the throat, and where X is invoked by the deictic directed character of her gestures towards the two throats that are concretely present (the picture and her own throat). Interestingly, in comparison to the two previous examples where the children were younger, Tea is not here referring to the object as such, but a more abstract property of this object, namely its length (cf. the *decomposition effect*, McNeill, 2005, p. 184).

Tea 28:19 also performs a few instances of another variant of TRACING pointing gestures (not shown in any figure here). In this variant, the finger moves around along the surface of a picture, but does not follow a specific shape. It appears in the context of pointing to a group of objects where the objects are all tokens of the



same type and these gestures are typically performed by revolving around the center of gravity of the group of objects that are being pointed out, rather than by creating an outline of the area populated by the group of objects. In all but one of these examples Tea also uses plural forms of her spoken words, such as when she points to a group of ants that are depicted on one of the pages in a book and says “black-PLURAL” (*svarta*). Hence, this variant of tracing seems to be more oriented to plurality than to tracing shapes.

To conclude, there seems to be at least two different types of “tracing”: One which has to do with tracing shapes and associated properties such as “extent”, and one which has to do with plurality. The latter variant is only performed by Tea, and appears towards the end of the studied period, and it is possibly inappropriate to talk about this as “tracing”. The variant that is indeed a matter of TRACING shapes is far more common, occurring in all five children, although it is still rare compared to the number of pointing gestures that do not involve TRACING. It may be recalled from Chapter 5 that a peak in tracing gestures was found at the core of Transition Period #3. This increase in TRACING gestures may possibly be understood as an increasing ability in children to separate out properties such as shape and extent from the blended richness of properties that are otherwise present in real world objects. Interestingly, most of the TRACING gestures were found specifically in the context of looking at or creating pictures.

Even though TRACING gestures vary from instance to instance regarding the exact shape that is traced, they still form a kind of recurrent practice, employed both by the children and their parents. It is therefore possible that they should be considered, not only as iconic, but also as a conventionalized gestural technique. That is, in contrast to emblems the TRACING gestures do not have a strongly conventionalized specific form, since there is an axis of variation that varies from situation to situation, depending on the particular shape that is being traced in each specific case. Nevertheless, one might still consider the possibility that the overall technique of TRACING is something that children typically learn, at least in part, from their parents.

That there may be conventionality on the level of certain gestural techniques, rather than only on the level of specific forms, is a possibility that has rarely been considered when the issue of conventionality has been discussed in gesture research.

One may note that Arendsen (2009) found that some less prototypical variations in the performance of normatively conventionalized signs of signed language were judged more acceptable than others, and that those variations that were judged to be more acceptable were typically those that were consistent with iconic aspects

of the signs. This is relevant to the present discussion because it shows that there is not necessarily a contradiction between conventionalized ways of gesticulating and potentials for “productive” variation along an iconic axis and I suggest that this may possibly be what is going on in the case of TRACING gestures too, although I do not claim that they are as strongly conventionalized as the more *normatively constrained* forms of signed language. The claim is rather that tracing may be a conventionalized in the sense of being a *typified* gestural technique.<sup>12</sup> On a more general level, pointing gestures can themselves also be considered as such a gestural technique, which can be conventionalized in form (e.g. Kendon & Versante, 2003; Wilkins, 2003) at the same time as they have a kind of built in “productive” potential due to variation along the deictic axis of directedness, as the directedness change from situation to situation depending on the target that is pointed out, a little like TRACING gestures vary from situation to situation regarding the more exact shape that is traced.

## 6.5 Whole hand pointing

In Chapter 5 it was found that the vast majority of the children's pointing gestures are in fact index finger pointing gestures. Nevertheless, the children also perform pointing gestures that do not have the index finger hand-shape. Other variants of pointing gestures range from those that do not really have a clearly articulated form, such as quick flicks with the arms towards something, to those that do have an articulated form, just not the index finger hand-shape. Some researchers make a distinction between index finger pointing and “whole hand pointing” (e.g. Leavens & Hopkins, 1999; Cochet & Vauclair, in press), but are all “whole hand pointing” gestures created alike? Do they form a natural and homogeneous category?

Figure 6.8 shows Tea and her mother, looking in a picture book. In Frame #1 the mother points to a hat which is depicted in the book and says “a hat+POINT...” (*en hatt...*), which is the first part of a single utterance. Frame #2 shows the second part of the same utterance, and although the utterance is split into two illustrations in the figure, there is no pause in between these two “parts” of the utterance. Hence, in Frame #2 the mother says “...that one can have on the head+WHOLEHANDPOINT” (*...som man kan ha på huvet*). The hand-shape involved in this pointing gesture seems to bear an iconic relation to the spatial extent of a hat. In Frame #3 then,

---

<sup>12</sup>Conventionalization in the sense of typified ways of acting (Conv#2) and in the sense of normatively constrained ways of acting (Conv#3) were defined and discussed in Chapter 2.

## 6.5. WHOLE HAND POINTING

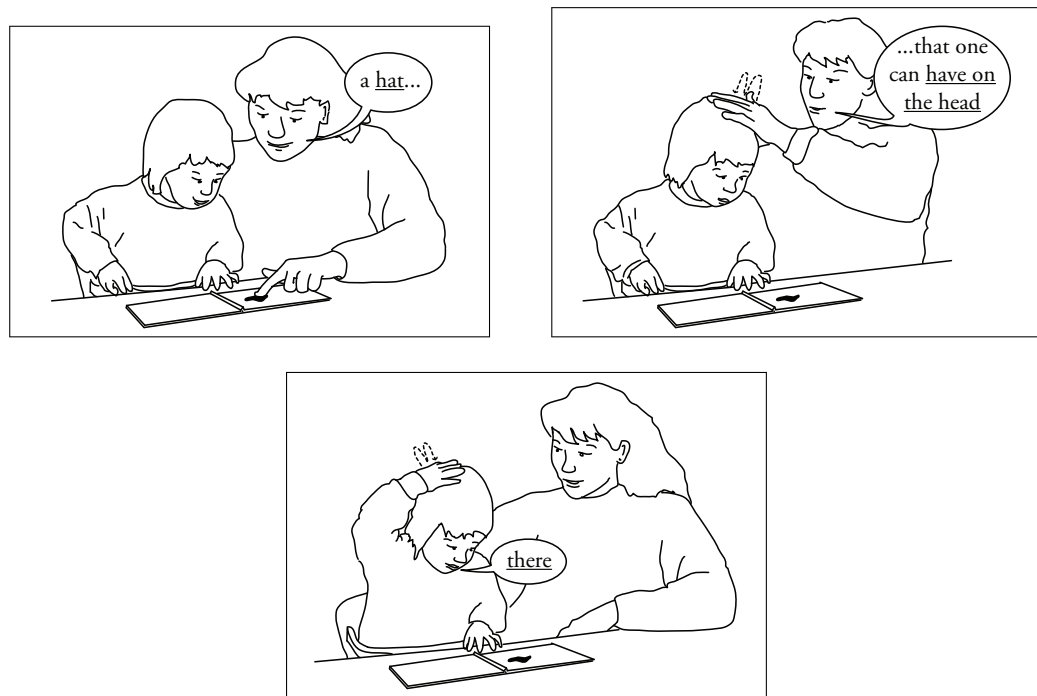


Figure 6.8: Pointing with the whole hand (Tea 20;26)

Tea makes a similar gesture, including the repeated tapping stroke, on her own head and she coordinates this with the deictic word “there+WHOLEHANDPOINT” (*dä*). Throughout Frame #1 to #3, Tea maintains a fixed gaze towards the hat picture in the book. Similar to the example in a previous section where Harry’s utterance “incorporated” his mother’s previous turn, Tea’s performance also incorporates her mother’s previous turn, so when she says “there+WHOLEHANDPOINT” it is really the hat that the mother has brought up that is supposed to be “there”, and of course, the gesture she performs is not just a gesture, but precisely a re-enactment of the very same gesture that the mother performed, i.e., imitation. It is interesting to note that if Tea recognizes her mother’s gesture, she is able to perform a cross-modal “translation” from what is primarily a tactile experience of the hand-shape and character of the stroke in the mother’s gesture into an own recognition and re-enaction of the gesture. She is quite obviously not able to see the mother’s stroke performance since it takes place on the top of her head, other than perhaps indirectly, such as seeing the movement of the mother’s arm in the periphery.

Figure 6.9 shows Harry sitting by the table together with his mother. Harry is drinking milk, but for a while he started playing with a teddy bear instead, and his mother eventually took it away. Harry didn’t seem to appreciate that, and in Frame #1 he says “a e sit here+WHOLEHANDPOINT” (*a e sitta hä*) in concert with



Figure 6.9: Pointing with the whole hand (Harry 29;17)

a flat hand pointing gesture towards the surface of the table. Just like the flat hand pointing gesture from the previous example, this gesture is performed with a stroke that consists in tapping the surface a few times.

This way of pointing to surfaces where something may be placed, as shown in the previous and in the present example, does indeed occur a few times in the data and it is perhaps one of the strongest candidate for a conventionalized way of pointing (cf. Kendon & Versante, 2003; Wilkins, 2003), apart from index finger pointing. Once again, the kind of conventionality that is suggested here is on the level of *typified* conventions (Conv#2), rather than normative conventions (Conv#3). It would not be considered “wrong” if Harry, instead of using a flat hand pointing gesture, used an index finger pointing gesture or some other form or variant, and it would not be considered “wrong” if Harry had omitted the repeated tapping of the target, instead only producing a stroke with single movement towards the target. That is to say, it would perhaps be too much to claim that it is a *normatively* constrained gesture (Conv#3) that must be performed in a certain way, but it is indeed a recurrent and recognizable typified way of pointing to surfaces where things may be placed, or similar things, such as where someone may sit.

Figure 6.10 shows Alice and her mother who are playing patient and doctor respectively. The mother (the doctor) has just asked if “there was something else?”, i.e., some more fictional health problem than the ones they have already dealt with. Alice has replied and answered “yes”, and in Frame #2 she elaborates on this by saying “my arms+WHOLEHANDPOINT” (*mina armar*) in concert with a flat hand pointing gesture. This gesture is performed by placing her left hand on the right arm and moving the hand back and forth one time. This gesture bears some resemblance to the TRACING gestures in that it is in a way a tracing of the extent of the arm, but it is also different. In fact, it looks more like the kind of self-touching that is associated



Figure 6.10: Pointing with the whole hand (Alice 28;01)

with pain, or rather what Heath (2002) has described as “demonstrative suffering” in studies of real patient-doctor encounters, where gestures are often used to “reveal emotional and personal experience” (*ibid.*, p. 597; see also Kendon, 2004, p. 30). Nevertheless, it is also still a pointing gesture in the sense that it points out to the mother which part of the arm that hurts.

Figure 6.11 shows Harry and his mother looking in a toy catalog. At one of the pages there is a tractor that Harry wants badly and he returns to this page several times, each time stating he wants the tractor with a voice loaded with desire. Several of these times, he also extends an open hand towards the tractor in a begging manner (cf. Bulwer, 1974 [1654]; Müller, 2004), as if trying to “get” the tractor. This is an ACTION\_BASED iconic aspect of the gesture. This is also what he does in the figure, and he coordinates this gesture with the word “mine+WHOLEHANDPOINT” (*min*). Even though there is an overall directedness towards the tractor in the performance of this gesture, there is a slight movement towards Harry’s body at the peak of its performance, that brings out the appearance of “getting” or “receiving” rather than “giving” or “putting”. The hand-shape bears an iconic resemblance to the actual hand-shape involved in receiving an object from another person, and it can be characterized as an *intention movement* where the initiation of a recognizable sort of action projects the continuation of the action even though the full action is not necessarily performed.<sup>13</sup>

The four whole hand pointing gestures that have been shown in this section are obviously different in character from each other, apart from the first two that were argued to be of the same general typified and conventionalized kind. Hence, “whole

<sup>13</sup>Intention movements were also discussed in Section 2.2.3 in relation to the Piagetian notion of differentiation. Intention movements are also discussed in several places in Chapter 9.

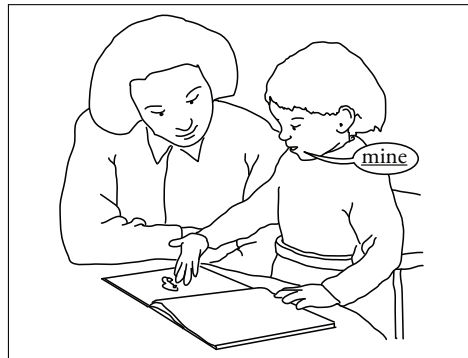


Figure 6.11: Pointing with the whole hand (Harry 21;15)

hand pointing” can really be a number of quite different things. Furthermore, none of these deictic gestures are sufficiently characterized by saying that they are simply “deictic”, since the way they are performed makes a difference to the meaning of these gestures so that, in effect, these gestures do not just refer to their targets, but they refer to their targets in a certain way (cf. Calbris, 1990; Kendon, 2004). As it was argued, there were iconic aspects in these gestures, as well as some conventionalized aspects.

## 6.6 Directedness in emblems and iconic gestures

In addition to various forms of pointing gestures, there are also gestures that may not primarily stand out to an observer as “pointing”, but which nevertheless have the character of being directed to a target. Figure 6.12 shows Harry making an “accusing” SCOLD\_POINT gesture towards a small toy figure that he is holding in his other hand. Even though the gesture employs the same hand-shape as index finger pointing, the orientation of the hand relative to the target of the gesture is different. That is, the way the SCOLD\_POINT is performed, the target is rather “hit” by the finger, as if hitting something with a stick, and there is no straight vector projected in the direction of the index finger. This “hitting” movement could be interpreted as an iconic aspect of this gesture although it is impossible to know whether Harry sees this iconic aspect or whether the gesture is simply learned and used as an “arbitrary” form with an arbitrary association to a specific meaning. Another difference between index finger pointing and the conventionalized performance of the SCOLD\_POINT is that in the SCOLD\_POINT the target is not touched by the hand,

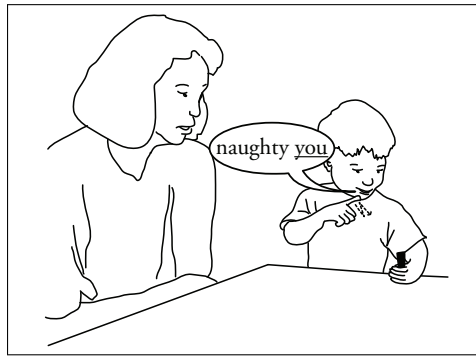


Figure 6.12: SCOLD\_POINT (Harry 26;18)

while this is optional in index finger pointing.<sup>14</sup>

When Harry performs this gesture, he also says “naughty you+SCOLD\_POINT” (*fy dej*). The directedness of the gesture qualifies who the deictic personal pronoun “you” is referring to. If Harry was to direct the gesture towards his mother rather than towards the toy figure, it would rather have been the mother that would have been accused of mischief. Another interesting aspect of the use of this gesture in relation to speech is that, in the data studied here, the children always use it with the Swedish exclamative word *fy*, which means roughly “naughty” or “bad”. This is the case in the example presented here too, but in this specific case, the gesture stroke is rather on the word “you” (*du*) whereas all other instances have the stroke coordinated with the word *fy*. As it happens, many other recurrently used gestures in the children’s repertoires also tend to be coordinated with specific words or expressions in this way, as a kind of multimodal recurrent *construction*. The issue of recurrent combinations of specific gestures or actions with specific words in children is discussed further in Chapter 8 in terms of *multimodal constructions*.

SCOLD\_POINT is clearly a conventionalized gesture, and as such it may be expected to be used only in one or a few cultures or in a restricted geographical area. It is therefore interesting to note that it seems to be a very widespread gesture. Zlatev & Andr n (2009, p. 390) found it in use by Thai children and Pizzuto et al. (2005, p. 232) reports it in Italian children. Indeed, conventional gestures do not always necessarily seem to be restricted to more specific cultures, such as the community of people that speaks a certain language, or the community of people that lives in a certain country (cf. Morris et al., 1979). Yet, this is of course not to say that SCOLD\_POINT is a universal gesture, but merely to point out that some emblems

<sup>14</sup>Earlier in this chapter it was stated that 90% of the children’s index finger pointing gestures to targets that are within reach involve touching of the target.

seem to have reached many parts of the world.

Some other emblems, such as HELLO and BYE-BYE, also have the character of being directed towards something. Both of these gestures are repeatedly used, not just directed to other persons, but also to referents such as characters in picture books. One may distinguish between two forms of directedness (this is also discussed in Section 9.5.1). First, there is a general directedness of many communicative acts towards the *recipient*, as if to show who the intended recipient of a communicative act is. This sort of directedness does not necessarily make the “target” of the directedness a referent in the content of the speech or gesture. Second, there is another kind of directedness which is rather directedness towards things that the communication is *about*, i.e., referents, such as the target of pointing gestures. It seems that gestures such as SCOLD\_POINT, HELLO, and BYE-BYE have elements of both these kinds of directedness. On the one hand they are often used in direction to the recipient of the communicative act, but on the other hand they also have a referential character since in the case of these gestures, the recipient is the referent. The referential character is especially salient when the communicative act is performed with a communicative orientation to the parent, although the gesture is directed towards an object in the world, as in Figure 6.12, so the referent aspect and the recipient aspect of directedness become clearly dissociated. The referential character of these gestures is perhaps also more salient when comparing them with a conventional gesture such as the GONE gesture where the palm(s) of one or two hands are held out, facing upwards (sometimes also with shrugged shoulders), since in a gesture like that the direction of this gesture relative to things in the environment is much less crucial than in the case of SCOLD\_POINT, HELLO, and BYE-BYE.

Gestures which primarily appear as ACTION\_BASED iconic gestures to an observer are also often directed towards a target, such as pretending to pour coffee in a real present cup or pretending to feed a doll by holding a spoon towards the mouth of the doll. In these gestures, the performance of the gesture does not only invoke the concept of the action such as “POUR COFFEE” or “FEED”, but rather “POUR COFFEE into (this specific) CUP” or “FEED (this specific) DOLL”, where the cup and the doll are specified by the directedness of the gesture. Gestures with both ACTION\_BASED iconic aspects and aspects of directedness are analyzed in more detail in Chapter 9.





Figure 6.13: Mediated pointing with an unrelated object.

## 6.7 Object-gesture variants of pointing

Sometimes children use objects to point at something, rather than pointing with their bare hands. Pointing with objects can be done in some principally different ways, and that is the topic of this section. Figure 6.13 shows Alice (36 months) performing a deictic gesture involving the use of a pen.<sup>15</sup> Alice’s mother first asks “where is the bathroom then?” (*var e badrummet då?*). Alice then points towards the bathroom with the pen she holds in her hand, and at the same time she says “it’s there+OBJECT\_POINT” (*de e där*). In this pointing gesture, the specific identity of the object does not seem to play any particular role in the gesture. There is no specific relation between pens and bathrooms that affect the meaning of this gesture in some specific way. Instead it is rather only the physical shape of the pen that affords its use as an extension ‘lever’ (cf. Butterworth, 2003).

Similar forms of mediated pointing gestures can also be found in the context of interacting around computers where it is a common practice to point to something on the computer screen by using the mouse arrow. For example, Ivarsson (2010) argues that “The cursor [the mouse arrow] is itself interesting as an indexing resource because, unlike say an outstretched index finger, the cursor is always present in the visual field defined by the screen.” Since the mouse arrow is always present, the person who performs the act of pointing with the mouse arrow must somehow do something in order to turn the mouse arrow into a communicative device, to provide what Kendon (2004) would call features of manifest deliberate expressive-

<sup>15</sup>This example is taken from Andrén (in press a), which was not constrained to the age range studied in this thesis. However, since the example is rather shown to make a principal point about one possible way to point with objects, this example is included anyway, even though Alice is 36 months old in this example. There are a fair amount of similar instances within the age range studied in this thesis too.

ness. In this regard Ivarsson mentions zigzagging of the mouse arrow as one of the available methods for achieving a communicative appearance.

The gesture shown in Figure 6.14 is different. Here, the identity of the object used for the pointing gesture *is* related to the target that is pointed out. Bella (22 months) is doing jigsaw puzzles together with her mother and in the course of action Bella grabs a piece of the puzzle and holds it over a location where some piece eventually will have to fit, but where the currently held piece does not. She does not attempt to put the piece in place, but instead stops her downward movement toward the location and simply holds the piece over the puzzle. At the moment in time when the downward movement of the hand has reached its destination, about one inch above the table, she utters “*there*+OBJECT\_POINT+HEAD\_SHAKE” (*dä*), and at the same time she also performs a lateral HEAD\_SHAKE which has the effect of negating the other components of her multimodal utterance, as if to say “this piece won’t fit there”. The HEAD\_SHAKE continues for a short while after the utterance is done. After this she goes on to put the piece at the correct place instead.<sup>16</sup> The specific identity object is relevant to the gesture, because Bella is not merely pointing to an empty location in the puzzle with “some object” (as in the previous example with Alice). Bella is rather establishing a relation between the object held in her hand and the target that she is pointing to. Hence, Bella is drawing on the particular knowledge of how to make a jigsaw puzzle in this gesture. Streeck (2008a, p. 297, my italics) captures this nicely when he writes that “minimal configurations and simple strokes suffice to evoke *things and events of the kinds that everyone knows*, that are part of the participants’ common ground, either because of their membership in a culture or because of the shared understandings that the discourse so far has yielded, or both.” The quote from Streeck also highlights the third-person intentionality that is involved — what anyone would intend when directing a piece of a puzzle towards the location where it may potentially fit, in relation to the typified relevancies of the actions involved in doing puzzles.

Another example in the corpus of a similar phenomenon is when Bella holds a memory card, directed to the other, matching, memory card. In this case too a relation is established between the object held in the hand and the target of the pointing gesture. Returning briefly to the previous example where Alice used an object (the pen) whose identity was not related to the target, one may note that an observer of this gesture implicitly draws on the knowledge that the pens are *not* related to bath-

---

<sup>16</sup>These sorts of actions are also analyzed in Section 9.5.1 on page 295, in terms of HOLD\_TOWARDS.

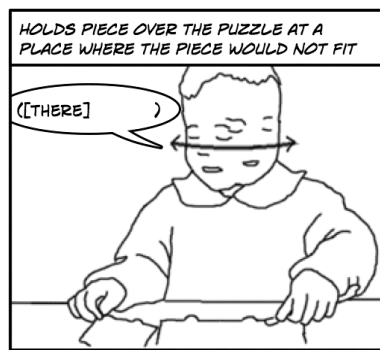


Figure 6.14: Mediated pointing with a related object.

rooms, and that the gesture can be interpreted without reference to such knowledge without missing out on the meaning of the gesture. Hence, the identity of the object is something that should be disattended (cf. the *disattend track*, Goffman, 1974). In yet other cases, there is an object held in the hand which is not used in the pointing gesture. For example, sometimes there is an object held in the hand, such as a bottle, and then the index finger is stretched to point to something. In this case it is not only the identity of the object that is to be disattended, but the object altogether. It may seem almost trivial that one should disattend the object in such cases, but at second thought, it is not. There must be some sort of knowledge or ability involved that allows an observer of such a gesture to make the interpretation, no matter how “effortless”, that the object that “happens to be held in the hand” is not part of the expression.

To summarize: In this section three principally different ways of pointing with objects have been presented. First there are those pointing gestures where the object should in fact not be attended to, as was discussed last. Second, there are those pointing gestures where the identity of the object should not be attended to, but merely the physical shape of the object and the way that this shape is directed towards a target, as in the example with Alice who pointed to the bathroom. Third, there are those pointing gestures where the identity of the object is of central relevance to the meaning of the gesture, as in the example where Bella used a piece of a puzzle to point to an empty location in the puzzle.

Arguably, there is also a fourth variant, which is gestures where the action of holding an object towards another object is relevant in such a specific way as to count as a recognizable kind of action in itself. An example of this is when holding a spoon towards the mouth of a doll, as if feeding the doll. However, in cases like that, the recognizable character of the action gestalt tends to take over, and the gesture is rather recognized as a gesture with ACTION\_BASED iconic aspects instead of a

pointing gesture, and the directedness of the act rather tends to appear as a sub-part of this overall action-scheme of feeding. Nevertheless, the directedness is still there, in the sense that it shows who it is that is being fed.

## 6.8 Discussion and conclusions

As mentioned in the beginning of this chapter, most examples that were shown were concerned with variants of pointing. Other deictic forms of gestures, such as GIVE, SHOW, and PLACE, are analyzed in more detail in Chapter 9.

One general point to emerge from the examples in this chapter is how it is impossible to treat “deictic gestures” in isolation from iconic and conventionalized aspects (cf. Zlatev & Andrén, 2009), even though this has been the dominant practice in research on children’s gestures at least up to now. The examples showed both iconic and conventionalized aspects to be present in pointing gestures. As a consequence of the very same argument, it has also been argued that gestures which may not primarily tend to be described as “deictic” gestures, such as the emblem HELLO or many ACTION\_BASED iconic gestures, are also often performed in direction towards something in the physical environment, and that they may also serve to ground referents. That is, the “dimensional” rather than “categorical” treatment of gesture that has been recommended by Kendon (2004) and McNeill (2005) ought to find its way into research on children’s gestures too.

In fact, most of children’s gestures at the ages studied here are directed to elements in the physical environment (i.e., the concrete *action-space*) in some way or another, either by being directed towards elements in the environment, or by incorporating objects in the performance of the gesture itself, as in object-gestures. Gestures that take place in an abstract *gesture-space* are rare, and typically appear toward the end of the studied period. The fact that children’s actions and thinking is to such an extent here-and-now oriented goes a long way in explaining the dominance in deictic aspects of gesture in their repertoires.

Another general point to emerge from the examples analyzed in this chapter is how pointing can be performed in a plethora of different ways. Hence, in research dealing with children’s pointing gestures or deictic gestures in general, a productive way of proceeding may be to make more detailed observations of the nature of, and complexity in, children’s pointing gestures, to see how the children continue to develop refined skills in using pointing gestures after the initial emergence of “pointing”, or perhaps some slightly more specific use of pointing such as “declar-

ative pointing” (Bates et al., 1975). Pointing gestures are clearly not just of “one kind”, as there is a lot of internal variation within this “category”, and there is clearly a further developmental story to be told regarding children’s pointing as a matter of a development “within” the category of pointing. For example, the “whole hand pointing” gestures that were analyzed in this chapter turned out to be a very heterogeneous set of gestures, which indicates that the use of such broad categories in research on children’s gestures may be questionable. The pointing gestures that involved objects were also of several different kinds in the sense that the relevance of the object that was involved in the gesture could be quite different in different instances. The TRACING variants of pointing that were analyzed were also quite heterogeneous, although in this case they still shared some basic properties, at least if disregarding the tracing gestures that seemed to be more related to pointing to a set of objects rather than to the shape of an object. The TRACING gestures in the data analyzed here all appear in contexts of interacting around pictures. The TRACING constitute one example of a more particular way of performing pointing gestures that seems to emerge during the 18–30 month period of children’s lives.

One may also consider other aspects of deictic gestures, apart from their functions of invoking targets of various kinds, such as the *responsive* and *response eliciting* functions of pointing gestures, and more generally, how the pointing gesture functions in the *operation-act-activity interdependence* nexus (see Chapter 3). For example, the sustained character of some pointing strokes was discussed in this chapter. It was argued that sustained strokes served to elicit responses from parents (cf. Andrén, in press b).<sup>17</sup> This means, among other things, that sustained pointing gestures are not only about achieving joint attention to a pointed out target as such, because the children are typically not satisfied with mere acknowledgments that the pointed out target has been noticed, but they keep on sustaining the gesture until the parent responds in a way that is relevant to the ongoing activity. On a more abstract level, these sustained strokes may be understood as a modulation of the pointing gestures on the *operation level* which has an effect on the overall way that the pointing gesture on the *action level* functions in the interaction on the *activity level*, i.e., a matter of *operation-act-activity interdependence*. It was also argued that the ways in which the sustained strokes turn out is a result of an unfolding and ongoing adaptation to the way the Other acts during the very performance of the gesture, which means that it must be understood as an inherently interactive phenomenon. This use of

---

<sup>17</sup>In a few fairly infrequent cases a “sustained” stroke may be nothing more than a “left over” stroke as the child becomes distracted by something in the middle of the performance of a pointing gesture, but this is not what the major bulk of sustained pointing strokes are about.

sustained strokes in some pointing gestures appears already in the first observations at 18 months in the data studied here, so the question of when this phenomenon might appear in development would have to be answered in future research.

Another finding presented in this chapter is how the children often construct their speech and (pointing) gestures in *response* to previous utterances from their parents in ways that incorporate elements from these previous utterances as part of their own current utterances. For example, it was shown how a child pointed to a location while saying “there+POINT”, and that this effectively meant “that piece of the puzzle should be at that location” due to the way the utterance built upon the mother’s previous turn. The exact same general phenomenon was involved when Alice used a pen to point to the bathroom, while saying “it’s there+OBJECT\_POINT”. In these cases the pointing gesture does not just specify a referent from the own utterance, but a referent that was introduced in the parent’s previous utterance. That is, the act of pointing becomes what it is due to the way it relates to previous actions, as a matter of act-activity interdependence. Being able to construct utterances in this responsive way obviously hinges on a certain degree of comprehension of other people’s prior utterances that the child’s present utterance builds upon, and one might therefore suspect that there is an interesting developmental story to be uncovered here. These aspects of act-activity interdependence in children’s gestures remain largely unexplored, with only a few rare exceptions (e.g. Bruner & Garton, 1978; Golinkoff & Gordon, 1988; Wootton, 1994, 1997; Zukow-Goldring, 2001; Corrin et al., 2001).<sup>18</sup> Most of the time children’s gestures are studied as isolated “expressions”, as if the meaning of the act was solely due to the act itself, without taking into account if and how these gestures serve to create a coherent interaction by being oriented to, and building upon, previous utterances in various ways.

None of these arguments relating to the operation-act-activity interdependence nexus are in themselves particularly new, as they have been central to interactionally and contextually oriented lines of inquiry, such as *conversation analysis*, for a long time. Nevertheless, it seems that they have so far not made it into “mainstream” research on children’s gestures. To be sure, most researchers would probably agree on general statements such as “communicative acts are context dependent” and similar, but yet it seems that many aspects of context are rather taken for granted than being

---

<sup>18</sup>One exception to this general situation is a number of studies that indicate that parents often respond to children’s pointing gestures by naming what they interpret the child to be pointing to (Bruner, 1978a; Hannan, 1992; Marcos, 1991; Marcos et al., 2003; Kishimoto et al., 2007), and decreasingly so after 20 months (Hannan, 1992), i.e., after TP#1. However, this is focused on how parents respond to their children, rather than the other way around. Caselli (1990, p. 65)

included in the analyses as a central topic of inquiry. I hope that these examples have shown that act-activity interdependence is an essential factor in the specification of how it is that gestures come to mean the things they do.<sup>19</sup>

---

<sup>19</sup>As suggested in Section 2.2.3, for the most part, the treatment of “context” in research on child gesture is treated only in relation to processes of *decontextualization* (cf. Werner & Kaplan, 1963, p. 94), such as children’s increasingly “autonomous” use of various gestures and words outside the quite restricted contexts and activities in which they are first mastered (e.g. Volterra et al., 1979; Goodwyn & Acredolo, 1998). There is certainly nothing wrong with studying such processes of decontextualization as such. The point I wish to advance here is just that treating context only in terms of decontextualization processes is somewhat one-sided and that more attention ought to be paid to aspects of act-activity interdependence in children’s gesture. There is certainly also developmental processes that are best characterized as an increasing *contextualization*, since children not only learn to abstract from specific activities in development, but also become increasingly able to handle various specific activities in more complex and well-informed ways. In Section 2.2.3 it was also suggested that the term *trans-contextualization* may be a better overall term, since it captures both aspects of decontextualization, generalization, and abstraction in development as well as the fact that communicative acts are also always contextualized in various ways. One reason for the relative neglect of aspects of context, act-activity interdependence, and interaction in child gesture research may be the dominant focus on quantitative research, although it must be pointed out that quantitative research need not be insensitive to these things *per se* (e.g. Bavelas et al., 2002; Andrén, in press b).





# CHAPTER 7

## Iconic aspects of the children's gestures

*As it occurs in philosophy, similarity tends under analysis either to vanish entirely or to require for its explanation just what it purports to explain.*

Goodman (1970, p. 29)

*[...] it is only by recognizing the reality of iconic motivation that iconicity can be opened up as a domain for semiotics.*

Sonesson (2001, p. 86)

### 7.1 Questions asked

The two main questions addressed in this chapter are: *What sort of gestures with iconic aspects do the children perform? What provides for the “transparency” of meaning in iconic gestures?* These two questions will be addressed in three main sections of this chapter. First there is a conceptually oriented section which deals primarily with the second question. Then there are two empirically oriented sections that deal with both of the questions. The first of those two sections provides examples of gestures with ACTION\_BASED iconic aspects and the second section concerns gestures with INDIRECT iconic aspects.

Although TRACING is also an iconic aspect, gestures that involve aspects of TRACING have already been discussed in Chapter 6, since gestures with aspects of TRACING appear exclusively in gestures that also have deictic aspects in the data studied here. For this reason TRACING is not discussed further in the present chapter.

## 7.2 What provides for the “transparency” of iconic gestures?

In Section 3.3 it was argued that people tend to see “through” gestures. More specifically, when someone sees a gesture, they typically do not attend explicitly to the movements of the articulating body parts as such, but rather experience some sort of meaning (cf. Kendon, 2004, p. 358; Streeck, 2009b, p. 139). In most cases, this goes without effort and without explicit inferential reasoning and this property was therefore called *gestural transparency*. How does this transparency come about in the case of iconic gestures?

### 7.2.1 Different views on the nature of iconicity in gesture

Perhaps in part due to the effortless experiential quality of gestural transparency, gesture has sometimes been claimed to be a kind of “universal language”, assumed to be a kind of communication with “natural” meaning that is understandable across cultures. When gesture is thought of in this way it is often the *iconic* (and *indexical*) aspects of gestures that are assumed to provide this naturalness of meaning in gesture, as opposed to the conventionalized aspects which are per definition culturally dependent and learned. Wundt (1973 [1921], p. 56) wrote:

For the practical purpose of communication, each language appears to be a conventional system of signs, the use of which requires special practice and instruction. With gestural communication the case is substantially different [...]. In its most important and prevalent forms, even if not in all, it is a kind of universal language having many components in common, frequently despite the most varied conditions of origin. Notwithstanding different forms of its development, comparable to the ‘dialects’ of a spoken language, communication is often possible. [...] gestures gain an originality and naturalness such as speech neither possesses today nor has ever had in any forms hitherto uncovered by linguistics.

From this point of view, gestural communication is considered to be a “natural” form of expression, largely due to its iconic features. This can be idealized into the theoretical position that the transparency of iconic gestures is due to a universal human trait which can be explained directly with reference to species-general human cognitive mechanisms, without reference to any specific socio-cultural practices/processes and conventions. This can be (and has been) problematized in two main ways.

First, one may point out that there are indeed plenty of conventionalized aspects, not just in language, but in gesture too (Kendon, 1984). This means that there are many aspects of gestures that are not universal.<sup>1</sup> However, this has no direct bearing on the question of whether *iconic* aspects of gesture, as opposed to conventionalized aspects, may somehow have the potential to be transparent in a universal and natural way.

Second, one may put into question the very idea that all iconic aspects of gesture are per definition “naturally” transparent (e.g. Streeck, 2008a, 2009b; Bouissac, 2008). Streeck (2009b, p. 120) argues that there is “a widely held belief, shared by ordinary interactants and researchers, that there is a *natural* relationship (usually of similarity) between the gestures that our hands make and the visual features of the objects and events that they depict” — a view that Streeck is skeptical towards, at least when formulated as a general fact that would pertain to all iconic gestures. According to the position taken by Streeck (2008a, p. 285), “The view that ‘iconic’ gestures uniformly function by way of some resemblance between signifier and signified is rejected”. It is not crystal clear whether Streeck wants to reject the notion of iconicity altogether, or whether he is mainly aiming to provide an alternative more multifaceted way of thinking about iconicity.<sup>2</sup> Nevertheless, in strong(er) versions of this position, the idea of iconicity as a sufficient ground for establishing the relation between expression and content/referent in a semiotic sign is doubted as such. Two proponents of this strong version are Goodman (1970) and Eco (1976). Both of them argue that the presumed iconicity of any mode of sign-production must instead be understood as a matter of *conventionalized* modes of sign production.<sup>3</sup>

---

<sup>1</sup>The issue of conventionality in gesture was discussed in relation to the issue of the upper limit of gesture in Section 2.3.

<sup>2</sup>After listing a number of quite disparate iconic gestures, Streeck (2008a, p. 285) writes that “Maybe in some language of theory, ‘similarity’ can be given a definition that is so abstract that it is possible to accommodate even these visually dissimilar signifieds rendered visible by two hands.” Instead he focuses his analysis on uncovering various techniques for achieving appearances of similarity in gesture.

<sup>3</sup>Streeck (2008a, 2009b) cites Goodman and draws on Goodman’s arguments against the soundness of the notion of iconicity.

This view on iconicity can be idealized into the theoretical position that the transparency of iconic gestures is mainly due to various conventionalized practices of sign production and sign interpretation, and that it is not a direct consequence of a self-sufficient cognitive capacity that would make (all) iconic gestures “naturally” transparent.

To conclude: Ideally speaking, there are those who argue that the transparency of iconicity comes naturally as a direct result of similarity and the capacity of human cognition to recognize similarity (*natural transparency*) and there are those who argue that it is rather a matter of conventionality, either wholly, or in part (*convention-based transparency*).

### 7.2.2 Mirror-neurons: A natural or convention-based affair?

The main proposal of a purely cognitive mechanism to be underlying the transparency of both action understanding in general and gesture understanding more specifically are the so-called *mirror neurons*.<sup>4</sup> Mirror neurons have been shown to fire both when macaques see an action and when they perform the action themselves (Rizzolatti et al., 1996; Gallese et al., 1996) and a number of brain imaging studies have suggested that parallel mechanisms exist in the human brain (cf. Gallese & Goldman, 1998; Rizzolatti & Craighero, 2004) and that they may explain the human ability to understand actions, gestures, and the signs of signed language (McNeill, 2005, Chapter 7; Emmorey, 2006). The question is, is the presence of mirror neurons enough to provide action or gesture understanding, or is something more required? In addition, one may ask whether mirror neurons are somehow innate, or whether they are rather a result of participation in various (partly conventionalized) activities, so that their organization must in part be understood as a the result of learning?

In order to work this out one must distinguish clearly between cognitive abilities in the sense of the *general capacity* of members of a certain species, and cognitive abilities in the sense of *actual abilities* of specific individuals with a certain life-history of learning. Even though the presence of, or possibility to develop, mirror neurons is no doubt an important piece of the general *capacity* for interpersonal action/gesture understanding, it is a mistake to assume that the mirror neurons by themselves provide the *substance* of this understanding.<sup>5</sup> As argued by Arbib et al.

---

<sup>4</sup>A number of authors have pointed out that action understanding is not necessarily equivalent to gesture understanding in all respects (Gallagher, 2005; Zlatev, 2008b).

<sup>5</sup>In addition, several authors have pointed out that the mirror neurons are only one part of the

(2005, p. 239), “mirror neurons are not innate but instead correspond to a repertoire of learned actions and learned methods for recognizing those actions” (see also Heyes, 2010). It has been shown, for example, that such obviously culturally defined activities as ballet and capoeira dancing shape mirror neurons differently in practitioners of each of those two cultural practices (Calvo-Merino et al., 2005).<sup>6</sup> When capoeira dancers and ballet dancers view others that engage in capoeira dancing or ballet dancing, their mirror neurons respond differently, depending on which of the two activities that they are familiar with. That is to say, on the level of actual abilities, the mirror neurons of capoeira dancers and ballet dancers respond differently, so that partly different actual abilities are involved in the two groups. Though, on the level of general capacity, it is safe to say that both capoeira dancers and ballet dancers “use” mirror neurons to make sense of what they see. Hence, when someone makes the statement that “mirror neurons provide a general mechanism for action understanding”, the truth of this statement depends on whether it is the level of potential capacity that is discussed, or the level of actual abilities.

This is consistent with a view of gesture and action as “skilled performance” (cf. Kendon, 1990, p. 4), which makes reference to more *particular* forms of knowledge and previous experience(s) as part of the constitution of gesture as an actual ability.<sup>7</sup> That is, the “iconicity” of content-loaded gestures that invoke interpretation in terms of typified action may very well be dependent on *socially* typified and praxis-based experiences of various sorts, both in their production and their comprehension. It should also be made very clear that many studies on mirror neurons have been concerned specifically with reaching for and grasping an object, which are both very generic aspects action, and this may be expected to develop early and in a quite similar way across individuals and across cultures. As a result, many of these studies have little to say about action understanding in terms of more particular kinds third-person intentionality (such as being familiar with ballet or capoeira). This may be deemed a shortcoming, since such typified and specific aspects of third-person intentionality are indeed characteristic of many types of action. One problem associated with mainly studying generic actions is that it often leads to the fal-

---

neural structures that are involved (Arbib, 2005, p. 110; Csibra, 2007, p. 435; Gallese, 2008, p. 325). That is, even if one would maintain a pure version of the cognitive transparency hypothesis, mirror neurons are not in themselves enough to explain interpersonal understanding.

<sup>6</sup>Even the most famous proponent of imitation of others’ actions as an innate ability writes “Are mirror neurons innate? This may be the case, but the role of experience in forming mirror neurons deserves more consideration than it has been given” (Meltzoff, 2005, p. 69) and “young infants cannot imitate the full range of gestures copied by older children, and there is development in imitation.” (ibid., p. 71).

<sup>7</sup>A similar argument was put forward in relation to the notion of Theory of Mind in Chapter 3.

lacy of treating objects solely as “physical objects”, without a cultural identity — what Rodríguez & Moro (2008, p. 89) calls the “naturalistic view of the object”, according to which, “children give meaning to objects in a natural, direct, and spontaneous manner, without the need of others” (see also Sinha, 2005; Rodríguez, 2007; Rodríguez & Palacios, 2007; Sinha & Rodríguez, 2008). This is problematic since it amounts to a neglect of the role of conventionality in semiosis. The issue of typified third-person intentionality in children’s iconic gestures will be further discussed in the analysis later in this chapter.<sup>8</sup>

To conclude: On the level of general capacities of *homo sapiens*, action understanding may well be characterized in general and generic terms. However, on the more specific level of the actual abilities involved in producing and understanding particular actually occurring actions, it is not sufficient to explain action understanding with recourse only to general neural mechanisms. For a full account of action understanding, on the level of actual abilities, one must also include the conventionalized structures of the social activities that the subject must have become familiar in his or her life history. Conventionalized practices typically exist prior to the individual’s first acquaintance with them, and for that reason these structures cannot be said to be a direct product of this individual’s cognitive abilities, even though the individual must of course have the required cognitive prerequisites (on the level of general capacities).

What are the implications for the idea of iconicity in gesture as a “universal language?” The gist of the arguments above is that at least some aspects of iconicity are rather a matter of typified third-person intentionality, which requires familiarity with specific actions and objects as connected to specific practices. As a consequence, in order to see the “iconicity” of many ACTION\_BASED gestures, one must be familiar with the practices and ways of acting that the gesture draws on. Sometimes the recognizability of these gestures may require very specific experiences, such as knowing the specific movement that one would perform to insert a 5.25” floppy disk in the old model of the 1541 disk drive for the Commodore 64 home computer and close the lid, which is detailed enough to be different from the movement one would make to do the same thing on the subsequent models of the 1541 disk drive. Action understanding relating to such specific actions quite obviously would not make sense in the same way to someone who does not know what a Commodore 64 computer is. In other cases the actions that are depicted may be very generic in nature, such as grabbing, putting, and giving/taking objects in gen-

---

<sup>8</sup>Cognitive aspects of third-person intentionality are also discussed in Section 9.8.3.

eral, as argued above. In the case of more generic ways of acting, there are no reasons why they may not be recognizable across cultures, in the guise of a “universal language”. This points to the need of a more differentiated view on the on the factors that may underlie the transparency of iconic gestures.

### 7.2.3 Different forms of iconicity

Sonesson (1992; 1994; 1997; 2008; 2010) makes a distinction between what he calls *primary iconicity* and *secondary iconicity*. As it happens, this distinction is quite similar to the distinction between natural transparency and convention-based transparency in iconic gestures. In the case of primary iconicity, someone first perceives a similarity between an expression and some content/referent, and is therefore able to see the expression as the signifier in a semiotic sign (Sonesson, 1994, p. 281). In the case of secondary iconicity, the situation is the reverse: The knowledge about the existence of a sign function between an expression and a content/referent makes it possible for someone to see the similarity between the expression and the content/referent (Sonesson, 1997, p. 741; Sonesson, 2008, p. 51). In Sonesson’s account, the knowledge that enables someone to see the iconicity in a semiotic sign is a matter of conventionality.

Primary iconicity first and foremost obtains in depictions where the depicted motive is sufficiently similar to something that is well known and typical in the Lifeworld, and perhaps human bodies and faces are among the most prototypical phenomena in this regard (cf. Sonesson, 2010, p. 24). The illustrations used to show various gestures in this thesis may serve as an example of this (see for example Figure 2.2 on page 58). These illustrations are (hopefully!) sufficiently *similar* to real people for an observer to spontaneously see a resemblance to people, without anyone having to tell the observer about it. At the same time, the illustrations are also sufficiently *different* from real people, so that an observer would not mistake the marks on the paper (or on the computer screen) for real people.<sup>9</sup> In an experiment carried out by Hochberg & Brooks (1962) it was found that a child with a history of close to zero acquaintances with pictures and no explicit instruction at all of how to interpret pictures was able to interpret both photographs and line-drawings precisely as pictures when first confronted with them at 19 months. This suggests that it may

---

<sup>9</sup>Whether an observer would be able to see more particular things, such as what sort of clothes the depicted people are actually wearing, and so forth, may not be equally obvious. Hence, it should be noted that when I discuss whether it is possible to see “people” in the pictures, the discussion is relative to assumptions of a certain level of specificity in the required interpretations. This is discussed in more detail later in this section.

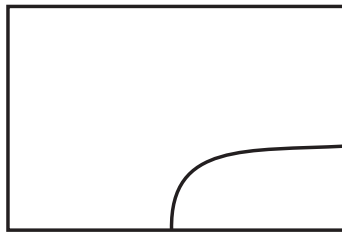


Figure 7.1: A doodle

indeed be possible to see pictures as a semiotic signs without prior experience with pictures and without any familiarity with conventionalized interpretative schemes and modes of picture production. In the case of gesture, some ACTION\_BASED iconic gestures are perhaps the main candidates for primary iconicity (or natural transparency). The reason for this is that *some* ACTION\_BASED gestures may be so similar to the way the body ordinarily acts in the world so that an observer may be able to see this similarity spontaneously. Indeed, as pointed out in both Chapter 2 and Chapter 5, ACTION\_BASED gestures may sometimes be so similar to ordinary actions that it is unclear to an observer whether they are gestures or not.

When it comes to secondary iconicity, Sonesson (1989, p. 223) provides the example of so-called *doodles*.<sup>10</sup> Doodles are images that are ambiguous in meaning and that can be seen as depicting several kinds of contents/referents. In a sufficiently abstract and ambiguous doodle it may not even be clear to the observer that the doodle is an iconic semiotic sign. It could perhaps also be seen an arbitrary symbol, or just a graphical pattern. Figure 7.1 shows an example of a doodle made up by myself. Possible interpretations of this doodle may be that it depicts a shoulder or that it depicts the front of a car. It could also seen as a curve in a mathematical diagram, and many other things as well. More crucially, it could perhaps also be seen as an arbitrary symbol of some kind (with unknown meaning), and thus not an iconically motivated sign at all. However, if the caption text below the doodle in Figure 7.1 would read “A shoulder” or “A logarithmic function” instead of “A doodle”, most people would probably not have any difficulties in seeing the figure as depicting a shoulder or a mathematical function in the form of a diagram.<sup>11</sup> Hence, when some sort of knowledge about the status of the doodle as an iconic sign is provided, an observer may have no problems in seeing the iconicity that is nevertheless there, without anyone having to explain to the observer how to

---

<sup>10</sup>The term ‘doodles’ stems from a popular book by Roger Price (1953).

<sup>11</sup>See Mittelberg (2008) for a discussion of diagrammatic iconicity in gesture.



## 7.2. WHAT PROVIDES FOR THE “TRANSPARENCY” OF ICONIC GESTURES?

---

go about interpreting various parts of the picture itself. That is to say, secondary iconicity is also a genuine form of iconicity. It is not just an illusion of iconicity. It is just that something more is required before an observer will be able to see this iconicity in a sufficiently constrained way.

Just like doodles, many gestures are also underdetermined in meaning in themselves if considered in isolation from co-occurring spoken utterances or other “outside” factors that may contribute to the meaning of the gesture. In the words of Kendon (2004, p. 161):

In many cases [...] gestural actions that are interpreted as being representational are drawn from repertoires of forms of action that are widely used. These actions have very general meanings — perceived on their own, the user may be recognized as ‘pointing to something’, ‘describing the shape of something’ or ‘acting out an action of some sort’. When the way such actions are used is taken in conjunction with the utterance’s verbal component, however, they are then understood in a much more specific way.

However, in the case of gesture, it may not be so much the status of the gesture as a semiotic sign that is unclear. As Kendon (1978, 2004) has argued, most gestures are identifiable as gestures directly on basis of the character of the movements themselves. One may therefore have to distinguish between several levels of ambiguity. The ambiguity that is involved may be found on different levels in different cases:

- (A) For some iconic expressions, such as more abstract variants of doodles, it may be unclear whether it is a sign at all. For example, a doodle may be mistaken for a decorative graphical pattern that is not supposed to be interpreted as a semiotic sign.
- (B) At other times it may be clear that an iconic expression is an expression of some sort, but it may not be clear that it is an iconic expression specifically or whether it is an arbitrary symbol of some kind.
- (C) At yet other times, as in the gestural movements described in the quote from Kendon above, it may be clear that something is a semiotic sign, and more specifically, that it seems to be a deictic sign, or an iconic sign, and so forth, and if it appears to be an iconic sign of some sort, it may only be the more specific meaning of the sign that is unclear.

Accordingly, an expression may be ambiguous in different ways and to different extents, and instead of a binary distinction between primary and secondary iconicity, one may have to consider several possible levels of ambiguity in iconic expressions.<sup>12</sup> Formulated otherwise, iconic expressions may vary to the extent that the expression is itself enough to provide the basis for a sufficiently constrained interpretation (“primary iconicity”), and the extent that other “external” *enabling factors* (such as a convention) play a part for an observer to arrive at a sufficiently constrained meaning (“secondary iconicity”). As Calbris (1990, p. 38) argues, “simply finding one motivation (an analogical link) among several possible ones does not constitute a successful interpretation.”

At this point one may also ask why the enabling factor behind the recognition of a resemblance between the expression and the content/referent would necessarily have to be a convention. As far as I can see, it does not have to be that way — at least not when conventionality is defined the way it is defined in this thesis.<sup>13</sup> In principle, any sort of factor that helps constraining the possible interpretations of an iconic gesture may be relevant, and that may also include indexical features (cf. Streeck, 2008a, p. 285) as well as other contextual factors overall (cf. Ahlner, in press), such as what has happened so far in the situation of a social encounter.<sup>14</sup>

Following Schutz & Luckmann (1973) one may also note that what counts as a sufficiently constrained interpretation of something is *always* relative to the purposes at hand in a specific situation.<sup>15</sup> Schutz & Luckmann (1973, p. 10) writes:

[...] it may not be enough for me to recognize a plant as a mushroom if I anticipate picking it, for the subordinate typifications ‘palatable’ or ‘poisonous’ are relevant for me. On the other hand, I may, while taking a walk, simply perceive ‘mushrooms,’ without my being motivated to an explication of ‘edible mushroom,’ ‘poisonous mushroom.’

Hence, as an analogy to the mushroom example, one may say that it is not only the

---

<sup>12</sup>In fact, different doodles may be ambiguous at different levels. Some may be hard to identify as signs at all, whereas some others rather suggest a choice between a more limited number of interpretations, all of which may be iconically motivated.

<sup>13</sup>Conventionality was discussed at length in Section 2.3.3.

<sup>14</sup>For example, one could imagine holding a paper copy of the doodle shown in Figure 7.1 just above the shoulder. If someone does this, an observer may also be able to see the similarity between the shape in the figure and the real shoulder, partly because of the contiguity (i.e., a form of indexicality) between the doodle and its content/referent, and partly due to the communicative appearances of holding a picture in this way (a variant of *SHOW*, where two objects are related to each other as exemplars of the same class, by holding them close together).

<sup>15</sup>This is not to deny that some levels of interpretation may be more typical than others.

case that something outside the iconicity in a doodle or a doodle-like gesture may be required to make interpretation of the iconic aspects possible — it is also the case that factors outside the expression will motivate what counts as a sufficiently constrained interpretation in a given situation. For example, in the case of gesture, the extent to which it may be sufficient to see that some specific gesture might signify something like “playing tennis” may vary. In some contexts “playing tennis” may be precisely what is signified. In another context, the same gestural movements may have a more specific meaning, such as being an enaction of a one-handed backhand as opposed to a two-handed backhand. From this point of view, no interpretation of a sign is ever entirely context-independent.

All in all then, a better way to think of the issue of primary and secondary iconicity, at least for the purposes of this thesis, may be to view them in the larger perspective of what factors that contribute to the recognition of a *specific* iconically motivated meaning in a sign. This includes (a) recognizing it as a sign, (b) recognizing it more specifically as an iconically motivated sign, and (c) recognizing a specific iconically motivated meaning in the sign that is sufficiently constrained for the purposes of a specific situation. Instead of formulating the question in terms of primary and secondary iconicity directly, one may therefore instead talk about two different aspects, that are present to different extents in every iconic sign. For gestures then, these two aspects can be understood as:

**DIRECT ICONICITY:** The extent to which a directly perceivable resemblance between the movements that are involved in the gestural performance (the expression) and some content/referent contribute to the recognition of a specific iconically motivated relation between expression and content/referent that is sufficiently constrained for the purposes at hand in a specific situation.

**IMPLIED ICONICITY:** The extent to which other factors than a directly perceivable resemblance between expression and content/referent contribute to the recognition of a specific iconically motivated relation between expression and content/referent that is sufficiently constrained for the purposes at hand in a specific situation. “Other factors” can be, but do not have to be, conventionalized factors.

Finally it should also be pointed out that failure to perceive a resemblance between expression and content/referent may either be due to *too little* similarity between expression and content/referent (as in very abstract doodles, and gestures that are like doodle-like in this respect) or due to *too much* similarity between expression and content/referent (as in the case of ACTION\_BASED gestures that are too similar

to actual actions to be recognized as having gestural qualities) (cf. Sonesson, 2010, p. 28). In both these two cases, the reliance on other factors for the expression to appear as a iconically motivated sign will be particularly pronounced.

### 7.2.4 Children and the transparency of iconic gestures

Namy et al. (2004) performed a study of children's abilities to learn gestures from others. Among other things they found that there was no advantage for signs that had an iconic aspect for children at 26 months. In a similar way, Tomasello (2008, p. 149) reports that:

we have found that young children have much difficulty understanding [...] creative iconic gestures, not only when they are used to request objects [...], but also when they are used to simulate for infants what action they need to perform to solve a problem [...].

In a case-study of one child, Caselli (1990) found that most early "iconic" gestures seemed to be learned from the parents. It is therefore unclear to what extent very young children are able to see the iconic aspects in these gestures, as they may simply be learned as "arbitrary" gestures. At least it indicates that very young children may not themselves be able to invent gestures with ACTION\_BASED iconic aspects to any considerable extent.

In a similar way, Nicoladis (2002, p. 244) argues that children start using iconic empty-handed gestures after they have learned to speak, in conjunction with increasing morphosyntactic complexity in development, and that earlier use of "iconic" gestures, such as the "symbolic gestures" described in infants by Acredolo & Goodwyn (1985), are probably best explained as a kind of conventional gestures that are learned from adults.

This should not necessarily be interpreted in an either-or sense, as if children of, say 18 months, are entirely incapable of dealing with iconicity in gesture. Nevertheless, it does suggest that children's abilities to understand and produce iconic gestures are limited and that they certainly do not come for free as an effortless natural transparency. Instead, learning of typified action-schemes from adults seems to be an important factor behind the emergence of early iconic gestures, which points in the direction of a learned and convention-based sort of transparency for many of these gestures. Formulated otherwise, it suggests that third-person intentionality may be an important factor in many of children's iconic gestures.

## 7.3 Gestures with ACTION\_BASED iconic aspects

In Chapter 5 it was shown that ACTION\_BASED iconic aspects are the most common kind of iconic aspect in the children's gestures. The purpose of this section is to take a closer look at some of these gestures.

### 7.3.1 Transparency in ACTION\_BASED iconic gestures

Figure 7.2 shows Alice (24;27) and her mother, playing with a toy figure (held by the mother) and an eraser (held by Alice). Before the gesture shown in the figure is carried out, Alice's mother holds up the toy figure, just as shown in the figure, while saying "I will disappear if you rub me out" (*ja kommer försvinna om du suddar mej*). The mother does not say this with her usual voice, but with a theatrical tiny little voice, as if it was the figure she holds in her hands that is talking. In response to this, Alice joyfully performs the ACTION\_BASED rubbing object-gesture that is shown in the figure, with a real eraser held in her hand, while saying "rub+RUB rub+RUB" (*sudda sudda*). The stroke of the rubbing is repeated twice, once for each of the two repetitions of the word "rub". Reduplication in both speech and in the gestural strokes is a common practice in situations like this, not only in the children, and it seems to be a way to indicate that the enacted action has a certain temporal extension — as an ongoing process.

Alice does not actually touch the toy figure with the eraser while performing the gesture, and therefore the gesture obtains a certain as-if quality. She also performs the stroke in a much quicker and approximate way than would be required when rubbing something out for real. This also contributes to the as-if quality of the gesture. Therefore, even though there is an object involved here and even though the eraser is not used as if it was something else (i.e. "using a banana as if it was a telephone") than an eraser, there is clearly differentiation between expression and content here. The practical act of rubbing is implied, but not actually performed. All in all then, the action is both sufficiently similar to the kind of act that serves as the content of the gesture to be readily recognizable as an instance of rubbing and sufficiently different from it not to be mistaken for an actual attempt to rub it out for real, which would be inappropriate in this pretense play context.

It is clear that the action that Alice performs in a gestural manner here is a typified action. There are several indications of this: First, Alice's gesture invokes the third-person intentionality of "using an eraser", i.e., using the eraser *according to*

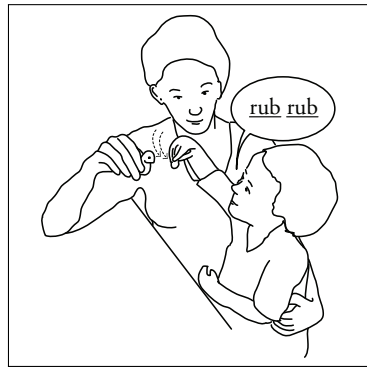


Figure 7.2: ACTION\_BASED as-if rubbing with an eraser (Alice 24;27)

*convention* (to rub something out) rather than some other more improvised and creative usage.<sup>16</sup> Second, Alice performs this gesture as a response to the mother's utterance, and it is in the mother's previous utterance that the idea of "rubbing" the toy figure is introduced. Hence, Alice's gesture displays an understanding of her mother's utterance, including the verbally encoded (lexicalized) idea of "rubbing", much in the same way as was discussed in Chapter 6.<sup>17</sup> Third, Alice also uses the verb "rub" (*sudda*) herself in coordination with the (as-if) act of rubbing, which shows that she knows what kind of action this is and that the action does not just happen to look like rubbing from an adult's point of view, or similar.

Another characteristic property of this gesture is that it is performed with an orientation to the material environment. The gesture is not performed in an *abstract space*, but is rather directed towards the toy figure as the target of her action in the concrete *action space*. That is, in addition to the ACTION\_BASED iconic aspects in this gesture, the directedness of this gesture towards something in the environment is a crucial part of how the gesture comes to have the meaning that it has.

In relation to the discussion about the transparency of iconic gestures in the previous section one may ask what factors that contribute to the meaning of this gesture. On the one hand the rubbing gesture is indeed quite similar to to ways the body normally acts in the world, especially since a real eraser is held in the hand. It could be argued that the movement itself is certainly constraining enough for the act to appear as a gesture of some kind (especially because there is no touch of the target), and even more specifically as an ACTION\_BASED iconic gesture of some kind (among other things because an object is involved and seems to be used). This

<sup>16</sup>Rubbing out toy figures is of course a less typical kind of usage of erasers, but the use of the eraser is nevertheless in accord with its conventional use.

<sup>17</sup>It is also apparent that Alice understood that the mother was not referring to herself when using the words "I" and "me" in her previous utterance, but to the toy figure.



Figure 7.3: Empty-handed ACTION\_BASED horse feeding (Harry 24;16)

means that there is certainly some factors of *direct iconicity* involved here.

On the other hand, there are also many other factors that help constraining the more specific interpretation of the movement performed with the hand, factors that help bringing out the more specific iconicity that is to be seen in the gesture. These are matters of implied iconicity. Perhaps the main constraining factor is the use of the verb “rub” together with each of the two strokes. Then again, the gesture does not simply evoke the concept “RUBBING” in a detached manner, since the gesture rather appears to be about “RUBBING *something*” out (the toy figure in this case). Hence, the indexical directedness of the gesture towards a target in the environment also contributes substantially to the meaning of the gesture. Since erasers are normally used precisely to rub *something* out, rather than some sort of detached target-less rubbing, it could be argued that the directed character of the action contributes to the recognizability of the iconic motivation of Alice’s gesture. The way the gesture responded to the mother’s utterance and displayed an understanding of this utterance is also a relevant factor for the meaning of the gesture since the mother’s utterance serves as a framing for the relevance of the gesture (if rubbing the toy, it will disappear). All in all, many factors contribute to the meaning of the gesture, some of which are a matter of direct iconicity, and others which are a matter of implied iconicity.

In the situation shown in Figure 7.3, Harry (24;16) and his mother are playing with toys. They have just finished assembling a plastic little fence around a toy horse and the mother concluded this by saying “now he stands there in the enclosure” (*nu står han där inne i hagen*) and Harry agreed to this by saying “mm” (*mm*). Then the mother said “he should eat a little” (*får han äta lite*) and “can you get him some hay?” (*kan du hämta lite hö te honom?*). Again, Harry agrees with his mother on this by saying “mm” (*mm*) and during this utterance the preparation phase of the gesture shown in Frame #1 has its onset.

In Frame #1 then, after his spoken utterance has finished, Harry makes a grabbing movement with the hand at an empty location on the table, as if grabbing something. In Frame #2 he puts this something down on the table within the enclosure. Then, as shown in Frame #3, he proceeds directly to hold his hand towards the mouth of the horse, as if feeding the horse, and the hand remains held there for a brief moment. This hold is arguably an instance of a *stroke hold*, as discussed in Section 3.3, where the hold is “an information-laden component” (McNeill, 2005, p. 283). That is, the duration is here iconically motivated as part of the act of feeding (on the content level) and it is not primarily used to coordinate the temporal flow of the gesture performance with a spoken utterance, or similar. Indeed, there is no speech in this case.

The “lack” of speech during the performance is in line with the finding from Chapter 5, according to which iconic gestures are the kind of gesture that is most frequently performed without speech in the children.<sup>18</sup> Harry’s iconic performance in Frame #1 to #3 therefore seems to “speak for itself”, as an autonomous performance that does not require any co-occurring speech to be comprehensible as a matter of getting some hay and feeding the horse with it. Thus it may at first glance seem to be a matter purely of direct iconicity, where the iconicity is direct and obvious from the performance itself. It is true that the movements involved in the performance are probably self-sufficient to the extent that they may be recognized as some sort of acts of bodily signification (gestures that qualify as semiotic signs, cf. 2.2.5), and more specifically as an iconically motivated performance, involving things like GRAB and PUT and so forth. However, it could be argued that if it was not for the way that Harry’s gesture is performed in response to his mother’s utterance (“can you get him some hay?”), the gestures shown in the figure would have been much more ambiguous. It would not be clear what it was that was grabbed in Frame #1

---

<sup>18</sup>The finding was that 24.0% of the gestures with iconic aspects are performed without speech, whereas 16.6% and 7.6% of the conventional and deictic aspects respectively were performed without speech.



and what was put down in Frame #2. Even though the gesture performed in Frame #3 might have been seen as an act of feeding, it would certainly not be clear that it was hay specifically that was being served.

Another factor that is involved in the recognizability of the more specific iconicity that is involved in Harry's handling of "hay" is that it draws heavily on the ways that his gestures are indexically directed towards different locations (just like in the previous example with Alice's rubbing gesture) — as a matter of implied iconicity. The GRAB performed by Harry in Frame #1 is directed towards an empty location on the table. The table thus serves as a kind of *supporting space* in the sense that the meaning of the gesture does indeed draw on its directedness towards the material environment, but only indirectly so.<sup>19</sup> There is no real hay there to be grabbed and the table is not treated as a table, but rather as if it was the ground in the "world" of the horse and the fence. The same goes for the PUT in Frame #2. Furthermore, the gesture in Frame #1 and the gesture in Frame #2 form a tightly coupled unit together, because they form a kind of contrast to each other. The location involved in Frame #1 is relevant in because it is *outside* the fence and the location involved in Frame #2 is relevant because it is *inside* the fence. Together they fulfill the request from the mother's previous utterance to get the horse some hay. That is to say, the relevance of the locations involved are also related to the "world" of the horse in the fence in the sense that they are relative to the position of the fence. The FEEDING gesture performed in Frame #3 has indexical properties too. The hand-shape involved is indeed very generic and non-specific to the act of feeding in itself, but through its directedness towards the mouth of the horse, it becomes clear that the gesture is an act of feeding. Just like in the previous example then, the recognizability, or "transparency", of the more specific iconicity that is involved in the gesture is due to several factors, both direct iconicity and implied iconicity.

#### 7.3.2 The continuity between typified actions and typified gestures

This section takes a brief look at actions that do not quite qualify as semiotic signs (Sem#3) in themselves, but which are rather a matter of typified *count-as* actions (Sem#2) with no clear differentiation between expression and content. Since there is no clear differentiation between expression and content in Sem#2 actions, most

---

<sup>19</sup>The distinction between action space, supporting space, and abstract space was introduced in towards the end of Section 2.2.3.

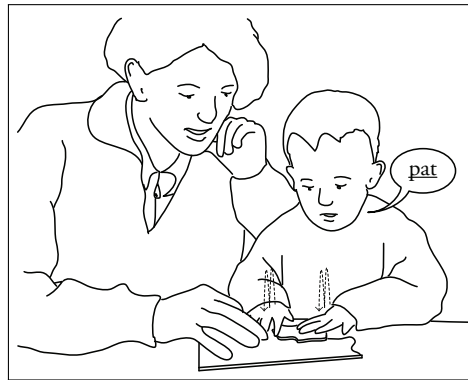


Figure 7.4: “ACTION\_BASED” in a literal sense (Bella 22;04)

researchers do not consider them to be gestures, which is to say that they fall below the lower limit of gesture as discussed in Chapter 2. This is clearly justified to the extent that there are differences between Sem#2 actions and Sem#3 actions. Yet there are also often many similarities between the way many Sem#2 actions are performed and acts of bodily signification on the level of Sem#3, and this section will highlight some of these similarities.

Even though Sem#2 do not qualify as acts of bodily signification (i.e Sem#3), they may still qualify as acts of bodily communication (Comm#3). The communicative character of the action may come about through “additions” of factors outside the action itself, such as being performed in coordination with a spoken utterance.

Figure 7.4 shows Bella (22;04) and her mother. They are doing a jigsaw puzzle together and Bella has troubles in making a piece fit into the puzzle. The piece is at the overall correct place, but needs to be slightly adjusted before it will fit precisely. Bella first tries to attach the piece by pressing it hard, but does not succeed. The mother then says “one more time!” (*en gång till!*) and Bella makes another brief attempt at squeezing the piece in place, but fails again. Directly after that, Bella performs the action shown in the figure. Instead of pressing the piece she now tries to attach it by hitting it with both hands. In coordination with the performance of this action, Bella also says “pat+PAT” (*klappa*). The act she performs is not a semiotic sign (Sem#3) that represents patting. Rather, the act she performs *is* patting, which is to say that it counts-as (Sem#2) a real instance of patting something. The mother then responds to this by saying “you pat like that yeah so it will work out fine” (*du klappar så ja så går de bra*). The mother thus picks up on Bella’s categorization of this act as an instance of patting.

This example serves as an illustration of how many of children's actions are not just spontaneous actions in general, but actions that count-as instances of recognizable and conventionalized types of action.<sup>20</sup> Just like acts of bodily signification, these count-as actions may be said to have a typified content, although there is no differentiation between the expression and the content. This highlights the continuity between typified actions and gestures that serve to express typified action. However, the main point here is not just that ACTION\_BASED gestures may emerge out of action, because that is a commonplace observation in the gesture literature. What I want to highlight here, more specifically, is how this example also illustrates the common practice among children to *name their non-gestural actions and comment upon them in coordination with the performance of the actions*.

A central finding within gesture studies is that many gestures are both *temporally* and *semantically* coordinated with co-occurring speech (Kendon, 1980b, 2004; McNeill, 1985, 1992, 2005). The “Thought-Language-Hand Link” that McNeill (2005, p. 233) talks about may thus not be unique to the empty-handed gestures that gesture researchers are typically concerned with when they study gesture, as it may be better understood as a much more general ability of coordinating speech with all sorts of bodily action. As far as I know, there are no other studies of the ways children coordinate speech with action in a more general sense, that employ the detailed kind of analysis that gesture researchers do (but see xAndrén, in press c). Even though there are studies that include object-gestures as part of the analysis these studies are generally restricted to object-gestures that qualify as semiotic signs, such as acts of symbolic play (e.g. Iverson et al., 1994; Capirci et al., 2005), i.e., actions that are in many ways “already gesture”. Perhaps the most relevant study in this regard is a study performed by Eckerman (1993), who found that around the end of the second year, children often start providing something like a running commentary on their own actions as they perform them.<sup>21</sup> For example, “sometimes toddlers map their verbal utterances onto the non-verbal co-ordinated action in a redundant fashion, as when the child both imitates jumping off a box and describes her action as ‘I am jumping’ “ (*ibid.*, p. 129).<sup>22</sup> This is precisely what is going on

---

<sup>20</sup>Of course, even if the action is typified, it is always also adapted to the local circumstances (i.e., Sem#1).

<sup>21</sup>See also Hendriks-Jansen's (1996, p. 308, p. 336) discussion of Eckerman's findings for an elaborated discussion on the possible implications of children's explicit labeling of their own actions.

<sup>22</sup>Garton (1986) also notes that people often say “this” in coordination with the act of grabbing an object. That is another variant on the theme that is discussed here, but noted by Garton in the context of deictic aspects of action rather than iconic aspects. Uses of GRAB in coordination with speech are analyzed in Chapter 9.

when Bella hits the piece of the puzzle and names the action. It was also what Alice did when she performed her rubbing gesture in coordination with the words “rub rub”, as shown in the previous section. Hanna’s act of drawing in coordination with the utterance “I so draw air balloon+DRAW”, as discussed in the previous chapter, was also an example of this and several other examples to be presented below also involves naming of the act that is performed.

In the quote above, Eckerman describes this coordination between action and speech as “redundant”, but one may want to argue that the explicit naming of the actions is far from redundant. Through the naming of the actions, the process of typification gains its publicly available and intersubjective nature, as a kind of *implied iconicity*. In this sense, one can speak of typification as something that is at least in part *achieved* in and through interaction, and not just a individual cognitive act of categorization.<sup>23</sup>

I would argue that if communicative actions that involve handling of objects may sometimes seem less well timed with spoken language than empty-handed gestures, this is not because they are radically different from empty-handed gestures, but rather (A) because these acts are not only coordinated in concert with speech and interaction but also with the practical requirements of manipulating a physical object, and (B) because communicative actions involving objects may not, in general, be as dependent on spoken language timing to be understandable as some gestures are, due to the nature of these acts as often being visibly rational in themselves, in part due to the contextual support offered by the objects. One may therefore expect to find both similarities and differences between the ways that gestures and count-as acts on objects may be coordinated with language, but at a more fundamental level I would claim that the same kinds of abilities are involved in both cases.

### 7.3.3 ACTION\_BASED gestures from an alter-centric perspective

In this section it will be argued that not all ACTION\_BASED gestures are readily understandable as stemming from one’s own experience in performing various actions. Sometimes ACTION\_BASED gestures may rather be a matter of taking someone else’s perspective.

Figure 7.5 shows a book-reading situation that involves Tea (23;07) and her

---

<sup>23</sup>See Section 3.2 for a discussion of the importance of typification and third-person intentionality in human interaction inspired by the theorizing of Alfred Schutz.

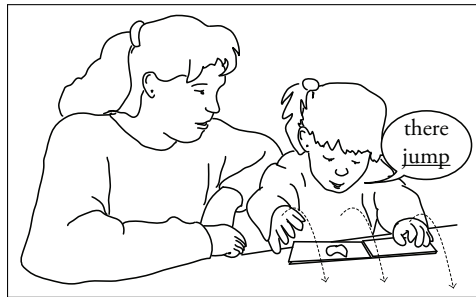


Figure 7.5: ACTION\_BASED but alter-centric frog jump (Tea 23;07).

Mother. Before the gesture shown in the figure is performed, the following events take place: Tea flips the page in a book and points to a frog in the book while saying “*there+POINT*” (*dä*). The mother responds to this the way that mothers often do, by naming the thing that the child has pointed out, but not categorized in a content-loaded manner. The mother says “a frog” (*en groda*). The child then performs the gesture shown in the figure, which is a forwards thrust of the body, with both arms stretched out in front of her, while saying “*there jump+FROGJUMP*” (*dä hoppa*). This gesture was once shown to a seminar audience without sound enabled when the video clip was played, and although several persons saw this as some sort of iconic gesture, nobody in the audience was able to see that it was a matter of a jumping frog. Hence, there are once again many aspects of implied iconicity involved in this gesture. The frog theme has been actualized in the previous turns of the interaction and Tea’s own explicit categorization of her act as an act of jumping also contributes to the meaning of this gesture as a frog jump.

The interesting thing here is that the kind of “jump” that is enacted by Tea hardly corresponds to the way she would have jumped herself, unless perhaps if she would have been diving into water, but she is too young for that. This means that even though this is an ACTION\_BASED gesture, it does not rest on previous experiences in carrying this sort of action out herself. What she performs here is rather an enactment of the way that “any frog would jump”, i.e., adopting the perspective of a kind of typified third-person intentionality. This issue was also discussed in Section 5.7.2, and the point here is that one may want to distinguish between the developmentally simpler form of ACTION\_BASED gestures that rest primarily on the child’s own experience in acting in the world, and the developmentally later variant of ACTION\_BASED gestures that involve taking someone else’s perspective. As argued in Section 5.7.2, one must distinguish between the technique of realization that is involved in the gesture (ACTION\_BASED or INDIRECT) on the one hand,

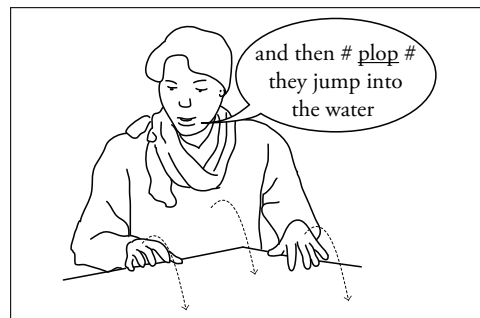


Figure 7.6: Tea's mother doing the same jump-gesture.

and the psychological perspective that may be involved (acting as self or as alter) on the other hand (see also Müller et al., manuscript).

One may ask whether this gesture is Tea's own creation, or if it may somehow be learned from an adult or a sibling. This is of course hard to know with certainty, but it is notable that a couple of years earlier, before Tea was born, Tea's mother performs almost the exact same gesture when talking about some turtles (not frogs) that jumped into the water at an earlier occasion. This gesture is shown in Figure 7.6 as it is performed in coordination with the utterance "and then # plop+JUMP # they jump into the water" (*å så # plupp # så hoppa dom i vattnet*). The "#" sign is used here to denote a pause in the speech. Thus it is not entirely unreasonable to assume that Tea may have learned this gesture from her mother. As noted previously in this chapter, Caselli (1990) has argued that children seem to learn many of their iconic gestures from their parents, and the observation here may lend some support to this view. It may also be argued that many of the other gestures discussed so far in this chapter would hardly look the way they do unless there was a sociogenetic process involved. Just to take one example, Alice's canonical use of the eraser more or less entails an imitative process. It is simply unlikely that children would happen to invent a way of using an ambiguous object such as an eraser that happened to match the conventional use of this object exactly, unless another person were involved in one way or another.<sup>24</sup>

<sup>24</sup>The notion of imitative process was defined in Section 2.3.2, and it is a wider concept than imitation as such. Imitative processes include all the possible routes for social spreading of conventionalized knowledge and skills (cf. *assisted imitation*, Zukow-Goldring, 2006; Zukow-Goldring & Arbib, 2007), whereas imitation refers more narrowly to the act of repeating the action that another person performed just before.

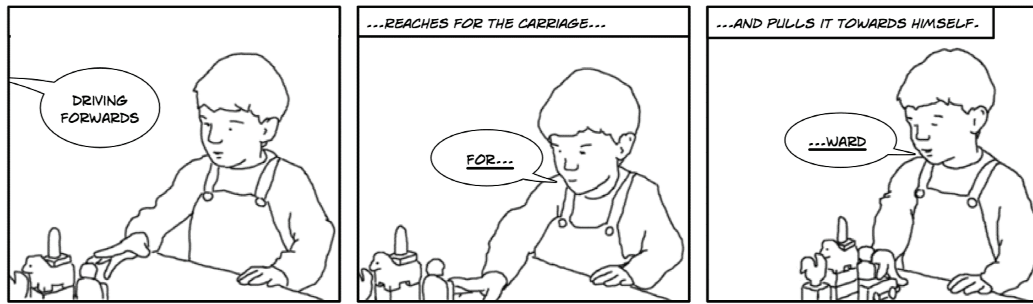


Figure 7.7: INDIRECT object-gesture (Harry 23;18).

## 7.4 Gestures with INDIRECT iconic aspects

In Chapter 5 it was found that gestures with indirect ICONIC aspects are less common than gestures with ACTION\_BASED iconic aspects. It was also found that gestures with INDIRECT iconic aspects were almost non-existent before around 23 months and that *empty-handed* gestures with INDIRECT aspects hardly existed at all. Nevertheless, in this section some of the few existing examples of empty-handed gestures of the INDIRECT kind will be shown, as well as some INDIRECT object-gestures.

### 7.4.1 From INDIRECT object-gestures to INDIRECT empty-handed gestures

As mentioned before, it is sometimes argued that symbolic play actions may be seen as iconic gestures (e.g. Volterra, 1981; Iverson et al., 1994; Capirci et al., 2005). As a further elaboration on the parallelism (or equivalence) between gestures and some play actions one may employ the distinction between ACTION\_BASED iconic aspects and INDIRECT iconic aspects on play acts too.

The first example (in Figure 7.7) shows Harry (23;18) who is playing with a toy carriage. Before the episode shown in the figure, he has been moving the toy carriage backwards, while saying “*reverse reverse reverse+REVERSE\_DRIVE there*” (*back back back dä*). Then, while Harry reaches for the carriage again, the mother says “driving forwards” (*köra framåt*), as shown in Frame #1. When he grabs the front of the carriage (Frame #2), he says “*forward+FORWARD\_DRIVE*” (*framåt*) while simultaneously dragging the carriage towards his own body (Frame #2 and #3). Thus, the move of the carriage is coordinated temporally and semantically with speech and the character of the movement is explicitly categorized through

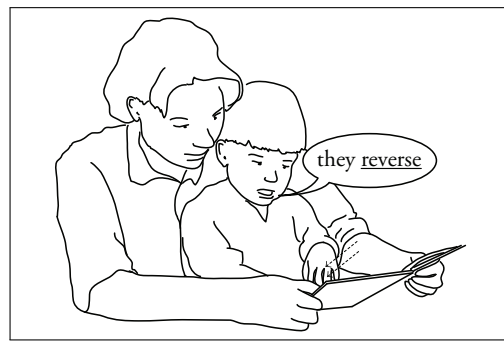


Figure 7.8: INDIRECT gesture showing a reverse movement (Harry 28;23).

the conventionalized label of the speech.

Harry's hand or body is not part of the world projected by this movement. That is, the involvement of the body in this case does not have anything to do with moving "forwards" in an ACTION\_BASED manner. The hand is only involved in so far as to realize an "image" which is out there "in the world". In fact, what would be forwards for Harry happens to be the opposite direction from what is forwards for the carriage. This shows how Harry rather takes the alter-centric perspective of the carriage than his own perspective.<sup>25</sup> The term "alter-centric" is used in a wide sense here, mainly as a contrast to ego-centric, and not necessarily in the sense of taking the psychological perspective of someone else. Another example of this sort of gestures is when a child holds a toy figure towards some toy food while saying "eat". In this case too the movements of the hands have nothing to do with movements involved in eating as such, and again, the body is rather used in an indirect way to realize an "image" which is out there "in the world". In these cases, the children almost always attend closely to what they are doing, which is also support for the view that they are rather using their bodies to realize some sort of imagistic scenery, and the movements of the body are indeed only indirectly related to these scenarios.<sup>26</sup>

Someone may now object: Are such actions really gestures? The answer depends on whether one would like to accept that gestures may sometimes in part be realized (or "articulated") through artifacts rather than directly with the body or if one would prefer to call that something else. In either case, the overall "scene" that is realized through this move of the carriage in coordination with the utterance is semiotically complex. The action is indeed coordinated both temporally and semantically with speech, much like more prototypical gestures often are.

<sup>25</sup>This phenomenon can also be observed in other similar actions when Harry is playing with toy



Harry likes to play with toy vehicles of various kinds, and there are a number of instances in several different recordings at different occasions where he moves these vehicles back and forth while spelling out their movements in speech like this. With only one exception, his use of the words “reverse” and “forward” appear only in this specific type of context in the recorded data (which does not involve contexts like sitting in a real car and so forth). The one case in the data where his use of the word “reverse” is not used in association with moving a toy vehicle is shown in the next example.

Figure 7.8 shows one of the few empty-handed INDIRECT gestures in the data, and here Harry and his mother are engaged in book reading. The mother first asks “what are they doing now?” (*va gör dom nu?*) with reference to the events depicted in the book. Harry responds by pointing to the depicted event while saying “like-that+POINT” (*så*). It is not unusual for the children studied here to reply in this “content-less” and referential way, by directing the attention of the Other with deictic means so that the Other may look for herself, instead of providing a content-loaded answer. In many contexts this type of response is not accepted by the parents, who rather want the child to spell out the answer in a more content-loaded way (cf. Murphy, 1978; Golinkoff & Gordon, 1988, p. 115). Hence, the mother asks “what did you say?” (*va sa du?*) even though one gets the impression that she did not in fact have any problems at all hearing what Harry said. Now Harry responds in a more content-loaded way. As shown in the figure, he says “they reverse+REVERSE” (*domme backa*) in coordination with a movement of the hand “backwards” towards his own body.

The movement he performs is strikingly similar to the many movements that Harry performs on all those other previous occasions when he pronounces the word “reverse” (*backa*) in coordination with moving a toy car back and forth. One may therefore speculate that this gesture is grounded in his previous experience of moving the toy cars. He is not looking at his hand when he performs this gesture — the gesture rather seems to come from “within”. This stands in contrast to events earlier in his life when he started to master his back/reverse toy vehicle movement abilities, where he usually attends closely to the scenarios realized by his hands. That is, what once was a movement which was part of realizing “scenes in the world” through handling of objects, from an alter-centric perspective and in an INDIRECT manner, now seems to come out as a spontaneous empty-handed imagistic gesture

---

vehicles, so it is not just a strike of luck.

<sup>26</sup>In this case the dissociation between the operation-level and the action-level is indeed in line with Leontiev’s (1981) ideas of a functional dissociation between the levels.

with INDIRECT iconic aspects.<sup>27</sup>

This line of reasoning is admittedly speculative. Yet one may note that children often perform a specific action or gesture in coordination with a specific word — this will be discussed in Chapter 8 — so that the word and the action or gesture seem to go together as a multimodal unit, rather than than a multimodal combination of two distinct units, i.e. as [gesture+word] rather than [gesture]+[word]. If this is what is going on with Harry's REVERSE gesture in this case, then it may perhaps strengthen the interpretation offered here a bit further. Now even if it is indeed true that Harry's INDIRECT empty-handed gesture has a pre-history in object manipulation (and typified sorts of scenes such as moving “forward” or “reverse”), this must not necessarily be the case for all empty-handed gestures that may emerge further on in the children's life. Nevertheless, it is notoriously hard to study the emergence of particular gestures in children's repertoires since it requires a lot of recordings that happen to capture a similar activity many times during the child's life. In this respect, the emergence of this gesture is interesting, since it does indeed seem to emerge out of an activity that happens to be captured in many of the recordings.

#### 7.4.2 INDIRECT gestures and imitation

As discussed in the analysis of ACTION\_BASED gestures, many ACTION\_BASED gestures seem to be learned, to some extent, from people in the children's surroundings. Could this also be the case for gestures with INDIRECT iconic aspects?

Figure 7.9 shows an example of one of the other few instances of empty-handed INDIRECT gestures that exist in the data. The episode shown in the figure follows directly after the frog-jump gesture that Tea (23;07) performed in Figure 7.5 (on page Figure 233). In Frame #1, Tea has just finished the stroke of her ACTION\_BASED frog-jump gesture. Her mother then says, with a questioning intonation: “the frogs jumped like that?+FROG\_JUMP” (*hoppa grodorna så?*) while performing a gesture with INDIRECT iconic aspects. That is, instead of jumping as if she herself was a frog, the mother rather acts as if her hand was a frog, i.e., an INDIRECT iconic aspect. To this Tea replies “mm” (*mm*) in agreement. Then Tea performs an almost exact imitation of her mother's gesture, as shown in Frame #2. When she performs this gesture, she also says “jump jump jump+FROG\_JUMP”. Once again the act that is involved in the gesture is named explicitly, as in many of the previous

---

<sup>27</sup>What would be classified as an “observer viewpoint” gesture according to McNeill's (1992; 2005; 2004) classification scheme.

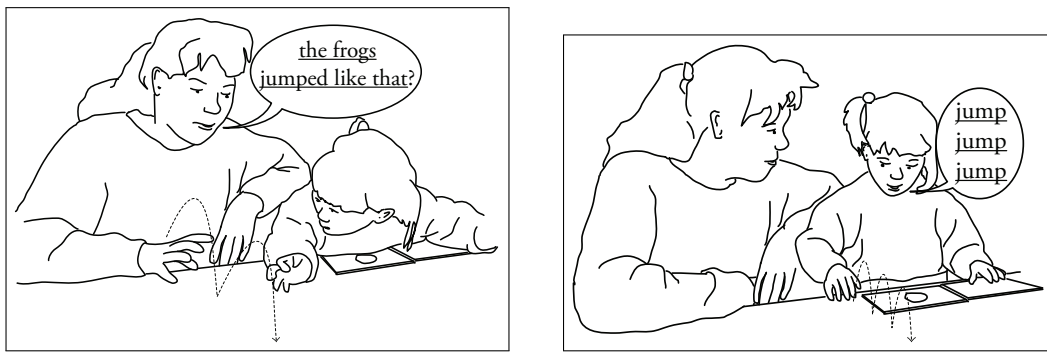


Figure 7.9: INDIRECT frog-jump (Tea 23;07).

examples. Tea looks at her hand when she performs this gesture, as if to monitor its outcome as an image “in the world”.

In this case then, the indirect empty-handed gesture is quite obviously an imitation of a gesture initially performed by the mother. (Recall that it was also suggested that the ACTION\_BASED frog-jump gesture was likely to be imitated.)

### 7.4.3 ACTION\_BASED and INDIRECT aspects in the same gesture

As argued in Chapter 4 where the distinction between ACTION\_BASED and INDIRECT was introduced, some gestures are rather hybrids between ACTION\_BASED gestures and INDIRECT gestures, and this often happens in cases where the hands “become” something that would otherwise be held in the hand, such as a pair of scissors, a toothbrush, or a gun.

Figure 7.10 shows Alice sitting in front of a plate which has a slice of bread on it. Her father has just asked “how many slices of cheese do you think would fit on the bread?” (*hur många ostskivor tror du får plats på brödet?*). In response to this, as shown in the figure, Alice first places one hand on the slice of bread while saying “one there+PLACE ...” and then she places her other hand next to the first while saying “...and one there+PLACE”. After this, she holds her hands still for a brief moment, before reorganizing her gesture so that the hands are not crossed anymore (not shown in the figure), as if to tidy up a gesture that turned out slightly messy.

On the one hand, this gesture seems to have an ACTION\_BASED component to the extent that it seems to correspond to an act of placing something on the slice of bread (i.e., slices of cheese). On the other hand, the spatial extent of the hands in this gesture seems to correspond to the spatial extent of the cheese, which seems to



Figure 7.10: An iconic gesture performed towards an object.

suggest that there is an INDIRECT aspect involved too.

To elaborate: The hand-shape employed does not seem to have an ACTION\_BASED motivation, because she does not seem to hold cheese in her hands. The hands become flat already during the preparation (the preparation of the gesture in Frame #2 can be seen in Frame #1). She rather seems to employ a cheese-formed hand-shape as if to measure how many slices of cheese that may fit on the bread, as asked by her father. Again, the indexical directedness of the gesture towards an object in the material environment is a crucial part of this gesture, i.e., a crucial part of what makes the hands seem like cheese.

## 7.5 Conclusions

In this chapter it has been argued that the effortless and seemingly “natural transparency” of iconicity in iconic gestures is to a large extent, although not wholly, recognizable through the contextual embedding of the iconic gesture, as kind of *implied iconicity*. Furthermore, to the extent that the iconicity is directly perceivable from the movement involved in the gestural performance (i.e., *direct iconicity*) this is to a large extent based on the typified and familiar character of the act that is involved. Hence, even in the case of direct iconicity, the recognizability of the iconicity often rests on familiarity with various conventionalized practices. This conventionalization is not necessarily a matter of a conventionalized gesture *per se*, but it is rather the content of the gesture that is conventionalized. However, in the case of ACTION\_BASED gestures, the medium involved in the expression, on the side of the signifier, is close to standing in a 1:1 relationship with the action that is signified in the sense that the body acts as if it was a body. Therefore, the conventionalized character of the meaning on the content-side “bleeds over” to the

expression-side, so that the form of the gesture exploits conventionalized aspects of form on the content side.

It was also argued that the factors that resulted in an implied iconicity were not only conventionalized factors, but also factors such as indexicality. Most of the gestures analyzed in this chapter involve some sort of directedness towards the material environment. The children's gestures are very rarely taking place in an abstract gesture space which is detached from the material environment (*abstract space*), but instead most gestures orient to the material environment either as a concrete action space, or at least as a kind of supporting space. In the case of supporting space, the material environment is not oriented to in a literal fashion, but it is still used as a kind of ground for the gestures. For example, a table may be treated as if it was an ocean where boats may sail.

Another important factor behind the implied iconicity was the way that the children's gestures responded to a previous utterance. In many cases the children's gestures would not be fully comprehensible without knowing what the parent said in a previous turn. Even in the case where a child performs a gesture "autonomously", without speech, and the gesture would thus seem to be self-sufficient it was shown that the recognizability of the more specific iconically motivated meaning that was involved was to a large extent due to the relation between the gesture and the mother's previous utterance.

The child's own utterance is also crucial in many cases for constraining the recognizability of a sufficiently specific meaning in the gesture (cf. Singleton et al., 1995; Goldin-Meadow, 2007a). Consequently, it was argued that the tendency of children to label their own actions — both Sem#2 and Sem#3 actions — in an explicit manner is not a matter of "redundancy". Speech seems to provide a very effective medium for constraining the possible interpretations of an iconic gesture. Even in cases where the gesture is interpretable in itself to a large extent, the addition of a single word may make the meaning of the gesture much more specific and precise (cf. Hirsch, 1996, 2009).<sup>28</sup>

It was also argued that gestures with INDIRECT iconic aspects are often much more open-ended in their meaning than gestures with ACTION\_BASED iconic aspects. Therefore gestures with INDIRECT iconic aspects seem to be even more dependent on factors of implied iconicity.

---

<sup>28</sup>On a lower level, combining vocal sounds with bodily movement may provide a signal to infants that help them segment the flow of bodily movement into bounded action gestalts (cf. Baldwin et al., 2008).

The meanings of the gestures that are performed all seem to have sociogenetic origins in various ways. Either they seemed to be imitated from the gestural expressions of the parents (cf. Caselli, 1990) or they seemed at least to rely on familiarity with conventionalized practices, as argued above. In surprisingly many cases, the children were able to name their actions explicitly, which means that they did not just happen to perform actions that could be interpreted as actions of this or that type, but that the children themselves construed their own actions as actions of this or that type by means of the conventionalized labels of spoken language (cf. Eckerman, 1993). Therefore, it does not seem to be sufficient to understand the processes of typification involved in the genesis of these gestures purely as sensorimotor schemes that emerge in the interaction with objects in the world, as many theories within cognitive linguistics would put it (e.g. Johnson, 1987). The embodied aspects are of course there, and the claim here is certainly not that they are unimportant, but there is also an irreducible role of the Other and of conventionality involved in the genesis of these gestures (cf. *mimetic schemas*, Zlatev, 2005). One may also suspect that spoken language, with its many labels for various kinds of actions, may play a role in this process since the children are obviously often able to name their own actions.

Around 18 to 21 months, there are considerably more actions performed by the children which are hard to understand overall. This goes not only for the researcher, but parents also often have trouble understanding what is going on. This means that there are by all means also a share of non-standardized and non-typified actions performed by the children. The suggestion is certainly not that everything is typified. However, since the non-typified, or less typified, actions are not possible to interpret in terms of “already known” typicalities with familiar rationalities, they tend to appear “uninterpretable” or “meaningless”. In cases where a child repeats a single non-typified action many times, the parents typically quickly lose interest (“Why hit the toy telephone with a pencil 15 times!?! It doesn’t make sense!”) and try to “move on” in the activity, or to establish a new activity which is more “standardized” in various ways. One may thus suggest that typification in action and the means for making the typified character of action publicly recognizable are not just a luxury to make interpersonal action understanding run more smoothly, but in many cases a prerequisite to make the actions understandable at all. Indeed, iconicity, by its very nature, implies that something is similar to something else, so that the expression is somehow an *instance* of this “something else” that is signified. That is to say that iconicity itself is strongly related to things like category membership, i.e., typification. If it is not possible to relate the gesture to something that is not “already

known”, in some sense, the gesture also loses its interpretability in iconic terms.

All in all then, there are many aspects of the children’s iconic gestures that do not seem to be invented on the spot, and to the extent that there is a certain history of experiences behind many of the gestures, this history is to a substantial degree a matter of sociogenetic processes of conventionalization. This may seem to go against McNeill’s (1992; 2005) conception of iconic gestures as idiosyncratic, creative, and dynamic inventions on the spot, where speech represents the conventionalized and “analytic” pole and gesture represents the spontaneous and “dynamic” pole. Maybe it does to some extent, but it should also be noted that McNeill himself does not believe that gestures start to have this idiosyncratic and dynamic character before what he calls the “gesture explosion” around the age of 3–4 years. What is clear, however, is that there are many iconic gestures even before this so-called “gesture explosion” (cf. Zlatev & Andrén, 2009) and that many of them are strongly tied to the handling of objects or in other ways directed towards the material environment, in ways that are often conventionalized. Hence, what is witnessable in the examples shown in this chapter is the pre-history of the more detached and abstract gestures that appear later in development.

Despite the emphasis on various factors relating to implied iconicity in this chapter, the purpose has not been to reject direct iconicity as such, the way that strongly convention-based accounts of iconicity would have it (e.g. Goodman, 1970; Eco, 1976), but rather to provide a more differentiated view of iconicity that acknowledges the importance of both direct iconicity and implied iconicity. I would say that iconicity is a very real phenomenon, but perhaps not just that “naturally transparent” as one might believe at first glance.





## CHAPTER 8

# Conventionalized aspects of the children's gestures

*...any term is capable of functioning as a designator of any meaning. That it does not is problematical.*

Garfinkel (1952, p. 338)

### 8.1 Questions asked

In Chapter 2 it was argued that one cannot speak of the presence or non-presence of conventionalized factors in gesture simply as an either-or issue. A more sophisticated understanding of conventionalization in gesture requires that at least two things are acknowledged.

First, conventionality may enter into gesture in multiple ways. There may be conventionality in the form of the expression, the content/referent may be conventionalized, and the relation between expression and content/referent may be conventionalized (cf. McNeill, 2005, p. 31). In addition, a gesture's contexts of use may also be conventionalized (Kendon, 2004).

Second, there are different degrees of conventionalization. The very lowest level of conventionalization (Conv#1) is really only noticed when it is violated. That is, when someone performs gestures in an odd way, such as making too big gestures or performing gestures too close to the face of an interlocutor, the prevailing order of "normal appearances" is disturbed. A more explicit form of conventionality is present when some sort of categorization has taken place (Conv#2), so that an ac-

tion does not just appear to be normal, but is also recognizable as an instance of a familiar *type* of action, commonly known in a group. At this level, the action may sometimes have a lexicalized name in spoken language. A yet more explicit form is present when the gesture is not just recognizable as a familiar type of action, but is also supposed to be performed in a quite specific way (Conv#3). If the performance deviates from its typical form too much it is not only considered atypical, but also *incorrect*. This is conventionality in the sense of normativity and the way the term emblem is used in this thesis refers only to gestures that are conventionalized on this level. Finally, one can also speak of an even more explicit form of conventionality (Conv#4), which is present when a competent member of the community is not only able to pre-reflectively act in accord with the norms of the community and to recognize deviant performances as incorrect, but when the norm, or rule, is itself spelled out explicitly. At this level, the conventionalized character of an action becomes explicit enough to gain an existence that is in part independent from the act itself. It sometimes happens that a parent remarks explicitly on the children's use of conventionalized gestures such as HEAD\_SHAKE and NODDING: a matter of explicit teaching.

Some conventionalized aspects of the children's gestures have already been illustrated in the analysis presented in Chapter 6 and Chapter 7. The present chapter is, however, mainly concerned with gestures that are conventionalized on the level of Conv#3, both with respect to form, meaning, the relation between form, and the gestures' contexts of use. Hence, the conventionalized gestures discussed in this chapter correspond to the kind of gestures that have been called *emblems*.

Since conventionalized gestures are recurrently used in ways that are particular to each specific type (HELLO, NODDING, and so forth), each of these types must to some extent be understood as a unique phenomenon. Even though one can discuss "conventionalized gestures" (or "recurrent gestures") at a more abstract and general level, in terms of "how children use conventionalized gestures", there is also a need to look at each of these gestures individually — to flesh out their particular ways of doing communicative work. Thus, the class of gestures that may be characterized as conventionalized is quite a heterogeneous "category" and all conventionalized gestures are certainly not learned at the same time. Some conventionalized gestures are commonly learned very early, even before the period studied here (e.g. Guidetti, 2002), and some are typically not learned until later (e.g. Graziano, 2009). It is also clear from the children studied here that they do not all use the same set of conventionalized gestures, apart from the most common ones. Some particular gesture may be used somewhat frequently by one of the children, but may not appear in any

of the recordings of the other children, and so forth.

In Chapter 5 it was found that the most common two conventionalized gestures used by the children were HEAD\_SHAKE and NODDING. These gestures are also special in other ways. Therefore, the first part of this chapter will focus specifically on these two gestures. In that regard, the following questions will be addressed: *How are NODDING and HEAD\_SHAKE used by the children? How does the use change over time? To what extent are they combined with other gestures?*

In addition there is also a second section in this chapter which is concerned with the relations between conventionalized gestures and speech. This does not only involve NODDING and HEAD\_SHAKE, but children's use of conventionalized gestures in coordination with speech more generally. The question that is asked is: *To what extent can gesture+speech utterances that involve conventionalized gestures be understood as a matter of combining separate units and to what extent may the "combinations" rather be rote-learned multimodal constructions?*

## 8.2 Head-shakes and nodding

### 8.2.1 Introduction

Except for various deictic gestures, NODDING (for affirmation) and HEAD\_SHAKE (for negation) are the most common specific types of gestures performed by the five Swedish children (see Chapter 5). In another study that included Thai children between 18 and 28 months in addition to three of the Swedish children studied here, we found that NODDING and the HEAD\_SHAKE were also the most common conventionalized gestures used by the three Thai children (Zlatev & Andrén, 2009). Nodding and head-shakes are also reported in American (US) children (Goldin-Meadow & Butcher, 2003; also in children who speak American Sign Language, Anderson & Reilly, 1997), Italian children (Capirci et al., 1996), French children (Guidetti, 2005; Kochan, 2008), and they have also been reported previously in Swedish Children (Eriksson & Berglund, 1998; Berglund, 1999; Allwood & Ahlsén, 1999). In studies of adults, these gestures have been shown to exist in many cultures across the world (Darwin, 1872; LaBarre, 1947; Brewer, 1951; Jakobson, 1972; Morris et al., 1979; Kendon, 2002; McClave, 2007; Harrison, 2009a). They do not, however, seem to exist in all cultures. Charles Darwin (1872, pp. 273) wrote:

With infants, the first act of denial consists in refusing food; and I repeatedly noticed with my own infants, that they did so by withdrawing

their heads laterally from the breast, or from anything offered them in a spoon. In accepting food and taking it into their mouths, they incline their heads forwards. [...] Nevertheless [...], these signs are not so universally employed as I should have expected; yet they seem too general to be ranked as altogether conventional or artificial.<sup>1</sup>

That is, even though these gestures are conventionalized and not used in all cultures (LaBarre, 1947; Darwin, 1872; Morris et al., 1979), and even though they may not be used in the exact same way in all cultures where they exist (Kita & Ide, 2007), this specific pair of gestures is very wide-spread. It has also been suggested that Bonobos may shake their heads in contexts where they seem to aim to prevent someone from doing something, and that this would be “a primitive precursor of the negatively connoted head-shaking behavior in humans” (Schneider et al., 2010, p. 199), although this conclusion seems to beg for a lot more evidence. Apart from being spread out over the world in an extraordinary way, this pair of gestures is special in many other ways.

First, they are the only gestures that are *routinely articulated with the head*. It seems as if their functions are important enough to be assigned to an articulator of their own. Since people primarily look at each others' faces when talking to each other in face-to-face interaction (Kendon, 1990; Gullberg & Holmqvist, 1999; Bavelas et al., 2002) one might suspect that their uses may be strongly integrated with speech. Indeed, people use these gestures in a multitude of ways together with speech (Kendon, 2002; Harrison, 2009b). They often serve a modal function that provides an interpretative frame for a spoken utterance, related to how facial expressions is sometimes used to signal an attitude towards what it said.<sup>2</sup> In a way, one could say that these gestures are conventionalized (also on a normative level) and

---

<sup>1</sup>There are several instances in the Swedish children studied here of head-shakes in the particular context of refusing food. In addition to Darwin's observations one may also add that they often seem to be motivated in part by the fact that the children may already have food in the mouth, which may make a “non-verbal” expression more useful for purposes of communication than speech, since the mouth is partly occupied. That is, in these contexts, the use of the HEAD\_SHAKE often takes the form of a gesture-only utterance, even in children that do not usually tend to use this gesture without speech in other circumstances.

<sup>2</sup>Even though facial expressions are also articulated with the head, they do not involve movement of the head itself, but rather movements in the face. The face has an interesting evolutionary story, highly intertwined with the functioning of the body as a whole. In association with the development of the forelimbs and upright posture, new functions emerged (Cole, 1998). What was once the front end of of the fish, with very limited possibilities of movement, evolved into an organ of complex displays, and while it was once used for “military” purposes of self-defense and attack, apart from its sensory, respiratory, and ingestive functions, it later became “demilitarized” (McNeill, 1998) and adopted for social and communicative functions. Even though humans share some aspects of facial expression with primates (Darwin, 1872), including facial expressions related to tastes such as bitter

integral part of the activity of speaking, on equal footing with, for example, other response morphemes. Hence, it is dubious to say that these gestures are *not* part of language, even though they do not appear in more abstracted linguistic contexts such as writing (cf. Linell, 2005).

Second, these gestures are *routinely performed in coordination, not just with speech, but also with other gestures* (cf. Kendon, 2004, p. 310). That is, they are not only coordinated in multi-gesture utterances of the kind where several gestures follow after one another, but they may also be used simultaneously with other gestures (see also Harrison, 2009a). A few examples of this has already been given in earlier chapters, such as the examples in Figure 6.3 and Figure 6.14. This feature of the *head-gestures*, as one might call them, is in part a consequence of the fact that these gestures are articulated with the head, which makes their use less conflicting with other gestures that are performed with the hands, even if it is also possible to perform two gestures as once with two hands.

Third, these two gestures *deserve to be called morphemes in a stronger sense* than many other conventionalized gestures. The reason for this is that they are not only lexicalized in character, but they also form *a contrasting pair* of gestures, both on the level of form and on the level of meaning. This is a matter of systematicity. The head-gestures are therefore language-like in more than one sense.

Fourth, they are often used *while someone else is speaking* to provide feedback to the speaker (Yngve, 1970; Allwood & Ahlsén, 1999; Bavelas et al., 2002), and not only as a turn in their own right in the interaction. These sorts of communicative moves are often called “back-channeling”. Back-channeling uses of the head-gestures are perhaps the most frequent way of using these gestures in adults and it can often be observed within seconds when watching people who are engaged

---

and sweet (Steiner & Glaser, 1995), there are also many differences. The hair of primates and other vertebrates is almost gone (especially in women), as part of a transition from touch and smell as the dominant sensory function of the face to vision, and the muscles that were previously controlling large facial hairs instead became available for finely controlled facial movement (Cole, 1998). The nudity of the human face also makes its movements clearly visible to peers. Lacking the ability to move facial muscles, as in the case of Möbius syndrome, can have severe effects on social life as the emphatic functions of the face become unavailable (Cole, 2001, 2009). Wundt (1973 [1921], p. 83) points out that “while the hand is able to mime every conceivable external circumstance, the face is only capable of reproducing itself in the different conditions of affective expression.” This is only partly true though, since it is now, for example, well known that the face may serve functions of grammatical marking in signed languages, modulating the meaning of manual expressions. There are also cultural differences in facial expressions (Jack et al., 2009). Another difference in the faces of humans with respect to primates is that the sclera in humans has become white, which makes it a lot easier to see the direction of gaze in humans than in primates Kobayashi & Kohshima (2001).

in conversation in a cafeteria or similar.<sup>3</sup> Most other gestures, although not all, are used in some kind of association with the act of performing a turn in the interaction. This is not to say that uses of the head-gestures as part of a turn in the interaction are uncommon though.

Fifth, the multifaceted use of head-gestures with respect to the affirmation and negation makes them different from many other, but not all other, conventionalized gestures used by children. Many of the other emblems used by children have much more restricted contexts of use. For example, the SCOLD\_POINT emblem (as discussed in Chapter 6) has a meaning which is only invoked in contexts when someone has done something bad. It is clear that this is a much less frequently occurring context than the general contexts of affirmation and negation covered by the head-gestures. The same restrictions on usability applies to emblems such as HELLO, BYE\_BYE, and BOW\_THANKS, which are only relevant in quite specific contexts.

### 8.2.2 When do the head-gestures emerge?

Alice performed both of the gestures at 12 months (and possibly earlier), but even though there are some clear articulations of these gestures already at that age, they were at that time typically less clearly articulated overall than at later ages. All children studied here perform these gestures in the first recording, except for Hanna, who does not use these gestures until the second recording (at 19;16). This may very well be a co-incident, since the rate of these gestures is quite low at 18 months in the children compared to later months (see Section 5.6.5).

Guidetti (2002, p. 275) argues that the head-gestures typically emerge in the beginning of the second year, and in her own study she finds them in 16 month old children, which is the youngest age group she studied. Since NODDING and the HEAD\_SHAKE are specific kinds of gestures, rather than general "classes" of gestures (such as gestures with deictic, iconic, conventionalized aspects and so forth), one may expect that they should be expected to emerge at somewhat different ages in different individuals, just like more particular aspects of spoken language may emerge at different times in different individuals. Indeed, in a large study, Berglund (1999, p. 69) found that HEAD\_SHAKE was performed already by 7% of the 8 month old children and 23% of the 10 months old children, which effectively places

---

<sup>3</sup>The justification for this claim are my own observations of people's use of these gestures in daily life. Systematic studies would be required to clarify this.

it among the very earliest gestures in use by children—at least for some children.<sup>4</sup> At 16 months, 80% of the children in Berglund’s study used the HEAD\_SHAKE.

The head-gestures are used in many different ways (Kendon, 2002) and it is clear that they continue to develop after their initial emergence in the repertoire (Kochan, 2008), in the sense that they come to be used in different and increasingly more complex ways later on in children’s development. As remarked by Gullberg (2009, p. 118), “targetlike forms do not necessarily have adult- or targetlike meaning”.

### 8.2.3 Nodding and head-shakes as back-channeling?

As it turns out, the five Swedish children do not use HEAD\_SHAKE and NODDING as feedback to the speaker while the speaker is talking (*back-channeling*). This use of the head-gestures is entirely absent in the data. All instances of NODDING and HEAD\_SHAKE come after the parents utterance, as an own turn in the interaction. This is so despite the fact that the parents often use these gestures while the children are talking, as back-channeling, and despite the fact that both gestures are in the children’s repertoires. That means that the children neither lack the ability to perform the gesture as such nor do they lack input of this use of the gestures. Why would back-channeling be absent in children? It is hard to say, but I can think of at least three possible explanations, though any conclusions would have to await further research.

First, back-channeling is often used, not to answer yes/no questions and other explicit response-demanding utterances like that, but rather to provide general indications that one has heard and understood what the other person is saying. It may very well be the case that the children have not yet learned that such feedback may be relevant to the parents. Hence, they may not yet be able to see this function and they may therefore not be sensitive to the cues to feedback elicitation (such as various gaze patterns) that the parents provide.

Second, child-directed speech (so-called *motherese*) is often rather short as parents adapt their speech to the children. Indeed, it is not only the children’s MLU that increases gradually over the period studied here, as shown in Chapter 5. The parents’ MLU also increase steadily across the period from 4 at the beginning of the period to 5 in the end of the period. This means that the average parental ut-

---

<sup>4</sup>NODDING did not appear in at least 20% of the children at 8 or 10 months though, and was therefore not mentioned by Berglund.

terance may not be long enough to provide a “slot” for back-channeling, and the children may therefore not develop this sort of use until later, when they are faced with longer, and more narrative-like utterances, on a more frequent basis.

Third, back-channeling requires a kind of real-time response which may be difficult for the children to manage in time. Conversation analysis research has shown that people (adults) are typically very quick and apt at placing their back-channeling words and gestures at specific relevant places in the speaker's stream of speech, which means that people seem to know slightly beforehand when a relevant “slot” for a back-channeling signal is coming up (cf. Ochs et al., 1996; Streeck, 2009a). That is, acting in “real time” often rather means that one has to be able to anticipate the flow of the situation to some extent. The children may not yet have become fluent enough in the dynamics of dialogue to be able to anticipate what might come up next in an utterance in this way. The children may rather be reacting to their parent's utterances than anticipating them.

#### 8.2.4 From responses to initiatives

At 18 months the five Swedish children's main use of NODDING and HEAD\_SHAKE is to respond to their parents utterances. They may either be using the gestures together with a single response morpheme such as “yes+NODDING” (*ja* and *jo*), “mm+NODDING” (*mm*), “no+HEAD\_SHAKE” (*nej* and *nä*) or alone, as a gesture-only utterance.<sup>5</sup> Figure 8.1 shows two curves. The curve labeled “responsive” shows head-gestures that are used in this primarily responsive way. Included in this category are also a few instances like “yes they do+nodding” which contains other words in addition to the utterance initial response morpheme, but these amount only to 4.7% of these “responsive” uses and the rest are either one-word utterances or gesture-alone utterances.

From around the onset of TP#1 and onwards another way of using the head-gestures emerges and increases in frequency over time.<sup>6</sup> This is shown in the other curve in the graph, and is labeled “initiating”. This refers to uses of the head-gestures that are not used together with response morphemes, but rather together with other kinds of utterances/words such as:

- “she can not sleep+HEAD\_SHAKE” (*hon kan inte sova*)

---

<sup>5</sup>The word *jo* is much less frequent than the other response morphemes and “the use of *jo* [...] is contingent upon the presence of syntactic negation in the utterance to which *jo* is given as a feedback item” (Plunkett & Strömquist, 1992).

<sup>6</sup>“TP#1” refers to Transition Period #1 as defined in Section 5.1.



## 8.2. HEAD-SHAKES AND NODDING

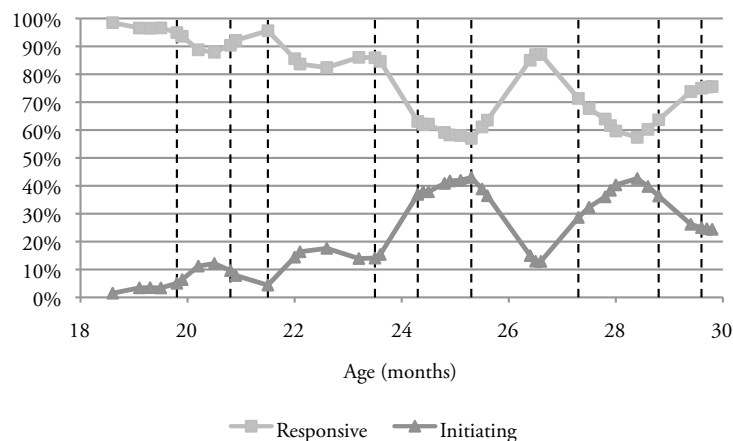


Figure 8.1: Head-gestures as response and initiative.

- “that is not yours+HEAD\_SHAKE” (*de e inte din*)
- “I want to taste one more+NODDING” (*ja vill smaka en till*).

In line with Harrison’s (2009a; 2009b) findings regarding gestures of negation in adults, the gesture stroke (including any occurring post-stroke holds) are frequently performed in a way that corresponds to the scope of the negation in the clause. Even though utterances like this are often uttered in response to a previous utterance too, just like the “responsive” uses of the head-gestures described above, they also have a stronger initiating aspect in the sense that they also contribute with something new to the dialogue (cf. Linell & Gustavsson, 1987; Linell et al., 1988). The degree to which they are also responsive may vary. They are therefore called “initiating” here, even though it may be more correct to call them something like “initiating(+responsive)”.

What Figure 8.1 shows is that *there is a progression from the initial exclusive use of these gestures in responding contexts at 18 months to increasingly more initiating contexts at later ages*. This can also be understood as a development from a more *reactive* use of these gestures, since the responding uses of these gestures are typically elicited and framed by the parent’s use of a yes/no question or similar, to a more self-induced and autonomous use of these gestures in the initiating utterances.

The profile is similar for NODDING and for HEAD\_SHAKES (see figure 8.2 and 8.2), although there are more initiating uses of SHAKE\_HEAD than there is of NODDING. Why should one expect to find more initiating uses of the HEAD\_SHAKE than NODDING? One possible explanation may be that affirmation is, as if it were, the default mode of communication. When someone utters something, this is typ-

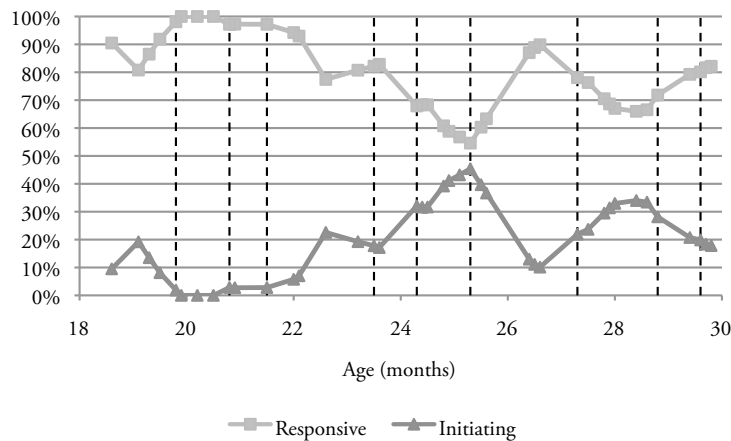


Figure 8.2: NODDING as response and initiative.

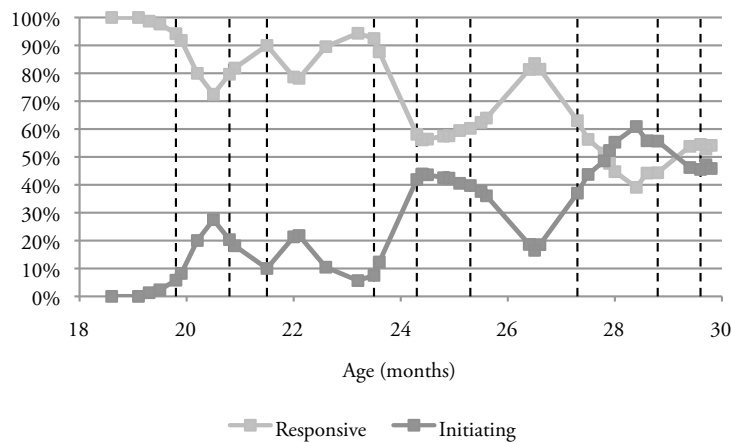


Figure 8.3: SHAKE\_HEAD as response and initiative.

ically not only an account *of* something—it is also an account *for* something (cf. Heritage, 1984), in the sense that speaker's are for the most part held accountable for what they are saying. For this reason, one may expect to find less explicit marking of affirmation both in speech and gesture than of negation, since negation is more like an exception (cf. Croft, 2003). However, this may not be the case to the same extent for the context of responding to yes/no questions, where the answer may be more open-ended and may for that reason have to be marked explicitly by a response morpheme or a gesture, or both. If this line of reasoning is true, it would explain why there are fewer initiating uses of NODDING than there is of the HEAD\_SHAKE.

A closer look at the initiating utterances that are coordinated with nodding seems to confirm this hypothesis. It turns out that most of the NODDING gestures that

## 8.2. HEAD-SHAKES AND NODDING

---

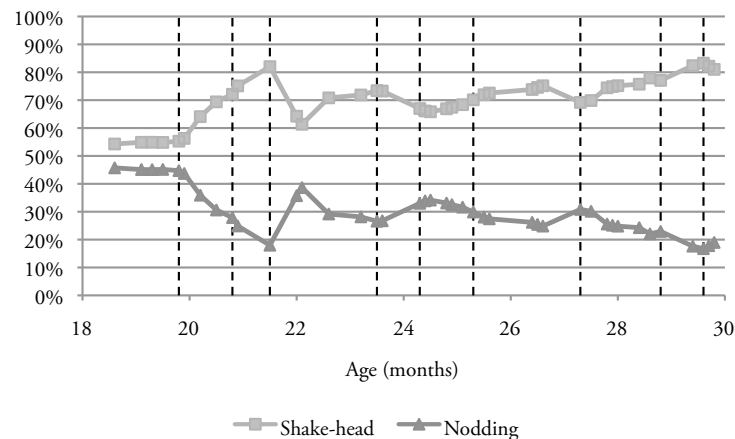


Figure 8.4: Proportions of NODDING and HEAD\_SHAKE over time.

are used in initiating contexts are not just coordinated with any affirmative utterances in a general sense, but primarily with affirmative utterances where the truth of the utterance or the parent's compliance with the utterance is somehow *at stake*. There seems to be two main relevant contexts in this regard. First, initiating uses of NODDING often occur in *imperative* contexts where the child expresses a desire for something. An example of this was shown in Figure 6.3 (page 188) where Bella is trying to convince her mother to “borrow” a piece of plastic toy food. Second, initiating uses of NODDING occur often in more “epistemic” contexts, where the child suggests that “something is the case” and the truth value of the statement is to some extent at stake. Typically, the parents know best and have the role of the one who knows what is actually the case and it is usually the parent that have the final word (at least potentially) when it comes to concluding whether something was correct, true, appropriate, or similar, or if it was not. In both these cases, it is as if the child wants to show the preferred answer or interpretation to the utterance by marking the affirmative aspects explicitly.

Figure 8.4 shows that the proportion of the head-gestures that consist in uses of the HEAD\_SHAKE as opposed to NODDING, increases over time. That is, after a brief period where the two head-gestures are used close to equally often, there is then a quite sudden divergence at TP#1 where the HEAD\_SHAKE becomes much more common. The fact that the relative frequency of the HEAD\_SHAKE increases over time is consistent with the interpretation given above according to which there is a shift from responding to initiating uses of the head-gestures over time, and according to which negation may have to be marked explicitly more often than affirmation since affirmation can be seen as the default mode of communication.

These findings may seem to go against the findings of Guidetti (2005), who studied agreement and refusal in children at 16, 24, and 36 months, both in gesture and speech. She found that there was a general development towards more compliant modes of communication across time, characterized more and more by agreement than by refusal. However, these findings need not be contradicting at all. First of all, she studied both gesture and speech and found that agreement and refusal was most frequently expressed in speech (except in the youngest age group). This means that the pattern may look slightly different when the analysis is restricted to the use of certain gestures, as it is in the present chapter. Second, negative expressions such as the HEAD\_SHAKE need not signal disagreement. They are also used to agree with someone that something is not the case (A: “You don’t want that apple, right?”, B: “no”). In the initiating uses of the HEAD\_SHAKE described above, the connection to refusal is even weaker. The negative component in an utterance such as “she can not sleep+HEAD\_SHAKE” have very little to do with disagreement with the parent as such. It can equally well be a kind of agreement, and it can also be a more purely initiating kind of utterance where the child simply tells the parent something. Hence, there are no direct reasons to believe that Guidetti’s (2005) findings would not be generalizable to the children studied here. Indeed, just by watching the recordings of the five Swedish children it seems clear that there is a lot more disagreement in earlier sessions, around 18 months, than in later sessions, around 30 months, despite the “reverse” pattern in the frequencies of NODDING and HEAD\_SHAKE over time.

To summarize: In this section it was shown that there is a general developmental pattern from more responding (or reactive) to more initiating kinds of uses of NODDING and the HEAD\_SHAKE, and that uses of the HEAD\_SHAKE became more frequent than uses of NODDING across time.

### 8.2.5 Head-gestures in combination with other gestures

In Section 8.2.1 it was pointed out that the head-gestures can be combined not only with speech, but also with other gestures that are simultaneously performed by the hands. The reason for this is precisely that the head-gestures are performed with the head, with their own articulator. In addition, the head-gestures may of course be performed in sequential combination with other gestures.

Figure 8.5 shows the number of head-gestures that are performed without any other gesture or action (and with or without speech) and head-gestures that are performed with another gesture, simultaneously or sequentially within the same

## 8.2. HEAD-SHAKES AND NODDING

---

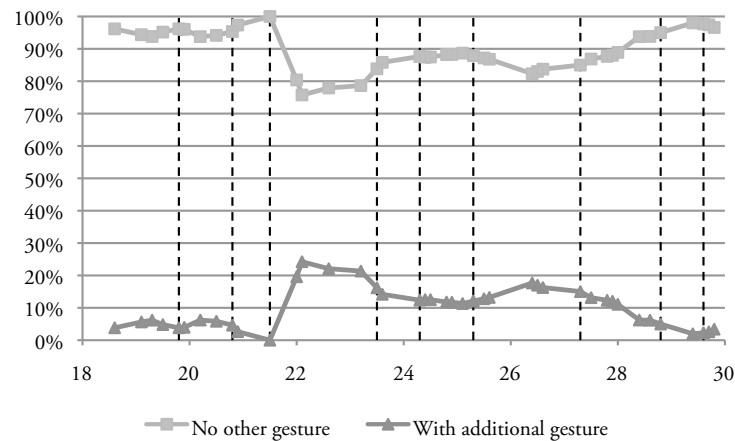


Figure 8.5: Head-gestures in combination with other gestures.

utterance. As it turns out, the period from 18 months to the end of TP#1 is vastly dominated by head-gestures that are not performed in coordination with another action or another gesture. Then, at the end of TP#1, there is a sudden increase in the tendency for the head-gestures to be combined with other gestures. SHAKE\_HEAD is more frequently combined with other gestures (16.4%) than NODDING (6.9%). The most common other gesture to go together with a head-gesture is, perhaps not surprisingly, pointing (78.1%). Almost all of the remaining ones are other gestures which are primarily deictic. There is only one instance of a truly iconic gesture being performed in combination with a head-gesture and there are a few instances of gestures such as SCOLD\_POINT and GONE being used together with HEAD\_SHAKE.

A curious fact about the sudden increase of the tendency for head-gestures to be combined with other gestures after the end of TP#1, apart from the sudden character of the increase itself, is that the children have both the head-gestures and the pointing gestures in their repertoires before this. That is, very low frequency of head-gestures combined with other gestures before the end of TP#1 cannot be explained by a lack of ability to perform these gestures as such. Instead, it suggests that the increase in gesture+gesture combinations at the end of TP#1 lies more precisely in the emergence of an ability to combine these gestures with each other, as illustrated in several of the examples in this thesis such as Figure 6.3 Figure 6.14, and Figure 8.8.

Another interesting thing about the curves in Figure 8.5 is that the eventual decrease of the tendency to combine head-gestures with other gestures from 26.4 months and onwards coincides precisely with the point in time where there was a

sudden increase in the number of multi-word utterances that were longer than two words (see Figure 5.8 at page 130). Therefore, one may suspect that at the point when children become more apt in constructing multi-word utterances (starting from 26.4 months in the children studied here) speech may begin to take over the role of gesture in the formulation of more complex multi-unit utterances.

## 8.3 Multimodal constructions

### 8.3.1 Introduction

In Section 8.2.4 I argued that there is a progression from responsive (or reactive) uses of the head-gestures to more initiating uses of these gestures in the children. An alternative way to understand this progression is to understand it as a progression from a more *fixed* kind of use of these gestures only together with a very restricted set of words (one or a few response morphemes) to a more *flexible* use of these gestures in combination with all sorts of units of speech, such as words and clauses.

Let us then return to Section 2.3.4, where I argued that one may distinguish between (A) combinations in the sense of [unit]+[unit], combinations, where the combined units (gesture+gesture, speech+speech, or gesture+speech) are truly separate units that are creatively or at least productively combined with each other, and (B) combinations in the sense of [unit+unit] combinations, where the “combined” units are rather a kind of “holophrases” than a matter of explicit combination. That is, *in some cases the “combinations” may be rote-learned rather than creatively or productively combined*, and this may apply not only to [speech+speech], but also to [gesture+speech] and [gesture+gesture] combinations. Henceforth I will speak of the first kind of combination as *flexible constructions*. The second kind of combination will be called *item-based constructions*. The term item-based refers to the fact that it is based directly on particular concrete expressions that the child has seen and heard (cf. Tomasello, 2003; MacWhinney, 2005).

Accordingly, during the initial period around 18 months where NODDING and the HEAD\_SHAKE are only combined with a restricted set of words, one could argue that this is a matter of multimodal item-based constructions. As these gestures gradually start to be used in more flexible ways, together with other words and clauses, this is a matter of multimodal flexible constructions. Hence, item-based constructions and flexible constructions should be understood as poles on a continuum rather than two discrete categories. Now consider the possibility that

this developmental pattern could be a much more general phenomenon that may apply to many other conventionalized gestures in addition to NODDING and the HEAD\_SHAKE.

At this point it will be useful to introduce some ideas commonly associated with the notion of “construction grammar” (e.g. Goldberg, 1995; Croft, 2001; Tomasello, 2003), since it is these ideas that have inspired the line of thinking to be presented in this section.<sup>7</sup> Some of the central ideas in construction grammar approaches to language are:

- An open-minded attitude towards what constitute the linguistic units that children employ when engaged in communication. For example, what may appear to an adult as a “combination” of two words may be a rote-learned and non-productive utterance that the child has learned as a whole.
- Grammatical “competence” is not operating on a single fixed level of generality or abstractness, as if it was entirely distinct from a lexical level of organization. According to construction grammar approaches, a “construction” can be anything from a single lexeme or a rote-learned collocation, to slightly more flexible constructions that can vary only to a limited extent, to more abstract and rule-like patterns.
- When it comes to children’s development, this is understood as beginning in the concrete and rote-learned end of the scale, and then it proceed towards more abstract and rule-like ways of organizing speech. However, more concrete kinds of formulaic expressions and collocations are never entirely abandoned, and remain an important part also in the linguistic competence of adults.

For example, in early language, an aspect of grammatical competence such as inflection may not be learned as a general rule. Just because a child uses a word like “walking”, it is not necessarily the case that the child has first learned the word “walk” together with a general rule that specifies how to apply the suffix “-ing” to an equally general category such as “verbs”. Instead, the child may simply have learned the word “walking” as a whole, or the “-ing” suffix may have been learned in a “local” sense, so that the child applies it only to a few specific verbs, but not to other verbs which are nevertheless in the child’s repertoire (e.g. Tomasello, 1992).

Even though Tomasello has written extensively about ideas of construction grammar in relation to the development of spoken language (Tomasello, 1992, 2003),

---

<sup>7</sup>Some of these ideas have already been touched on in Chapter 2 and Chapter 5.

and even though he has written extensively about gesture (Call & Tomasello, 2007; Tomasello, 2008), he has not applied the logic of the theorizing that he applies to the development of *spoken language*, i.e., the construction grammar approach, to the development of *gesture*.<sup>8</sup> In his studies of gesture, gestures have instead mainly been treated from the point of view of how they may constitute a forerunner to language, both in children's development and on an evolutionary time-scale. More specifically, he has treated gesture primarily in terms of the potential of deictic gestures (primarily pointing in his case) to establish joint attention, and in terms of the "natural" motivation of ACTION\_BASED iconic gestures (as discussed in Chapter 7). He has paid less attention to conventionalized aspects of gestures and to the potential similarities there may be between the development of conventionalized gestures and conventionalized speech as well as the fact that they may form expressive units *together*, i.e., multimodal constructions.

Accordingly, in this section I intend to explore precisely this connection that is "missing" in Tomasello's work. I will try to apply some of the ideas from construction grammar approaches to the use of conventionalized gestures and their coordination with speech. As far as I know, nobody has made this connection before. The possibility that I will suggest is that the openness with respect to what constitutes the "units" that children employ in their communicative performances that construction grammar approaches have endorsed could also be generalized to a multimodal analysis.

A drawback of construction grammar approaches to communicative development is that it is only focused on the child development as a matter of extracting expressive structures out from the perceptual "input", in a way which could be criticized for being overly static. There are often little considerations of things like intersubjectivity, the dynamics of interaction, or aspects of cognition other than extraction of structure on basis of the perceptual "input" such as how utterances are understood and so forth. Therefore, the use of various ideas from construction grammar approaches are not endorsed as a complete framework for the understanding of communicative development — it has instead mainly served as a source of inspiration regarding the particular aspects of the development that are discussed here.

---

<sup>8</sup>It may be interesting to note that Tomasello (personal communication, the 6th of August 2008) stated that the reason for this is not that he has considered this possibility and then rejected it, but rather that he had not thought of it before.



### 8.3.2 Item-based multimodal constructions

In Ekman & Friesen's (1969, p. 63) definition of "emblems", they pointed out that emblems are word-like in the sense that they have a specific form that is paired with a specific meaning, and that they are often "translatable" into one or a few words. *Less noted is that many of children's conventional gestures are not only "translatable" into one and a few words — they are very often also performed in concert with specific words.* To see this, one has to look at each child and gesture individually.

#### Harry's and Tea's SCOLD\_POINT

Harry performs the SCOLD\_POINT (abbreviated as S\_P below) a number of times in the data (see Figure 6.12 at page 203 for an illustration). In the first session at 17;26 he does not yet know how to perform this gesture properly. The gesture itself becomes a focalized topic in the interaction and his mother grabs his hand and tries to help him shape the gesture correctly.<sup>9</sup> Hence, at this point in time his use of this gesture has not yet crystallized. The next time he performs this gesture is when he is 2 years old (24;16), and in this recording there are two occurrences of the gesture. Both of these times he coordinates the gesture with the exclamative word *fy!* There does not seem to be any directly corresponding translation of this word in English, and therefore it is translated as "naughty". The first time he says "naughty!+S\_P" (*fy!*), which is a direct imitation of what the mother did in the previous turn. That is clear evidence that this precise multimodal construction is available in Harry's "input". The second time he says "said naughty+S\_P" (*sa fy*). Then the gesture appears again, four times, at 25;10. Three of these times he says "naughty+S\_P dog" (*fy hånne*). The fourth time he says "naughty+S\_P xxx" (*fy bana*) and it is unclear what *bana* means as it is not a proper Swedish word. (Therefore this word is not translated but simply marked as "xxx".) Judging from the context, it may have something to do with someone who is taking a bath (*bada*) with the clothes on. Finally, there is also one occurrence of this gesture at 26,18. This occurrence was shown in Figure 6.12 (see page 203). Harry says "naughty you+S\_P" (*fy dej*). Thus, in this case too Harry says *fy*, but in contrast to the other occurrences, the stroke of the gesture does not fall on the word *fy*.

Tea also use the S\_P gesture. At 23;07 she says "naughty naughty+S\_P" (*aj aj*). Note that *aj aj* is another expression than *fy*, but that it is nevertheless translated to

---

<sup>9</sup>That is to say, the normativity involved in the performance of this gesture is made explicit by the mother, as a matter of Conv#4.

“naughty naughty” since there seems to be no corresponding exclamative in English. Both *fy* and *aj aj* are conventionalized expressions that are recognizable as typical in coordination with this gesture. At 23;07 she also performs the SCOLD\_POINT two times together with the word “dä+S\_P”. In these two instances the gesture is combined with speech much like a “normal” pointing gesture. At 25;17 she uses the gesture twice while saying “xxx+S\_P” (*döti* — not a recognizable word) and “doll naughty+S\_P” (*docka fy*).

### Bella's GONE gesture

The GONE gesture is not performed by any of the children except for Bella, and she performs it in four different sessions. The first three occurrences appear at 20;23. She says “x xxx gone+GONE” (*e chitt båtta*), “and gone+GONE” (*å botta*, illustrated in Figure 8.6), and “x it gone+GONE” (*a de botta*, illustrated in Figure 8.7).<sup>10</sup> Then comes two instances at 23;17. Bella says “now finished+GONE” (*nu slut*) and “x finished+GONE” (*a a slut*). At 26;13 there are four instances of the GONE gesture: “gone+GONE+HEAD\_SHAKE” (*borta*, illustrated in Figure 8.8), “it finished+GONE” (*de slut*), “finished+GONE” (*slut*), and one instance without speech. At 30;17 there are two instances: “I don't know+GONE” (*ja vet inte*) and “I don't know+GONE” (*ja vet inte*). These two instances follow directly after each other.

Thus, until 30 months there is no evidence that Bella uses this gesture with any other words than “gone” (*borta*) and “finished” (*slut*). At 30 months, the somewhat formulaic expression “I don't know” (*ja vet inte*) is also used with this gesture (this exact utterance occurs many times in the data). Just like in the case of Harry's SCOLD\_POINT, all of these three precise combinations are performed by Bella's mother at different occasions in the data, including the one with the HEAD\_SHAKE, i.e., “gone+GONE+HEAD\_SHAKE” (*borta*).

### The HELLO gesture in four children

The HELLO gesture occurs in four of the five children. Bella performs this gesture once in the data, at 19;28, and says “jajaj hello+HELLO” (*jajaj hej*). The first part of the utterance (*jajaj*) is a singing-like vocalization. Harry performs this gesture three times. At 17;26 he performs it without speech. At 20;26 he says “hello hello+HELLO” (*hej hej*, illustrated in Figure 8.9). At 24;16 he says “hello+HELLO” (*hej*). Hanna

---

<sup>10</sup>Again non-recognizable “words” are simply marked with “x” or “xxx” in the English translation.

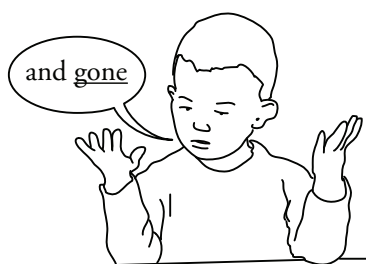


Figure 8.6: GONE (Bella 20;23)



Figure 8.7: GONE (Bella 20;23)

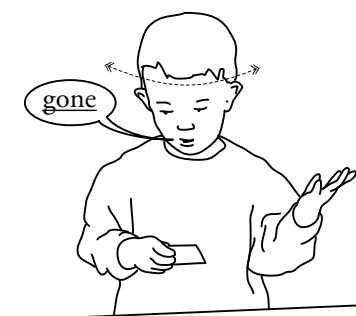


Figure 8.8: GONE + HEAD\_SHAKE (Bella 20;23)

uses this gesture three times in a single session at 29;22. Two of these times she performs it without speech and the third time she says “Hanna waves+HELLO” (*Hanna vinkar*, illustrated in Figure 8.10) which is more like the naming of the acts that were found in many of the gestures with iconic aspects that were analyzed in Chapter 7. Tea also use this gesture without speech a few times: Twice at 24;25 and once at 30;23. At 19;15 she says “grandma+HELLO” (*mormor*), at 25;17 she says “grandma+HELLO” and at 26;12 she says “hello hello grandma+HELLO” (*hej hej mormor*). In all of Tea’s performances of HELLO, the gesture is directed to her grandmother, who is also present in the room but not involved in the interactions except as a recipient for these occasional HELLO gestures.

Thus, in 10 out of 13 instances the HELLO gesture is either performed without speech, or specifically with the word “hello” (*hej*). Both of these variants are also performed by the parents in the data. Tea’s use of the word “grandma” (*mormor*) together with the HELLO gesture is potentially a bit more “productive” in the sense that an open-class word is used to name the recipient of the act, and it seems possible that she may be using this gesture together with different names depending on the recipient in a given situation. Yet it is also striking that the recipient is in fact precisely the same person each time, and in a similar kind of situation, which



Figure 8.9: HELLO (Harry 20;26)

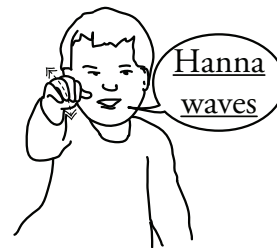


Figure 8.10: HELLO (Hanna 29;22)

suggests that this “wave hello game” with the grandmother may be something like a recurrent ritual. Also, as mentioned above, Tea does in fact use the “hello hello” (*hej hej*) expression in one particular case.

### Bella's BYE\_BYE gesture

The BYE\_BYE gesture differs in form from the HELLO gesture in that the waving motion is performed in a larger fashion, involving the whole arm, whereas the HELLO gesture is typically articulated mainly by wiggling the fingers up and down. It is unclear whether these differences mainly reflect differences in how the children happen to perform these gestures in the data, or whether there is a distinct difference between the form of HELLO and the form of BYE\_BYE that is established with cultural standards. My own intuitions tell me that the exact forms involved in saying hello and bye bye gesturally are quite variable overall, and sometimes personal. In either case, the BYE\_BYE gesture occurs in four of Bella's sessions. At 18;09 there are four instances where she says “bye bye+BYE\_BYE” (*hejdå*, one of them is illustrated in Figure 8.11). At 19;28 there are two more instances of the same multimodal utterance and at 20;23 there is yet another one. At 25;03 Bella says “fly bye bye+BYE\_BYE” (*fluga hej då*).

### Discussion of item-based multimodal constructions

This section has shown that conventionalized gestures in children often seem to be performed in coordination with a quite limited set of expressions in children between 18 and 30 months. Those of the gestures that were performed by more than one child are often performed with the same words in different children. This is evidence for wide-spread conventions regarding some typical ways of organizing

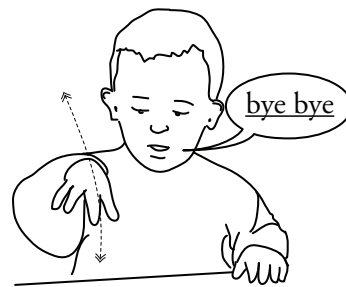


Figure 8.11: HELLO (Bella 18;09)

these gestures in coordination with speech. In many cases it was also possible to find instances in the data where the parents used the same gestures in coordination with the very same words.

The stroke of the gestures almost always co-occurred with the part of the utterance that corresponded to the recurrent spoken expression. Even though the expressions seem to be wide-spread and conventionalized, there was also evidence for personal profiles of use. For example, some children used some of the gestures more with one expression whereas another child may use the same gesture more with another expression. This is in accord with the item-based logic, since it serves to acknowledge that each individual child has to learn various conventions anew, and that not every child may pick up on the same set of conventions, especially in the initial stages of their acquisition.

Some of the gestures seemed to be used without speech more often than other gestures — at least by some of the children. Note that this is not evidence against the idea of item-based gesture+speech constructions, since this idea does not predict that there cannot also be rote-learned speech-only and gesture-only constructions. In a similar way, some of the gestures seem to recur in coordination with more than one spoken expression, sometimes in a single child. The idea of item-based gesture+speech constructions does not predict that there cannot be more than one rote-learned way of combining a certain gesture with spoken expressions. There was also one instance of a construction that involved two gestures and one spoken word: “gone+GONE+HEAD\_SHAKE” (*borta*, illustrated in Figure 8.8). It was notable that this exact multimodal configuration was also performed by the parent in at least one occasion, in a different session, which suggests that there may in some cases be more than one gesture involved in the construction.

It should also be said that the exact form involved in the articulation of a particular gesture often varied a little between the instances. For example, Bella’s GONE

gesture is performed in a one-handed way at some occasions (e.g. Figure 8.8) and in a two-handed way in other occasions (e.g. Figure 8.6 and Figure 8.7). Do these variations correspond to differences in meaning? I would argue that although differences of this kind may be indicative of a difference in meaning in adults (cf. Harrison, 2009b), in these children they are typically not. First, the recurrent use of certain expressions together with these gestures is evidence that despite the slight variance in form on the articulation of the gestures across instances, it seems to be the same gesture. At least that is how it seems to an observer. In accord with the operation-act-activity interdependence idea, it is not only operation factors (form) that contributes to the identification of these gestures, but also other factors such as co-occurring utterances. Second, the one-handed and the two-handed performance of the GONE gesture does not, for example, correspond to the two different expressions — “finished” (*slut*) and “gone” (*borta*) — used by Bella together with these gestures, but rather seems to cut across the use of these expressions. Furthermore, the children are in the process of learning to use these gestures and one may neither expect an entirely consistent use of certain forms, nor that the children would use a certain form in the exact same contexts that adults may use them (cf. Kendon, 2007, p. 12). It should be recalled that when it comes to speech, words are not always pronounced in the exact same way every time either (also in adults). Nevertheless, they may still correspond in a recognizable way to the “ideal type”.

There does not seem to be many studies that may throw light on the question of whether adults also sometimes use concrete item-based gesture+speech constructions. In some cases it is more or less obvious, such as the use of NODDING and HEAD\_SHAKE that often co-occur with response morphemes in adults too. The fact that the parents produced many of the particular speech+gesture constructions analyzed in this section is also indicative. In an unpublished study of German recurrent gestures it seemed to be the case that a gesture called the ABWINKEN gesture frequently occurred with the specific expression “ist egal”, which means “doesn't matter” (Jana Bressemer, personal communication, 7th of December 2008). Then again, many recurrent gestures are used in much more flexible ways together with speech in adults (cf. Kendon, 2004) and highly recurrent gestures such as POINTING, NODDING, and HEAD\_SHAKE come to be used in flexible ways already in young children. This is in line with the general logic of construction grammar, since it is expected (A) that development will begin in the concrete and more item-based end, and then proceed towards more flexible and abstract ways of using and organizing the constructions, and (B) that some constructions will remain more concretely organized even later on in development.

Finally, it is important to note that the phenomenon may be more general than only involving those actions that correspond to a category of emblems from an adult's point of view. For example, it was mentioned in Chapter 7 that Harry frequently says things like “forwards” (*framåt*), “reverse” (*backa*), and “brrrm brrm” (*brrm brrm*) when making certain movements with a toy car, and in this case too there are a number of instances in the data where his mother does the same things. Nevertheless, since conventionalized gestures and conventionalized multimodal constructions are among the kinds of actions that are recurrent, they will be subject to this logic, among other recurrent actions.

#### 8.3.3 Flexible multimodal constructions

More flexible ways of using recurrent and conventionalized gestures with speech have already been observed in association with NODDING and HEAD\_SHAKE earlier in this chapter. Graziano (2009) studied the development of the gesture that has been referred to as the GONE gesture in this thesis, as it is used by Italian children in narrative discourse.<sup>11</sup> She found that it was used with more restricted meanings at 4 years (which was the earliest group of children investigated), such as together with utterances such as “I don't know” (*non lo so*), “It's finished” (*basta*), and “I don't remember” (*non mi ricordo*), which corresponds very well to some of the utterances coordinated with this in the Swedish children as reported earlier in this chapter. The 4 year olds in her study also sometimes used the gestures in more flexible ways with expressions such as “and then there's the tree” (referring to the last image of a cartoon). She also found that this gesture was used in a variety of more complex functions later on at 4–6 years. All in all then, there are reasons to believe that there may be a similar kind of development from initial item-based constructions that involve the GONE gesture to later uses of the GONE gesture that stands in a more flexible relationships to the spoken utterance.

Is it possible that a similar developmental logic underlies the coordination at different ages between pointing gestures and speech too? That is, is there any evidence that pointing is coordinated with a more restricted subset of the vocabulary than what would be expected only by judging from the child's general productive vocabulary? Of course, at the ages studied here, pointing is already used in quite flexible ways with a huge array of open-class words such as nouns. Nevertheless, in Zlatev & Andr en (2009, p. 393) we found a substantial increase at 26–27 months

---

<sup>11</sup>She refers to this gesture using Kendon's terminology, as PL (palm with lateral movement).

in the number of pointing gestures that co-occur with activity verbs, such as “run” (*springa*), and other motion words, such as “away” (*bort*) in the sense of “moving away”. For example, a child may say “many horses run fast there+POINT” (*många hästa springe fort dä*) (Tea 28;19) or “they run+POINT” (*dom springe*) (Tea 30;23) and coordinate the pointing gesture with the verb (as in both examples) and possibly also with the verb only (as in the latter example). This suggests that it may be the case that pointing gestures come to be used in even more flexible ways, together with new kinds of words, around this age.<sup>12</sup>

### **Another possible interpretation of gesture in the transition from one-word speech to two-word speech**

There are a number of studies that have investigated the role of gesture in the transition from one-word speech to two-word speech and related phenomena (e.g. Capirci et al., 1996; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005b; Iverson et al., 2008; Özçalışkan & Goldin-Meadow, 2009). This was briefly discussed in Section 2.3.4, but in short: In this line of research it has been shown that there is a stage when children tend to use gestures in “redundant” ways so that the meaning of the gesture roughly corresponds to the word that is used together with the gesture. At a later stage, children may use gestures together with speech that expresses a different meaning than the gesture, so that the combination of the gesture and the speech function similar to a two-word utterance. This later “non-redundant” use of gesture and speech also happens to be a good predictor for the emergence of two-word speech, which comes some time *after* the “non-redundant” gesture+speech combinations.

Similar results were also presented earlier in this chapter (see Section 8.2.5) where it seemed to be the case that there is a stage from around the end of TP#1 and roughly up to TP#3 when there is an increased tendency to combine gestures such as NODDING and HEAD\_SHAKE with other gestures (with or without speech) and that this period preceded the period when multi-word utterances with more than two-words eventually become more common.

One may now ask whether these findings could be interpreted, not just as a development from “redundancy” in the gesture+speech combinations to a later stage of “non-redundancy”, but as a development from item-based multimodal constructions to flexible multimodal constructions. A possible problem for this proposal

---

<sup>12</sup>Note that 26–27 months was also the point in time when the number of 3- and 4-word utterances increased quite sharply (see Chapter 5).



is that some examples provided of “redundant” combinations in some of the studies mentioned above involve not only a more constrained range of words such as “this”, “there”, and “here” and so forth, but also such things as pointing to a chair and using a content-loaded open-class word such as “chair+POINT”. One may still hypothesize that there may be an even earlier stage in the beginning of the “redundant” period where the first words used together with pointing would rather be a more restricted set of words such as the deictic “there”, “that”, “here” and so forth or that the gestures are performed alone, without speech. This cannot be investigated further here since this would require data from younger children. Nevertheless, it does at least seem to be a possibility, which is all that is claimed here. More precisely, it may be the case that early gestures such as pointing are initially used either without speech or with a quite restricted set of words, before they start to be combined with a larger range of words, and that this may not only be due to the small vocabulary that children have at that time, but that it could also be a matter of learning concrete item-based multimodal constructions that gradually come to be used in more flexible ways later on in development.

## 8.4 Conclusions

The first part of this chapter was concerned with NODDING and HEAD\_SHAKES, or the head-gestures, as one may call them. It was shown that although the children initially use these gestures exclusively to respond to parents, they do not use them as back-channeling, as feedback to the interlocutor while the interlocutor is speaking. At around 22 months, there was a marked increase in the tendency for the head-gestures to be combined with other gestures. From around 20 to 21 months, the children used the head-gestures increasingly often for other functions than responding to parents — as part of more initiating aspects of their utterances.

In the second part of this chapter it was argued that emblems are not only word-like in the sense that there is a conventionalized form with a conventionalized meaning (cf. Ekman & Friesen, 1969); there was also a strong tendency for the children to use the conventionalized gestures together with specific utterances. Therefore it was argued that in some cases, gesture+speech “combinations” are not combinations of separate units, but rather a matter of multimodal wholes. This phenomenon was dubbed *item-based multimodal constructions*, inspired by terminology from construction grammar approaches to language development (e.g. Tomasello, 2003). In item-based multimodal constructions, gesture and speech do not seem to consti-

tute opposite modes of representation (cf. McNeill, 2005, p. 18), but rather seem to form multimodal wholes.

## CHAPTER 9

# Communicative action gestalts in the manipulatory area

*The stone which the builders rejected, the same is become the head of the corner*

Matthew 21:42

### 9.1 Questions asked

In Chapter 2 it was suggested that there is nothing that prevents an action gestalt that is semiotically complex from including the handling of an object. It was also argued that there is nothing that prevents an action gestalt that is not semiotically complex from nevertheless being explicitly communicative. The analyses presented in Chapter 6 and Chapter 7 were conducted in accord with these suggestions, and I hope to have shown that it is productive to view the nature of gesture in this way — as a matter of family resemblance, involving a range of different qualities that may vary with respect to each other such as different levels of semiotic complexity, communicative explicitness, and conventionality, as well as different forms of intentional perspectives. The present chapter provide an in-depth empirical exposé of a large number of action gestalts that (A) involve objects and (B) have some sort of communicative appearances, as well as considerations of how these communicative action gestalts come into being.

Even though the analysis in this chapter is based on observations of the actions that appear in the data, the analytic focus rather lies on more generic aspects of the

communicative action gestalts in the manipulatory area. That is, while analysis in this chapter is concerned with quite detailed phenomena, the analysis nevertheless has a generalized character. Less attention will therefore be paid to which child did what, at what age, and so forth, and rather than presenting each example in detail, with pictures and the original Swedish utterances, the examples that are brought up are treated in a somewhat idealized way.

The starting point of the analysis is the observation that a very large amount of the utterances in the children observed here are performed in temporal and semantic coordination with some sort of action. A mildly conservative coding of the corpus reveals that, on average, 51% of the utterances involve some sort of act of bodily communication, and for four of the five children the rate lies over 70% in one or more sessions. The maximum rate observed in one session is 84%.<sup>1</sup> This includes both empty-handed gestures and acts that involve handling of objects, but in either case, they are all actions that are performed in systematic coordination with speech.

The aim of the present chapter is to bring some further clarity into the nature of those gestures that involve handling of physical objects, and how these are coordinated with other semiotic resources — mainly with regard to speech. The overarching questions asked are:

1. *What sort of acts on objects have communicative appearances?*
2. *What is it that makes these acts recognizable as certain kinds of action gestalts?*
3. *What makes these action gestalts distinguishable from each other?*

As the hands move along — approaching objects, grabbing them, picking them up, using and exploring them, holding them towards other objects, locations, and recipients, putting them down, and eventually withdrawing from them — the actions they perform can take on communicative appearances in various ways. This may in part result from the ways they themselves are performed — although many gestures appear a quite unspecific in meaning when seen in isolation (Kendon, 2004, p. 161) — or it may be due to their orchestration with other semiotic resources such as gaze orientation (Streeck, 1993; Gullberg & Kita, 2009) and speech (Kendon, 1980b; McNeill, 1985), as well as the ways the acts are embedded in, and create, the sequential organization of social activities. As argued in section 3.4 (p. 86), a full non-reductionist understanding of the meaning of an action requires that attention is paid both to operation-act interdependence and to act-activity interdependence. The meaningful appearances of an action cannot be understood without

---

<sup>1</sup>See Table 5.2 (p. 120) for descriptive statistics.

taking both its form features and its embeddedness in social activity into account.

Semiotic resources differ from each other, already on a general level, in the ways they become communicatively highlighted (Kendon, 2004, p. 13):

[...] some actions are almost invariably assigned main-track status, simply by virtue of what sort of actions they are. Vocalization and speech are specially oriented to and take a kind of first place in the attentional hierarchy. However, certain patterns of bodily movement are also given main-track status simply because of the character they have as movements.

Speech is therefore, in this sense, different from gesture. Harding (1983) has shown how mothers consistently endow the vocalizations of infants' with communicative significance. This means that vocalization may in itself be enough for parents to interpret the child's acting as communicative, even though it may be unclear *what* the child may be trying to communicate. Research on speech intelligibility has demonstrated that it is often hard to identify words in children's speech on the basis of sound only, and that *video* data, which includes visible aspects of where children look and what they do *etc*, provides a substantial source of help in identifying words in the "audio" signal (Vihman & McCune, 1994). In one experimental study that was based on recordings of 24 month old children's speech, 76% of the words investigated were unintelligible from audio recordings alone although they were selected for the experiment in accord with the criterion of being fully intelligible in context (Navarro et al., 2005, p. 1696). Part of this "context" was also other words in the same utterances, but then again, many of the utterances are one- or two-word utterances.<sup>2</sup> In another study, Thompson & Massaro (1986) designed an experiment where it was found that pointing gestures could influence the interpretation of speech sounds (/ba/ versus /da/), and that this effect was strongest when the speech sounds were ambiguous. In yet another study, on adults, Berger & Popelka (1971) found that speech intelligibility was highly improved when speech was combined with gesture (well-known emblems), in the experimental task they studied. That is, the orchestration of speech and bodily movement is not only co-determinative on the level of semantic processes such as precisification, elaboration, vaguification

---

<sup>2</sup>As pointed out by Margaret Fleck (personal communication, 5th of February 2010) one should not commit the mistake of assuming that *adult* speech is always intelligible without contextual support either. Nevertheless, there is still a quite notable difference between the intelligibility of the speech of children between 18 to 30 months and of adults. In the study by Navarro et al. (2005, p. 1696) the intelligibility rate for adult speech was around 93%.

of what is said (Kendon, 2004, chapter 9; Hirsch, 2009), but to some extent this orchestration may even provide for the intelligibility of speech and identification of certain words in the first place — especially in children.<sup>3</sup> More generally, there is also evidence from neuroscience that corroborates socio-cultural ideas about perceptual processes, finding that hearing a word indeed transforms what people are able to see (e.g. Lupyan & Spivey, 2010).

No matter if the performance of an action becomes communicatively highlighted as a result of its movement dynamics, through its orchestration with speech and other semiotic resources, contextual factors, or a mixture of these, the net effect is that some stretch of movement suddenly appears foregrounded as part of a communicative gestalt — i.e., ending up in the “main track” of the interaction (Goffman, 1974). As such, it stands out, as a stroke, from other movements that may have been going on before, and movements that may follow after it (Kendon, 1978, 1980b, 2004; Arendsen, 2009).<sup>4</sup> This means that the distinction between movement that is communicatively highlighted, and such that is not, does not simply correspond to a distinction between movement and non-movement. Apart from the interplay between the bodily movement and other semiotic resources and contextual factors, it is also a matter of *how* the hand, or some other articulator, is moved, shaped, oriented, and located (cf. ten Holt et al., 2009; Stokoe, 1960; see also Stokoe, 1960).

The full set of communicative acts that involve objects is, however, quite heterogeneous and in some sense “messy” since such acts are fused with the logic of practical action. Streeck (2009b, p. 23) points out that “in the context of conversation, it is usually not difficult at all to identify gesture units and their boundaries. In contexts of work, however, gestural communication may consist in nothing more than a repetition or a slight embellishment or exaggeration of an instrumental act.” These acts vary from clearly articulated communicative acts, which may have communicative appearances in themselves, to more or less instrumental acts that are only indirectly assigned some sort of communicatively directed attention in the form of a loosely based comment upon the action, after it was performed. An example of the latter is when someone smashes an irritating mosquito which was sucking blood from the skin of the arm, due to purely private reasons, and only after this turning to an interlocutor to say “ha! I got it!”. In such situations, the interlocutor is expected to understand the utterance in relation to the previous (“private”) act, including among other things the deictic reference of the word “it”, which is to be

---

<sup>3</sup>In this regard, it is also interesting to note that the motor systems in the brain plays an important role also in perception of speech (D’Ausilio et al., in press).

<sup>4</sup>The concepts of action gestalts and gesture phases are described in Section 3.3 (p. 79).

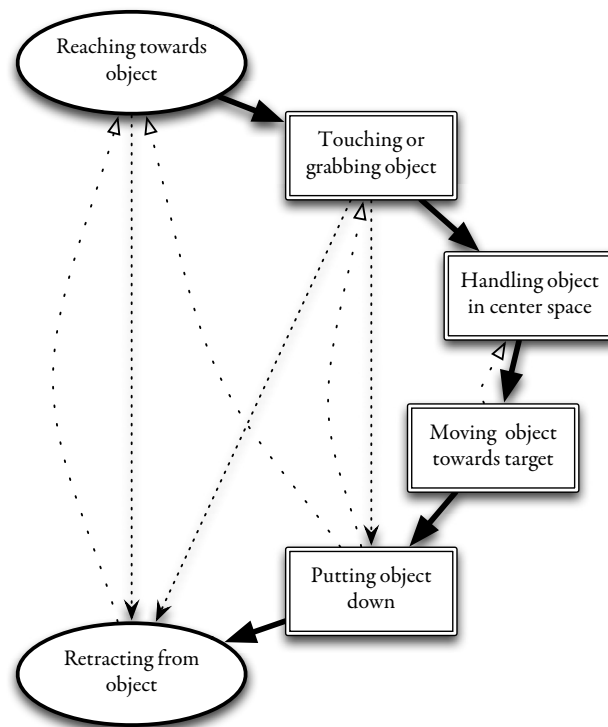


Figure 9.1: Idealized scheme of common trajectories through action excursion space.

understood with regard to what the action was directed towards.

The treatment of the relation between action and speech in the present analysis is, however, restricted to coordination between action and speech in the sense of direct temporal (and semantic) coordination with action and movement. More indirect relations between acting and speaking such as the mosquito example are left aside. This does not mean that the remaining set of acts is very homogeneous. It is rather characterized by family resemblance, as a result of the interplay between a complex set of factors, and it would not be possible to articulate this analytically in terms of a set of clear-cut typology of action types. On the other hand, ordinary language teaches us that there are indeed typified sorts of action, such as “throw”, “give”, “pick up the phone”, and so forth, and as competent members of various cultural communities, we identify acts in terms of such categories on a routine basis. Although such ordinary language labels for actions are often quite vague with respect to the actual performances involved, they also indicate how we experience action in a typified way (cf. Section 2.2.3). How to approach such action in a systematic and principled way?

Rather than starting out from a particular categorical typology of pre-established

gesture types, the analysis presented in this chapter starts out from the unifying fact that all sorts of actions that involve handling of objects can all be viewed as instantiations of *manual excursions* — beginning with the approaching of an object, and ending, sooner or later, with the disengagement from the object and a retraction of the hand/arm. That is, the underlying idea is that it is possible to postulate an *action excursion space* that covers most of the possible communicative acts that *involve objects* in a principled way (see Figure 9.1, p. 275).<sup>5</sup>

The analysis is performed in the manner of dividing this space of possibilities into a number of subspaces, and then investigating the possibilities for communication that arise in each of those. As the analysis will show, approaching the activity in this way reveals a structured set of possibilities and contingencies for acting and communicating that are frequently exploited by children (and adults). Each of the subspaces present different opportunities and constraints for communicative action.

Even though there are of course a vast array of differences in the ways that gestures are used in different cultures, it is my conviction that this represents an aspect of action that should be expected to exhibit a number of similarities across most cultures, since it constitutes a relatively fundamental condition of the acting human body. The ways in which the hands are able to gear out into the intersubjective sphere of the manipulatory area, achieving communicative significance, are neither unconstrained nor arbitrary.

The division into subspaces can be done with different degrees of granularity, but for the purposes of the present analysis, the following divisions, or “slots”, were chosen: Reaching towards the object, touching the object, grabbing the object, handling the object in center space, moving the object towards a target, putting the object down, and finally, withdrawing from the object, as shown in Figure 9.1.

Different actions differ in the way they extend over one or more of these subspaces (see Figure 9.2). That is, what is generally perceived as “one action” — an action gestalt — varies with respect to how many of the subspaces of action excursion space are subsumed by the action. For example, in the case of TRANSPORT, where an object is seamlessly picked up, moved to another location, and finally put down, several steps are involved. In the case of PUT though, only the last part of TRANSPORT is involved, i.e., putting an object down.

---

<sup>5</sup>Of course, the analysis will not be, and can never be, exhaustive. The types of action reported here reflects what is available in the data. Still, the approach in itself is framed in a way that theoretically allows for analysis of all relevant actions.



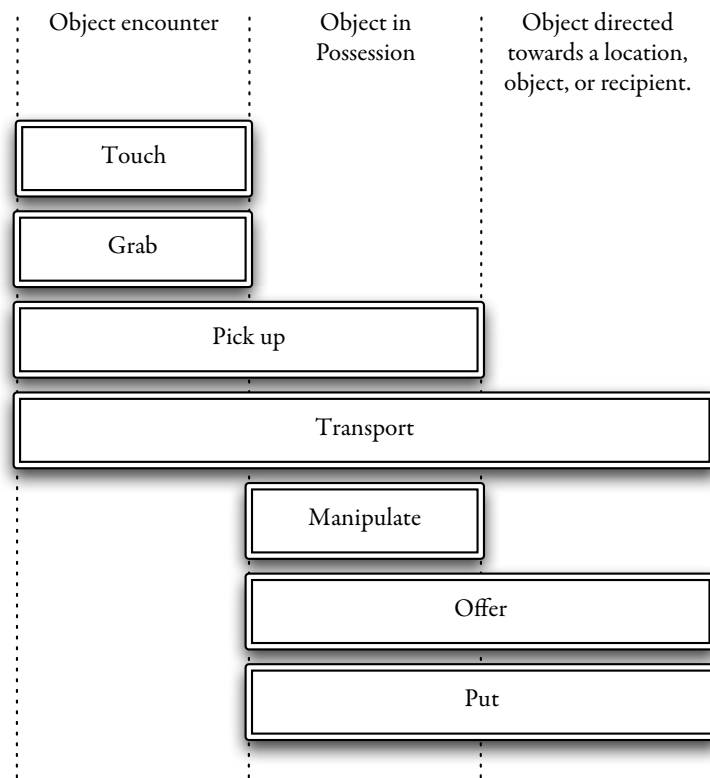


Figure 9.2: Simplified view of the extent of some common action gestalts.

The next sections consists of an analysis in the form of a systematic “walk” through the various communicative manifestations that tend to appear in each slot of action excursion space, beginning with REACH and ending with WITHDRAW. It is an attempt at carving out the intersubjective and communicative structure of manual action in the manipulatory area at its joints, at the “emic” operational level (rather than an “etic” level), where differences in performance usually make a distinct difference to the overall meaning of the act. All presented types of communicative actions appear at least a few times in the corpus — in most cases dozens, or even hundreds, of times — and they appear *in different children in different situations*. Highly singular types of communicative actions, which happen only once or maybe twice in the data analyzed here, are left out from the presentation provided here. The analytic focus does not, however, lie on quantification of how common various types of actions are, but rather on how the acts are constituted. Note that this analysis is quite dense, and readers who are mainly interested in the overall points that emerge in the course of this analysis may want to skip to the summary and conclusions provided in Section 9.8 on page 315. Readers who are interested in the detailed workings of each particular subspace are encouraged to continue reading.

## 9.2 Reaching towards the object

### 9.2.1 Definition

REACH consists of moving an empty hand towards an object, but still not touching it. Since the hand is still empty, it is yet possible to perform empty-handed gestures such as pointing towards the object. Excursions can also be interrupted in the middle of the REACH, turning into WITHDRAW, or into a new REACH to another target or an empty-handed gesture. The movement may also stop, resulting in a hold, without being abandoned, and it can be part of a more specific content-loaded performance. When the phase of reaching towards an object receives main-track status as a foregrounded communicative action gestalt, the communicative function that emerges is often similar to that of pointing, also in cases which do not, in fact, turn into actual pointing gestures such as the prototypical index finger pointing gestures. That is, such acts often appear to indicate, locate, and/or single out, directly or indirectly, an item of relevance in the current activity. The analysis only includes actions that do involve a hand-shape associated with grabbing though, and while pointing gestures do, in one sense, also belong to this slot in action excursion space, they are instead analyzed in Chapter 6. The same goes for empty-handed iconic gestures that are performed towards an object, which are treated in Chapter 7.<sup>6</sup>

Reaching often has future-oriented qualities. Studies of reaching in adults has shown that the movement is generally adapted, already during the reaching movement, to the goal of the act to be performed once the object has been picked up (Johnson-Frey et al., 2004). In a similar vein Sudnow (1978, p. 34) notes how his hands are “going for” particular sounds already in the phase of reaching towards a particular key on a piano keyboard. This is also similar to preparation phases of empty-handed gestures, which are embarked upon before the actual gesture is about to be performed in a timely manner together with some particular part of the ongoing flow of speech (Kendon, 2004, p. 116). Also, in the context of dancing, movements of the body need to anticipate the moment when the next beat is about to come, since this typically involves a stroke-like movement in coordination with the

---

<sup>6</sup>Empty-handed gestures can also be performed towards objects and in one sense such gestures also fit in this slot of action excursion space because they can be initiated directly from a home position without first picking some object up. On the other hand, empty-handed ACTION\_BASED gestures that involve the use of pretended objects also share a number of semiotic properties with the acts analyzed in the section titled “Moving the object towards a target” (page 295), where a real object is held in the hand. This goes both for ACTION\_BASED iconic gestures where the hand appears to hold a non-existing object and ACTION\_BASED iconic gestures where the hand itself “embodies” an object usually held in the hand such as a scissor or a gun.

beat which requires preparatory movement. Apart from often being coordinated with future oriented speech, REACH towards objects is often also *visibly* future oriented, having a visible character of being “already planned”. This can be quite subtle, but it certainly forms a part of what makes REACH appear the ways it does, also in children. It can be revealed by the hand-shape, movement dynamics, and the direction of the movement. Other studies have shown that a lot of the “information” available in gestures and signs is perceived early, already around the time when the preparation phase has just turned into a stroke, and that even seeing the preparation (or retraction) phase alone may yield surprisingly good rates of recognition (in the case of signed language) (Arendsen, 2009; ten Holt et al., 2009). This rhymes well with the thesis argued for by Sudnow (1972, p. 260), who states that “for many interactional sequences there seems to exist a requirement for interpretation with ‘no more than a glance.’”

### 9.2.2 Analysis

Sometimes, during an *ongoing* and uninterrupted reaching movement, a spoken utterance highlights the movement (REACH [ONGOING]). Since REACH is, in such cases, part of an ongoing movement, the movement generally ends with, or at least includes, a GRAB or some manipulation of the object.<sup>7</sup> Furthermore, the act of grabbing generally appears more stroke-like and highlighted than the preceding REACH, and in such cases REACH [ONGOING] tends to dissolve into the larger action gestalt of GRAB, as a mere operational part of the GRAB.

Still, there are some systematic features to be found in the ways that speech is coordinated with REACH [ONGOING]. In most cases observed in the data highlighting of ongoing REACH movements happens in the case of multi-word utterances where onset of the first part of the utterance co-occurs with the *onset* of the reaching movement, and the second part with the moment of grabbing the object. Examples of utterances include “take+REACH orange+GRAB” when reaching for, and grabbing, an orange pencil, “should have+REACH that+GRAB”, and “and then+REACH such there+GRAB”. While the children mainly coordinate GRAB with speech that refers to objects, speech during the preceding REACH is often future-oriented, either through future tense, progressive aspect, or by referring to past acts that are about to be performed again (“more+REACH such+GRAB”), or by referring to a next step of a recognizable procedure (“and then”).

---

<sup>7</sup>GRAB is further analyzed in the next section.

Sometimes a reaching movement is *stopped* before the hand touches the target, similar to a hold of an empty-handed gesture (REACH [HOLD]). This in itself produces much stronger appearances of a communicative foregrounding than REACH [ONGOING], and REACH [HOLD] acts generally attain main-track status through the very nature of the movement. The significances that appear in such situations are partly different in cases where the object is out of reach and when it is within reach.

In cases of REACH [HOLD] where the object is *within reach*, the arm is typically not being fully extended. Since the object is visibly within reach of the child, but still not attempted to be taken, these movements are generally not interpreted as a imperative “pointing” requests to get hold of an object, which is what would generally happen if the object was out of reach. Sometimes they do nevertheless appear in an imperative function, but then generally as part of a request for the parent to do something with the object, rather than a request for the object itself. In such cases the halted reaching movement appears more like a non-touching pointing gesture that draws on the qualities of non-involvement and passivity associated with non-touching despite a visible potential to actually touch the object since it is within reach. The imperative request for the parent to do something with the object is often made more explicit through a spoken utterance, such as “mommy take+REACH [HOLD]” or “you do+REACH [HOLD]”. To the extent that there is an about-to-grasp hand-shape employed in the act, this has the effect of construing the object as an object which is relevant to GRAB by means of the iconic resemblance of grabbing.

At other occasions, REACH [HOLD] towards an object that is within reach may produce appearances of searching for an object among a set of possible candidates. For example, the child may look around at the table in front of him/her, while holding out a hand, ready to seize a suitable object once it has been identified. Examples of this include searching for a particular piece of a puzzle, or searching for a toy building block of a certain color. Spoken utterances such as “uuhm+REACH [HOLD]” or “such piece+REACH [HOLD]” may also contribute to the foregrounding of such movements. Such searching movements are not always strictly speaking “holds” though, because there is sometimes sideways movement, as part of the sweeping movement of searching the area. Also, the child may be “looking around” in a way that also contributes to the appearances of searching for something. In cases where there is no sweeping movement, the hold rather appears as a consideration of one specific object rather than a search — “Should I take this object? Is it suitable?”.

Halting an action “in the middle” of its performance, as in the case of REACH [HOLD], produces an effect which has a logic that is common to many sorts of halted

actions, not only the ones in this particular slot of action excursion space. The effect that appears is that of projecting the possibility of performing some sort of action. This is captured in the ethological concept of *intention movement* (Lorenz, 1957 [1939]; Tinbergen, 1951; see also Darwin, 1872; Mead, 1934, pp. 42; Tomasello, 2008, all of which are presumably rooted in Engel's concept of "intentional expressions", cf. Kendon, 2004, p. 87). An intention movement is present in cases where the initiation of an action is sufficient to project the possibility of the continued line of action without actually performing the action (at least not yet).<sup>8</sup>

In cases of REACH [HOLD] where the object is *out of reach*, the action tends to function more like a pointing gesture. It is not unusual for the index finger to be extended during parts of the REACH. Because the object is not reachable by the child the REACH is often treated by parents as an imperative pointing gesture where the object is requested, again due to intention-movement-like projections although in this case it is not possible to actually perform the act since the object is out of reach.<sup>9</sup> In contrast to the appearances of halted reaching towards objects within reach, the imperative function in cases where the object is *not* within reach generally appear as requests for objects rather than requests for the parent to do something with the

---

<sup>8</sup>Tomasello (2008) uses the concept of intention movement in a way that makes a distinction between symbolic gestures (of humans) and intention movements (of apes). I use the concept of intention movements in a more general sense, referring to simpler forms of intention movements found in various animals as well as to more explicitly "symbolic" gestures (i.e., semiotic signs) that rest on initiation of a movement to project an implied continuation. In some cases, this communicative effect may be a mere by-product of a visible hesitation in some movement (on the level of Comm#1 or Comm#2), and in other cases this may be more explicitly exploited as part of a communicatively organized act (on the level of Comm#3). The examples analyzed here, however, are generally part of communicatively organized intention movements in the sense that they are coordinated with co-expressive speech. Müller (2004, p. 237) discusses the same phenomenon in terms of *metonymy*, where "the stretching out of the hand can act as a request for an object because it is perceived as a contingent part of the action of receiving something." The notion of metonymy as such does not require that the part of the projected action that is presented must necessarily be the *first* part of the action, and it is therefore closer to Wundt's notion of *connotative gesture*, Wundt, 1973 [1921], where any aspect of an action that is drawn out from the action as a whole may stand for the action as a whole. Metonymic and connotative gesture are therefore more general notions, and intention movements is a special case.

<sup>9</sup>One of the suggested origins of pointing gestures in the literature is precisely ritualization from such reaching movements (see for example the sharp transitions at 9 months in the development of give-take events in Bruner, 1978b, pp. 77–79). In this regard it is surprising that Tomasello (2008) suggests two strictly separate lines of development for iconic gestures (from ritualized intention-movements in apes) and for deictic gestures (from "attention-getters" in apes such as HAND\_CLAP and GROUND\_SLAP), which seems to rule out the possibility of pointing as emerging (at least in part) from ritualized intention-movements of reaching, at the same time as stating that "no one knows whether pointing is somehow ritualized by infants from some other behavior" (p. 112). Also, it is not clear why the origins of pointing has to be phrased as an either-or question of ritualization or imitation — as far as I can see, the development of pointing may well include both of these aspects.

object (other than handing it over). Imperative appearances are often associated with an about-to-grasp hand-shape, or opening and closing of the palm towards the target (cf. RITUALIZED\_REACHING in Bates et al., 1979 and REACHING in Caselli, 1990, p. 56), or similar variants. In a few cases the hand-shape oscillates between an open palm for (“*give me that one*”) and an index finger pointing hand (“*give me that one*”), constituting some sort of mix, or combination, of imperative and declarative appearances. The type of object reached for also affects the appearances of the action; a cookie may be more likely to be “wanted” (imperative), and a lamp may be “looked at” (declarative).

Reaching movements can also be stopped in the sense of being interrupted and abandoned, either changing into a withdrawal, a reach for something else, or into a preparation phase for an empty-handed gesture. Even though such movements generally do not take on the status of a “proper action”, they are still often salient for co-participants. Therefore they sometimes enter the main track despite being on the verge to the disattend track. Such shifts in movement are treated as indicative of cognitive activity (“changing one’s mind”, “disliking something”, *etc.*), and parents (in the data analyzed here) regularly comment upon them. In the case of changing into a withdrawal, the object generally appears to be considered inappropriate/non-desired/non-interesting by the child and an example of an adult’s comment in this case may be “nope, that’s no good” (in agreement) or “Don’t give up. Try again!”. In the case of changing into a reaching towards another object, it tends to contrast the “old” no longer reached for object with the “new” (presumably more appropriate/desired/interesting) object. An example of a parent comment in such cases is “Yeah, that’s better!”, that in this case includes an explicit *comparative* component which contrasts the non-performed and projected action with the act that is performed instead. The children themselves sometimes coordinate “negative” speech with such actions, such as saying “no” or “don’t fit” at the moment of interruption, perhaps combined with mutual gaze with the parent, which may serve to push the action into the main track. The effect is even stronger if the act constitutes a refusal to perform an action that was suggested to the child by the parent in a previous turn.

## 9.3 Touching or grabbing the object

### 9.3.1 Definition

The moment that the hand makes contact with an object is one of the points in action excursion space that most frequently takes on communicative appearances.

“Contact” may refer to a number of different types of action such as the grabbing involved in picking something up, or grabbing in the sense of seizing a grip of an object without picking it up, or it may refer to touching the object as part of some form of pointing-like action or exploration of the object as the object is now available to tactile experience. In the case of GRAB, the object appears “selected” for further action.

#### 9.3.2 Analysis

In many cases, touching an object or grabbing it functions much like pointing gestures, including the ways in which it is coordinated with speech. This is true both in the case when the movement is halted at the moment of touching (TOUCH [HOLD]) or seizing a grip around an object (GRAB [HOLD]), and in the case where the grabbing is part of a continuous movement of picking something up (GRAB [ONGOING]). Typical examples of utterances that are semantically and temporally coordinated with GRAB and TOUCH make reference to the object being grabbed, either by means of specific open-class words (content-loaded) such as “thumb”, “card”, “cookie” or more generic expressions that fit many kinds of objects such as “this”, “here”, “such”, “mine”, etc. There can also be future-oriented speech that refers to upcoming action when grabbing the object, stating what to do next with the object. An example of this is a child saying “I’ll+REACH paint it+GRAB” at the moment of grabbing/taking a pencil, which is used for drawing a moment later. In cases where speech is prospectively oriented to future action, the grab appears as a preparatory act, thus projecting further action or perhaps even a whole activity.

When the movement of the hand *stops* while being extended towards an object, either at the moment of touching the target (TOUCH [HOLD]), or when seizing a grip around the object (GRAB [HOLD]), there is a strong tendency for the act to enter the main track. The interrupted flow of action tends to capture attention in itself. This is consistent with research stating, more generally, that holding an empty-handed gesture tends to capture the interlocutor’s attention towards the speaker’s gesture, whereas interlocutors otherwise spend most of their time attending to the speaker’s face (Gullberg & Kita, 2009).<sup>10</sup>

---

<sup>10</sup>However, in contrast to the dyadic face-to-face situations studied by Gullberg and Kita, the situations studied here are for the most part side-by-side or L-formed formations (Kendon, 1990, p. 213) of object-oriented triadic interaction. Accordingly, in the situations studied here, participants attend a lot more to the actions of the hands and the objects involved in the activity, and less to the face, compared to the results of Gullberg’s and Kita’s study of a face-to-face situation.

TOUCH [HOLD] is most similar to pointing, sometimes appearing more specifically as POINT\_FLATHAND which is sometimes more or less equivalent with TOUCH (although TOUCH does not refer to a specific hand-shape). Sometimes it may be clear that there is no intention to grab the object, as in the case of pointing gestures, and sometimes it is less clear whether this will develop into a GRAB. For example, the nature of the object itself may make it clear that no grabbing is intended, such as when pointing to a table or to oneself. An example of an utterance in this context is “dolly sit there+POINT\_FLATHAND”, pronounced while pointing to the table with a flat hand, with a tapping stroke, as an explanation to the mother where the child wants the mother to place a doll. GRAB [HOLD] on the other hand, tends to be more action-related. When it is combined with speech, such as “this+GRAB one”, it does not only indicate or point out some object, but makes it appear “selected” in preparation for further action involving this object.

The frozen movement involved in GRAB [HOLD] and TOUCH [HOLD] often appear to “demand” a response from the interlocutor, and indeed, in many cases the hold is sustained precisely until some sort of response is received from a parent (either verbally, or through a requested action, or something else), similar to the organization of pointing stroke endpoints in cases where the stroke is still held after the child’s own utterance has ended (cf. page 189 and Andrén, in press b).

In the analysis of REACH it was mentioned that REACH sometimes gives the impression of making a selection between objects among a set of candidates. TOUCH [HOLD] and GRAB [HOLD] often achieve a similar effect, but in the case of GRAB [HOLD] the emphasis often appears to lie more on whether the child should *proceed* to perform the projected sort of action with the specific object involved, and in the case of TOUCH [HOLD] more on the issue of selecting *what* object (among other candidates) to use for the performance of an action. It is common for children to establish mutual gaze with their parents at this point, as a form of social referencing. Again, these sorts of held actions have elements of intention-movement logic, since they project a potential future line of action (picking the object up to do something with it). More precisely, the held movement makes the act appear as a *choice* between proceeding along the lines of the projected action or abandoning it, turned into an issue of joint decision making. Parents also often respond to such actions in ways that orient to the projection of a possible continuation, such as confirming that some object is indeed suitable for the purposes at hand. An example of an utterance in this context is “this one?+TOUCH [HOLD]”.

A second type of contact with an object, apart from more or less static holds, consists in doing something with an object “on-site”, without picking it up. Two



variants will be discussed here: *haptic exploration* of the object and *adjustments*. Cadoz & Wanderley (2000) refer to “sensory” activities like haptic exploration of objects as the *epistemic* function of the “gestural channel” where the hand functions as a perceptual organ. The epistemic function may co-occur with instrumental and communicative functions of manual action, and with visual inspection of objects. A clear example of this is when a child and her parent read a book where each page contains a new kind of texture or material that the child is supposed to touch and feel with the hand and the child performs the exploration with an index finger pointing hand-shape at the same time as naming the entity being pointed to (and explored), perhaps ending the act by engaging in a mutual gaze with the parent. The act of exploring an object is not only a private affair in terms of “sensory input”, because the explorative quality of such movements is generally visibly recognizable for interlocutors too, no matter if this is intended or not.<sup>11</sup> Sometimes the children coordinate tactile exploratory acts temporally and semantically with speech that relates to the quality of the sensed object, thus making the exploration explicitly communicatively foregrounded. Examples of this are “soft+TOUCH [EXPLORE]” and “yuck+TOUCH [EXPLORE]”. Parents often comment upon, and respond to, such explorative acts, treating them as communicative turns in their own right — especially when the exploration is part of a joint activity and especially when the children perform them together with speech and vocalizations.

Adjustments — another type of on-site handling of objects — consists in actions where an object is not really “used”, with reference to some canonical and typified form of “usage”, but rather “adjusted” or “corrected” to be in better position for the upcoming line of action. Such actions are often characterized by being an apparent lack of consequentiality. A typical example is when a child first puts something down, and a moment later grabs the object again to perform some minor fine-tuning of its position or orientation. In some cases, the adjustment may be explainable with reference to some sort of general geometrical harmony, such as correcting the perceivable unevenness of a row of objects that are only almost well lined up, but for the vast majority of cases, such adjustments exhibit a clear sensitivity to canonical use and handling of objects (Rodríguez & Moro, 2008; Sinha & Rodríguez, 2008; Sinha, 2005, 2009b). A toy figure that ends up slightly to the side of a plate placed

---

<sup>11</sup>Empty-handed gestures sometimes exploit this sort of public visibility of “tactile” sensing, for communicative purposes. For example, conductors sometimes perform manual movements that describe a sensory tactile quality of something when guiding the orchestra (Bräm & Bräm, 2000, 2004). The conductors use “tactile” gestures in metaphorical ways, where *tactile* qualities signify the desired quality of the *sound* produced by the orchestra. However, non-metaphorical uses of such tactile gestures also exist.

on a toy table may be adjusted to end up directly in front of the plate, since this is the canonical order in the context of eating. Adjusting objects in this way amounts to treating such mismatches as “repairables” and the practice of adjusting positions (and other forms of state) of objects bears some resemblance to that of *repair* in speech which may also be self-initiated and other-initiated and so on (cf. Schegloff, 2007). Sometimes adjustments seem to be more locally oriented to mismatches between *expected results* and actual outcomes of an action (even though it may be a situation-specific non-canonical action), rather than mismatches between *typified canonical orders* and outcomes of an action. The “proper” state of an object is often motivated in terms of the canonical next step of an activity. In this regard, Clark’s (2003, p. 261) “accessibility principle” is of relevance here (and also later in the chapter):

*Accessibility principle:* All other things being equal, an object is in better place for the next step of action in a joint activity when it is more accessible for the vision, audition, touch, or manipulation required in the next step.

There is a tendency for such adjusting manipulations to appear as “mere adjustments”, on the operational level, rather than to appear as communicative action gestalts, but there are nonetheless numerous examples of communicatively highlighted ADJUSTS that qualify for the action level. The children commonly coordinate such ADJUSTS with an utterance such as “like that+ADJUST” (*Sw. så*), highlighting the fact that the object is now in “proper” position, which may sometimes turn the act into an “official” turn in the interaction.

A third type of encounter with objects consists in cases where GRAB is part of a more or less continuous movement flow involved in picking an object up (GRAB [ONGOING]). The main reason that this particular moment of grabbing in such movements appears communicatively foregrounded, despite the seamless continuity of the movement, is due to “external” means such as simultaneous speech or vocalizations (or mutual gaze) being temporally coordinated with the GRAB. Coordinated speech and gazes and co-text are important factors in making other types of action communicatively foregrounded too, but speech appears to be an especially important factor in making “continuous” actions communicatively foregrounded. Without speech or mutual gaze, GRAB [ONGOING] appears as mere instrumental and non-communicative grabbing, unless it is performed in some sort of marked way, or as an explicit response to a request from a parent to pick something up or similar contextual embeddings.

Many types of acts can be performed as soon as an object is grabbed, and due to the continuous movement involved in GRAB [ONGOING] it is often subsumed in other actions as a sub-operation rather than an action gestalt in itself. As pointed out before (recall Figure 9.2, p. 277), it may constitute the first part of TRANSPORT (grabbing an object, lifting it, moving it to another location, and putting it down) or PICK\_UP (grabbing an object and lifting it). However, even if the action extends over several subspaces in action excursion space, speech is frequently coordinated with the precise moment of grabbing.

Some actions appear more *generic* than others, in the sense that they can be performed productively with many sorts of objects (GRAB, SHOW, GIVE, MOVE, PUT, PICK\_UP, *etc.*) In this regard they are similar to deictic words and gestures, which can be used for many types of referents, *even without specific knowledge of the sort of object involved*. Other actions appear more *specific*, in the sense that they have a much more particular meaning and a more particular relation with regard to the object involved. Such actions often have a corresponding verb in ordinary language (SIT, CUT, DRINK, *etc.*) and they could be characterized as more content-loaded than generic actions.<sup>12</sup> Even when such acts are performed in purely instrumental ways, they have the character of being experienced as tokens of some sort of *type*. They are not just “actions”, but *typified* actions, made with reference to conventionalized typifications. As such, they require previous experience with the sort of objects they involve, and familiarity with their canonical usage (cf. Chapter 2, 3, and 7).

An example of an action that “objectively” involves grabbing, but which would generally not be experienced in generic terms such as “grabbing” or “moving” something is flipping a page in a book while saying “next+FLIP\_PAGE”. Saying “next” also renders visible the child’s knowledge of pages in a book (as a cultural artifact) as arranged in a serial order that makes typified concepts like “next” and “previous” page relevant. The utterance also transforms the action into something that is made manifest to the interlocutor, i.e., the action and the speech forms a communicative and co-expressive whole. Another example is when a toy is grabbed and pushed aside while saying “clean+GRAB/PUSH”. Through the utterance it becomes visible that the child is not just simply moving the object aside as part of some emotionally driven aversion, but that this action is performed as a token of the culturally defined typification of “cleaning up”. Yet another example, involving TOUCH instead of GRAB, is the comforting “patting” of a toy animal, which is an act typically

---

<sup>12</sup>Generic and specific actions should be considered two poles on a scale.

performed by touching some animate, or pretended to be animate, object. Accordingly, even though such acts are not governed by normative criteria of correctness of *form*, in the way emblems are, they are still supposed to be performed in some particular way in order to *be* such an action — as a token of a canonical type. Either grabbing the object (the step “after” in the excursion) or not touching it (the step “before” in the excursion) would constitute a deviation from canonical “patting”.

An example of a more content-loaded instance of GRAB [HOLD] (working as an intention movement) is when a child performs the typified action of grabbing a (toy) coffee pot without picking it up, establishing mutual gaze with her mother and asking “coffee?+GRAB [HOLD]”. The parent does not mistake this for a question of whether the held object is a coffee pot, or whether it contains coffee, or something similar. When the parent responds (“yes please”) the child progresses through the action sequence, serving her mother coffee, and then offering some sugar, and so on. That is, the action projected by this intention movement is the more specific culturally defined and content-loaded act of “serving coffee”.

Yet other examples include when an object that is supposed to be squeezed is grabbed in a way that achieves a SQUEEZE while saying “squeeze+SQUEEZE”, and when a relatively large doll is “grabbed” by performing HUG on it and lifting it up in this way while vocalizing “mmm+HUG”, or somewhat more indirect variants of “grab” such as bringing a glove into a state of possession by sticking the hand inside while saying “like that”, rather than performing a more generic GRAB where it is grabbed as if it was any sort of object.

Cases where objects are grabbed in a relatively prototypical ways (form-wise) can also exhibit subtle hints of typified knowledge of the object and its canonical uses. In the words of Milner & Goodale (2006, p. 229): “the very way one grasps an object such as a screwdriver or a banana reflects what one intends to do with it”, such as grabbing a screwdriver by its handle (above chance). For this to be possible, “there needs to be top-down modulation of actions by previously acquired knowledge” (ibid.). That which happens to tickle the retina at the moment is not enough to explain the systematic orientation to canonical use of various objects in grasping actions by the children.

Sometimes grabbing somehow appears to be difficult and non-successful, much more so than REACH and TOUCH. Even though not intended as such<sup>13</sup>, problems involved in grabbing often give rise to similar projections of not-yet realized actions as intention movements. Problems of this kind will usually give rise to explicit con-

---

<sup>13</sup>At least if it is assumed that people generally seek to avoid problems.

cern from both child and parent. In many cases, parents immediately express various interpretations of the trouble, helping the children in various ways, or describing the problem verbally. Children also frequently make use of negative emotional vocalizations in temporal coordination with the attempts, even though emotional vocalizations are not that common for children of the ages studied here. The children also very frequently establish mutual gaze with the parents in this context, and in some cases they also ask for help explicitly (often with a negatively loaded emotional voice). Thus, parental responses seem to be sought routinely in cases of trouble and the expressions of negative emotions associated with the trouble.

It is interesting to note that there are several instances in the data where GRAB proves to be troublesome and a co-occurring spoken utterance is restarted each time a new GRAB attempt is performed. Although such restarts of an action in coordination with a corresponding restart of the spoken utterance do not always occur (the spoken utterance is not always restarted), the fact that it happens in a number of cases is a strong indication that there may indeed be some form of coordination going on between speech and action that involves objects which is not a mere coincidence. In a few particularly noteworthy cases, there are even up to 4–5 restarts of the GRAB before it succeeds, each with a corresponding restart of the same utterance.

One study has shown that children aged 8 months still miss the target altogether in about 11% of the attempts to grab an object, which is still a large improvement from 4 months, where they miss about one third of the attempts (Blake et al., 1994). At the ages studied here (18–30 months) problems are less common, but it is still interesting to note that grabbing is apparently not a trivial issue, despite its very mundane character. Grabbing takes time to learn and master.

As the children grow older (18 < 30 months), GRAB and some of the other acts analyzed here are increasingly often (although still not very frequently) coordinated temporally, but *not* semantically, with speech. Early examples of such “dissociation” commonly include routine production of response morphemes (“yea”, “no”, “m^hm”, “mm” etc) in response to parent utterances, in *temporal* coordination with a GRAB, but where the GRAB bears no relation to the current topic of the talk.<sup>14</sup> From about 24 months (i.e., TP#2) and onwards the children become markedly more capable of routine communication about matters beyond the here-and-now

---

<sup>14</sup>Note that response morphemes are relatively weak in terms of initiative (Linell & Gustavsson, 1987; Linell et al., 1988) — they introduce relatively little content — and in light of this it perhaps not surprising that such explicit dissociation between simultaneous action and speech appears first with these kinds of utterances.

of the current situation. Around this age, the children in the data also appear to be able to act simultaneously in two tracks at once, that is, the “main track” (Goffman, 1974) of the joint activity is split into what might be called one *activity track* and one *communication track*. For example, parents and children may play with toy building blocks in a jointly coordinated way, giving blocks back and forth and jointly arranging the blocks in a particular pattern, while speaking of something completely unrelated, such as what they ate at a party last week. The activity track is different from the tracks discussed by Goffman (1974), such as the disattend track (things that should not be paid attention to), the directional track (things that serve to regulate the communication), and the — for the present purposes — more marginally relevant overlay tracks and concealment tracks. For example, it is not considered strange to insert a comment on the activity track in the middle of turns related to the communicative track, nor does it cause disruption or surprise (cf. Kendon, 1992, p. 327). The phenomenon of split tracks is more like a splitting of the main track into two parallel tracks.

## 9.4 Handling the object in center space

### 9.4.1 Definition

This part of action excursion space consists in movement in or through central space, or any other intersectional area between the locations of picking an object up and putting it down. At this point, the moving hand no longer appear to be “directed towards” the object, the way it is for REACH, TOUCH and GRAB. Instead, the object is now held in the hand and the movements are no longer a question of object indication or selection.

McNeill (1992, pp. 86) uses the term “center space” to refer to the part of *gesture space* which is right in front of the body of the speaker, but below the head, and as McNeill (ibid.) writes, “The gesture space can be visualized as a shallow disk in front of the speaker, the bottom half flattened when the speaker is seated”. The way McNeill characterizes gesture space, it appears to be a kind of detached space that is defined only with reference to the speakers body, and not to the material surroundings. I refer to this sort of gesture space as *abstract space*. Some gestures do indeed seem to take place in such a space, but there are also all those gestures that are somehow directed to the material surroundings (cf. Chapter 6 and Chapter 7). Therefore, I suggested in Chapter 2 that the complementary notions of action space (relating to the present environment of the speaker in a concrete way) and support-

ing space (relating to the present environment of the speaker, but in an indirect way), are useful complements for the description of the kind of space that some gestures inhabit. Hence, when the term “center space” is used in the present section, it rather refers to a sector of action space and not to a detached kind of space.

### 9.4.2 Analysis

The hand is moving in or through center space, or any other intersectional area between the locations of picking up and putting down an object. In cases when utterances are coordinated with *ongoing* non-interrupted movement in or through central space — such as in the middle of a TRANSPORT action, which involves grabbing and object, lifting it, moving it, and putting it down — the utterance tends to be oriented towards some act that is about to be performed, and to which the current movement is preparatory. This is similar to some of the future oriented utterances found in coordination with REACH. However, on the whole, coordination of an utterance in the middle of TRANSPORT is rare. When speech comes at these “in the middle of transportation” movements, it generally does *not* appear “coordinated” with the movement — at least not strongly so. Recall that speech that was coordinated with REACH movements was for the most part coordinated with the onset of the REACH, but in the case of CENTER\_MOVE [ONGOING] the “onset” is rather the moment when an object is grabbed, which is outside the present slot in action excursion space. Indeed, in the vast majority of the cases where peaks in the intonation contours of the co-occurring speech are coordinated with ongoing movement through center space it tends to be coordinated with the moments in the transportation maneuver where the object is grabbed and/or put down. In cases where both grabbing and putting down are coordinated with speech the overall act also tends to appear as a *sequence* of two more or less distinct units/“parts” of GRAB/PICK\_UP and PUT (see page 305) rather than as a single whole action gestalt (TRANSPORT). Movement dynamics are also involved in the contrasting appearances between either the act as a whole or as consisting in a series of two action gestalts. A fluent and quickly performed transportation tends to appear as a whole (TRANSPORT), and a more hesitating or “step-wise” performance may rather give rise to the appearances of a sequence of TAKE and PUT. TRANSPORT is not the only type of action that fits into the category of CENTER\_MOVE [ONGOING] though.

Another type of maneuvers performed in this slot of action excursion space are those that involve either changing hands (CHANGE\_HAND), and/or changing the type of grip of the object (CHANGE\_GRIP). They may also be performed in a rela-

tively fluent and ongoing way, but since “something happens” along the way, they create something like a node in the experiential structure of the action to which an utterance can be attached or coordinated. A typical example of communicative highlighting of such acts is when the act is highlighted through a running commentary such as “take with the other+CHANGE\_HAND hand then”. CHANGE\_GRIP may also come with utterances that comment on the more specific act of changing the way the object is held in some way, such as “I turn it+CHANGE\_GRIP”. In most cases though, the utterances combined with CHANGE\_HAND and CHANGE\_GRIP are more generic, such as “that”, “there”, or “like-that”. These maneuvers are sometimes similar to what was described as ADJUST in the analysis of GRAB/TOUCH.

In cases where the hand *stops* moving in center space, or in any other intersectional area between the locations of picking up and putting down an object, the stopped movement often appear to mark the *end of an action* rather than a hold of an action in progress. Cases in point are endpoints of PICK\_UP or REMOVE.<sup>15</sup> There is sometimes speech associated with such endpoints of action. In the case of PICK\_UP, the relevance of the action often lies in the achieved possession of the object, and indeed, sometimes possession/ownership (“mine+PICK\_UP [END]”) is precisely what is at issue in the interaction when PICK\_UP is performed. For the most part the speech coordinated with PICK\_UP is similar to what is typically co-occurring with pointing and other deictic gestures (“that one” etc.). In the case of REMOVE the relevance lies in the achieved removal of the object from its previous location and the current possession of the object may appear more like an incidental by-product of the action. Whenever an action appears as an instance of REMOVE, the object is often actually moved sideways from center space to peripheral space, rather than into or through center space. In the data studied here, children’s use of negation words (“no”, “not”) between 18–24 months (approximately) frequently occur in the specific context of coordination with some form of “undoing” of the effects of some act performed by the parent the moment before. REMOVE cover a large share of those cases, but it can also be other forms of “undoing” actions.

Yet another type of actions that fits into this slot of action excursion space, and which also contain stopped movement is SHOW\_UNCOVER, which consists in actions that serve to show a  $T(\textit{arget})$  to someone by removing something that covers  $T$ . In such cases, the object held in the hand is uninteresting and the gaze is often

---

<sup>15</sup>Note that the very fact that the movement stops in central space contributes to the appearances of PICK\_UP, rather than something else, having been performed. In the absence of a stopped movement, the action gestalt appearing may well be one of an ongoing TRANSPORT instead. This was what was shown in Figure 9.2 on page 277.



strictly focused on *T* instead. As an example, the act may be combined with a deictic word like “there+SHOW\_UNCOVER”, uttered precisely at the moment when the movement stops, and the “cover” is removed. It bears some resemblance to the game of PEEKABOO, but is also different from it, since SHOW\_UNCOVER is a matter of triadic interaction around objects rather than the dyadic engagement of PEEKABOO. It also requires the child to master *object constancy* in a way which makes it different from the case of PEEKABOO. Similar to the game of PEEKABOO, SHOW\_UNCOVER often produces a dramatic effect.

In other cases CENTER\_MOVE [HOLD] appears as a movement that is *stopped in the middle of its performance*, rather than at the end of an action. This is similar to the intention movements discussed previously, which also appeared precisely when some sort of movement was held “in the middle” of its performance. An example of this is when the movement stops briefly, while the child appears to be searching for a suitable location for a PUT action while vocalizing “*e::b*”.

In yet other cases CENTER\_MOVE [HOLD] appears as the performance of a communicative action in itself, neither being interrupted “in the middle” nor appearing to be a mere “hold” after “the stroke” of an action that would appear to constitute a hold at the end of an action. Examples of such acts are SHOW and INSPECT, both of which appear to be “ongoing” rather than “paused” despite the hand being held still.<sup>16</sup> This is not surprising, since both of these acts are organized around visual inspection (by the self or by the other, respectively), and for this reason they have to be held relatively still to be “inspectable”. As previously noted, holding a communicative movement also captures attention in itself (see page 283).

SHOW is typically used together with utterances of the kind seen together with pointing: “this one”, “such”, “pencil” (naming the object), “look this”, and so forth. A slightly more complex example is “use this+SHOW+HEAD\_SHAKE”, which also illustrates how head-shakes can be used to modulate the meaning of speech+gesture (Kendon, 2002; Harrison, 2009a). It should be noted that SHOW is somewhat fleeting (but not arbitrary) when it comes to its place of articulation. In some cases it is articulated by holding an object upward. In other cases, however, SHOW is rather performed by holding an object towards a recipient in some way.<sup>17</sup>

---

<sup>16</sup>Note that this is different from gesture “holds” that serve to make sure that the gesture aligns well with some word in the spoken utterance (cf. Kendon, 2004, p. 127). In the case of SHOW and INSPECT, the “hold” is rather part of the performance of the act itself, even though the duration of the “hold” may be varied, and is varied, depending among other things on the response received from a parent. In many cases they are held until a response is received.

<sup>17</sup>See page 298 for an analysis of SHOW in the case of holding objects towards someone.

Appearances of showing can also be achieved by means of a kind of “extra grasp” of the objects already held in the hand, as if to invoke the indicative and selective powers of GRAB although the object has already been picked up. A similar example can be found in Goodwin (2007, pp. 196), where a man says “she sold me this+CENTER\_MOVE [GRAB]” while noticeably “grasping” an object already held in the hands. In some cases SHOW involve bringing an object forth from some location where it was hidden into an area where it is being visible (often precisely center space). In such cases there are some similarities between SHOW\_UNCOVER (where the removed covering object held in the hand is to be disattended, cf. Goffman, 1974, pp. 202) and SHOW (where the object held in the hand is the object to be attended to) in that a dramatic effect may be achieved and in that the shift from hidden to visibility may become part of the more specific meaning of the communicative act.

Now, consider INSPECT instead. The inspection of an object held in central space, can be everything from a brief glance towards the object to more extensive visual exploration involved in inspecting the object while turning it around to view it from various angles. Along these lines, inspect may involve both a (perhaps very brief) static hold of the object (as in the case of a quick glance) or active manipulation of the object (to view it from various angles) which may also involve action gestalts such as CHANGE\_GRIP and CHANGE\_HAND, although the location of the object is often relatively fixed during inspection. INSPECT can also be fused with haptic exploration (EXPLORE), as analyzed in the previous section.

Verbal utterances used with INSPECT often categorize or describe the object in some way: Telling what type of object it is, whose object it is, what color the object has, what can be done with it or what to do next with it. In cases of categorizing or describing the object, adjectives and other descriptive terms appear more often in temporal coordination with INSPECT than it does with almost all of the other acts analyzed here, apart from EXPLORE. Some examples include; “it’s red+INSPECT”, “there’s a little left+INSPECT” (looking inside a container held in the hand), “yummy+INSPECT” (while looking at a cookie held in the hands, before eating it), “I’ve got something+INSPECT to eat” (while holding a citrus fruit in the hands). In some cases, the act of contemplation or inspection is itself made manifest through vocalizations or speech. Examples involve making manifest the “covert” act of thinking by saying “uuehm+INSPECT” (when searching for the word for the object held in the hand), or the act of looking by saying “look one-such+INSPECT”. The latter example also illustrates how INSPECT sometimes becomes fused with SHOW, that is, it is often a matter of joint inspection, rather than inspection only

on behalf of the child.

In cases where INSPECT and SHOW are performed, the holding of an object often serve as a ground for other communicative acts, such as pointing to the object as a whole, or perhaps some particular part of the object, while commenting on it. In either case, it should be remembered that INSPECT, just like all other acts analyzed here refers to cases where INSPECT is done in a way that is made communicatively manifest. It does not refer to inspection in a private sense. At the same time, INSPECT is indeed one of the action gestalts analyzed here that often is somewhere in the borderland between the main track and the disattend track. In some cases it is more like a non-salient background to the utterances without apparent temporal coordination between the act and the speech apart from going on at roughly the same time, in other cases the act of inspection is actively made visible to a recipient. Part of the explanation may lie in the tendency to look at the object in cases of INSPECT than at a recipient, but of course, there can also be shifts back and forth between looking to the object and looking to the other.

Returning now to the distinction between generic (often “deictic”) and specific content-loaded (often “iconic”) appearances of action. In the case of specific and content-loaded actions that takes place in center space, or some other intersectional area between picking up and putting down, they are often oriented to some specific part of an object, such as pressing a button on a toy telephone held in the hand, as a token of the culturally established type “calling a number”. Some instances of CHANGE\_GRIP, that involve changing to a grip which is more appropriate for the canonical use of the object while saying things like “*like that*+CHANGE\_GRIP”, also orient to, and make manifest, content-loaded knowledge of the particular object involved.

## 9.5 Moving the object towards a target

### 9.5.1 Definition

In this slot of action excursion space, the object is still held in the hand, but in contrast to the actions gestalts analyzed in the previous section, the hand now moves the object *towards* some target location or object, or towards another social actor — a recipient. This has the effect of creating a vector of directedness, similar to that of pointing gestures, and similar to the actions that fall in the REACH and TOUCH/GRAB slots of action excursion space. Whereas REACH, TOUCH, and

GRAB consisted in directedness towards one object the actions described in this section often have the effect of *relating* the object held in the hand *with another object or location*.

Acts can be directed (in space) towards “targets” in two main ways: Either the act is directed towards a communicative *recipient*, or the act is directed towards an entity which the communication is *about*. As the analysis will show, both forms of spatial directedness occur in the data. In some special cases, these two forms of directedness collapse into one and the same “target”, such as when pointing to someone while saying “you+POINT”, but generally they are separate. Spatial direction towards a recipient is a special case of the more general fact that utterances are generally addressed to someone. Addressing an utterance to others can be done in many ways though, and many of these do not involve spatial direction. Spatial direction is a distinct feature of communication that involves bodily movement. Communicative acts — either performed empty-handed or with an object held in the hand — that are directed towards a target object or location, or a recipient, are also discussed and analyzed in chapter 6.

### 9.5.2 Analysis

The first class of actions to be analyzed are those when the hand stops moving on its way towards a target location or object, or towards a recipient (MOVE\_TOWARDS [HOLD]). One type of acts that appear here are, again, those that have intention movement character, appearing to be interrupted in the middle of the action, before being fully executed. For the most part they work similar to what has been described earlier, where a stopped movement is recognizable as a “not yet fulfilled” action, thus projecting the possibility of performing precisely this action. One example is when a child holds the lid of a coffee pot towards the coffee pot while saying “there+MOVE\_TOWARDS [HOLD]”, without releasing the lid (the lid fits, so that is not a problem), but she apparently changes her mind and removes the lid again while saying “no+REMOVE”, again illustrating how intention movements often indicate where alternative lines of continuations can be chosen.

Generally, objects appear as “moved towards” something until they are actually released and as long the object is still held, the action appears unfinished despite the fact that the object has reached a (potential) destination. An apparently unique property of the action gestalts that appear in this slot of action excursion space, at least in the data studied here, is that some of them *require* complementary collab-

oration from an interlocutor in order for the action to be completed.<sup>18</sup> The most common example of this is probably OFFER, which requires the parent to take the object being offered in order to complete the action as an instance of GIVE. This is a joint action by design. Most research on child gesture labels events of giving as GIVE, rather than OFFER, even though it is the subsequent response of the parent that makes the difference between events in which an object was given to a parent and those where it wasn't. Thus, GIVE is something like an after-the-fact label of a communicative event that involves both the "speaker" and the recipient and I would argue that OFFER is a better label. Blake (2000, p. 80) does distinguish between OFFER and GIVE, but she defines the difference between them in terms of the end result — the criterion being whether the parent takes the object or not. It is generally inappropriate to categorize communicative acts in terms of their appearances as "products" instead of in terms of the communicative properties they have while they are ongoing performances that are doing communicative work (Schutz, 1945, pp. 542; Garfinkel, 1952, p. 364). The fact that parents sometimes refuse to take an object offered to them does not take away the original act of OFFER performed by the child, even though it transforms them into a product — a "refused offer". Of course, OFFER is generally performed in order to bring about an event of giving, and this is how it appears to interlocutors — otherwise they would not reach out to take the objects offered to them — but that is another matter. There are instances in the data where the children coordinate speech (such as "there") with the actual moment of achieving a GIVE, that is, when the *parent* grabs the object offered by the child, thus highlighting the GIVE event as an accomplishment — a product.

Other examples, which may require collaboration from the parent in specific instances, are when a child leans forwards towards a parent, holding a hair brush in the hand, and the parent has to lean forwards for the child to be able to reach the hair of the parent to brush the hair, or when a child indicates that she wants to feed the mother ("as if") with an empty bottle by holding a bottle towards the mouth of the mother, and the mother has to lean down towards the child in order for the child to be able to actually reach the mouth. In such cases, the acts are produced through joint efforts that are only possible through shared orientation to the same, or similar, typified orders.

OFFER often appears as a request for the parent to do something specific with the

---

<sup>18</sup>Parents regularly assist their children in the performance of all kinds of actions (Adamson & Bakeman, 1985; Adamson et al., 1990; Zukow-Goldring, 1998, 2006), but these types of actions do not *require* collaboration from others in order to become recognizable instances of their types.

object, not just to take it, typically as part of asking for help with something.<sup>19</sup> This is for the most part recognizable due to the previous turns of the interaction (both speech and action) and the spoken utterance (or vocalizations with negative affect). Other times OFFER rather appears to be a form of practical disengagement, without any other reason for giving the object to the parent than simply getting rid of the object. A typical characteristic of such acts is that the child attends to something else than the parent, giving the impression that something other than the act of giving is the current focus of the activity for the child.

As the children grow older, they may also bring the typification of OFFER into more abstract contexts. There are a few examples (at 26–28 months) where children play with dolls and perform both the OFFER towards a doll with one hand *and* the part usually played by an interlocutor — taking the object that is offered — by putting the object in the hands of the doll, which is controlled with the other hand, thus enacting the whole interactive event of GIVE themselves. Another type of action gestalt that appears here is SHOW, which has already been discussed in relation to the slot of action excursion space called “handling objects in center space” (see page 292). Different from the types of SHOW discussed before, the SHOW actions that appear here are of the more prototypical kind where an object is held *towards a recipient* for inspection rather than upwards in upper space. However, sometimes when an object is shown by holding it upwards, this can often be interpreted as holding the object at the level of the adult interlocutor’s eyes, and thus these acts too are at least in some cases “held towards”, rather than being cases of non-directed handling in center space. In any case, the directedness of the movement is in this case not directed to what the communicative act is *about*. The “indicated” object is rather that which is held in the hand.

Sometimes the way SHOW is performed makes a particular *part* of an object relevant, rather than showing the object as a whole. An example of this is when a child holds a plate in front of her, remarking on a small spot on the plate by saying “a spot+INSPECT there” and then while saying “there+SHOW” she flips the plate so

---

<sup>19</sup>Thus, while the child “gives” the object, it may still serve as a request. This is relevant to research on the (empty-handed) Palm Up Open Hand gesture (Müller, 2004, pp. 236) where it has (plausibly) been proposed that the PUOH gesture stems from practical actions of giving and receiving. Müller proposes that, among other things, giving is extended to signify proposals, and receiving is extended to signify requests. However, this all focuses on the object. In some of examples analyzed here, the act of offering and giving an object functions as requests to the adult, not for the object, obviously, but for an action. I think this specific function (offering something as a request for action) fits well with examples such as using the PUOH gesture to hand over the turn to someone (i.e., OFFER), while saying “...or what do *you think*+PUOH?” in order to *request* a response from the interlocutor.

the mother can see the spot too. Due to this flipping, it becomes clear to the parent that the relevant side of the object is the one flipped towards the parent (and the spoken utterance makes it clear that it is rather a spot on the plate than the plate itself that is relevant). In this case the change in orientation of the object is enough to change the appearances of the object as first being “not directed towards the mother” to “being directed towards the mother”, even though the object is held out in front of the child’s body in roughly the same way during both *INSPECT* and *SHOW*. The different appearances in these two cases is also due to difference in gaze behavior: in the case of *INSPECT* the child looked at the object and in the case of *SHOW* she also looked at the mother’s face. Another strategy exploited by children to indicate that a particular part of an object held is relevant is to hold the object towards the recipient while pointing to some part of the object with the other hand, using both hands to combine *SHOW* with *POINT*.

*SHOW* is highly similar in form to *OFFER* and one may ask what it is that makes an act of *SHOW* recognizably distinct from *OFFER* and vice versa. In both cases an object is held towards a recipient (disregarding the *SHOW* gestures performed by holding the object upwards for the moment). In fact, in some cases the distinction does not seem to be well specified for the participants either, and sometimes it appears to be a relatively open matter whether the parent will decide to take the object (as if it was *OFFER*) or comment on it (as if it was *SHOW*). Thus, this demonstrates a case where there is vagueness in the “categoriality” of the phenomenon itself — a less specified definiteness of intention (Næss, 1953; Hirsch, 1996, 2009). It may be noted that most gesture researchers proceed methodologically by treating every gesture as being of one type or another, taking the definiteness of intention, or the categoriality, of the communicative acts for granted — at least on a methodological level. The typical solution, when approaching gestures in this way is to apply inter-rater coding measures and procedures to ensure that most of the coding is “right”. This stands in contrast to allowing for the possibility that vagueness and ambiguity may sometimes be part of the phenomenon studied, rather than a methodological “problem” to be “remedied” or rationalized away by the researcher (cf. Garfinkel & Sacks, 1970, pp. 339). As shown by Wootton (1994) acts related to object transfer between persons are also subject to mutual processes of negotiation and repair — the same kind of partial open-endedness that characterizes other types of communicative acts.

Nevertheless, in many, or even most, cases the intent *is* relatively clear. First of all, the coordinated utterance often makes it clear whether the parent is supposed to take the object (*OFFER*) or whether the object is displayed for inspection (*SHOW*).

Also, SHOW tends to be performed toward the face of the adult, or otherwise placing the object in the line of the adult's vision, whereas OFFER tends to be performed towards the hands of the adult (or towards the mouth in cases of offering food). Most of the time, OFFER is also performed closer to the body of the recipient than SHOW (seemingly orienting to the *accessibility principle*). Blake (2000, p. 80) notes that SHOW is often being performed with bent elbow. This is not a *necessary* criterion of SHOW but still rather typical. In some cases the vagueness mentioned above results in misunderstandings, and there are a few examples where the parent interprets an object held towards the parent as OFFER, but the child does not want to give it away, or inversely, situations where a parent interprets an object held towards them as SHOW by commenting on it, but the child "upgrades" the performance and makes it clear that the object is supposed to be taken. Crucially, in the cases where such misunderstandings occur, the object is often held somewhere between the eyes and the hands of the adult, which provides for ambiguous appearances.<sup>20</sup> SHOW and OFFER are generally coordinated with the type of words associated with other deictic gestures such as POINT. In cases where OFFER is performed in response to a request for an object on behalf of a parent ("would you give me that one please?") it is less inclined to occur with speech. In some cases it is so evident that the child wants the parent to take something that the offer almost disappear and end up outside the main track, a little like the previously mentioned separation between an activity track and a communication track (see page 290).

Another major class of actions that occur in this particular slot in action excursion space are OBJECT\_POINT actions, which consists in holding an object towards a target object or location rather than towards a recipient. OBJECT\_POINT was also analyzed in Section 6.7 (page 205), where it was mentioned that it can be performed both with related and unrelated objects. In some cases the object held in the hand, used for pointing purposes, appear to be incidentally held in the hand and *not* related to the target, such as pointing towards the bathroom using a pencil. When OBJECT\_POINT is performed with an object with some sort of relation to the target object, this "relatedness" generally consists of a canonical and typified relation which existed before the particular event in which the pointing gesture was produced. Such pointing can actualize all sorts of relations. For example, when holding a Memory<sup>21</sup> card towards the other card in the pair of two identical cards the act will actualize their similarity, which has a special canonical and typified relevance in the particular context of playing Memory — the aim of the game is precisely to

---

<sup>20</sup>See Steffensen et al. (2010) for a related type of analysis of an ambiguously held object.

<sup>21</sup>This game is also known as "Concentration" in English.



find all the pairs of identical cards. An act of holding a toy figure towards a toy chair actualizes typified schemes of “sitting”. Pointing with a piece of a puzzle to a location in the puzzle as part of a suggestion (expressed in speech) that the piece may fit there actualizes schemes associated with doing puzzles. That is, the canonical schemes associated with the objects involved is an important part in establishing the specific communicative effect that appears and in order to see whether an object is related to the target or not, one has to be a competent member, armed with cultural knowledge. However, such pointing gestures do not only make reference to pre-existing knowledge. To some extent, holding an object towards another object may of course also take part in the *creation* of a relation between the two objects and sometimes the effect of the act is more locally related to some particular situation than a canonical type. Sometimes the OBJECT\_POINT functions like an intention movement, as if the child was asking “Should I proceed to do X to the target object with the object held in the hand?”, often involving gaze towards the parent for purposes of social referencing. It also appears in contexts of asking for support with the action projected by the objects involved. The example with the toy figure and the chair is a case in point. The child holds a small toy figure towards a small toy chair and alternates gaze between the toy chair and the parent while saying “can you help this-one sit there?+OBJECT\_POINT”.

OBJECT\_POINT is more commonly performed without touching the target than index finger pointing gestures are. This is presumably because the hand-shape of index finger pointing gestures often makes it clear that the act is one of indication, rather than some sort of use or handling of the object, despite touching the target. For OBJECT\_POINT, holding the object at a small distance from the target may be more important in order to make it clear that the act is a communicative one and not an attempt to perform an instrumental act.

Sometimes, objects are moved in ways that negate a relation between the object held in the hand and some target or recipient. An example of that is when a parent asks “can I have that?” (asking for the object held by the child), and the child responds by holding the object away from the parent (a form of REMOVE). In such cases, acting according to the inverse of the accessibility principle makes a negation visible and in some instances such actions are also combined with a HEAD\_SHAKE and a negative word such as “no”. There are also some instances where a pointing gesture such as index finger pointing is performed while an object is incidentally held in the hand, and in such cases too the object appears unrelated to the target pointed to.

Apart from content-loaded variants of OBJECT\_POINT (with canonically re-

lated objects), a lot of other content-loaded acts figure in this slot of action excursion space. These acts differ from OBJECT\_POINT in that they perform a more specific type of action towards the object. In fact, this is the slot where most of the content-loaded acts are found in the data studied here and content-loaded acts that are *not* oriented to the physical surroundings are actually relatively rare. One reason for this is found in the ways that objects differ in how they are used. Some artifacts are used by doing something with the object itself, such as pushing a button on the artifact to make a sound — which can be performed in center space without moving the object towards something. Other artifacts are used by directing the object towards something, such as putting a toy phone towards the ear. Yet other artifacts involve both manipulation and directedness, such as pressing the trigger of a toy gun while directing it towards someone. In the data studied, most artifacts involved are of the kind that only requires directedness (sometimes in combination with some specific type of movement) in order to be used in canonical recognizable types of ways.

Some actions involve *touching* the target (such as combing the hair of a doll using a comb), and others do not (such as pouring pretended milk into a toy coffee pot, from a slight distance above the coffee cup). Thus, in cases of content-loaded performance, the issue of touching or not touching the target to which the action is directed is often determined by the typified act which is instantiated or signified. In cases of “as if” performances, the “as if” aspect of the act is sometimes achieved by means of not touching the target, even though the “real” act would have included touching.<sup>22</sup> The general rule of thumb is, however, that children touch objects if they can (if appropriate), as part of their communicative acts. This also applies to empty-handed index finger pointing, where the children in the data studied here touch the target 90% of the time (and where the non-touching is sometimes motivated by the current frame of meaning, such as pretending that the target of the pointing gesture is a hot frying pan).

Some actions, such as phoning, are performed by *holding* an object (such as a toy phone) statically towards something (the ear), without moving the hand during the stroke. Other actions, such as walking with a toy figure against the floor or a table, are performed as *ongoing movements*. This means that whether the object is held still or not, during a stroke, is again largely specified by the typified action which is instantiated or signified. This is particularly true for the ACTION\_BASED

---

<sup>22</sup>Bouissac (ms. in prep.) notes that we “always know if a gesture involves a contact”, and even if the particular word “always” may be in need of further qualification, this at least points toward the importance of including the haptic sensory channel in the understanding of gesture. Streeck (2009b, 2002a) also emphasize the haptic aspect in gestures.

communicative acts discussed here because the signifier of ACTION\_BASED communicative acts remains congruent with the temporal logic and movement dynamics of the signified and that is different from some other types of iconic gestures.<sup>23</sup> In both of these two types of strokes, the stroke appear to be “in play”. Such “holds” are distinct from what is sometimes also called gesture holds, which is rather concerned with keeping a gesture held for a more or less extended moment in order to coordinate the performance temporally with speech (Kendon, 2004). However, it is also true that in many cases such communicative performances of typified actions actually change their temporal characteristics into more snappy performances (see also Andrén, in press c). Returning to the example mentioned above of pouring (pretended) milk into a coffee cup, the temporally halted movement that would be required to actually pour some milk in the coffee pot is rationalized away. The result is that the act becomes more clearly temporally organized with respect to the spoken utterance, just like Kendon describes.

Content-loaded communicative acts that involve objects may also be directed towards recipients. An example would be when a child holds a toy figure directed towards a recipient such as the parent, while moving the arm of the toy figure up and down to make it perform a WAVE\_HELLO gesture. That is, the act which is originally a dyadic communicative act is now transformed into a triadic form of interaction that involve objects.

Finally, an interesting class of action gestalts that fit into this slot of action excursion spaces deserve a special mention are actions that involve drawing (see also Section 6.4). As is well known, one type of iconic gestures are those that consist in drawing a shape of some sort (Kendon, 2004; Streeck, 2008a; Müller, 1998b; Wundt, 1973 [1921]) and it is unclear whether such gestures emerge ontogenetically out of actual drawing or not.<sup>24</sup> In either case, what *is* clear from the data studied here is that actual drawing is in many cases coordinated temporally and semantically with speech, just like empty-handed gesture are. Such actions are not only directed towards a target surface, but they also leave a mark on this surface (cf. Goodwin, 2003; Streeck & Kallmeyer, 2001). Interestingly, in the case of drawing, the spatial

---

<sup>23</sup>See chapter 7 for a discussion of other types of “iconicity” than action-based iconicity.

<sup>24</sup>Interestingly, my own daughter started producing this sort of gestures (with an index finger hand-shape) during the same weeks in her life when she also started to draw recognizable shapes in a markedly more creative way (about three years old), i.e., not only reproducing shapes that she had learned to draw as a result from drawing together with an adult. Furthermore, at one occasion she also referred to the gesture she had just produced by saying “but it was smaller than I drew it”, that is, by spontaneously using the verb “draw” to refer to the outlining gesture she had been performing, showing her own understanding of the mode of representation involved in the gesture she had produced.

and temporal domain of the signified is potentially incongruent with the spatial framework in which the act of drawing is performed. A child may well draw an object whose size and orientation bear no relation to the here-and-now context of the child — the domain of meaning only make sense “within the paper” — and the temporal order of the drawing movements may bear no relationship to some actual temporal order from the domain of the signified.<sup>25</sup> Furthermore, the manual action involved in the act of drawing does not correspond to what is signified. The hand-shape, for example, is instead motivated by holding a pencil (generally with an index finger extended in the case of empty-handed gestural DRAW). Granted, the act of drawing could itself be considered a form of action, but in the meaning of the act is not (in the examples discussed here) “drawing” or some object related to drawing such as “pen”, but rather what is drawn. Almost anything could be drawn, without being related to the body that produces the drawing in any way. Thus, it is a productive *mode of representation* (to use Müller’s term, see also Müller, 1998b, 2004; Kendon, 2004; Streeck, 2008a, 2009b for similar ideas) in the sense that the same overall technique can be used to represent a large number of referents. This conventional, but still open-ended, gestural technique also stands as a correction to the widespread idea that conventional gestures are only those that map a highly determined form to a certain meaning (emblems, Ekman & Friesen, 1969). In this case, it is rather the mode of representation as such that is conventional, than the exact performance and/or referent in each actual case. Thus, it is more like a conventional technique of representation than an “action” of the kind that figures as the signified in action-based gestures. For example, a child may draw a round shape in temporal and semantic coordination with the word “balloon+DRAW” (referring to a hot air balloon). Younger children (closer to 18 months than 30 months) also coordinate DRAW with speech, although in their case the motives drawn are often more concrete and the speech that goes with the the act is often deictic (“there+DRAW there+DRAW there+DRAW” when drawing a series of dots) or referring to what is drawn in a more concrete way such as saying “a line here+DRAW” when drawing a line on the paper, which is not incongruent with the spatial framework of the child when producing the act. A line on a paper *is*, in one sense, a line (as a token of a

---

<sup>25</sup>Sonesson (2008, p. 67) also notes this, stating that “It is important to note the difference between pictographs and kinetographs: in the first case, the expression, which is a sequence of movements, is iconic for the content, which corresponds to the limits, or some other static property, of the object rendered; in the second case, however, both expression and content are temporal sequences, and may thus possibly be mapped onto each other iconically in every detail. In fact, there is of course every intermediate case, from the direct quotation of a non-gestural, or even gestural, movement, to some rather schematic correspondence, as the transposition of the movement of the legs to the fingers.” According to this terminology, DRAW is to be understood as a ‘pictograph’.

type), and not just a “depiction” of it.

## 9.6 Putting the object down

### 9.6.1 Definition

In this slot of action excursion space the object is no longer held towards a target object or at a target, but the hand has also released the grip of the object previously held in the hand. This results in the performance of PUT. Together with grabbing and moving an object towards, putting something down is among the most common subspaces of action excursion space to be communicatively highlighted. Clark (2003, p. 243) has also made this point, even claiming that “pointing is often thought to be the only, or prototypical, way to anchor communication, but it is neither” and that such anchoring can also be achieved by “by *placing* things in just the right manner”. He makes a distinction between (A) acts that consist in *directing-to* — directing attention to a target by means of a directional vector, and the prototypical example of this is pointing — and (B) acts that consist in *placing-for* — directing the target to attention, and the prototypical example of this is PUT (what Clark calls “placing”).

The main benefit of this analysis is that it puts the spotlight at the fact that there is a whole range of actions that may serve to ground utterances in communication — also those that involve handling of objects. The distinction between *directing-to* and *placing-for* captures some important aspects of the ways in which such anchoring may work. However, many of the actions analyzed here do not fit well into the binary distinction between *directing-to* and *placing-for*, as presented by Clark. A great deal of the actions apart from the prototypical examples of pointing and PUT (“placing”) are rather characterized by a mixture of the properties that Clark either ascribes to *directing-to* or to *placing-for*.

### 9.6.2 Analysis

Putting objects down may be done in a variety of ways and it may have a variety of communicative functions. As argued by Clark (2003), putting often functions more or less like a deictic gesture. Prototypically, it serves to present the object which is put down (“this+PUT one”) or to highlight the location where it is placed (“there+PUT”), and as Clark points out, this is indeed a very common type of communicative act, both in children and adults. Still, the mere fact that an object is put

down is not enough in itself to make it communicatively foregrounded. Sometimes the act of putting an object down appears as mere disengagement from an object, after the performance of some other more foregrounded or essential act. In such cases, putting becomes part of the disattend track, and the act is often performed while gaze is directed elsewhere and without coordinated utterances.

Even though PUT [DISENGAGE] is for the most part not communicatively foregrounded, it certainly can be, in various ways and to various degrees — especially in cases where the accomplishment of the previously performed act is highlighted (Andrén, in press c). For example, a child may say “like that+PUT” (past-oriented speech) in temporal coordination with a PUT [DISENGAGE] (putting a comb on the table) after having performed an act of combing the hair of a doll with a comb while saying “combing+COMB”. In such cases PUT constitutes the endpoint of the previous action gestalt, constituting the shift from ongoing action to *accomplished* action, as a kind of real-time aspectual profiling of the ongoing action.

Sometimes PUT appears as putting *away* — away from the center of the manipulatory area — rather than just putting *down*. In such cases the object, such as a book, may be put down in the periphery of the manipulatory area in order to indicate that an activity, such as book reading, is to be considered finished. Putting something away sometimes becomes a two step procedure. That is, sometimes the object is first placed on a table (PUT), and then pushed away (REMOVE). In either case, acts of putting an object in the periphery of the manipulatory area gains their meaning due to making the object *less accessible* for further action, that is, an inverse orientation to the accessibility principle (Clark, 2003). The foregrounding of putting objects away *as* putting objects away may be strengthened through utterances such as “not that one+REMOVE”. They may also be foregrounded through the ways they respond to previous utterances. An example of this is when a parent asks “should we read that book?” and the child responds by putting the book away, as a manifest denial of the parent’s suggestion, despite not saying anything.

An object may also be put down in a way that makes it appear as if the object is put *forth*, in preparation for further activity. The book reading activity mentioned above started out when the the child picked up a book which was lying in the periphery of the manipulatory area and then put it in the center of this area, in front of herself and her mother, while saying “like that+PUT read the-book mommy”. That is, the object is put down at a location, and in an orientation, which is suitable for a specific form of further activity. A variant of such preparatory putting is when the position of an object is slightly adjusted — by means of grabbing the object, moving it slightly, and then releasing it — as a little tweak to make the object *even better*

suited for further activity.

There is also a variant of OFFER which is performed, not through the prototypical holding of an object towards the hands of an interlocutor, but by placing the object on a surface close to the hands of the interlocutor. This is called PUT\_GIVE here and it is distinct from OFFER in that it does not require collaboration from the interlocutor to turn into GIVE. In fact, sometimes the children first perform OFFER directed towards the hands of the interlocutor, but when the interlocutor does not take the object, the child instead puts the object on the table in front of the interlocutor to ensure that GIVE is achieved despite the lack of collaboration from the interlocutor. Another variant of PUT which is related to OFFER/GIVE consists in putting an object in the hand of a parent as a response to the parent when the parent holds an empty hand towards the child, asking for a specific object, as yet another possible version of an overall GIVE event. This is of course different from prototypical OFFER, which is typically initiated by the child.

In some cases PUT\_GIVE appear as a request or a suggestion that the parent should do something specific with the object (just like OFFER does). For example, when a piece of a cake is put before the parent, the type of object itself may suggest that the child gives the object to the parent in order for the parent to eat it, or something similar. In this sense knowledge of the objects involved take part in the more specific appearances of PUT\_GIVE. Still, the very act of placing something before another person tends to appear as some sort of giving in itself, even in cases where the more specific intention of what should be done with the object remains unclear to the interlocutor and/or the observer.<sup>26</sup> Whether PUT\_GIVE appears as a request for the interlocutor to do something specific with the objects or not is also highly dependent on the context of the preceding activity. That is, specifications of the more exact intentions involved in PUT\_GIVE (and most other acts discussed here) is not only a matter of operation-act interdependence, but also a matter of act-activity interdependence (“framing”).

In some cases PUT becomes tricky for the child, similar to the trouble that is sometimes involved in GRAB. Putting a piece of a puzzle in place is a typical example of this, since it requires some precision in how the object is put, or rather attached, to other objects. In such contexts there are often turn-expansions. An example of this is when a child says “here+PUT”, but when she has some trouble attaching the object held in the hand to another object, she expands the turn into “here+PUT

---

<sup>26</sup>This is easily demonstrated by turning the audio off and by isolating the act from the context of the preceding activity, when playing the video recording of the interaction.

should be” by adding “should be” to the previously pronounced first part of the utterance (“here+PUT”). That is, the trouble of the performance may itself invoke further communication.

The analysis so far has shown that the specific *place* where an object is placed is often significant for the communicative appearances of PUT. When a child puts a piece of a jigsaw puzzle on a place in the puzzle where the piece fits, as just mentioned, this is not just a matter of putting the object down. Instead it is a more specific content-loaded and situated act that appears as just this: putting a piece of a puzzle in a place in the puzzle where the piece fits (cf. Livingston, 2008). This may be performed with a deictic word such as “there+PUT” or with a content-loaded word such as “fits there+PUT”. With or without content-loaded words, such acts typically appear content-loaded anyway, in themselves, being visibly oriented to some sort of typified cultural knowledge. Deictic words referring to objects such as “this” or “that” are rather infrequent here. Instead, the majority of the deictic words that appear here are oriented to *locations* (“on there+PUT”) or the *actions* performed (“like this/that+PUT”) (Sw. *så här/så där*).

Another example of PUT that displays an orientation to knowledge of the objects involved, without involving “pretense”, is when a child who attaches the cap of a pen to the pen says “there+PUT”. The pen is held by the parent, but the action is initiated by the child. The cap involved is selected among a set of other possible candidates, but the one which she picks up happens to be the one that has the same color as the pen. Thus, this acts displays knowledge of “caps should be placed on top of pens” and “the cap attached to a pen should have the same color as the pen”.

Sonesson (1994, p. 279) points out that the significance of objects is partly dependent on where they are located. In semiotic terms, there is a difference between a car which is parked on the street and a car shown on an exhibition, even though the very same car may be involved in both cases. When a car is displayed on an exhibition it becomes an explicit sign in the sense of being an *exemplar*, displayed as a representative of a class of objects.

Such effects of transforming objects into exemplars can also be brought about through the ways bodily actions are performed, in integration with other semiotic resources. Streeck (1996, p. 371) provides an example where SHOW works like this, and sometimes PUT also work like this. Clark (2003) provides the example of a customer who places an object at a counter in a shop, to indicate what the customer wants to buy. In this case it is not primarily a matter of constructing the object as a representative exemplar of a class. Although the customer may or may not be satisfied with buying any other member of this class, the act also indicates that the



customer plans to buy this specific object. However, in this case too the particular location where the object is placed makes the object significant in a certain way. This is in fact a version of the previously discussed *PUT\_GIVE*, where the object is placed for someone else to take, although there are in this case also specific connotations regarding what the salesman is supposed to do with the object (telling the customer how much to pay for it). Such expectations are of course also dependent on other factors. If the customer instead says “this banana was rotten” when putting a banana on the counter, it would be quite clear that the intention is not to buy the object in question.

In the previous analysis of *OFFER* and *SHOW* (p. 298) it was also mentioned that directing an object towards the hands (*OFFER*), towards the eyes (*SHOW*), or towards the mouth (*FEED*) of an interlocutor often makes a difference regarding the type of communicative act that results. That is, the specific place of articulation of a communicative action may have all sorts of communicative effects, although the movement of the hand itself may be similar. This is of course also true for *PUT*, even particularly true, since *PUT* always involves placing the object *somewhere*. Furthermore, the exact relevance of this “somewhere” is a complex result of many factors. For example, putting something on the particular location of a counter in a store includes a range of different aspects, including “putting something away” from the own body (disengaging from it), placing it at a particular place (the conventions surrounding the counter, also in relation to the type of object; money vs. an item to purchase), putting it within reach for someone else (as a form of *GIVE\_PUT*), putting it so the interlocutor sees it (locations as specified by sensory availability) and so on, all as part of the very same act of *PUT*.

The extent to which an object may or even should be placed at a specific place varies. For example, there are strong expectations that food should be put on a plate or in the mouth, rather than on the table beside the plate. In other cases, it may be considered proper to put an object in a specific place, but not necessarily improper to put it somewhere else. For example, when a piece of a puzzle is put at a location in the puzzle where the piece fits, this amounts to a proper and content-loaded use of the object, whereas putting pieces of a puzzle on a table is not necessarily improper. In this way, the act of putting an object at this or that location may display content-loaded knowledge of the canonical use and placement of the object involved. In many other cases, suitable locations for putting may be a more open question, rather being a matter of practicality. In such cases, *PUT* appears in more generic way, as general handling of the object.

Placing objects may also be part of more abstract signifying actions such as sym-

bolic play. An example of this is a child who says “may eat a pen+PUT” while placing a pen close to the mouth of a toy figure held by the mother. While the mother holds the toy figure she also makes sounds of eating as if the sounds were coming from the toy figure. Some moments later, a similar act of feeding is again performed with another pen, with a similar utterance “he wants a pen+PUT”. In these cases, the act of PUT becomes an act of feeding, partly due to being directed precisely towards the “mouth” of the toy figure, and partly due to the utterance, in the first case partly due to the orchestration with the content-loaded verb “eat”. The pens are thus treated as if they were “food”, even if they are referred to using the noun “pen” rather than some other word that would refer to something that would normally be edible. That is, although pens and mouths are not normally related, a specific construal (pen -> food) may indeed make them related in the sense of drawing on previously established typifications — procedures of eating and feeding, in this case.

Even though the object which is put down is generally highly thematic in acts of putting something down, there are also a few cases where the object is less thematic. One example of this is when a child says “hide the trousers+PUT” while she covers another object (which is thematic) with a pair of trousers (she does not hide the trousers). As a consequence, the PUT results in hiding *another* object, and the object which was put down is only relevant on a very generic level — it is only relevant to the extent that it has the physical properties that is required for covering another object.

Sometimes the hand releases the grip of an object that has just been put down, but then the movement is stopped and the hand remains close to the object. This usually has the communicative effect of “trying an act out”, which means that the act appears to be finished in the sense of having shifted from ongoing action to performed action (a product), but still ongoing in the sense that its *consequences* appear to be (visibly) (re)considered and possibly subject to change. This is similar to acts described in the previous section where HOLD\_TOWARDS [HOLD] was performed without releasing the grip of the object. These two types of action are similar in that both appear to “try an act out”, as if asking: Was that the correct object? Was the object placed at the correct/suitable location? Does it fit? But in cases where the grip of the object is not released (as in the case of HOLD\_TOWARDS [HOLD] described in the previous section) the act does not appear finished, and the visible (re)consideration of the act takes place already before the act has become a “product”. In most cases this may not make a big difference apart from a difference in degree regarding how “sure” or “unsure” the act appears, but the difference is still perceivable. I was unable to find any instances in the data where these two variants

are treated in markedly distinct ways by the participants, but more generally it may be noted that in the game of chess, for example, there is a rule stating that whenever a piece is touched, it has to be moved although it may still be changed into some other valid move using this piece, and whenever a piece is put down somewhere (in the sense that the grip of the object is released), the move may not be changed.

So, at least in some actually occurring contexts, the difference between actually releasing the grip of the object or not may be of importance, precisely because releasing the grip of the object produces the appearances of PUT as a finished “product”. The halting of a movement produces some intention movement-like appearances. In this case, the continuation appears like a point of *choice* between redoing, adjusting, or changing the act in some way, or withdrawing from the act/object and thus letting the act remain as it was. This visible appearance of a point of choice is also what makes the act is an important part of what produces appearances of “consideration”.

## 9.7 Withdrawing from the object

### 9.7.1 Definition

Once the object is released at the location where it was placed or attached, the hand may start to withdraw from the object. This generally means that the character of the hand as being directed towards something ceases, unless the withdrawal is immediately halted through a hold close to the object. In most cases withdrawal from an object does not become highlighted as a communicative act in itself. In such cases the withdrawal may of course still be informative to interlocutors, since it often marks the end of some previously performed action.<sup>27</sup> It makes the previous act appear as a finished *product* of some sort, rather than an *ongoing* action (Schutz, 1945, pp. 542; Garfinkel, 1952, p. 364). An action that was being “in progress” a moment ago now becomes transformed either into some sort of accomplishment, rather than ongoing performance, or in some cases into some sort of failure. This is not entirely unlike how the retraction phase of an empty-handed gesture works. Such retractions also do not signify in the referential sense of gesture strokes, but they do contribute on an operational level to the appearance of an action gestalt as

---

<sup>27</sup>Acts may of course appear “finished” before the hand withdraws from the object. An act of picking an object up appears finished at the moment when the object is held in center space, and so on. The point is rather that there are few actions associated with the withdrawal slot of action excursion space, so this aspect of action is the dominant one here.

now being *finished* and in this sense they are not entirely devoid of meaning.

In some cases, withdrawal from an object does become communicatively highlighted (Andrén, in press c), typically through concurrent speech, but sometimes also through the addition of 'effort' and 'shape' to the withdrawal, which is what normally characterizes a stroke (Kendon, 1980b, 2004) rather than the withdrawal involved in retraction phases of empty-handed gestures. In such cases, the highlighting of the withdrawal often has the effect of making the accomplished character of the previous action explicit. Thus, such highlighting of withdrawal works as a kind of real-time aspectual profiling of the ongoing action. Highlighting of the accomplished character of an action also often goes hand in hand with some sort of evaluation — positive in cases of “success” or negative in cases of “failure” — of the action that was just performed.

### 9.7.2 Analysis

In some cases the withdrawal of the hand from an object co-occurs with an utterance that makes the completion of some previous act a communicated and intersubjectively established fact. Examples include: “like that+WITHDRAW”, “there+WITHDRAW”, or “done+WITHDRAW”. The movement dynamics of such withdrawals may be segmented in a way that make the coordination between speech and movement clear, as well as providing for the difference between “mere retraction” and *marked retraction*. For example, the tone nucleus of the utterance may co-occur with the moment the hands fall on the lap.<sup>28</sup> Letting the hands “fall”, rather than “moving” them to a position of rest, is one of the characteristics that may make withdrawal more salient, especially when the falling hands produce a clapping sound in temporal coordination with the tone nucleus of the utterance, which happens from time to time. In some cases there is in fact a proper hand clap performed instead. That is, the withdrawing hand does not end up in a position of rest, but in a hand clap in center space. Examples of this is when some notable achievement has just taken place, such as putting the last building block on top of a tower of building blocks, thus completing the construction of a tower, and the child’s withdrawal of the hand from the building block ends up in a hand clap — “yeah+CLAP \*laughter\*” as a form of positive evaluation of the act just performed. Hand clapping generally occurs in this specific context in the data, as positive evaluations of the act that was just performed. The clapping sound appears to occur in contexts where the semantic theme

---

<sup>28</sup>Jana Bressem has also noted how retractions after empty-handed gestures sometimes have this peculiar character of “letting the hands fall” (Bressem, pers. comm., December 2008).

of marking the successful transition from “ongoing action” to “accomplished action” is actualized.<sup>29</sup>

Apart from hand claps, there are also other types of marked retractions, which are more common. In cases where a notable accomplishment of some sort has just been performed and achieved (such as some relatively tricky multi-step procedure; an activity) there is often a markedly quick, almost forced, withdrawal from the object. Such types of marked retractions will be referred to as *snappy retractions*. They may end up in a position of rest, but in many cases they rather become part of a two step/phase withdrawal procedure, where the hand first quickly moves back towards the body, and only then moves to a position of rest, which typically means that the hand is moved downwards. Snappy retractions also often come with utterances that refer to the “correctness”, or more emotionally loaded positive evaluations of “success”, of the previously accomplished act or activity. In some cases, what is notable about the previous action is that it failed in some way, and here too the previous ongoing trouble has now been transformed into “failure” — as a product — and the evaluation may be loaded with negative emotion, maybe combined with some request for help or support from the parent.

The Swedish word *så* deserves a special mention in relation to WITHDRAW.<sup>30</sup> In the contexts analyzed here *så* corresponds roughly to “like this” or “like that”, similar in many respects to some uses of German *so* (Streeck, 2002b; Golato, 2000), English *like* (Coseriu, 1988; Streeck, 2002b), and Estonian *nii* (Keevallik, 2005, 2009). More generally though, *så* is a highly multifunctional word in Swedish that appears as a form of deictic word, discourse marker, adverb, conjunction, and more (Lindström & Londen, 2008; Norrby, 2002; Ottesjö & Lindström, 2006; Ekerot, 1988). There is also a long row of variant forms that reside somewhere between collocations and compounds, such as *sådär* (so-there, often referring backwards in time), *såhär* (so-here, often used prospectively), *sådärja* (so-there-yeah, which is more explicitly evaluative), *så att* (so-that, which is often pronounced as a single word; “*s’att*”), *också* (also), *sån(t)* (such), and many more.

In relation to gesture, this type of word has mainly been described in contexts where it is used for demonstration or “body quotation” (Streeck, 2002b). In such cases it is generally coordinated with the *performance*, that is, with the stroke of an action or a gesture. However, at least the Swedish counterpart (*så*) is also of-

---

<sup>29</sup>A number of examples involving clapping sounds and hand claps are analyzed in Andrén (in press c).

<sup>30</sup>See also Andrén (in press c) for an analysis of a number of examples that involve uses of the word *så* in the context of being finished of some previous action.

ten used *after* the act, when the act is *accomplished* rather than during its ongoing performance. In such cases it is used to indicate that something was *properly* accomplished, according to norms and conventions, or, in a few cases, in relation to locally established expectations. This particular use of *så* is frequently used by the children in this data and all children studied use *så* both in the accomplishment sense and in the ongoing performance sense, and in almost all cases it is clear which of those two uses that is currently instantiated. It is thus part of the conventionalized use of this word to be used in certain temporally organized ways with respect to the ongoing activity. For the present analysis, it is important to note here that when *så* is used in its accomplishment sense, it often co-occurs with the endpoint of a retraction and these retractions may, or may not, be marked with ‘effort’ and ‘shape’ in their performance.

In this regard it is also interesting to note that the word *så* — in the accomplishment sense, after an action — is frequently used by *parents* as a form of evaluation of *children’s* acts, thus making the *typified and recognizable character of the children’s acts manifest for the children*. That is, it also figures as part of processes of educating children’s recognition of their own actions as content-loaded *types* of action. Along the lines of Sartre, we may say that “language teaches me the structure which my body and my actions have for the Other” (cf. Schutz, 1948, p. 192). In a related study by Tomasello & Barton (1994) it was found that 24-month-olds were sensitive to whether adults said “whoops!” or “there!” after the performance of an action (by the parent) which had been labeled with a nonce verb before the performance of the action — the children were less inclined to take the verb as a label for the act that was performed. This means that subsequent speech may change the experience of what a previous act “really was”. Maguire et al. (2006) report on the results of a study by Behrend and Wittek, where it was shown that 30-months old children often took the presence of a verb as a sign that a novel action was being named, even if the act had the character of unintentional movement (although the apparent degree of volition also played a role in the strength of the effect).

The word *så* and the marked retractions all have in common that they have the effect of making the “aspectual profile” of the ongoing action intersubjectively manifest (cf. Andrén & Zlatev, accepted). What this means is that they mark the shift from “ongoing performance” to “proper accomplishment” (or failure) of a content-loaded action, and they do this at the precise point in time when the shift occurs. To some extent, they could even be said to *constitute* these shifts. When, for example, a child closes a book, and then, in coordination with a marked retraction, says “done+WITHDRAW”, the closing of the book is part of a multimodal ensem-

ble which highlights, and constitutes, the fact that an interactive sequence has now come to an end. Thus, the child displays specific content-loaded knowledge of this particular activity of book reading through this act, such as “book reading is finished after the last page is read” and “closing the book amounts to being done with this specific type of activity”. In this way, the state of “being done” with something can be communicated and made intersubjectively established.

Marked retractions do not, in general, appear as independent *actions* in the same way as picking something up, offering something to someone, performing a content-loaded act with an object, or putting something down. The hand claps are the closest to appear as “communicative gesture” among the marked retractions observed here, but even these mean very little on their own, and the meaning of these hand claps is only understandable with reference to the “main act” which has just been performed. In this way, marked retractions are more like *operations*, forming a part of an action instead of appearing as actions in their own right. That is, even though WITHDRAW may be communicatively highlighted, it still rarely takes the form of a referential communicative act. Their meaning rather has the form of contributing to making those “main acts” into “what they were”. On the other hand, uttering the referential word *så* also means relatively little on its own, even though saying it amounts to an action on the action level (especially when it is used on its own as a one-word utterance). This word too depends on other “contextual factors”, such as the performance of a bodily act, in order to gain its meaning, and its exact meaning is also dependent on its temporal coordination with action.

## 9.8 Summary and conclusions

With respect to the first of the three questions posed before embarking on the analysis — “*What sort of acts on objects have communicative appearances?*” — this stroll through the various subspaces of action excursion space has identified a relatively large variety of action gestalts with many different kinds of communicative properties. This goes, not only for empty-handed gestures and “as if” acts involved in symbolic play, but also for more situation-specific types of action and *typified* actions, which may be communicatively highlighted through the way they are performed and/or due to their orchestration with other semiotic resources such as speech and gazes. The aim of this summary is not to recapitulate those in their detailed specificity, but instead to explicate some overall patterns and characterizations emerging from the analysis as a whole.

One basic conclusion to be drawn from the analysis, as framed in terms of the subspaces of action excursion space, is that communicative action gestalts appear in all of the subspaces, and when they do, their meaning is motivated in terms of the relevancies prevailing in the particular “ecology” of the subspace in which they occur. For example, when reaching for an object, without actually taking it, one distinction that systematically makes a fundamental difference to the meaning of the act is whether the object is within reach or not. This distinction is, however, completely irrelevant in other subspaces. For example, in the subspace where objects are put down, the object is of course already held in the hand when the act begins. When putting objects down, the main factor in providing the meaning of the act is instead *where* the object is placed. In this way, *each subspace comes with its own logic and affordances for signification*. At the same time, there are also properties of the ways in which the actions are performed that produce similar effects across all of the subspaces, at least on an abstract level, such as the distinction between *ongoing* and *stopped* (statically “held”) movement, and the distinction between acts with *generic* appearances and those with more specific *appearances*. Both of these distinctions are treated below.

With respect to the second question — “*What is it that makes these acts recognizable as certain kinds of action gestalts?*” — a number of factors involved the communicative foregrounding of these acts were identified on all three levels of the operation-act-activity interdependence nexus:

1. *The operation level:* This includes various aspects of the movements that make up the acts, and how the acts are orchestrated temporally with other semiotic resources: primarily speech, but there were also some observations on gaze orientation. The analysis in terms of an action excursion space, and subspaces therein, also belongs to this level since the action excursion space can be understood as a kind of form feature.
2. *The action level:* There belongs the question of whether the action gestalt is recognizable as having a specific relation to the object involved — such as the object-specific type of movement involved in brushing teeth with a toothbrush — or if it appears as a more generic type of action of the sort that can be performed with many sorts of objects — such as GIVE, SHOW, PUT, PICK\_UP, TRANSPORT, *etc.* Semantic coordination between action and speech also belongs to this level.
3. *The activity level:* Various sorts of relations between the act and the ongoing activity. The act may be recognized in a specific way due to its “framing” (Goff-



man, 1974)<sup>31</sup> in an activity type or initiation of a new activity (creation of framing such as saying “we read this+PUT [BOOK]” when putting a book in front of the mother to initiate an activity of book reading), or it may be foregrounded because it is performed as a direct response to an utterance from the interlocutor, i.e., attaining communicative status due to its sequential relevance. Responsive aspects of action involve both acting in compliance and performing conflicting actions such as “undoing” the result of a previous act by a parent.

With respect to the third question posed — “*What makes these action gestalts distinguishable from each other?*” — the answer is to some extent given as a result of answering the previous two questions, both of which specify what is unique in the constitution of various sorts of acts, but in addition to this, some analytical attention was also paid to ambiguities. More precisely, it was noted that action gestalts that are sometimes mistaken for each other are precisely those that are similar to each other in terms of their organization in action excursion space, e.g. the action gestalts of SHOW and OFFER.

### 9.8.1 Ongoing vs. stopped movement

For all of the subspaces, a major distinction is whether the movement becomes communicatively highlighted when it is ongoing (moving) or stopped (statically held). In cases where the movement of an action is *stopped*, at least three different basic experiential structures are possible. The stopped movement may constitute (A) the *end* of an action (i.e., stopping the movement of the hand in center space after having picked something up), (B) a *pause* in the middle of a yet unfinished action, or (C) the *performance* of an action that contains stopped movement as part of its canonical performance (i.e., the “ongoing” holding of a telephone towards the ear to do “phoning”). For both (B) and (C) the movements involved are often enough in themselves to produce communicative appearances, even though speech is of course often coordinated with the action too. As pointed out in the analysis, the holds of empty-handed gestures have been shown to capture interlocutors’ attention (cf. Gullberg & Kita, 2009). Also, the kind of holds that serve the specific purpose of aligning and coordinating speech and gesture with each other, as de-

---

<sup>31</sup>Note that Goffman’s use of this concept is relatively static. My understanding of this term is more in line with the work of Linell (1998, 2009) who treats this concept in a more dynamic way, noting how several frames may be actualized at the same time as well as being modified by the acts that realize the activity (act-activity interdependence, see Section 3.4).

scribed by Kendon (2004, p. 127), could be added to this scheme as a fourth type of holds. When the stopped movement constitutes the end-point of an action, as in (A), the action is for the most part communicatively foregrounded as a result of coordination with speech or mutual gaze, rather than through manual movement alone. In such cases the speech is often either of a deictic kind (“this”, “there”) or oriented to the achievement of the action (“mine”, “like that”). The fact that stopped movement is sometimes perceivable as the endpoint of an action requires, among other things, that the consequences of the movement are recognizable as possible stopping conditions for some sort of act, i.e., being able to see a purpose in the act. This is also true for (B) and (C), but in these cases the presumed stopping conditions are perceived as not yet being met or as in the process of being met.

In the case of (B), the acts work much like intention movements (Lorenz, 1957 [1939]; Tinbergen, 1951; Darwin, 1872; Mead, 1934). That is, the movement involved in an action is stopped before the act is fully realized — thus projecting a possible future performance of the act, as a suggestion. However, it is not only the case that these acts suggest the possibility of a future action, they rather imply a *choice* between continuation or interruption (and possibly also modification) of the act — “Should I proceed with this act or not?”. These stopped movements are almost invariably attended to by the interlocutors and the children routinely seek feedback from the interlocutors, either by means of looking towards the interlocutor, asking for a response or for help<sup>32</sup>, and/or by holding the action until a spoken response is received. The latter holds actualize yet another aspect of the organization of stopped movement, the fifth one, similar to how other gestures such as pointing are often held until the precise moment when a response is received.

The more specific relevance of a stopped movement is also due to the subspace of action excursion space in which it occurs. For example, when stopping a movement in the subspaces of REACH and MOVE\_TOWARDS the stopped movement projects a vector of directedness towards a target object, which is often similar to pointing, although in the latter case, the effect is often a more specific one of actualizing a particular relation between the object held in the hand and the target object or location which the object is held towards (OBJECT\_POINT). If stopping a movement as part of moving an object towards a recipient, the appearances are rather that of OFFER or SHOW depending on the more exact place of articulation in space. When putting something down (PUT), the issue evoked by the stopped movement is generally the target location, and often the specific relation between the target location

---

<sup>32</sup>This is sometimes done through a negatively loaded emotional vocalization rather than speech.

and the object held in the hand. When movement is stopped in center space, there is generally no projection of a vector of directedness since the hand is generally not going towards something. In this way the various subspaces differ when it comes to the more specific effect evoked by stopped movement.

In cases where the movement of an action is not stopped, but *ongoing*, there are two main types in the data. The first one is when a movement is foregrounded in the middle of a seamless ongoing movement by means external to the movement itself (primarily speech). Such movement generally appears to be a mere part of an action excursion and is generally not particularly foregrounded in itself. Typical examples are the REACH which precedes an action of picking something up, or the MOVE\_TOWARDS that precedes an act of putting something down. In fact, in such cases the moment when something is grabbed or put down are more salient and foregrounded than the preceding ongoing movement. Still, the speech which concurs with such movement often have a future-oriented character towards the act which is to be performed as soon as the hand reaches its goal such as “I’ll take” in “I’ll take+REACH that+GRAB one”.

The second type of foregrounded ongoing movements is much more salient than the previous type, and it is also more common. This type consists in moments of ongoing excursions that appear as a joint or node in the overall flow of movement. Examples include the moment when an object is grabbed or put down, or when some more specific action is performed, such as in the middle of pulling a lever. Even though such moments may be performed without stopping the movement, they still appear as moments of particular relevance — “something happens”. In contrast to the first type, this type is more highlighted due to the character of the movement itself, although co-occurring speech is a very important factor in producing an overall impression of a communicative act for this type too. Examples include “that+GRAB”, “there+PUT”, and “open+PULL\_LEVER”. Such moments are of particular relevance no matter if they appear as an action gestalt in themselves (such as putting an object down), or if they appear as a subsumed part of a more encompassing action gestalt (such as TRANSPORT which subsumes several operational phases of grabbing an object, lifting it, moving it, putting it down).

### 9.8.2 Coordination with speech

The fact that speech is semantically and temporally coordinated not only with empty-handed gesture, but also other sorts of actions, is such a mundane, commonplace, and inconspicuous phenomenon that it is easily overlooked despite its theoretical

import. The insight that empty-handed gestures are temporally and semantically organized with respect to speech is perhaps the single most important factor behind the rapid expansion of interest and research in this area (Kendon, 1980b, 2004; McNeill, 1985, 2005), but what is the scope of this skill of multimodal orchestration? Sometimes it seems to be assumed that there is a ‘Thought-Language-Hand link’ (McNeill, 2005, pp. 233) which pertains specifically to empty-handed gestures and not to acts that involve objects (implicitly equated with “instrumental movement”, *ibid.*, p. 245):

In terms of brain function, [the study of IW] implies that producing a gesture cannot be accounted for entirely with the circuits that perform instrumental actions; at some point, the gesture enters a circuit of its own and is tied there to speech.

While it is true that communicative action is not completely reducible to the capacities involved in instrumental action (Gallagher, 2005), I hope to have shown in this chapter, and this thesis as a whole, that (A) coordination of speech with action is not restricted to empty-handed gestures (on the Sem#3 level) that take place solely in abstract space. Neither is it restricted to empty-handed gesture plus “as if” acts of the specific kind that is called symbolic play. There is also an abundance of situation-specific acts (level Sem#1) and typified acts (level Sem#2) coordinated with speech in the data analyzed here. Furthermore, I hope to have convinced the reader that (B) (the conceptually distinct) communicative and instrumental dimensions of action often co-exist in one and the same actually occurring action, and that (C) even if a communicative act lacks instrumental properties, it may still involve handling of objects. This is not to say that there are not also generally some differences between those communicative acts that are empty-handed and those that involve handling of objects. For example, they differ in the way they require explicit guidance by means of visual pathways for their performance, but it should be clear that this does not justify a one-sided analysis only in terms of differences between empty-handed gesture and acts that involve objects, and it certainly does not justify an exclusive association of speech neither with acts that are exclusively communicative nor with empty-handed acts (and/or “as if” acts).

One of the most central conclusions from the analysis in this chapter is therefore that *all the three levels of semiotic complexity may be part of communicative multimodal acts*. This is related to Tomasello’s (2008, p. 69 and p. 218) argument (with reference to Leslie, 1987), that communicative acts need to be “quarantined” from purely instrumental purposes in some manifest way in order to avoid misinterpre-

tation.<sup>33</sup> However, Tomasello, like many other gesture researchers, seems to assume that this quarantining needs to be implemented by the movement involved in the action itself, and the possibility that acts may become communicative by means of its orchestration with speech or due to contextual factors does not seem to be taken into account. Thus, many gesture researchers contradict the otherwise dominant idea that the parts of multimodal orchestration are to be regarded as a communicative whole, since they disregard a large class of actions/movements because they may not have appeared communicative when isolating them from speech and other semiotic resources, as well as their context.

The analysis also showed that the speech which goes along with the handling of objects may be *past-oriented*, *now-oriented*, and *future-oriented*. Now-oriented speech is related to what happens at the very moment the act is performed, and this is by far the most common type of speech-coordination in the data studied here. The next most common type is future-oriented speech, and this is related to what will happen next. Typically, this involves saying something about what the object currently being grabbed will be used for or how it will be used. That is, while the speech is future-oriented, the expression also relates to the act on the object performed in the *present*. In this way, several temporal orientations may co-exist in the same multimodal utterance. Past-oriented speech is less common in temporal and semantic coordination with acts, since it generally (but not necessarily) relates to something that has just happened, before the present act. In cases where past-oriented speech is indeed pronounced with relation to the presently performed act, the speech is generally oriented to the accomplishment (or failure) of an action, such as saying “mine” at the endpoint of an act of picking something up, or “like that” together with a marked withdrawal movement from the object, after having succeeded in putting a piece of a jigsaw puzzle at a location where it fits.

Another central conclusion from this chapter is that the meaning of the acts in various subspaces of action excursion space are to a large extent tied to the specific “ecology” of each subspace. This is also reflected in the speech that is coordinated with the actions. To idealize a bit (see Figure 9.3), one could speak of four overarching phases in action excursions: An initial phase of being *directed towards the*

---

<sup>33</sup>It could be argued that in some situations it does not necessarily matter even if an act is understood as a communicative or not, as long as the intended effect appears. For example, if I put a toy in front of an infant while saying “here you go”, in order to give her something to play with, it does not necessarily matter to me if the infant understands the communicative aspects of this acts, as long as she notices that toy and start playing with it and seems to be satisfied with that. However, on the whole it is of course often crucial whether an act is understood in communicative ways or not.

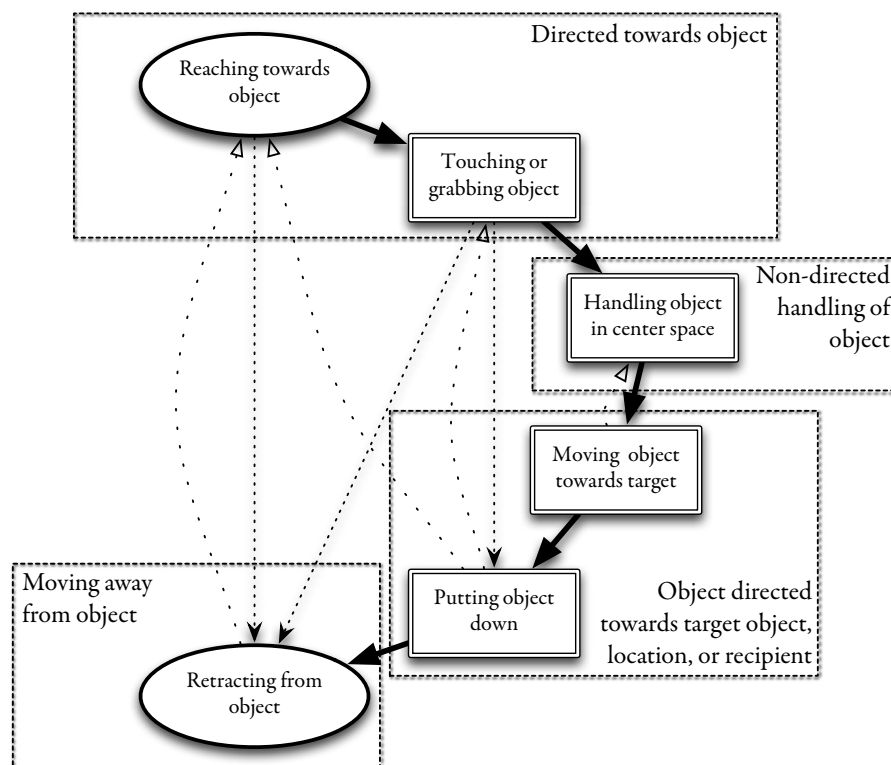


Figure 9.3: Idealized types of relevance in action excursion space.

*object* (REACH and TOUCH/GRAB), a phase of *non-directed handling of the object* (CENTER\_MOVE), a phase where the object is directed towards a target object, location, or recipient (MOVE\_TOWARDS, PUT), and finally a phase of moving away from the object (WITHDRAW). Whenever speech contributes to the highlighting of any of those four phases, there are some quite clear patterns in the semantics of the speech which goes along with the actions, at this level of idealized analysis. This is not to say that there are no exceptions to these general patterns, and the discussion should be considered a description of what is *usually* going on in each of the subspaces. It should not be considered a strict rule or a necessity.

Table 9.1 summarizes the dominant patterns of semantic focus in spoken expressions coordinated with action in each of the four overarching phases of action excursions. Simply put, the speech during first phase is focused on the object that is being approached and “selected” (deictic words relating to objects, naming the objects, verbs denoting acts specific to this part of action excursion space). Speech during the second phase is rather focused on doing things with the object (non-directional actions), and formulating propositions *about* the object (for example by using adjectives). Speech during the third phase is generally either focused on the

Movement phase	Dominant semantic focus of speech
<i>Directed towards the object:</i>	Mainly now-oriented speech. Deictic word relating to objects ( <i>this, that, such</i> ), naming of objects using nouns ( <i>ball, doll</i> ), verbs (mainly generic) pertaining specifically to the current subspace ( <i>take, have</i> ), but also more specific verbs ( <i>pat</i> ).
<i>Non-directed handling of the object:</i>	Mainly now-oriented speech. Predication <i>about</i> the object ( <i>red, empty, mine, too little, such one</i> ). Verbs/adverbs relating to non-directional acts on the object ( <i>open, fold this, like this, (I) hold like this</i> ), and to inspection/showing ( <i>look</i> ).
<i>Object directed towards target object, location, or recipient:</i>	Mainly now-oriented speech. Deictic words relating to target locations ( <i>here, there</i> ) when putting objects down. Verbs, verb phrases, and vocalizations that supplement the performance of content-loaded acts that are directed towards a target object, location, or recipient ( <i>tickle tickle, myam yam yam</i> when pretending to eat toy food held towards the mouth, <i>hello?</i> when holding a toy phone to the ear). OFFER and SHOW are often combined with deictic words or nouns that relate to parts of the objects or the objects as whole ( <i>this, a spot (on the plate), plate</i> ), and they are also often combined with verbs pertaining to each of the acts ( <i>look this, mommy take</i> ).
<i>Moving away from object:</i>	Mainly past-oriented speech referring to accomplishments or failures ( <i>like-that, done!, yeah!, there, doesn't fit</i> ), marking the transition from ongoing action to accomplished (or failed) action.

Table 9.1: Idealized semantic focus of action-coordinated speech.

object held in the hand and the recipient (in cases of showing and giving) or on the target location, target object (another object than the one held in the hand), or the relation between the object held in the hand and the target location/object (in cases of performing acts with the objects or pointing gestures that include the object held in the hand in the articulation).

### 9.8.3 Generic and specific aspects of action

Some acts have *generic* appearances, in the sense that they can be performed with many sorts of objects (TAKE, OFFER/GIVE, THROW), and other acts have more *specific* and content-loaded appearances in the sense that they involve specific sorts of objects (BRUSH [brush], EAT [food], PHONE\_CALL [phone]). To be sure, no action is completely “generic” in the sense that it is completely devoid of features that are particular to just this type of action. This is partly analogous to how generic words such as “here”, “there”, “mine”, “yours”, and “such” are relatively open-ended in their meaning with respect to the particular objects (referents) involved, while they still do have semantic content.

Within linguistics, the distinctions between *function words* versus *content words* and between *closed-class words* versus *open-class words* share some similarities with the distinction between generic and specific action. Although it is not trivial to provide a once-and-for-all definition of these distinctions in a way that makes its application crystal clear for each particular word in a language — it is rather a continuum — these distinctions still serve to highlight the difference between words that have a relatively specific meaning and/or referent such as “windmill”, “thumb”, “right-click” (on a Computer mouse), and “David Bowie”, and those that have a more generic meaning such as “this”, “do”, “she” and “and”. The latter type of words tends to be part of the situation-transcending “core” of a language, and their meaning tends to be abstract rather than concrete, schematic (desemanticized) and grammatical rather than referential, and the words themselves tend to be short rather than long (Heine & Kuteva, 2007, p. 45; see also Croft, 2003; Tomasello, 2008, p. 234). In a similar way, Tomasello (1992, p. 115) notes how verbs that denote acts on objects also lie on a continuum from verbs that apply to specific objects (hammering requires a hammer, buttoning requires a button) and actions that involve non-specific objects (bite and throw).

Actions gestalts can also be recognizable as specific and content-loaded in several ways, even though the main focus in this analysis lies on how acts are specific and content-loaded in relation to the use of particular objects. Actions can be recognizable as some relatively specific sort of action due to (A) the *form* of the act (movement, hand-shape, orientation, the location of articulation, the objects involved *etc.*), (B) the particular type of *result* of the action (including acts that serve to “correct” or “repair” the result of a previous action), (C) co-occurring speech (or other semiotic resources) that contribute to making it clear what sort of action a certain movement is, (D) a locally established “convention” in a specific situation



(without a general typified content-loaded meaning outside the situation), (E) its framing in an overall schematized activity (situation-transcending activity types), (F) the way the act responds to a previous content-loaded imperative (for example) which requires that the child is able to “decode” the content of the previous utterance in a somewhat proper way. These different sources of content-loaded appearances may also appear together in various constellations, and it is possible to extend this list quite a lot more with yet other possible sources of content-loaded appearances. A and B are most directly related to the nature of the movement itself, C is also due to performance, but rather due to the multimodal utterance as a whole than due to the movement alone, and D–F are related to the contextual embedding of the act.

An interesting pattern that emerges from the analysis in this chapter is that acts that have *generic* appearances (such as GRAB and TRANSPORT) are almost always those that are coordinated with the sorts of speech that are typically found together with *deictic* gestures, and that acts that have more specific and content-loaded appearances are coordinated with speech much like *iconic* gestures are. This suggests that there is a partial correspondence between ‘deictic gestures’ and ‘iconic gestures’ and acts that appear to be either more generic or specific, both in relation to the acts themselves and in relation to the ways that the acts are coordinated with speech.

Why do some acts appear as mostly deictic, and others mostly iconic? Even though a single act may include both deictic, iconic, and conventional aspects, one of them generally dominates. Somehow most gestures seem to be *primarily* deictic, *primarily* iconic, or *primarily* conventional (which is probably why many researchers have been relatively satisfied with gesture typologies that have mutually exclusive categories after all). One result from the analysis in this chapter is that it seems as if, as a rule of thumb, “content-loaded” aspects of action generally take precedence over deictic/generic aspects (even if they are present) in communicative acts if such content-loaded aspects are present. Content-loaded acts generally have more complex meanings, such as “relate object A to object/location B by means of action of type C” whereas deictic action tends to be closer to “relate to object A”. Supporting results comes from a study by Nicoladis et al. (1999, pp. 523) where it was found that utterances produced by 2 to 3.5 years old children with (empty-handed) iconic gestures tend to contain more morphemes and that they tend to be combined more often with verbs than with nouns, indicating that multimodal utterances that contain iconic gestures generally actualize more complex semantic relationships. That is, in cases of directional actions (putting a toy phone to the ear) the directed aspect of the act is not experientially thematic (most people would

not experience this as a “pointing gesture”) in the same way as the content-loaded aspect, even though the directedness is of course seen and apprehended and even though the child may well say “there” when putting the phone to the ear in way that is similar to how speech is used with pointing gestures. If the toy phone would instead have been held towards a table, the act would not appear to be a (pretended) act of making a phone call. Such recognition obviously requires a culturally competent member in order to be able to produce, and see, the specific content-loaded aspects of the communicative act, as it is situated in the operation-act-activity nexus.

Jeannerod (1996, p. 41) points out that fundamental aspects of our behavior, like the ability to use tools, originate from neural specialization for “perceiving, grasping, *recognizing and categorizing* objects” (my italics). For example, there are specific neural impairments (visual form agnosia) where a person may have no trouble grasping an object in a generic sense, but may fail to grab the object in ways that are suitable for the subsequent *use* of the object, such as grabbing a screwdriver in a way that allows it to be used immediately (Carey et al., 1996). Jeannerod et al. (1994) demonstrated the opposite in an optic ataxic patient who could not accurately grasp a neutral object such as a small cylinder, but whose grasping became much improved when the object was replaced with a familiar (cylindrical) object such as a tube of lipstick (see also Jakobson et al., 1991). In a similar vein, but in a study of healthy persons, Creem & Proffitt (2001) demonstrated that when people have to perform an additional semantic task at the same time as grabbing an object, the visomotor system can direct an effective grasp of the artifacts involved, but not in a manner that is appropriate for the *use* of these artifacts, and the reversed result when the subjects performed additional visuospatial tasks. Péran et al. (2010, p. 89, my italics) also report results in favor in of an interpretation of separate parts of the brain being related to “the grasping and manipulation of *any* object” versus “*specific* object-related gestures” and actions. The very understanding of these content-loaded or “iconic” acts is a result of learning in the context of *social* activity. It is not a nut that can be cracked solely by means of content-free inferential processing and innate capacities alone (Calvo-Merino et al., 2005; Catmur et al., 2007; Heyes, 2010),<sup>34</sup> even though appropriate biological capacities are obviously required too (else it would be hard to explain differences between species). It should never be forgotten that “cognitive processes” and “social cognition” rely

---

<sup>34</sup>I would refrain from using the term “associative learning” though, as employed by Catmur, Walsh, and Heyes, due to the unnecessary connotations of behaviorist and reductionist theorizing implied by the use of this term. The crucial point, in my view, is that there is a necessary component of learning involved, and that this is often a matter of social learning.

## 9.8. SUMMARY AND CONCLUSIONS

---

to a large extent on *actual* experiences and knowledge (“the stock of knowledge at hand”, in Schutz terms) which have their basis in a previous sedimented history of actual engagement in various activities.

Once again we find that the theoretical ideas and concepts from research on empty-handed gestures are not limited to the scope of empty-handed gestures. Instead, they apply much more generally to several kinds of ‘doing and speaking’ in social settings.



## Part III

# Rounding up



# CHAPTER 10

## Summary and conclusions

This thesis investigated the gestures of five Swedish children between 18 and 30 months of age. This concluding chapter summarizes the main contributions, both on an empirical and on a conceptual level. The will be done by offering answers to the initial questions posed in Chapter 1.

### 10.1 What sorts of gestural actions do the children perform?

The overarching typology of gestures described in Chapter 4 came into being through a balance act of sorts. On the one hand, it was (of course) inspired by other researchers' attempts to classify gestures (see Kendon, 2004, Chapter 6). On the other hand, it was a result of an inductive approach: no types were added to the typology unless these were motivated by observations of gestures that actually occurred in the data. In line with the ambition to provide an account of the whole gestural repertoires of the children, and not just some selected parts, there was an ambition to create a typology that would apply, as far as possible, to *all* observed performances that seemed to have gestural qualities. This resulted in a typology with categories on two levels of abstraction: main categories and sub-categories.

The three main categories *deictic*, *iconic*, and *conventionalized* are aspects of gestures. The term “aspects” is a key word here since these three categories have not been treated as mutually exclusive types — unlike in nearly all other research on children's gestures.<sup>1</sup> This is surprising, given that there is practically a consensus

---

<sup>1</sup>One type of exception to this can be found in some studies of children's gestures that do not

among researchers of adult gestures that deictic, iconic, and conventionalized semi-otic motivations are better understood as dimensions that may co-occur in one and the same gesture (Kendon, 2004; McNeill, 2005), than as distinct types of gestures. This *dimensional approach to gesture*, as one may call it, has been employed consistently throughout the thesis: both in the quantitatively oriented parts and in the qualitatively oriented ones.

The typology also included a number of sub-types: e.g., if a gesture was annotated as having iconic aspects, a further annotation, on a more specific level, concerned the kind of iconic aspects involved. The deictic sub-types were INDEX-FINGER POINTING, OTHER POINTING (pointing gestures without the index-finger pointing hand-shape), and a category called OTHER. The latter category included gestural actions such as GIVE, SHOW, and PLACE. These bring an object to someone's attention (cf. Clark, 2003) rather than bringing someone's attention to the object, as is the case in pointing gestures. The iconic sub-types were ACTION-BASED aspects (the gesture bears a resemblance to ordinary ways of acting in the world), INDIRECT aspects (some part of the body figures in the gesture as if it were something else than an acting body), and TRACING aspects (a shape or path is traced in space). For conventionalized gestures, the sub-types were specific: NODDING, HEAD\_SHAKE, HELLO, GONE, and so forth. Since this typology was developed at least in part on an inductive basis it can be regarded as a research finding in itself, and not as an *a priori* stipulation of annotation categories.

These annotations were then used in two main ways: as *coding* and as *indexing*. In Chapter 5 (and to some extent in Chapter 8) the annotations were used as codes: they served as the basis for quantitative measurements such as calculating Gestures per Minute (GPM) and Gestures per multimodal Utterance (GPU). In Chapter 6 to Chapter 9 the annotations were rather used as a form of indexing: they were used as a starting point for further qualitative analysis, and more generally to help "navigate" in the data.

The quantitatively oriented analyses were all concerned with the analysis of developmental patterns, with findings summarized in Section 10.2. The more qualitatively oriented ones were less concerned with developmental patterns, and instead sought to understand the character of the children's gestures: how deictic, iconic,

---

employ typologies of this general kind at all (e.g. Wootton, 1994; Haviland, 1998), but these studies remain few. Furthermore, while such studies represent valuable research, they do not address issues relating to preparing a corpus of data for quantitative treatment. This requires some sort of typology. When it comes to studies that have nevertheless tried to apply a typology of this non-mutually-exclusive kind to children's gestures, I am only aware of one, and that is Zlatev & Andr n (2009) which presented an earlier version of the typology employed in this thesis.



and conventionalized aspects occur together in various sorts of gestures; how the more specific meaning of children's gestures emerges from an interplay between gestural form, coordination with speech, previous utterances — and other factors.

Chapter 6 focused on deictic aspects in children's gestures, and in particular on various forms of pointing. There, I demonstrated how children may employ different hand-shapes in pointing: some conventionalized and others more iconically motivated. Furthermore, I demonstrated how pointing gestures may be organized in ways that are sensitive to the overall activity: responding to a previous utterance in certain ways, and eliciting responses from interlocutors (parents). In other words, pointing gestures are typically not performed simply to refer to things as such, but to refer to things in certain ways and for certain purposes, in a manner relevant for the overall activity. I also investigated pointing gestures involving the TRACING of a shape. Interestingly, all of these appeared in contexts where the parent and the child look at, or produce, pictures. It was highlighted how actual acts of drawing are often coordinated with speech in similar ways as TRACING gestures, both by children and by their parents. There was also an analysis of pointing gestures performed with a held-in-hand object. It was shown that objects can be exploited in different ways in acts of pointing: sometimes the cultural identity of the object is used for purposes of signification, and in other cases the cultural identity is disregarded. In that case it is rather the physical characteristics of the object that are exploited. Finally, I argued that directedness towards the concrete environment is present in the majority of the children's gestures — including also gestures that may primarily tend to be perceived as iconic or conventionalized, rather than deictic. With respect to this, a distinction between three different kinds of gesture spaces was proposed: *action space* (where the gesture incorporates objects in the concrete material environment), *supporting space* (where the gesture is directed towards the material environment, but the latter is incorporated in the signification as something different from what it actually is), and *abstract space* (a space that disregards the actual surroundings of the gesture). The major finding was that most gestures in children between 18 and 30 months of age seem to be oriented to action space and supporting space, and much less so to abstract space.

In Chapter 7, iconic aspects of the children's gestures were submitted to scrutiny. There I concentrated on gestures with ACTION\_BASED and INDIRECT aspects. First, I offered a theoretical discussion of iconicity in gesture. What provides for the effortless recognition of a resemblance between expression and content/referent in many iconic gestures? Is it the gestural form itself? Is it the spoken utterance that co-occurs with the gesture? Is it mirror neurons or social conventions? I offered

the following answers: (A) the factors involved in constraining the possible interpretations of iconic gestures can vary, so that one cannot account for iconicity in a uniform manner, and (B) typified and conventionalized schemes of interpretation may be more important for iconicity in children's gestures than usually acknowledged, but conventionality should not be seen as a sufficient explanation. Examples were provided of both ACTION\_BASED and INDIRECT aspects of gesture, and hybrid variants of these. The vast majority of the iconic gestures analyzed involved directedness towards the environment, mentioned above: either in the form of action space or supporting space.

Chapter 8 dealt specifically with conventionalized aspects of the children's gestures. The discussion was restricted to strongly conventionalized gestures (those commonly referred to as *emblems*). The first part of the chapter focused on NODDING and HEAD\_SHAKE: the two most common emblem-like gestures in the data. It was shown how they first seem to function mainly as response "morphemes", and how they later develop into more flexible kinds of communicative resources that may serve to express affirmation or negation in more abstract ways. Since NODDING and HEAD\_SHAKES are articulated using the head, they are inclined to be performed in concert with other, manual, gestures. However, the tendency to combine NODDING and HEAD\_SHAKES with other gestures is not present from the start. Rather, an abrupt increase occurred around 22 months of age.

Chapter 9 concentrated on how communicative appearances may emerge from the manual handling of objects. I presented a lengthy and quite detailed exposé of communicative characteristics that may become foregrounded in different phases of object handling. Instead of treating the acts under consideration in terms of a pre-specified set of actions such as GIVE, SHOW, PLACE, and so forth, the analysis showed how various communicative action gestalts appear in the manual handling of objects within the overall activity.

One may also ask what kinds of gestures typical for older children and adults were *not* found in the children's repertoires. One such type are *beat* gestures (McNeill & Levy, 1982; McNeill, 2005), also known as *batons* (Ekman & Friesen, 1969). These are gestures that serve to segment, accentuate, or emphasize a particular syllable, word or phrase in the discourse. Children of the ages studied in this thesis rarely, if ever, produce clear-cut beats. Some instances of single nods seem to mark emphasis. However, these can still be seen as NODDING, reduced in form, but still compatible with the meaning of affirmation.<sup>2</sup>

---

<sup>2</sup>Özçalışkan & Goldin-Meadow (2009, p. 194) studied children between 14–34 months, and

## 10.2. WHAT ARE THE CHANGES IN THE GESTURAL REPERTOIRE OVER TIME, AS THE CHILDREN GROW OLDER?

---

Another kind of gesture not found in the data are iconic gestures that McNeill (1992, 2005) has characterized as *metaphoric*. McNeill (2005, p. 39) provides the example of a speaker that “appears to be holding an object, as if presenting it, yet the meaning is not presenting an object but rather that she is holding an ‘idea’ or ‘memory’ or some other abstract ‘object’ in her hand.” It is known that children understand more and more of creative (non-conventionalized) types of metaphor in spoken language much later than non-metaphorical use of the same words and expressions (Winner, 1988; Waggoner & Palermo, 1989). It is noteworthy that the same overall developmental pattern seems to hold true for gesture development too.

While pointing is one of the first kinds of gestures to emerge, there are more abstract uses of pointing that come late in development. The use of *abstract pointing*, where the gesture is not used to refer to a target in the present surroundings, but is rather used to create an imaginary target (cf. McNeill et al., 1993, p. 5), is one such case. McNeill (1992) argues that it is not until around 12 years of age that this sort of pointing gestures appear in the repertoires of children. Indeed, there are no instances of abstract pointing in the data analyzed. There are a few instances where non-present objects are pointed out, especially towards the end of the studied period, but all of those are still concretely anchored to the present situation: the gesture is performed in direction to a location where an object has been located previously, or where an object is supposed to be. That is, there are a number of pointing gestures that can be seen as taking place in *supporting space*, but not in *abstract space*.

## 10.2 What are the changes in the gestural repertoire over time, as the children grow older?

Developmental patterns were mainly treated in Chapter 5. These were investigated from the point of view of how some aspect of the children’s gestures varied at dif-

---

concluded that “the incidence of beat gestures was extremely rare in the data; this category was therefore excluded from the analyses”. Interestingly, Nicoladis et al. (1999, p. 520) found at least a few beat gestures in children, but not until the children’s MLU reached 2.7 or more. Other researchers have also related the use of beat gestures to more complex forms of discourse structures (Jancovic et al., 1975; Silverstein, 1984; McNeill, 1992, p. 93; Colletta et al., 2010). Hence, beat gesture, or the discourse marking function of gesture as a refined aspect of gesticulation on its own, does indeed seem to be a later attainment, related to the contexts of relevance that comes into existence when the children are able to produce longer utterances and more complex narrative structures. Possibly, the emergence of beat gestures could be understood as a refinement and purification of the more general discourse marking aspect of gestures.

ferent ages.

How many gestures did the children perform overall? The average gesture rate for the children in the whole corpus was 0.51 gestures per multimodal utterance and 7 gestures per minute. The pattern looked differently across ages when “gesture rate” was treated in terms of gestures per minute and in terms of gestures per multimodal utterance. Therefore, I suggested that gesture researchers ought to make an informed choice between these two measures. Roughly speaking, there were more gestures performed during the first half of the age range (18–24 months) than during the latter half (24–30 months), but for the “gestures per minute” (GPM) measure there was also a substantial rise from a lower rate to a higher rate between 18 and 21 months. This did not appear at all when gesture rate was measured with the “gestures per multimodal utterance” (GPU) measure.

Gesture+speech utterances were found to be the dominant utterance form up to 24 months. After that, speech-only utterances became most common. Gesture-only utterances were rare in comparison to both gesture+speech utterances and speech-only utterances.

Gestures with deictic aspects were by far the most common kind: around 5 times as common as gestures with iconic aspects and emblems (highly conventionalized gestures). Most gestures with deictic aspects were performed at 21 months. After that there was a steady decrease until the end of the period studied (30 months). For iconic gestures, the pattern was the reverse, as the rate of gestures with iconic aspects increased from around 20 months to 28 months. After that, there was a sharp drop in the number of iconic gestures again, between 28 and 30 months. The rate of emblems over time turned out to be close to equivalent to the rate of *NODDING* and *HEAD\_SHAKES*, since these two emblems were by far the most common ones in the children’s repertoires. For these two gestures, there was no clear overall developmental trend, but there seemed to be a slight increase with age.

For gestures with deictic aspects, index-finger pointing was about as frequent as the category *OTHER* (consisting in acts like *GIVE*, *SHOW*, and *PLACE*). Pointing gestures without an index-finger pointing hand-shape were much less frequent. Gestures with *ACTION\_BASED* iconic aspects were present throughout the period, with a maximum rate at 28 months. *INDIRECT* iconic aspects seemed to emerge quite suddenly around 24 months.

Gestures with deictic aspects were most frequently performed in coordination with speech. That is perhaps not surprising, since the potential for gestures to be explicitly directed towards the environment is absent in the speech modality. The sec-

ond most common kind of gestures to be coordinated with speech were emblems. Gestures with iconic aspects were least likely to be coordinated with speech, even though all kinds of gestures were much more often coordinated with speech than performed silently. Gestures handling (real) objects (called “object-gestures” in this thesis) were about as common as empty-handed gestures. Gestures with indirect iconic aspects were extreme in this respect, since they were almost exclusively performed with objects.

The longitudinal analyses were designed with the primary goal of providing a view on development with fine temporal resolution. This sort of analysis revealed that there were three points in time during the children’s development when there were many developmental changes going on at the time. These three points in time were called Transition period #1, #2, and #3 respectively. Transition period #1 occurred at around 20–21 months, and was characterized by a sudden increase in the number of communicative acts per minute (both in gesture and speech) and a peak in the number of deictic gestures. Transition period #2, around 24 months, corresponded to the point in time when mean length of (spoken) utterance (MLU) reached 2. At this point speech-only utterances became more common than gesture+speech utterances and the use of pointing gestures decreased. Gestures with INDIRECT iconic aspects emerged. Transition period #3, at around 28 months, corresponded to the point in time when mean length of utterance (MLU) reached 3 and there was a sharp increase in 3- and 4-word utterances (but a decrease in 2-word utterances). At this point too there was a decrease in the rate of pointing gestures.

To sum up, the major finding was that the changes in various gesture rates seem to be strongly related to changes in the organization of speech — highlighting the symbiotic relationship between gesture and speech in the communicative ecology.

## 10.3 The lower limit of gesture: action and gesture

A dimensional approach to gesture was employed in this thesis, not only with respect to deictic, iconic, and conventionalized semiotic motivations, but also with respect to different levels of communicative explicitness and semiotic complexity. Treating these phenomena in terms of different levels, rather than as binary distinctions, made it possible to make a more fine-grained analysis of various actions with gestural qualities than when “gestuality” is treated as an either-or issue, be it for methodological or theoretical reasons. This approach rested on three major conceptual conclusions.

First, a distinction was made between communicative explicitness and semiotic complexity. This dissociation makes it possible to see that an action may be clearly communicative without being semiotically complex, and semiotically complex without being communicative. This opens up for a comparative semiotic treatment of different configurations of gestural qualities in actions — many of which typically fall outside the scope of gesture studies — and to highlight similarities and differences between these actions.

Second, the concept of *typification* was a key to many of the analyses. It served to acknowledge a level of meaning that lies below that of semiotic signs. In semiotic signs, there is differentiation between expression and content/referent. In typified actions, this is not the case. Typified actions are rather those that *count-as* tokens of a type, rather than as representations of that type. Nevertheless, there are many similarities between the ways in which typified actions and semiotic signs may be employed. Both may be explicitly communicative and both may be coordinated with speech in similar ways. Whereas there is much research that discusses the relation between gesture and signed language (i.e., the upper limit), there is little that applies an equally comparative semiotic approach to the relation between those actions that clearly qualify as gesture and those that are somehow “too simple” for that.

Third, an action may involve the handling of an object and still qualify as both explicitly communicative and as a semiotic sign. Just about every chapter in the thesis reflected this view on what may constitute a “gesture”. The term “object-gesture” was used to refer to such gestures, and it stands in contrast to what is referred to with the term “empty-handed gesture”, which is what research on gesture is more or less exclusively concerned with.

The claim made in this thesis is that the dimensional, or comparative semiotic, approach allows for a better understanding of the relation between action and gesture.

## 10.4 The upper limit of gesture: gesture and language

The children studied in this thesis are not learning a signed language. Therefore the issue of the upper limit is not as directly relevant as the issue of the lower limit. Nevertheless, some themes that were addressed are still of relevance to the question of the relation between gesture and language (cf. Kendon, 2000, 2008). Simply put, it is the question of similarities and differences between gesture and language.

When it comes to similarities between gesture and language, four of the results from the thesis may be highlighted.

First, in Chapter 8 I suggested that some gesture+speech combinations in children are best understood as rote-learned wholes, rather than as a matter of “combining” separate units. This phenomenon was characterized in terms of *multimodal constructions*, and I distinguished between item-based multimodal constructions (which are more fixed in character) and flexible multimodal constructions (where conventionalized gestures are used in more creative and productive ways vis-a-vis spoken utterances). For some of the more frequently performed conventionalized gestures, there seems to be a developmental progression from an earlier stage where some specific gesture is used as an item-based multimodal construction, together with specific words and phrases, to a later stage where the gesture is coordinated in more productive and flexible ways with speech. The idea of item-based constructions is not itself new, but the novelty was to acknowledge that the “units” involved in these constructions may not only reside within a single modality, i.e., speech. They may also be a matter of multimodal units, where some specific configuration of gesture and speech constitute a unit as a whole, rather than a combination of two separate parts. The idea of item-based multimodal constructions implies that the relation between gesture and speech may sometimes, and in some respects, be characterized as being on equal footing.

Second, I have argued for a relatively strong presence of conventionality in children’s gestures, though not in the sense that all of children’s gestures are to be understood as emblems, i.e., as normatively constrained signs. Again, the notion of typification provided a middle ground, serving to acknowledge the existence of typified conventions. Typified conventions are not strong enough to be considered normative, i.e., atypical performances of typified actions are not perceived as incorrect. At the same time, typified conventions are not merely a matter of “normal appearances”. The performance of a typified action is recognizable as a token of a recognizable *type*. On basis of the examples that were analyzed, it was argued that many, if not most, of children’s gestures may be understood as resting on typified conventions. In the case of gestures with ACTION\_BASED iconic aspects, this convention may apply to the level of content. However, in the case of ACTION\_BASED gestures, there is a strong correspondence between the medium used for expression (bodily action) and the medium of the content/referent (also bodily action). This means that the form of the expression in ACTION\_BASED gestures draws on the form of the actions that are signified, including any conventionalized aspects of these actions. Thus, from a developmental point of view, it is valid to say that

many of the ACTION\_BASED gestures rest on familiarity with various conventions and socio-culturally structured practices and that they are not particularly creative in character.

Third, there did not seem to be any opposition between conventionalized aspects of gesture and speech. Even though some of the more strongly conventionalized gestures can be performed without speech, in an autonomous manner, it turned out that even the most conventionalized gestures were in practice more commonly performed together with speech than gestures with iconic aspects. Hence, the label “co-speech gesture” seems to be equally applicable to the conventionalized gestures as to the iconic gestures.

Forth, NODDING and HEAD\_SHAKES may be characterized as morphemes: not only in the sense that they have a conventionalized form that represents a conventionalized meaning, but also because they form a contrasting pair with respect to each other (both on the level of form and on the level of meaning). Nevertheless, this seems to be an exception to the more general “rule” that all, or almost all, other gestures used by the children are not contrastive with respect to each other in this manner.

The main *difference* between children’s use of gesture and spoken language rather seems to lie in the lack of systematicity in their use of gesture. Gesture+gesture combinations were found to be very rare in comparison to word+word combinations, which confirms findings from other studies (see Chapter 5). In addition, very few gesture+gesture combinations qualify as conventionalized gesture+gesture combinations. One of the possible exceptions in this respect is the combination of the GONE gesture and the HEAD\_SHAKE, which was found in the performances both of children and adults.

## 10.5 Final words

In this thesis, I have strived to provide a broad overview of children’s gestures between 18 and 30 months of age. Above all, I hope to have shown that gesture is a very fascinating topic — a topic that is of potential relevance to many scholarly disciplines. I might also add that I do indeed intend to continue the study of children’s gestures; partly because much research remains to be done and many questions remain to be answered, but also because I have been delighted by the positive spirit and genuine curiosity that characterizes the field. Writing this thesis would not have been as fun if there were no other gesture researchers to share the joy with.



# References

- Acredolo, L. & Goodwyn, S. (1985). Symbolic gesturing in language development. *Human Development*, 28, 40–49.
- Acredolo, L. & Goodwyn, S. (1990). Sign language among hearing infants: The spontaneous development of symbolic gestures. In V. Volterra & C. J. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children* (pp. 68–78). Berlin: Springer.
- Adamson, L. B. & Bakeman, R. (1985). Affect and attention: Infants observed with mothers and peers. *Child Development*, 56(3), 582–593.
- Adamson, L. B., Bakeman, R., & Smith, C. B. (1990). Gestures, words, and early object sharing. In V. Volterra & C. J. Erting (Eds.), *From gesture to language in hearing and deaf children* (pp. 31–41). Berlin: Springer.
- Ahlner, Felix & Zlatev, J. (in press). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies*.
- Aitchison, J. (1998). On discontinuing the continuity–discontinuity debate. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 17–28). Cambridge: Cambridge University Press.
- Alibali, M. W. & DiRusso, A. A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, 14(1), 37–56.
- Alibali, M. W. & Don, L. S. (2001). Children’s gestures are meant to be seen. *Gesture*, 1(2), 113–127.
- Allwood, J. (1976). *Linguistic Communication as Action and Cooperation: A study in pragmatics*. PhD thesis, Dept. of Linguistics, University of Gothenburg.
- Allwood, J. & Ahlsén, E. (1999). Learning how to manage communication, with special reference to the acquisition of linguistic feedback. *Journal of Pragmatics*, 31(10), 1353–1389.
- Anderson, D. E. & Reilly, J. S. (1997). The puzzle of negation: How children move from communicative to grammatical negation in ASL. *Applied Psycholinguistics*, 18(4), 411–429.
- Andrén, M. (in press a). Gesture and object manipulation. In C. Sinha, C. Rodríguez, & J. Vang (Eds.), *Companion to Cognitive Semiotics 1: Meaning and materiality*. Peter Lang.
- Andrén, M. (in press b). The organization of children’s pointing stroke endpoints. In G. Stam & M. Ishino (Eds.), *Integrating Gestures: The interdisciplinary nature of gesture*. Amsterdam: John Benjamins.
- Andrén, M. (in press c). The social world within reach: Intersubjective manifestations of action completion. In P. Demuru, R. Fusaroli, & A. Borghi (Eds.), *Companion to Cognitive Semiotics 2: The intersubjectivity of embodiment*. Peter Lang.
- Andrén, M., Sanne, J. M., & Linell, P. (2010). Striking the balance between formality and informality in safety-critical communication: Train traffic control calls. *Journal of Pragmatics*, 42(1), 220–241.

- Andrén, M. & Zlatev, J. (2007). Relations between mimesis and language in ontogeny: Studies in children's bodily communication. Lund University: Centre for Languages and Literature.
- Andrén, M. & Zlatev, J. (accepted). On the intersubjective construal of meaning: A multimodal analysis of Swedish children's use of *så*. In E. Zima & G. Brône (Eds.), *Interactional Discourse in Cognitive Linguistics: Theory, methods, and application*.
- Anisfeld, M., Rosenberg, E. S., Hoberman, M. J., & Gasparini, D. (1998). Lexical acceleration coincides with the onset of combinatorial speech. *First Language*, 18(53), 165–184.
- Anscombe, G. E. M. (1957). *Intentions*. Oxford: Blackwell.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(2), 105–167.
- Arbib, M. A. (Ed.). (2006). *Action to Language via the Mirror Neuron System*. Cambridge: Cambridge University Press.
- Arbib, M. A., Oztop, E., & Zukow-Goldring, P. (2005). Language and the mirror system: A perception/action based approach to communicative development. *Cognition, Brain, Behavior*, IX(3), 239–272.
- Arendsen, J. (2009). *Seeing Signs: On the appearance of manual movements in gestures*. PhD thesis, Industrial Design, Delft University of Technology.
- Ash, M. G. (1995). *Gestalt Psychology in German Culture, 1890-1967: Holism and the quest for objectivity*. Cambridge: Cambridge University Press.
- Auer, P. (2007). Syntax als prozess. In H. Hausendorf (Ed.), *Gespräch als Prozess: Linguistische Aspekte der Zeitlichkeit verbaler Interaktion* (pp. 95–142). Tübingen: Narr.
- Bach, K. & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge: MIT Press.
- Bakeman, R. & Gottman, J. M. (1997). *Observing Interaction: An introduction to sequential analysis* (2nd ed.). Cambridge: Cambridge University Press.
- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106(3), 1382–1407.
- Barker, R. G. (1978). *Habitats, Environments, and Human Behavior*. San Francisco: Jossey-Bass.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and Theory of Mind*. Cambridge: MIT Press.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, 21(1), 37–46.
- Bartok, P. J. (2005). Brentano's intentionality thesis: Beyond the analytic and phenomenological readings. *Journal of the History of Philosophy*, 43(4), 437–460.
- Bates, E., Bretherton, I., Camaioni, L., & Volterra, V. (Eds.). (1979). *The Emergence of Symbols: Cognition and communication in infancy*. New York: Academic Press.
- Bates, E., Bretherton, I., Shore, C., & McNew, S. (1983). Names, gestures, and objects: The role of context in the emergence of symbols. In K. Nelson (Ed.), *Children's language*, vol. IV (pp. 59–123). Hillsdale, NJ: Lawrence Erlbaum.
- Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*, 21(3), 205–226.

- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21, 85–123.
- Bavelas, J., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, 3, 566–580.
- Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction*, 27(3), 201–221.
- Bavelas, J. B. (2007). Face-to-face dialogue as a micro-social context. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language* (pp. 127–146). Amsterdam: John Benjamins Publishing Company.
- Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, 15, 469–489.
- Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, 8(6), 492–499.
- Bentham, J. (1907 [1780]). *An Introduction to the Principles of Morals and Legislation*. Reprint of 1823 edition. Oxford: Clarendon Press.
- Berger, K. W. & Popelka, G. R. (1971). Extra-facial gestures in relation to speech-reading. *Journal of Communication Disorders*, 3(4), 302–308.
- Berglund, E. (1999). *Early Communicative and Language Development in Swedish Children: Methods, results, clinical implications and prospects for the future*. PhD thesis, Uppsala University, Dept. of Psychology.
- Berglund, E., Eriksson, M., & Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scandinavian Journal of Psychology*, 46, 485–491.
- Bergman, B. (1982). Sign typology. In *Forskning om Teckenspråk*, vol. XI (pp. 1–21). Stockholm University: Department of Linguistics.
- Bickerton, D. (2003). Symbol and structure: A comprehensive framework for language evolution. In M. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 77–93). Oxford: Oxford University Press.
- Blake, J. (2000). *Routes to Child Language: Evolutionary and developmental precursors*. Cambridge, UK: Cambridge University Press.
- Blake, J., McConnell, S., Horton, G., & Benson, N. (1992). The gestural repertoire and its evolution over the second year. *Early Development and Parenting*, 1(3), 127–136.
- Blake, J., O'Rourke, P., & Borzellino, G. (1994). Form and function in the development of pointing and reaching gestures. *Infant Behavior and Development*, 17(2), 195–203.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge: Cambridge University Press.
- Blumer, H. (1969). *Symbolic Interactionism: Perspective and method*. Berkeley: University of California Press.
- Bouissac, P. (2008). The study of metaphor and gesture: A critique from the perspective of semiotics. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 277–282). Amsterdam: John Benjamins.
- Bouissac, P. (ms. in prep.). The multimodality of gestures: Problems of definition, description, and interpretation.
- Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.

- Brentano, F. (1995 [1874]). *Psychology from an Empirical Standpoint* (2nd ed.). Translated by A. Rancurello, et. al. London: Routledge.
- Bretherton, I., O'Connell, B., Shore, C., & Bates, E. (1984). The effect of contextual variation on symbolic play: Development from 20 to 28 months. In I. Bretherton (Ed.), *Symbolic play: The development of social understanding* (pp. 271–298). New York: Academic Press.
- Brewer, W. D. (1951). Patterns of gesture among the Levantine Arabs. *American Anthropologist*, 53(2), 232–237.
- Brown, R. (1973). *A first language: The early stages*. Cambridge: Harvard University Press.
- Bruner, J. S. (1978a). Acquiring the use of language. *Canadian Journal of Psychology*, 32(4), 204–218.
- Bruner, J. S. (1978b). Learning how to do things with words. In J. S. Bruner & A. Garton (Eds.), *Human Growth and Development* (pp. 62–84). Oxford: Oxford University Press.
- Bruner, J. S. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.
- Bruner, J. S. & Garton, A. (Eds.). (1978). *Human Growth and Development*. Oxford: Oxford University Press.
- Bräm, P. B. & Bräm, T. (2000). A pilot study of the expressive gestures used by classical orchestra conductors. In K. Emmorey & H. Lane (Eds.), *The Signs of Language Revisited* (pp. 143–167). Hillsdale, NJ: Erlbaum.
- Bräm, P. B. & Bräm, T. (2004). Expressive gesture used by classical orchestra conductors. In C. Müller & R. Posner (Eds.), *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998* (pp. 127–144). Berlin: Weidler Buchverlag.
- Bulwer, J. (1974 [1654]). *Chirologia: Or the natural language of the hand and Chironomia: Or the art of manual rhetoric*. Carbondale and Edwardsville: Southern Illinois University Press.
- Butcher, C. & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and Gesture* (pp. 235–257). Cambridge: Cambridge University Press.
- Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 9–33). Mahawa, NJ: Lawrence Erlbaum Associates.
- Butterworth, G. & Grover, L. (1989). Joint visual attention, manual pointing, and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Motor representation and control*, vol. 13 of Attention and performance (pp. 605–624). Hillsdale, NJ: Erlbaum.
- Bühler, K. (1982 [1934]). The deictic field of language and deictic words. In R. J. Jarvella & W. Klein (Eds.), *Speech, Place and Action: Studies in deixis and related topics* (pp. 297–313). Chichester: John Wiley & Sons.
- Cadoz, C. & Wanderley, M. M. (2000). Gesture-music. In M. Wanderley & M. Battier (Eds.), *Trends in Gestural Control of Music, IRCAM* (pp. 71–93). Paris: Centre Pompidou.
- Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington: Indiana University Press.
- Call, J. & Tomasello, M. (2007). *The Gestural Communication of Apes and Monkeys*. Mahawa, NJ: Lawrence Erlbaum Associates.
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15(8), 1243–1249.

- Camaioni, L., Aureli, T., Bellagamba, F., & Fogel, A. (2003). A longitudinal examination of the transition to symbolic communication in the second year of life. *Infant and Child Development*, *12*, 1–26.
- Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, *5*(1-2), 155–177.
- Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gesture and words during the transition to two-word speech. *Journal of Child Language*, *23*, 645–673.
- Carey, D. P., Harvey, M., & Milner, A. D. (1996). Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia. *Neuropsychologia*, *34*(5), 329–338.
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Sixteen-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, *21*, 315–330.
- Caselli, M. C. (1990). Communicative gesture and first words. In V. Volterra & C. J. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children* (pp. 56–67). Berlin: Springer.
- Catmur, C., Walsh, V., & Heyes, C. M. (2007). Sensorimotor learning configures the human mirror system. *Current Biology*, *17*, 1527–1531.
- Christensen, L. (2003). The acquisition of tense. In G. Josefsson, C. Platzack, & G. Håkansson (Eds.), *The Acquisition of Swedish Grammar* (pp. 31–74). Amsterdam and Philadelphia: John Benjamins.
- Chu, M. & 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.
- Cienki, A. & Müller, C. (Eds.). (2008). *Metaphor and Gesture*. Amsterdam: John Benjamins.
- Ciolek, T. M. & Kendon, A. (1980). Environment and the spatial arrangement of conversational encounters. *Sociological Inquiry*, *50*(3–4), 237–272.
- Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 243–268). Mahawa, NJ: Lawrence Erlbaum Associates.
- Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse studies*, *7*(4-5), 507–525.
- Clark, R. A. (1978). The transition from action to gesture. In A. Lock (Ed.), *Action, Gesture and Symbol* (pp. 231–257). London: Academic Press.
- Cochet, H. & Vauclair, J. (in press). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*.
- Cole, J. (1998). *About face*. Cambridge, MA: MIT Press.
- Cole, J. (2001). Empathy needs a face. *Journal of Consciousness Studies*, *8*, 51–68.
- Cole, J. (2009). Impaired embodiment and intersubjectivity. *Phenomenology and the Cognitive Sciences*, *8*(3), 343–360.
- Colletta, J.-M., Pellenq, C., & Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication*, *52*(6), 565–577.
- Condillac, E. B. d. (1771 [1756]). *An Essay on the Origin of Human Knowledge*. Translated by Thomas Nugent. Facsimile Reprint, Gainesville, Fla.: Scholars' Facsimiles and Reprints.
- Corballis, M. C. (2002). *From Hand to Mouth: The origins of languages*. Bloomington: Indiana University Press.
- Corrin, J., Tarplee, C., & Wells, W. H. G. (2001). Interactional linguistics and language development: A conversation analytic perspective on emergent syntax. In M. Selting & C.-K. E. (Eds.), *Studies in Interactional Linguistics* (pp. 199–225). Amsterdam: John Benjamins.

- Coseriu, E. (1985). Linguistic competence: What is it really? *The Modern Language Review*, 80(4), xxv–xxxv.
- Coseriu, E. (1988). *Sprachkompetenz: Grundzüge der Theorie des Sprechens*. Tübingen: Francke.
- Costall, A. (1995). Socializing affordances. *Theory and Psychology*, 5(4), 467–481.
- Creem, S. H. & Proffitt, D. R. (2001). Grasping objects by their handles: A necessary interaction between cognition and actions. *Journal of Experimental Psychology. Human Perception and Performance*, 27(1), 218–228.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, W. (2003). *Typology and Universals* (2nd ed.). Cambridge: Cambridge University Press.
- Csibra, G. (2007). Action mirroring and action understanding: An alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Attention and Performance: Sensorimotor foundations of higher cognition*, vol. 22 of *Attention and performance* (pp. 435–459). Oxford: Oxford University Press.
- Da Cunha Pereira, M. C. & De Lemos, C. (1990). Gesture in hearing mother-deaf child interaction. In V. Volterra & C. J. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children* (pp. 178–186). Berlin: Springer.
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.
- D'Ausilio, A., Craighero, L., & Fadiga, L. (in press). The contribution of the frontal lobe to the perception of speech. *Journal of Neurolinguistics*.
- De Jaegher, H. (2008). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition*, 18(2), 535–542.
- DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, 8(2), 66–70.
- Donald, M. (1991). *Origins of the Modern Mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, 23(2), 283–292.
- Duranti, A. (1991). Four properties of speech-in-interaction and the notion of translocutionary act. In J. Verschueren (Ed.), *Pragmatics at Issue*. Amsterdam: John Benjamins.
- Duranti, A. (2010). Husserl, intersubjectivity and anthropology. *Anthropological Theory*, 10(1), 1–20.
- Durkheim, E. (1965 [1895]). *The Rules of Sociological Method* (8th ed.). New York: Free Press.
- Eckerman, C. O. (1993). Imitation and toddlers' achievement of coordinated action with others. In J. Nadel & L. Camaioni (Eds.), *New Perspectives in Early Communicative Development* (pp. 116–138). London: Routledge.
- Eco, U. (1976). *A Theory of Semiotics*. Bloomington: Indiana University Press.
- Edwards, D. (1997). *Discourse and Cognition*. London: Sage.
- Efron, D. (1941 [1972]). *Gesture, Race and Culture*. Hague: Mouton.
- Eisner, W. (1985). *Comics and Sequential Art*. Tamarac, Florida: Poorhouse Press.
- Ekerot, L.-J. (1988). *Så-konstruktionen i svenskan: Konstruktionstypen "Om vädret tillåter, så genomföres övningen" i funktionellt grammatiskt perspektiv*. PhD thesis, Lund University.
- Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usages and coding. *Semiotica*, 1, 49–98.
- Embree, L. (Ed.). (2004). *Gurwitsch's Relevancy for Cognitive Science*. Dordrecht: Springer.

- Emmorey, K. (2006). The signer as an embodied mirror neuron system: Neural mechanisms underlying sign language and actions. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 110–135). Cambridge: Cambridge University Press.
- Enfield, N. J. (2001). 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture*, 1(2), 185–212.
- Eriksson, M. & Berglund, E. (1998). *Communicative Development in Swedish Children 8–16 Months old: The Swedish Early Communicative Development Inventory — Words and Gestures*, vol. 844 of Reports from the Department of Psychology. Stockholm University.
- Erting, C. J. & Volterra, V. (1990). Conclusion. In V. Volterra & C. J. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children* (pp. 299–303). Berlin: Springer.
- Firbas, J. (1971). On the concept of communicative dynamism in the theory of functional sentence perspective. *Philologia Pragensia*, 8, 135–144.
- Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72, 323–327.
- Fogel, A. & Hannan, T. E. (1985). Manual actions of nine- to fifteen-week-old human infants during face-to-face interactions with their mothers. *Child Development*, 56(5), 1271–1279.
- Franco, F. & Butterworth, G. (1996). Pointing and social awareness: Declaring and requesting in the second year. *Journal of Child Language*, 23, 307–336.
- Fricke, E. (2009). *Grundlagen einer Multimodalen Grammatik des Deutschen: Syntaktische Strukturen und Funktionen*. Berlin, New York: de Gruyter.
- Gallagher, S. (2004). Situational understanding: A Gurwitschian critique of Theory of Mind. In L. Embree (Ed.), *Gurwitsch's Relevancy for Cognitive Science* (pp. 25–44). Dordrecht: Springer.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. New York: Oxford University Press.
- Gallagher, S. (2008). Intersubjectivity in perception. *Continental Philosophy Review*, 41(2), 163–178.
- Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, 3(3-4), 317–333.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593–609.
- Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501.
- Ganger, J. & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, 40(4), 621–632.
- Garfinkel, H. (1952). *The Perception of the Other: A study in social order*. PhD thesis, Harvard University.
- Garfinkel, H. (1964). Studies of the routine grounds of everyday activities. *Social Problems*, 11(2), 225–250.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Garfinkel, H. (2002). *Ethnomethodology's Program: Working out Durkheim's aphorism*. Lanham, MD: Rowman & Littlefield.
- Garfinkel, H. (2006 [1948]). *Seeing Sociologically: The routine grounds of social action*. Boulder: Paradigm Publishers.
- Garfinkel, H. & Sacks, H. (1970). On formal structures of practical actions. In J. C. McKinney & E. A. Tiryakian (Eds.), *Theoretical Sociology: Perspectives and Developments* (pp. 337–366). New York: Appleton-Century-Crofts.

- Garton, A. F. (1986). The production of *this* and *that* by young children. *First Language*, 6(16), 29–39.
- Gelman, S., Coley, J., Rosengren, K., Hartman, E., & Pappas, A. (1998). Beyond labeling: The role of maternal input in the acquisition of richly structured categories. *Monographs of the Society for Research in Child Development*, 63(1), 1–157.
- Gerholm, T. (2007). *Socialization of verbal and nonverbal emotive expressions in young children*. PhD thesis, Department of Linguistics, Stockholm University.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Glenn, P. J., LeBaron, C. D., & Mandelbaum, J. (Eds.). (2003). *Studies in Language and Social Interaction: In Honor of Robert Hopper*. Mahawa, NJ: Lawrence Erlbaum Associates.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday Anchor.
- Goffman, E. (1963). *Behavior in Public Places*. New York: The Free Press.
- Goffman, E. (1971). *Relations in Public: Microstudies of the public order*. New York: Basic Books.
- Goffman, E. (1974). *Frame Analysis*. New York: Harper and Row.
- Golato, A. (2000). An innovative German quotative for reporting on embodied actions: *Und ich so/und er so* 'and i'm like/and he's like'. *Journal of Pragmatics*, 32(1), 29–54.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development*, 71(1), 75–84.
- Goldin-Meadow, S. (2003a). *Hearing gesture: How our hands help us think*. Cambridge, MA: Belknap Press.
- Goldin-Meadow, S. (2003b). *The Resilience of Language: What gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press.
- Goldin-Meadow, S. (2007a). Gesture with speech and without it. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language* (pp. 31–49). Amsterdam: John Benjamins Publishing Company.
- Goldin-Meadow, S. (2007b). Pointing sets the stage for learning language-and creating language. *Child Development*, 78(3), 741–745.
- Goldin-Meadow, S. & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85–107). Mahawa, NJ: Lawrence Erlbaum Associates.
- Goldin-Meadow, S. & Feldman, H. (1975). The creation of a communication system: A study of deaf children of hearing parents. *Sign Language Studies*, 8, 226–236.
- Goldin-Meadow, S. & Iverson, J. M. (1998). Editors' note. In *The Nature and Functions of Gesture in Children's Communication* (pp. 1–7). San Francisco: Jossey-Bass Publishers.
- Goldin-Meadow, S., McNeill, D., & Singleton, J. L. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, 103(1), 34–55.
- Goldin-Meadow, S. & Morford, M. (1985). Gesture in early child language: Studies of deaf and hearing children. *Merrill-Palmer Quarterly*, 31, 145–176.
- Goldin-Meadow, S. & Morford, M. (1990). Gesture in early child language. In V. Volterra & C. J. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children* (pp. 249–262). Berlin: Springer.



- Golinkoff, R. M. & Gordon, L. (1988). What makes communication run? Characteristics of immediate successes. *First Language*, 8(23), 103–124.
- Goodman, N. (1970). Seven strictures on similarity. In L. Foster & J. W. Swanson (Eds.), *Experience and Theory* (pp. 19–29). Amherst: Univ. of Massachusetts Press.
- Goodwin, C. (1995). Co-constructing meaning in conversations with an aphasic man. *Research on Language in Social Interaction*, 28(3), 233–260.
- Goodwin, C. (2003). Pointing as situated practice. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 217–241). Mahawa, NJ: Lawrence Erlbaum Associates.
- Goodwin, C. (2007). Environmentally coupled gestures. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language* (pp. 195–212). Amsterdam: John Benjamins.
- Goodwyn, S. W. & Acredolo, L. P. (1998). Encouraging symbolic gestures: A new perspective on the relationship between gesture and speech. In *The Nature and Functions of Gesture in Children's Communication* chapter 4, (pp. 61–73). San Fransisco: Jossey-Bass Publishers.
- Goudge, T. A. (1965). Peirce's index. *Transactions of the Charles S. Peirce Society*, 1, 52–70.
- Graham, T. A. (1999). The role of gesture in children's learning to count. *Journal of Experimental Child Psychology*, 74(4), 333–355.
- Grassmann, S. & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, 13(1), 252–264.
- Grathoff, R. (Ed.). (1978 [1940-41]). *The Theory of Social Action: The Correspondence of Alfred Schutz and Talcott Parsons*. Bloomington, IN: Indiana University Press.
- Graziano, M. (2009). *Rapporto fra lo sviluppo della competenza verbale e gestuale nella costruzione di un testo narrativo in bambini dai 4 ai 10 anni*. PhD thesis, Unpublished. SESA, Università degli studi "Suor Orsola Benincasa", Napoli & Université Stendhal, Grenoble.
- Gredebäck, G. & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition*, 114(2), 197–206.
- Green, J. R. (1968). *A Gesture Inventory for the Teaching of Spanish*. Philadelphia and New York: Chilton Books.
- Greenfield, P., Reilly, J., Leaper, C., & Baker, N. (1985). The structural and functional status of single-word utterances and their relationship to early multi-word speech. In M. D. Barrett (Ed.), *Children's Single-word Speech* (pp. 233–267). Chichester: Wiley.
- Greenfield, P. & Smith, J. (1976). *The Structure of Communication in Early Development*. New York: Academic Press.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–388.
- Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). One-year-olds' understanding of nonverbal gestures directed to a third person. *Cognitive Development*, 24(1), 23–33.
- Guidetti, M. (2002). The emergence of pragmatics: Forms and functions of conventional gestures in young French children. *First Language*, 22(3), 265–286.
- Guidetti, M. (2005). Yes or no? How young French children combine gestures and speech to agree and refuse. *Journal of Child Language*, 32(4), 911–924.
- Guidetti, M. & Nicoladis, E. (2008). Introduction to special issue: Gestures and communicative development. *First Language*, 28(2), 107–115.
- Gullberg, M. (2009). Gestures and the development of semantic representations in first and second language acquisition. *Aile...Lia (former Aile)*, 1, 117–139.

- Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language*, 28(2), 200–236.
- Gullberg, M. & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, 7(1), 35–63.
- Gullberg, M. & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, 33(4), 251–277.
- Habermas, J. (1991 [1981]). *Reason and the Rationalization of Society*, vol. 1 of *The Theory of Communicative Action*. Translated by T. McCarthy. Cambridge: Polity Press.
- Hall, E. T. (1963). A system for the notation of proxemic behaviour. *American Anthropologist*, 65, 1003–1026.
- Hannan, T. E. (1992). An examination of spontaneous pointing in 20 to 50-month-old children. *Perceptual and Motor Skills*, 74, 651–658.
- Harding, C. G. (1983). Setting the stage for language acquisition: Communication development in the first year. In R. M. Golinkoff (Ed.), *The Transition from Prelinguistic to Linguistic Communication*. Hillsdale, N.J.: Erlbaum.
- Harrison, S. (2009a). The expression of negation through grammar and gestures. In J. Zlatev, M. Andrén, M. Johansson Falck, & C. Lundmark (Eds.), *Studies in Language and Cognition* (pp. 421–435). Newcastle: Cambridge Scholars.
- Harrison, S. (2009b). *Grammar, Gesture, and Cognition: The case of negation in English*. PhD thesis, Université de Bordeaux 3.
- Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, 3(1), 3–45.
- Haviland, J. B. (1998). Early pointing gestures in Zinacantan. *Journal of Linguistic Anthropology*, 8(2), 162–196.
- Hay, D. F. & Murray, P. (1982). Giving and requesting: Social facilitation of infants' offers to adults. *Infant Behavior and Development*, 5(2–4), 301–310.
- Heath, C. (1992). Review of 'Conducting Interaction: Patterns of behavior in focused encounters' by Adam Kendon. *American Anthropologist*, 94(3), 705–706.
- Heath, C. (2002). Demonstrative suffering: The gestural (re)embodiment of symptoms. *Journal of Communication*, 3, 597–616.
- Heath, C., Hindmarsh, J., & Luff, P. (1999). Interaction in isolation: The dislocated world of the London underground train driver. *Sociology*, 33(3), 555–575.
- Heidegger, M. (1962 [1927]). *Being and Time*. Translation of the German original. New York: Harper & Row.
- Heider, F. & Heider, G. M. (1941). Studies in the psychology of the deaf, no. 2. *Psychological Monographs*, 53(5), 1–153.
- Heine, B. & Kuteva, T. (2007). *The Genesis of Grammar: A reconstruction*. Oxford: Oxford University Press.
- Hendriks-Jansen, H. (1996). *Catching Ourselves in the Act: Situated activity, interactive emergence, evolution, and human thought*. Cambridge, MA: Harvard University Press.
- Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.
- Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, 34(4), 575–583.
- Hirsch, R. (1993). The Gestalt Hypothesis: On the form and meaning of gestures. *Semio Nordica*, 1-2, 17–32.

- Hirsch, R. (1994). Gesture universals: Material and formal constraints. In T. A. Sebeok & J. Umiker-Sebeok (Eds.), *Advances in Visual Semiotics* (pp. 475–490). Berlin: Mouton de Gruyter.
- Hirsch, R. (1996). The act of speaking: Spoken language and gesture in the determination of definiteness of intention. In C. Pankow (Ed.), *Indexicality: Papers from the symposium Indexikala Tecken*, SSKKII Report 9604 (pp. 14–30). University of Gothenburg.
- Hirsch, R. (2009). The hand is quicker than the mind. In J. Zlatev, M. Andrén, M. Johansson Falck, & C. Lundmark (Eds.), *Studies in Language and Cognition*. Newcastle: Cambridge Scholars.
- Hochberg, J. & Brooks, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *American Journal of Psychology*, 75(4), 624–628.
- Hockett, C. (1966). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of Language* (2 ed.). (pp. 1–29). Cambridge, MA: MIT Press.
- ten Holt, G. A., van Doorn, A. J., de Ridder, H., Reinders, M. J. T., & Hendriks, E. A. (2009). Which fragments of a sign enable its recognition? *Sign Language Studies*, 9(2), 211–239.
- Husain, F. T., Patkin, D. J., Thai-Van, H., Braun, A. R., & Horwitz, B. (2009). Distinguishing the processing of gestures from signs in deaf individuals: An fMRI study. *Brain Research*, 1276, 140–150.
- Husserl, E. (1983 [1913]). *Ideas pertaining to a pure phenomenology and to phenomenological philosophy: First book*. Translated by F. Kersten. The Hague: Martinus Nijhoff.
- Ivarsson, J. (2010). Developing the construction sight: Architectural education and technological change. *Visual Communication*, 9(2), 171–191.
- Iverson, J. M. (1998). Gesture when there is no visual model. In *The Nature and Functions of Gesture in Children's Communication* chapter 6, (pp. 89–100). San Francisco: Jossey-Bass Publishers.
- Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, 9, 23–43.
- Iverson, J. M., Capirci, O., Volterra, V., & Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. *First Language*, 28(2), 164–181.
- Iverson, J. M. & Fagan, M. K. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Development*, 75, 1053–1066.
- Iverson, J. M. & Goldin-Meadow, S. (Eds.). (1998). *The Nature and Functions of Gesture in Children's Communication*. San Francisco: Jossey-Bass Publishers.
- Iverson, J. M. & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 368–371.
- Iverson, J. M. & Thelen, E. (1999). Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6, 19–40.
- Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19(18), 1543–1548.
- Jakobson, L. S., Archibald, Y. M., Carey, D. P., & Goodale, M. A. (1991). A kinematic analysis of reaching and grasping movements in a patient recovering from optic ataxia. *Neuropsychologia*, 29(8), 803–809.
- Jakobson, R. (1972). Motor signs for 'yes' and 'no'. *Language in Society*, 1, 91–96.
- Jancovic, M., Devoe, S., & Wiener, M. (1975). Age-related changes in hand and arm movements as nonverbal communication: Some conceptualizations and an empirical exploration. *Child Development*, 46(4), 922–928.

- Jeannerod, M. (1996). The neural control of object-oriented actions. In F. Lacquaniti & P. Viviani (Eds.), *Neural Bases of Motor Behaviour* (pp. 41–76). Dordrecht: Kluwer Academic Publishers.
- Jeannerod, M., Decety, J., & Michel, F. (1994). Impairment of grasping movements following a bilateral posterior parietal lesion. *Neuropsychologia*, 32(4), 369–380s.
- Johnson, H., Ekman, P., & Friesen, W. (1975). Communicative body movements: American emblems. *Semiotica*, 15, 335–353.
- Johnson, M. (1987). *The Body in the Mind*. Chicago: University of Chicago Press.
- Johnson-Frey, S. H., McCarty, M. E., & Keen, R. (2004). Reaching beyond spatial perception: Effects of intended future actions on visually guided prehension. *Visual Cognition*, 11, 371–399.
- Jones, S. E. & Zimmerman, D. (2003). A child's point and the achievement of intentionality. *Gesture*, 3(2), 155–185.
- Josefsson, G. & Platzack, C. (2003). Introduction. In G. Josefsson, C. Platzack, & G. Håkansson (Eds.), *The Acquisition of Swedish Grammar* (pp. 1–21). Amsterdam and Philadelphia: John Benjamins.
- Josefsson, G., Platzack, C., & Håkansson, G. (Eds.). (2003). *The acquisition of Swedish grammar*. Amsterdam and Philadelphia: John Benjamins.
- Junefelt, K. (1987). *Blindness and Child-Adjusted Communication*. PhD thesis, Stockholm University: Department of Nordic Languages.
- Keevallik, L. (2005). The deictic *nii* 'so, in this way' in interaction. In D. Monticelli, R. Pajusalu, & A. Treikelder (Eds.), *From utterance to uttering and vice versa: Multidisciplinary views on deixis* (pp. 109–126). Tartu University: Department of Modern Languages.
- Keevallik, L. (2009). Marking boundaries between activities: The particle *nii* in Estonian. *Research on Language and Social Interaction*.
- Kendon, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A. W. Siegman & B. Pope (Eds.), *Studies in Dyadic Communication* (pp. 177–210). Elmsford, NY: Pergamon Press.
- Kendon, A. (1978). Differential perception and attentional frame in face-to-face interaction: Two problems for investigation. *Semiotica*, 24(3-4), 305–315.
- Kendon, A. (1979). Some emerging features of face-to-face interaction studies. *Sign Language Studies*, 22, 7–22.
- Kendon, A. (1980a). A description of a deaf-mute sign language from the Enga Province of Papua New Guinea with some comparative discussion. Part II: The semiotic functioning of Enga signs. *Semiotica*, 32(1/2), 81–117.
- Kendon, A. (1980b). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship between verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.
- Kendon, A. (1984). Did gesture have the happiness to escape the curse at the confusion of Babel? In A. Wolfgang (Ed.), *Nonverbal Behavior: Perspectives, Applications, Intercultural Insights* (pp. 75–114). Lewiston, New York: C. J. Hogrefe.
- Kendon, A. (1985a). Some reasons for studying gesture. *Semiotica*, 62(1/2), 3–28.
- Kendon, A. (1985b). Some uses of gesture. In D. Tannen & M. Saviile-Troike (Eds.), *Perspectives on Silence* (pp. 215–234). Norwood, NJ: Ablex Publishing Corporation.
- Kendon, A. (1990). *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

- Kendon, A. (1992). The negotiation of context in face-to-face interaction. In *Rethinking Context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.
- Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, 23(3), 247–279.
- Kendon, A. (1996). An agenda for gesture studies. *Semiotic Review of Books*, 7(3), 8–12.
- Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and Gesture* (pp. 47–63). Cambridge: Cambridge University Press.
- Kendon, A. (2002). Some uses of the head shake. *Gesture*, 2(2), 147–182.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kendon, A. (2007). Some topics in gesture studies. In A. Esposito, M. Bratanic, E. Keller, & M. Marinaro (Eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue* (pp. 3–19). Amsterdam: IOS Press.
- Kendon, A. (2008). Some reflections on the relationship between ‘gesture’ and ‘sign’. *Gesture*, 8(3), 348–366.
- Kendon, A. & Versante, L. (2003). Pointing by hand in “Neapolitan”. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 109–137). Mahawa, NJ: Lawrence Erlbaum Associates.
- Kishimoto, T., Shizawa, Y., Yasuda, J., Hinobayashi, T., & Minami, T. (2007). Do pointing gestures by infants provoke comments from adults? *Infant Behavior and Development*, 30(4), 562–567.
- Kita, S. (1993). *Language and thought interface: A study of spontaneous gestures and Japanese mimetics*. PhD thesis, Department of Psychology and Department of Linguistics, University of Chicago.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge: Cambridge University Press.
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes*, 24(2), 145–167.
- Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer Interaction* (pp. 23–35). Berlin: Springer.
- Kita, S. & Ide, S. (2007). Nodding, Aizuchi, and final particles in Japanese conversation: How conversation reflects the ideology of communication and social relationships. *Journal of Pragmatics*, 39(7), 1242–1254.
- Kita, S. & Lausberg, H. (2008). Generation of co-speech gestures based on spatial imagery from the right-hemisphere: Evidence from split-brain patients. *Cortex*, 44(2), 131–139.
- Kobayashi, H. & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, 40(5), 419–435.
- Kochan, A. (2008). The acquisition of negation: A socio-pragmatic study of a bilingual child. Master’s thesis, Ecole Normale Superieure.
- Koffka, K. (1938 [1915]). On the foundation of the psychology of perception. A reply to V. Benussi. In W. D. Ellis (Ed.), *A Source Book of Gestalt Psychology* (pp. 371–378). London: Routledge & Kegan Paul.
- Köhler, W. (1938 [1920]). Physical Gestalten. In W. D. Ellis (Ed.), *A Source Book of Gestalt Psychology* (pp. 17–54). London: Routledge & Kegan Paul. .

- LaBarre, W. (1947). The cultural basis of emotions and gestures. *Journal of Personality*, 16(1), 49–69.
- Ladewig, S. H. (2010). “It has a certain [gesture]”: Syntactic integration of gestures into speech. Poster presented at the 4th conference of the ISGS. Frankfurt/Oder, Germany.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Langacker, R. W. (2004). Remarks on nominal grounding. *Functions of Language*, 11(1), 77–113.
- Leavens, D. & Hopkins, W. (1999). The whole-hand point: The structure and function of pointing from a comparative perspectives. *Journal of Comparative Psychology*, 113(4), 417–425.
- LeBaron, C. & Streeck, J. (2000). Gesture, knowledge, and the world. In D. McNeill (Ed.), *Language and Gesture* (pp. 118–138). Cambridge: Cambridge University Press.
- Lempers, J. D. (1979). Young children’s production and comprehension of nonverbal deictic behaviors. *Journal of Genetic Psychology*, 35, 93–102.
- Leontiev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The Concept of Activity in Soviet Psychology* (pp. 37–71). Armonk, NY: Sharpe.
- Lerner, G. H. & Zimmerman, D. H. (2003). Action and the appearance of action in the conduct of very young children. In C. D. LeBaron, J. S. Mandelbaum, & P. J. Glenn (Eds.), *Studies in Language and Social Interaction: In honor of Robert Hopper* (pp. 441–457). Mahawa, NJ: Lawrence Erlbaum Associates.
- Leslie, A. (1987). Pretense and representation: The origins of “theory of mind”. *Psychological Review*, 94(4), 412–426.
- Leung, E. H. & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology*, 17(2), 215–220.
- Levelt, W. J. M. (1980). On-line processing constraints on the properties of signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and Spoken Language: Biological constraints on linguistic form* (pp. 141–160). Weinheim: Chemie.
- Levelt, W. J. M. (1981). The speaker’s linearization problem. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 295(1077), 305–315.
- Levinson, S. C. (2005). Living with Manny’s dangerous idea. *Discourse Studies*, 7(4-5), 431–453.
- Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, 67, 427–446.
- Lindström, J. (2008). *Tur och Ordning: Introduktion till svensk samtalsgrammatik*. Nordstedts.
- Lindström, J. & Londen, A.-M. (2008). Constructing reasoning: The complex connectives *för att* (causal), *så att* (consecutive) and *men att* (adversative) in Swedish conversations. In J. Leino (Ed.), *Constructional reorganization* (pp. 105–152). Amsterdam: John Benjamins.
- Lindwall, O. (2008). *Lab Work in Science Education: Instruction, inscription, and the practical achievement of understanding*. PhD thesis, Linköping studies in arts and science, no. 426, Department of Theme Research, Linköping University.
- Linell, P. (1982). *The Written Language Bias in Linguistics*. Studies in Communication 2. University of Linköping: Dept. of Communication Studies.
- Linell, P. (1998). *Approaching Dialogue: Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins.
- Linell, P. (2005). *The Written Language Bias in Linguistics: Its nature, origins and transformations*. London & New York: Routledge.
- Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically*. Charlotte, NC: Information Age Publishers.

- Linell, P. & Gustavsson, L. (1987). *Initiativ och Respons: Om dialogens dynamik, dominans och koherens*. Linköping University: Department of Communication Studies.
- Linell, P., Gustavsson, L., & Juvonen, P. (1988). Interactional dominance in dyadic communication: A presentation of initiative-response analysis. *Linguistics*, 26(3), 415–442.
- Liszkowski, U. (2005). Human twelve-month-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. *Gesture*, 5(1-2), 135–154.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- & 18-month-olds point to provide information for others. *Journal of Cognition and Development*, 7(2), 173–187.
- Livingston, E. (2008). Context and detail in studies of the witnessable social order: Puzzles, maps, checkers, and geometry. *Journal of Pragmatics*, 40(5), 840–862.
- Lock, A. (1997). The role of gesture in the establishment of symbolic abilities: Continuities and discontinuities in early language development. *Evolution of Communication*, 1(2), 159–192.
- Lock, A. J. (1980). *The guided reinvention of language*. London: Academic Press.
- Lorenz, K. (1957 [1939]). Comparative study of behaviour. In C. H. Schiller & K. S. Lashley (Eds.), *Instinctive Behaviour: The Development of a Modern Concept* (pp. 239–263). New York: International University Press.
- Lupyan, G. & Spivey, M. J. (2010). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS One*, 5(7), 1–10.
- Maasen, S., Prinz, W., & Roth, G. (Eds.). (2003). *Voluntary Action: Brains, minds, and sociality*. Oxford: Oxford University Press.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2005). Item-based constructions and the logical problem. In p. (Ed.), *Proceedings of the Second Workshop on Psychocomputational Models of Human Language Acquisition* (pp. 53–68).
- Maguire, M., Hirsh-Pasek, K., & Golinkoff, R. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 364–392). Oxford: Oxford University Press.
- Marcos, H. (1991). Reformulating requests at 18 months: Gestures, vocalizations and words. *First Language*, 11(33), 361–375.
- Marcos, H., Ryckebusch, C., & Rabain-Jamin, J. (2003). Adult responses to young children's communicative gestures: Joint achievement of speech acts. *First Language*, 23(6), 213 – 237.
- Markova, I. & Linell, P. (2007). Coding elementary contributions to dialogue: Individual acts versus dialogical interactions. *Journal for the Theory of Social Behaviour*, 26(4), 353–373.
- Masataka, N. (2003). Index-finger extension to index-finger pointing: Ontogenesis of pointing in preverbal infants. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 69–84). Mahawa, NJ: Lawrence Erlbaum Associates.
- Masur, E. F. (1983). Gestural development, dual-directional signaling, and the transition to words. *Journal of Psycholinguistic Research*, 12(2), 93–109.
- Mauss, M. (1979 [1936]). Body techniques. In *Sociology and Psychology: Essays by Marcel Mauss* (pp. 97–135). Transl. Ben Brewster. London: Routledge and K. Paul.
- McClave, E. Z. (2007). Potential cognitive universals: Evidence from head movements in turkana. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language* (pp. 91–98). Amsterdam: John Benjamins Publishing Company.

- McCune, L. (1995). A normative study of representational play in the transition to language. *Developmental Psychology*, 31(2), 198–206.
- McCune-Nicolich, L. (1981). Toward symbolic functioning: Structure of early pretend games and potential parallels with language. *Child Development*, 52(3), 785–797.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92, 350–371.
- McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D. (1998). *The Face: A Natural History*. New York: Little, Brown.
- McNeill, D. (2000). Introduction. In D. McNeill (Ed.), *Language and Gesture* (pp. 1–10). Cambridge: Cambridge University Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: Chicago University Press.
- McNeill, D. (2007). Gesture and thought. In A. Esposito, M. Bratanic, E. Keller, & M. Marinaro (Eds.), *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue* (pp. 20–33). Amsterdam: IOS Press.
- McNeill, D., Cassel, J., & Levy, E. (1993). Abstract deixis. *Semiotica*, 95(1/2), 5–19.
- McNeill, D. & Levy, E. (1982). Conceptual representations in language activity and gesture. In *Speech, Place, and Action: Studies in Deixis and Related Topics*. Chichester: Wiley.
- McNeill, D., McCullough, K.-E., & Duncan, S. D. (2004). An ontogenetic universal and how to explain it. In C. Müller & R. Posner (Eds.), *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998* (pp. 157–171). Berlin: Weidler Buchverlag.
- McNeill, D. & Sowa, C. (2007). When gestures repeat: Imagery, morphology, priming, and catchments. Keynote presentation at ISGS'07. Northwestern University, IL.
- Mead, G. H. (1932). *The Philosophy of the Present*. LaSalle, IL: Open Court.
- Mead, G. H. (1934). *Mind, Self and Society*. Chicago: Chicago University Press.
- Mead, G. H. (1938). *The Philosophy of the Act*. Chicago: Chicago University Press.
- Mehrabian, A. (1971). *Silent Messages*. Belmont, Cal: Wadsworth.
- Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science*, vol. 2 (pp. 55–77). Cambridge, MA: MIT Press.
- Milner, A. D. & Goodale, M. A. (2006). *The Visual Brain in Action* (2nd ed.). Oxford: Oxford University Press.
- Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 115–154). Amsterdam: John Benjamins.
- Mondada, L. (2006). Participants' online analysis and multimodal practices: Projecting the end of the turn and the closing of the sequence. *Discourse Studies*, 8(1), 117–129.
- Morford, M. & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, 19, 559–580.
- Morissette, P., Ricard, M., & Gouin Decarie, T. (1995). Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *British Journal of Developmental Psychology*, 13(2), 163–177.
- Morris, D. (1977). *Manwatching*. New York: Abrams.
- Morris, D., Collett, P., Marsh, P., & O'Shaughnessy, M. (1979). *Gestures: Their origins and distribution*. New York: Stein & Day.



- Morris, P. H. & Lewis, D. (2010). Tackling diving: The perception of deceptive intentions in association football (soccer). *Journal of Nonverbal Behavior*, 34(1), 1–13.
- Murphy, C. M. (1978). Pointing in the context of a shared activity. *Child Development*, 49(2), 371–380.
- Månsson, A.-C. (2003). *The relation between gestures and semantic processes: A study of normal language development and specific language impairment in children*. PhD thesis, Dept. of Linguistics: University of Gothenburg.
- Müller, C. (1998a). Iconicity and gesture. In S. Santi, I. Guaitella, C. Cavé, & G. Konopczynski (Eds.), *Oralité et Gestualité* (pp. 321–328). Paris: L'Harmattan.
- Müller, C. (1998b). *Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich*. Berlin: Amo Spitz.
- Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family. In C. Müller & R. Posner (Eds.), *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998* (pp. 233–256). Berlin: Weidler Buchverlag.
- Müller, C., Fricke, E., Ladewig, S., Mittelberg, I., & Teßendorf, S. (manuscript). Gestural modes of mimesis: Mimetic techniques and cognitive-semiotic processes driving gesture creation.
- Müller, C. & Posner, R. (Eds.). (2004). *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998*. Berlin: Weidler Buchverlag.
- Namy, L. L. (Ed.). (2005). *Symbol Use and Symbolic Representation*. Mahawa, NJ: Lawrence Erlbaum Associates.
- Namy, L. L., Campbell, A. L., & 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.
- Navarro, A., Pearson, B. Z., Cobo-Lewis, A. B., & Oller, D. K. (2005). Differentiation in early phonological adaptation? In J. Cohen, K. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 1690–1702). Somerville, MA: Cascadilla Press.
- Neumann, R. (2004). The conventionalization of the Ring Gesture in German discourse. In C. Müller & R. Posner (Eds.), *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998* (pp. 216–224). Berlin: Weidler Buchverlag.
- 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(4), 241–266.
- Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology*, 35(2), 514–526.
- Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Gesture and Thought* (pp. 186–198). Chicago: Chicago University Press.
- Norrby, C. (2002). Svenska påhängsuttryck av typen *å så* och *eller nåt*. en diskussion av deras förekomst och funktion(er) i ett samtida ungdomsmaterial. *Språk & Stil*, 11, 183–210.
- Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.
- Næss, A. (1953). *Interpretation and Preciseness*. Oslo: Dybwad.
- Núñez, R. E. & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, 30(3), 401–450.

- Ochs, E., Schegloff, E. A., & Thompson, S. A. (Eds.). (1996). *Interaction and Grammar*, vol. 13 of Studies in Interactional Sociolinguistics. Cambridge: Cambridge University Press.
- O'Connell, B. G. & Gerard, A. B. (1985). Scripts and scraps: The development of sequential understanding. *Child Development*, 56(3), 671–681.
- Ottesjö, C. & Lindström, J. (2006). *Så som diskursmarkör*. *Språk & Stil*, 15, 85–127.
- Parrill, F. (2007). Metagesture: An analysis of theoretical discourse about multimodal language. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language* (pp. 83–89). Amsterdam: John Benjamins Publishing Company.
- Payne, T. R. (1968). *S. L. Rubinštejn and the philosophical foundations of Soviet psychology*. Dordrecht: Reidel.
- Payrató, L. (1993). A pragmatic view on autonomous gestures: A first repertoire of catalan emblems. *Journal of Pragmatics*, 20(3), 193–216.
- Peirce, C. S. (1931–35). *Collected papers*. Cambridge: Harvard University Press.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Petitto, L. A. (1988). “Language” in the pre-linguistic child. In F. Kessel (Ed.), *The Development of Language and Language Researchers: Essays in honor of Roger Brown* (pp. 187–221). Hillsdale, NJ: Erlbaum Associates.
- Piaget, J. (1962 [1946]). *Play, Dreams and Imitation in Childhood*. Translated by C. Cattegno & F. M. Hodgson. New York: Norton.
- Pinker, S. (1989). *Learnability and Cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Pizzuto, E. & Capobianco, M. (2005). The link (and differences) between deixis and symbols in children's early gestural-vocal system. *Gesture*, 5(1-2), 179–199.
- Pizzuto, E., Capobianco, M., & Devescovi, A. (2005). Gestural-vocal deixis and representational skills in early language development. *Interaction Studies*, 6(2), 223–252.
- Pizzuto, E. A. & Capobianco, M. (2008). Unraveling the complexity of indexes in spoken and signed discourse. *Gesture*, 8(1), 82–103.
- Plunkett, K. & Strömquist, S. (1992). The acquisition of Scandinavian languages. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition, Vol. 3* (pp. 457–556). Hillsdale, NJ: Erlbaum.
- Price, R. (1953). *Doodles*. New York: Simon & Schuster.
- Péran, P., Démonet, J.-F., Cherubini, A., Carbebat, D., Caltagirone, C., & Sabatini, U. (2010). Mental representations of action: The neural correlates of the verbal and motor components. *Brain Research*, 1328, 89–103.
- Ramachandran, V. S. & Blakeslee, S. (1998). *Phantoms in the Brain: Probing the mysteries of the human mind*. New York: William Morrow.
- Rein, R. & Holle, H. (2010). Assessing interrater agreement of movement annotations: A work in progress. Poster presented at the 4th conference of the ISGS. Frankfurt/Oder, Germany.
- Richthoff, U. (2000). *En svensk barnspråskorpus: Uppbyggnad och analyser*. PhD thesis, Department of Linguistics, University of Göteborg.
- Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131–141.

- Rodrigo, M. J., González, A., de Vega, M., Muñetón-Ayala, M., & Rodríguez, G. (2004). From gestural to verbal deixis: a longitudinal study with spanish infants and toddlers. *First Language*, 24(2), 71 – 90.
- Rodríguez, C. (2007). Object use, communication and signs. In J. Valsiner & A. Rosa (Eds.), *The Cambridge Handbook of Socio-Cultural Psychology* (pp. 257–276). New York: Cambridge University Press.
- Rodríguez, C. (2009). The ‘circumstances’ of gestures: Proto-interrogatives and private gestures. *New ideas in Psychology*, 27, 288–303.
- Rodríguez, C. & Moro, C. (2008). Coming to agreement: Object use by infants and adults. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on intersubjectivity* (pp. 89–114). Amsterdam: John Benjamins.
- Rodríguez, C. & Palacios, P. (2007). Do private gestures have a self-regulatory function? A case study. *Infant Behavior and Development*, 30(2), 180–194.
- Rolfe, L. (1996). Theoretical stages in the prehistory of grammar. In A. Lock & C. R. Peters (Eds.), *Handbook of human symbolic evolution* (pp. 776–792). Oxford: Oxford University Press.
- Rome-Flanders, T. & Ricard, M. (1992). Infant timing of vocalizations in two mother-infant games: A longitudinal study. *First Language*, 12(36), 285–297.
- Rosenfeld, L. B., Kartus, S., & Ray, C. (1976). Body accessibility revisited. *Journal of Communication*, 26(3), 27–30.
- Roth, W.-M. (2002). From action to discourse: The bridging function of gestures. *Cognitive Systems Research*, 3(3), 535–554.
- Rowe, M. L., Özçalışkan, S., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture’s role in predicting vocabulary development. *First Language*, 28(2), 182–199.
- Ryle, G. (1968). The thinking of thoughts: What is ‘le Penseur’ doing? In *University Lectures*, number 18. University of Saskatchewan.
- Ryle, G. (1999 [1968]). The thinking of thoughts: What is ‘le Penseur’ doing? *Studies in Anthropology*, 11.
- Sacks, H. & Schegloff, E. A. (2002). Home position. *Gesture*, 2(2), 133–146.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50(4), 696–735.
- Saitz, R. L. & Cervenka, E. J. (1962). *Colombian and North American Gestures*. Bogota: Centro Colombo Americano.
- Salgado, J. & Hermans, H. (2005). The return of subjectivity: From a multiplicity of selves to the dialogical self. *E-Journal of Applied Psychology*, 1(1), 3–13.
- Saussure, F. d. (1983 [1916]). *Course in general linguistics*. Translated by R. Harris. London: Duckworth and Co.
- Saxe, G. B. & Kaplan, R. G. (1981). Gesture in early counting: A developmental analysis. *Perceptual and Motor Skills*, 53, 851–854.
- Schegloff, E. (2007). *Sequence Organization in Interaction: A primer in Conversation Analysis*, vol. 1. Cambridge, MA: Cambridge University Press.
- Schegloff, E. A. (1984). On some gestures’ relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* chapter 12, (pp. 266–295). Cambridge: Cambridge University Press.
- Schegloff, E. A. (1986). The routine as achievement. *Human Studies*, 9, 111–152.

- Schegloff, E. A. (1987). Between micro and macro. In J. Alexander (Ed.), *The Micro–Macro Link* (pp. 207–234). Los Angeles: University of California Press.
- Scheler, M. (1954). *The Nature of Sympathy*. Translated by P. Heath. London: Routledge and Kegan Paul.
- Schneider, C., Call, J., & Liebal, K. (2010). Do bonobos say *no* by shaking their head? *Primates*, 51(3), 199–202.
- Schutz, A. (1932). *Der Sinnhafte Aufbau der Sozialen Welt: eine Einleitung in die verstehende Soziologie*. Wien: Julius Springer.
- Schutz, A. (1943). The problem of rationality in the social world. *Economica*, 10(37), 130–149.
- Schutz, A. (1944). The stranger: An essay in social psychology. *The American Journal of Sociology*, 49, 499–507.
- Schutz, A. (1945). On multiple realities. *Philosophy and Phenomenological Research*, 5(4), 533–576.
- Schutz, A. (1948). Sartre's theory of the Alter Ego. *Philosophy and Phenomenological Research*, 9(2), 181–199.
- Schutz, A. (1951). Choosing among projects of actions. *Philosophy and Phenomenological Research*, 12(2), 161–184.
- Schutz, A. (1953). Common-sense and scientific interpretation of human action. *Philosophy and Phenomenological Research*, 14(1), 1–38.
- Schutz, A. (1954). Concept and theory formation in the social sciences. *The Journal of Philosophy*, 51(9), 257–273.
- Schutz, A. (1955). Symbol, reality and society. In L. Bryson, L. Finkelstein, H. Hoagland, & R. M. MacIver (Eds.), *Symbols and Society: Fourteenth Symposium of the Conference on Science, Philosophy and Religion* (pp. 135–203). New York: Harper.
- Schutz, A. (1959). Type and eidos in Husserl's late philosophy. *Philosophy and Phenomenological Research*, 20(2), 147–165.
- Schutz, A. (1966). *Collected Papers III: Studies in phenomenological philosophy*. The Hague: Martinus Nijhoff.
- Schutz, A. (1967 [1932]). *The Phenomenology of the Social World*. Evanston, IL: Northwestern University Press.
- Schutz, A. & Luckmann, T. (1973). *The Structures of the Life-World*, vol. 1. Translated by R. M. Zaner & H. Tristram Engelhardt, Jr. London: Heinemann.
- Schutz, A. & Luckmann, T. (1989). *The Structures of the Life-World*, vol. 2. Translated by R. M. Zaner & D. J. Parent. Evanston, IL: Northwestern University Press.
- Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Minds*. Cambridge: Cambridge University Press.
- Searle, J. R. (1995). *The Construction of Social Reality*. London: Allen Lane.
- Seedhouse, P. (2007). On Ethnomethodological CA and “Linguistic CA”: A reply to Hall. *The Modern Language Journal*, 91(iv), 527–533.
- Seo, M.-S. & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, 42(8), 2219–2239.
- Seyfeddinipur, M. (2004). Meta-discursive gestures from Iran: Some uses of the ‘Pistol Hand’. In C. Müller & R. Posner (Eds.), *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998* (pp. 205–216). Berlin: Weidler Buchverlag.
- Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture*. PhD thesis, Radboud Universiteit Nijmegen.

- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4–13.
- Sharrock, W. & Coulter, J. (2009). ‘Theory of Mind’: A critical commentary continued. In I. Leudar & A. Costall (Eds.), *Against Theory of Mind* (pp. 56–88). New York: Palgrave Macmillan.
- Sherzer, J. (1972). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society*, 2(1), 117–131.
- Sherzer, J. (1991). The Brazilian thumbs-up gestures. *Journal of Linguistic Anthropology*, 1(2), 189–197.
- Shore, C., Bates, E., Bretherton, I., Beeghly, M., & O’Connell, B. (1990). Vocal and gestural symbols: Similarities and differences from 13 to 28 months. In V. Volterra & C. J. Erting (Eds.), *From Gesture to Language in Hearing and Deaf Children* (pp. 79–91). Berlin: Springer.
- Shuter, R. (1976). Proxemics and tactility in Latin America. *Journal of Communication*, 26(3), 46–52.
- Sidnell, J. (2005). Gesture in the pursuit and display of recognition: A caribbean case study. *Semiotica*, 156(1), 55–87.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children’s thinking*. New York: Oxford University Press.
- Silverstein, M. (1984). On the pragmatic “poetry” of prose: Parallelism, repetition, and cohesive structure in the time course of dyadic conversation. In D. Schiffrin (Ed.), *Meaning, form and use in context* (pp. 181–199). Washington, DC: Georgetown University Press.
- Singleton, J. L., Goldin-Meadow, S., & McNeill, D. (1995). The cataclysmic break between gesticulation and sign: Evidence against a unified continuum of gestural communication. In K. Emmorey & J. S. Reilly (Eds.), *Language, Sign, and Space* (pp. 287–311). Hillsdale, NJ: Erlbaum.
- Sinha, C. (2005). Blending out of the background: Play, props and staging in the material world. *Journal of Pragmatics*, 37(10), 1537–1554.
- Sinha, C. (2009a). Language as a biocultural niche and social institution. In V. Evans & S. Pourcel (Eds.), *New Directions in Cognitive Linguistics* (pp. 289–310). Amsterdam: John Benjamins.
- Sinha, C. (2009b). Objects in a storied world: Materiality, normativity, narrativity. *Journal of Consciousness Studies*, 16(6-8), 167–190.
- Sinha, C. & Rodríguez, C. (2008). Language and the signifying object: From convention to imagination. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on intersubjectivity* (pp. 357–378). Amsterdam: John Benjamins.
- Slama-Cazacu, T. (1976). Nonverbal components in message sequence: “Mixed syntax”. In W. C. McCormack & S. A. Wurm (Eds.), *Language and Man: Anthropological issues* (pp. 217–227). The Hague: Mouton.
- Sonesson, G. (1989). *Pictorial Concepts*. Lund: Lund University Press.
- Sonesson, G. (1992). *Bildbetydelser: Inledning till bildsemiotiken som vetenskap*. Lund: Studentlitteratur.
- Sonesson, G. (1994). Prolegomena to the semiotic analysis of prehistoric visual displays. *Semiotica*, 100(3/4), 267–232.
- Sonesson, G. (1996). Indexicality as perceptual mediation. In C. Pankow (Ed.), *Indexicality: Papers from the third bi-annual meeting of the Swedish Society for Semiotic Studies*, SSKKII Report 9604 (pp. 127–143). Gothenburg University.

- Sonesson, G. (1997). The ecological foundations of iconicity. In I. Rauch & G. F. Carr (Eds.), *Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS, Berkeley, June 12-18, 1994.* (pp. 739–742).
- Sonesson, G. (2001). From semiosis to ecology: On the theory of iconicity and its consequences for an ontology of the lifeworld. *Visio*, 6(2-3), 85–110.
- Sonesson, G. (2007). From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In J. Zlatev, T. Ziemke, & R. Frank (Eds.), *Body, Language and Mind: Embodiment*, vol. 1 (pp. 85–128). Berlin: Mouton de Gruyter.
- Sonesson, G. (2008). Prolegomena to a general theory of iconicity: Considerations of language, gesture, and pictures. In K. Willems & L. De Cuypere (Eds.), *Naturalness and Iconicity in Language* (pp. 47–72). Amsterdam: John Benjamins.
- Sonesson, G. (2009). New considerations on the proper study of man — and, marginally, some other animals. *Cognitive Semiotics*, 4, 133–168.
- Sonesson, G. (2010). Lecture 3: From the critique of the iconicity critique to pictorality. *Semiotics Institute Online*. <http://projects.chass.utoronto.ca/semiotics/cyber/Sonesson3.pdf>. October 7 2010.
- Sperber, D. & Wilson, D. (1995 [1986]). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell Publishers.
- Stefanini, S., Bello, A., & Caselli, M. C. (2009). Co-speech gestures in a naming task: Developmental data. *Language and Cognitive Processes*, 24(2), 168–190.
- Steffensen, S. V., Thibault, P. J., & Cowley, S. J. (2010). Living in the social meshwork: The case of health interaction. In S. J. Cowley, J. C. Major, S. V. Steffensen, & A. Dinis (Eds.), *Signifying Bodies: Biosemiotics, Interaction and Health* (pp. 207–244). Braga: Braga University Press.
- Steiner, J. E. & Glaser, D. (1995). Taste-induced facial expressions in apes and humans. *Human Evolution*, 10(2), 97–105.
- Stokoe, W. C. (1960). *Sign Language Structure*. Buffalo, NY: Buffalo University Press.
- Streeck, J. (1992). Previews: Gestures at the transition place. In P. Auer & A. D. Luzio (Eds.), *The Contextualization of Language* (pp. 135–157). Amsterdam: John Benjamins.
- Streeck, J. (1993). Gesture as communication i: Its coordination with gaze and speech. *Communication Monographs*, 60, 276–299.
- Streeck, J. (1996). How to do things with things: Objets trouvés and symbolization. *Human Studies*, 19, 365–384.
- Streeck, J. (2002a). A body and its gestures. *Gesture*, 2(1), 19–44.
- Streeck, J. (2002b). Grammars, words, and embodied meanings: On the uses and evolution of *So* and *Like*. *Journal of Communication*, 3, 581–596.
- Streeck, J. (2003). The body taken for granted: Lingering dualism in research on social interactions. In P. Glenn, C. LeBaron, & J. Mandelbaum (Eds.), *Studies in Language and Social Interaction* (pp. 427–440). Mahawa, NJ: Lawrence Erlbaum Associates.
- Streeck, J. (2008a). Depicting by gesture. *Gesture*, 8(3), 285–301.
- Streeck, J. (2008b). Metaphor and gesture: A view from the microanalysis of interaction. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 259–264). Amsterdam: John Benjamins.
- Streeck, J. (2009a). Forward-gesturing. *Discourse Processes*, 46(2), 161–179.
- Streeck, J. (2009b). *Gesturecraft: The manufacture of meaning*. Amsterdam: John Benjamins.

- Streeck, J. & Kallmeyer, W. (2001). Interaction by inscription. *Journal of Pragmatics*, 33(4), 465–490.
- Strömquist, S. (1997). Om tidig morfologisk utveckling [on early morphological development]. In R. Söderbergh (Ed.), *Från joller till läsning och skrivning* (pp. 61–80). Malmö: Gleerups.
- Sudnow, D. (1972). Temporal parameters of interpersonal observation. In D. Sudnow (Ed.), *Studies in Social Interaction* (pp. 259–279). Glencoe, IL: Free Press.
- Sudnow, D. (1978). *Ways of the Hand: The organization of improvised conduct*. London: Routledge & Kegan Paul.
- Söderbergh, R. (1969). Strukturer och normer i barnspråk. *Nordisk tidskrift*, 45, 67–82.
- Sørensen, J. (2010). *Action parsing in ritual and non-ritual behavior*. Presentation at the winter symposium of the center for semiotics at Aarhus University: Meaning and Interaction, January 2010.
- Thoermer, C. & Sodian, B. (2001). Preverbal infants' understanding of referential gestures. *First Language*, 21, 245–264.
- Thomas, W. I. (1928). *The Child in America: Behavior problems and programs*. New York: Knopf.
- Thompson, L. A. & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, 42(1), 144–168.
- Tinbergen, N. (1951). *The Study of Instinct*. New York: Oxford University Press.
- Tomasello, M. (1992). *First Verbs: A case study of early language development*. Cambridge: Cambridge University Press.
- Tomasello, M. (1996). Do apes ape? In C. M. Heyes & B. G. G. Jr. (Eds.), *Social Learning in Animals: The roots of culture* (pp. 319–346). San Diego: Academic Press.
- Tomasello, M. (2001). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, 11(1-2), 61–82.
- Tomasello, M. (2003). *A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2008). *Origins of Human Cognition*. Cambridge, MA: MIT Press.
- Tomasello, M. & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, 30(5), 639–650.
- Tomasello, M. & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4(12), 197–211.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–347). Cambridge: Cambridge University Press.
- Uebersax, J. S. (1987). Diversity of decision-making models and the measurement of interrater agreement. *Psychological Bulletin*, 101(1), 140–146.
- Valsiner, J. & van der Veer, R. (2000). *The Social Mind: Construction of the idea*. Cambridge: Cambridge University Press.
- Vihman, M. M. & McCune, L. (1994). When is a word a word? *Journal of Child Language*, 21, 517–542.
- Volterra, V. (1981). Gestures, signs, and words at two years: When does communication become language? *Sign Language Studies*, 22, 351–362.

- Volterra, V., Bates, E., Benigni, L., Bretherton, I., & Camaioni, L. (1979). First word in language and action: A qualitative look. In E. Bates, I. Bretherton, L. Camaioni, & V. Volterra (Eds.), *The Emergence of Symbols: Cognition and communication in infancy* (pp. 141–222). New York: Academic Press.
- Volterra, V., Caselli, M. C., Capirci, O., & Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello & D. I. Slobin (Eds.), *Beyond the Nature-Nurture: Essays in Honor of Elisabeth Bates* (pp. 3–40). Mahawa, NJ: Lawrence Erlbaum Associates.
- Volterra, V. & Erting, C. J. (Eds.). (1990). *From Gesture to Language in Hearing and Deaf Children*. Berlin: Springer.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in Society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Waggoner, J. E. & Palermo, D. S. (1989). Betty is a bouncing bubble: Children's comprehension of emotion-descriptive metaphors. *Developmental Psychology*, 25(1), 152–163.
- Watzlawick, P., Beavin, J. H., & Jackson, D. D. (1967). *Pragmatics of Human Communication*. New York: Norton.
- Weber, M. (1947 [1922]). *The Theory of Social and Economic Organization*. Translated by A. M. Henderson and T. Parsons. New York: Free Press.
- Wenger, E. (1998). *Communities of Practice; Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Werner, H. & Kaplan, B. (1963). *Symbol Formation*. New York: Wiley.
- Wertsch, J. V. (1981). The concept of activity in Soviet psychology: An introduction. In J. V. Wertsch (Ed.), *The Concept of Activity in Soviet Psychology* (pp. 3–36). Armonk, NY: Sharpe.
- Whiten, A., Horner, V., Litchfield, C. A., & Marshall-Pescini, S. (2004). How do apes ape? *Learning & Behavior*, 32(1), 36–52.
- Wikström, Å. (2008). *Den finita satsen i små barns språk*. PhD thesis, Dept. of Scandinavian Languages, Lund University.
- Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 171–215). Mahawa, NJ: Lawrence Erlbaum Associates.
- Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. Cambridge, MA: Harvard University Press.
- Wittgenstein, L. (1953). *Philosophical investigations*. New York: MacMillan.
- 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.
- Wootton, A. J. (1994). Object transfer, intersubjectivity and third position repair: Early developmental observations of one child. *Journal of Child Language*, 21, 543–564.
- Wootton, A. J. (1997). *Interaction and the Development of Mind*. Cambridge: Cambridge University Press.
- Wundt, W. (1973 [1921]). *The Language of Gestures*. The Hague: Mouton.
- Yngve, V. (1970). : (pp. 567–578).
- Zahavi, D. (2010). Empathy, embodiment and interpersonal understanding: From Lipps to Schutz. *Inquiry*, 53(3), 285–307.
- Ziemke, T., Zlatev, J., & Frank, R. (Eds.). (2007). *Body, Language and Mind: Embodiment*, vol. 1. Berlin: Mouton de Gruyter.



- Zinchenko, P. I. (1981). Involuntary memory and the goal-directed nature of activity. In J. V. Wertsch (Ed.), *The Concept of Activity in Soviet Psychology* (pp. 300–340). Armonk, NY: Sharpe.
- Zinober, B. & Martlew, M. (1985a). The development of communicative gestures. In M. D. Barrett (Ed.), *Children's Single-Word Speech* (pp. 183–215). Chichester: Wiley.
- Zinober, B. & Martlew, M. (1985b). Developmental changes in four types of gesture in relation to acts and vocalizations from 10 to 21 months. *British Journal of Developmental Psychology*, 3, 293–306.
- Zlatev, J. (1997). *Situated Embodiment: Studies in the emergence of spatial meaning*. PhD thesis, Department of Linguistics, Stockholm University.
- Zlatev, J. (2003). Mimesis: The “missing link” between signals and symbols in phylogeny and ontogeny? In A. Pajunen (Ed.), *Mimesis, Sign and Language Evolution*. Publications in General Linguistics 3. University of Turku, Finland.
- Zlatev, J. (2005). What's in a schema? Bodily mimesis and the grounding of language. In B. Hampe (Ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics* (pp. 313–342). Berlin: Mouton de Gruyter.
- Zlatev, J. (2007). Embodiment, language, and mimesis. In T. Ziemke, J. Zlatev, & R. M. Frank (Eds.), *Body, Language, and Mind, Vol. 1: Embodiment* (pp. 297–337). Berlin and New York: Mouton de Gruyter.
- Zlatev, J. (2008a). The co-evolution of intersubjectivity and bodily mimesis. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on intersubjectivity* (pp. 215–244). Amsterdam: John Benjamins.
- Zlatev, J. (2008b). From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology-Paris*, 102, 137–151.
- Zlatev, J. (2009). Levels of meaning, embodiment, and communication. *Cybernetics and Human Knowing*, 16(3-4), 149–174.
- Zlatev, J. & Andrén, M. (2009). Stages and transitions in children's semiotic development. In J. Zlatev, M. Andrén, M. Johansson Falck, & C. Lundmark (Eds.), *Studies in Language and Cognition* (pp. 380–401). Newcastle: Cambridge Scholars.
- Zlatev, J., Racine, T., Sinha, C., & Itkonen, E. (Eds.). (2008a). *The Shared Mind: Perspectives on intersubjectivity*. Amsterdam: John Benjamins.
- Zlatev, J., Racine, T. P., Sinha, C., & Itkonen, E. (2008b). Intersubjectivity: What makes us human? In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity* (pp. 1–14). Amsterdam and Philadelphia: John Benjamins.
- Zukow-Goldring, P. (1998). Perceiving eye-to-eye: Caregiver gestures cultivate the lexical development of Latino and Euro-American infants. *Infant Behavior and Development*, 21, 60.
- Zukow-Goldring, P. (2001). Perceiving referring actions: Latino and Euro-American infants and caregivers comprehending speech. In K. L. Nelson, A. Aksu-Koc, & C. Johnson (Eds.), *Children's Language: Interactional Contributions to Language Development*, vol. 11 (pp. 139–165). Hillsdale, NJ: Erlbaum.
- Zukow-Goldring, P. (2006). Assisted imitation: Affordances, effectivities, and the mirror system in early language development. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 469–500). Cambridge: Cambridge University Press.
- Zukow-Goldring, P. & Arbib, M. A. (2007). Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing*, 70(13-15), 2181–2193.

- Özçalışkan, S. & Goldin-Meadow, S. (2005a). Do parents lead their children by the hand? *Journal of Child Language*, 32, 481–505.
- Özçalışkan, S. & Goldin-Meadow, S. (2005b). Gesture is at the cutting edge of early language development. *Cognition*, 96, B101–B113.
- Özçalışkan, S. & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes*, 24(2), 190–217.

# TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND

FONDÉS PAR BERTIL MALMBERG

1. *Carl-Gustaf Söderberg*. A Typological Study on the Phonetic Structure of English Words with an Instrumental-Phonetic Excursus on English Stress. 1959.
2. *Peter S. Green*. Consonant-Vowel Transitions. A Spectrographic Study. 1959.
3. *Kerstin Hadding-Koch*. Acoustico-Phonetic Studies in the Intonation of Southern Swedish. 1961.
4. *Börje Segerbäck*. La réalisation d'une opposition de tonèmes dans des dissyllabes chuchotés. Étude de phonétique expérimentale. 1966.
5. *Velta Ruke-Dravina*. Mehrsprachigkeit im Vorschulalter. 1967.
6. *Eva Gårding*. Internal Juncture in Swedish. 1967.
7. *Folke Strenger*. Les voyelles nasales françaises. 1969.
8. *Edward Carney*. Hiss Transitions and their Perception. 1970.
9. *Faith Ann Johansson*. Immigrant Swedish Phonology. 1973.
10. *Robert Bannert*. Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. 1976.
11. *Eva Gårding*. The Scandinavian Word Accents. 1977.
12. *Gösta Bruce*. Swedish Word Accents in Sentence Perspective. 1977.
13. *Eva Gårding, Gösta Bruce, Robert Bannert* (eds.). Nordic Prosody. 1978.
14. *Ewa Söderpalm*. Speech Errors in Normal and Pathological Speech. 1979.
15. *Kerstin Nauclér*. Perspectives on Misspellings. 1980.
16. *Per Lindblad*. Svenskans sje- och tjeljud (Some Swedish sibilants). 1980.
17. *Eva Magnusson*. The Phonology of Language Disordered Children. 1983.
18. *Jan-Olof Svantesson*. Kammu Phonology and Morphology. 1983.
19. *Ulrika Nettelbladt*. Developmental Studies of Dysphonology in Children. 1983.
20. *Gisela Håkansson*. Teacher Talk. How Teachers Modify their Speech when Addressing Learners of Swedish as a Second Language. 1987.
21. *Paul Touati*. Structures prosodiques du suédois et du français. Profils temporels et configurations tonales. 1987.
22. *Antonis Botinis*. Stress and Prosodic Structure in Greek. A Phonological, Acoustic, Physiological and Perceptual Study. 1989.
23. *Karina Vamling*. Complementation in Georgian. 1989.
24. *David House*. Tonal Perception in Speech. 1990.
25. *Emilio Rivano Fischer*. Topology and Dynamics of Interactions — with Special Reference to Spanish and Mapudungu. 1991.
26. *Magnus Olsson*. Hungarian Phonology and Morphology. 1992.
27. *Yasuko Nagano-Madsen*. Mora and Prosodic Coordination. A Phonetic Study of Japanese, Eskimo and Yoruba. 1992.
28. *Barbara Gawronska*. An MT Oriented Model of Aspect and Article Semantics. 1993.
29. *Bengt Sigurd* (ed.). Computerized Grammars for Analysis and Machine Translation. 1994.
30. *Arthur Holmer*. A Parametric Grammar of Seediq. 1996.
31. *Ingmarie Mellenius*. The Acquisition of Nominal Compounding in Swedish. 1997.
32. *Christina Thornell*. The Sango Language and Its Lexicon (Sêndâ-yângâ tí Sängö).

- 1997.
33. *Duncan Markham*. Phonetic Imitation, Accent, and the Learner. 1997.
  34. *Christer Johansson*. A View from Language. Growth of Language in Individuals and Populations. 1997.
  35. *Marianne Gullberg*. Gesture as a Communication Strategy in Second Language Discourse. A Study of Learners of French and Swedish. 1998.
  36. *Mechtild Tronnier*. Nasals and Nasalisation in Speech Production. With Special Emphasis on Methodology and Osaka Japanese. 1998.
  37. *Ann Lindvall*. Transitivity in Discourse. A Comparison of Greek, Polish and Swedish. 1998.
  38. *Kirsten Haastrup & Åke Viberg* (eds.). Perspectives on Lexical Acquisition in a Second Language. 1998.
  39. *Arthur Holmer, Jan-Olof Svantesson & Åke Viberg* (eds). Proceedings of the 18th Scandinavian Conference of Linguistics. 2001.
  40. *Caroline Willners*. Antonyms in Context. A corpus-based semantic analysis of Swedish descriptive adjectives. 2001.
  41. *Hong Gao*. The Physical Foundation of the Patterning of Physical Action Verbs. A Study of Chinese Verbs. 2001.
  42. *Anna Flyman Mattsson*. Teaching, Learning, and Student Output. A study of French in the classroom. 2003.
  43. *Petra Hansson*. Prosodic Phrasing in Spontaneous Swedish. 2003.
  44. *Elisabeth Zetterholm*. Voice Imitation. A Phonetic Study of Perceptual Illusions and Acoustic Success. 2003.
  45. *Johan Frid*. Lexical and Acoustic Modelling of Swedish Prosody. 2003.
  46. *Anastasia Mukhanova Karlsson*. Rhythm and Intonation in Halh Mongolian. 2005.
  47. *Susanne Schötz*. Perception, Analysis and Synthesis of Speaker Age. 2006.
  48. *Victoria Johansson*. Developmental Aspects of Text Production in Writing and Speech. 2009.
  49. *Gilbert Ambrazaitis*. Nuclear Intonation in Swedish. Evidence from Experimental-Phonetic Studies and a Comparison with German. 2009.
  50. *Mats Andréén*. Children's Gestures from 18 to 30 months. 2010.