



LUND
UNIVERSITY

Image schemas, mimetic schemas, and the development of Swedish children's gestures

Klara Sif Skarphéðinsdóttir

Supervisor: Prof. Jordan Zlatev

MA Program in Language & Linguistics

Specialisation in Cognitive Semiotics

MA Thesis, 30 ECTS

October 2019

Abstract

Language and gesture are two semiotic systems that work together in close integration during communication. One issue in particular, that of *image schemas*, has been studied in relation to gestures, though most often in adult gesture and communication (Williams, forthcoming; Cienki, 2013; Mittelberg, 2006, 2018). This contrasts with *mimetic schemas*, which are thought to be more specific structures than image schemas, underlying children's first iconic gestures (Zlatev, 2014).

This thesis explored the gestures of three Swedish children in their fourth year. The aim was to discover whether there is a transition from more concrete modes of expression motivated by mimetic schemas, to more abstract modes of expression motivated by image schemas. This was achieved by analysing video-recorded data of three Swedish children and their caregivers in a natural environment using the CLAN software (MacWhinney, 2000). Their gestures and corresponding language were analysed, and in particular, iconic gestures and the corresponding language, with focus on the categories of *Viewpoint* and *Mode of Representation* (MOR).

The results showed that more concrete modes of expression in terms of Character Viewpoint and Acting MOR were more frequent at the beginning of the fourth year, while more abstract modes of expression in terms of Observer Viewpoint and Tracing/Drawing MOR are more frequent towards the end of the fourth year, which could be interpreted as an emergence of image schema-based gestures. Representing and Moulding gestures were intermediary. The overall analysis showed that there may be a trend towards abstractness that is more gradual than abrupt, though further studies would be needed to confirm this. Further, abstract gestures did not emerge until language had been firmly established, as well as its link to gesture. This finding points to a link between image schemas and language, directly or not.

Keywords: Gesture, Communication, Language, Image Schema, Mimetic Schema, Sign, Sign function, Cognitive Semiotics, Viewpoint, Mode of Representation, Semiotic Grounds.

Acknowledgements

Above all, I would like to thank my supervisor Jordan Zlatev for his incredible patience when I tended to rush, and support when he had a million things to do yet still made time to answer my many questions, even on the weekends and evenings. I am lucky to have had the pleasure to work with Jordan, and to learn more from him about not only cognitive semiotics, but also about the process of writing and thinking about complex subjects. *Mnogo blagodarya!*

A thank you to Sara Lenninger for her invaluable comments on the final version of this thesis.

I also want to thank my friends for taking the time to listen to me ramble on about this project, and even reading several, unfinished drafts. I especially want to thank my mother, Guðlaug, for chastising me when I needed the reality check, and for her words of encouragement when I needed the reassurance.

Finally, I want to thank my children, Dísa and Lillý, for accepting me even when I could not spend as much time with them as they wanted. I hope this inspires them when they decide to pursue their own passions.

Table of Contents

Chapter 1. Introduction	1
Chapter 2. Theoretical background	4
2.1 Introduction	4
2.2 Cognitive semiotics	4
<i>2.2.1 General characteristics</i>	<i>4</i>
<i>2.2.2 Signs and grounds</i>	<i>7</i>
2.3 Schemas	9
<i>2.3.1 Image schemas</i>	<i>9</i>
<i>2.3.2 Mimetic schemas</i>	<i>11</i>
2.4 Gestures	12
<i>2.4.1 Indexical gestures</i>	<i>14</i>
<i>2.4.2 Conventional gestures</i>	<i>15</i>
<i>2.4.3 Iconic gestures</i>	<i>16</i>
2.5 Gestures, schemas, and language	20
2.6 Summary and general hypotheses	24
Chapter 3. Methods	26
3.1 Introduction	26
3.2 Data	26
3.3 Procedure	27
<i>3.3.1 Transcription</i>	<i>27</i>
<i>3.3.2 Identification and semiotic ground</i>	<i>27</i>
<i>3.3.3 Coding</i>	<i>28</i>
<i>3.3.4 Statistics</i>	<i>30</i>
3.4 General analysis	30
<i>3.4.1 Iconic gestures</i>	<i>30</i>
<i>3.4.2 Language</i>	<i>32</i>

3.5 Summary and Predictions	33
Chapter 4. Results	34
4.1 Introduction	34
4.2 Gesture proportions	34
4.3 Hypothesis 1.....	36
4.4 Hypothesis 2.....	38
4.5 Hypothesis 3.....	39
4.5 Summary	42
Chapter 5. Discussion	43
5.1 Introduction	43
5.2 Viewpoint (H1)	43
5.3 Mode of Representation (H2)	45
5.4 Language (H3)	46
5.5 Other issues	48
5.6 Summary	51
Chapter 6. Conclusions	53
References	56
Appendix A: Language aligned with <i>Acting</i> gestures	63
Appendix B: Language aligned with <i>Representing</i> gestures	65
Appendix C: Language aligned with <i>Moulding</i> gestures.....	68
Appendix D: Language aligned with <i>Tracing</i> gestures.....	69
Appendix E: Language aligned with <i>Drawing</i> gestures	71

List of figures

Figure 1. The conceptual-empirical loop (adapted from Zlatev 2015b, p. 1058).....	5
Figure 2. An imagistic icon of the Icelandic President, Gudni Th. Johannesson, and a diagrammatic icon of a human body.	8
Figure 3. Representing: she <i>is</i> the glass that tips over. (A) is preparation, (B) is the stroke of the gesture. "Den ha vält ner på golvet" (see Appendix B).	18
Figure 4. Moulding: size (A) and distance/length (B) (see Appendix C).	18
Figure 5. Tracing (A): "Och i loftet!" (see Appendix D) & Drawing (B); "Leka med dom." (see Appendix E).	19
Figure 6. The gesture coding scheme in CLAN's "coder mode"	29
Figure 7. CHAINS %spa	30
Figure 8. Overall proportions of gesture types (dominant ground) over time.....	36
Figure 9. Character and Observer Viewpoint gestures over time.....	37
Figure 10. MOR over time.....	38
Figure 11a and Figure 11b: Viewpoints and MORs with speech.....	39

List of tables

Table 1. Methodological triangulation, as applied to the present study (based on Zlatev, 2012, p. 15).	6
Table 2. Dimensions of gestural meaning (adapted from Zlatev, 2014, p. 7).	13
Table 3. The data-set with 4 data points per child, with age given in months (m) and days (d).....	27
Table 4. The number of tokens for each gesture type (= dominant semiotic ground) and the corresponding percentage of gestures that belong to each type.....	35
Table 5. Tokens of gestures from the Character and Observer Viewpoints throughout the time period.....	37
Table 6. Tokens of gestures in the Acting, Representing, and Other MORs.	38
Table 7. Tokens of speech in regard to Viewpoint and MOR.....	40
Table 8. Tokens of linguistic expressions in regard to Viewpoint	40
Table 9. Tokens of linguistic expressions in regard to Acting, Representing, and Other Modes of Representation	41
Table 10. Tokens of linguistic expressions used with Other gestures	47
Table 11. Tokens and percentages of object-use during gestures with different Viewpoint .	49
Table 12. Tokens of object-use during Viewpoint and MOR combination gestures	49

Chapter 1. Introduction

Language and gesture are two closely connected semiotic systems. When people speak, they usually gesture, engaging in polysemiotic communication (Zlatev, 2019; Louhema et al., under review). The nature of these gestures and their relationship to speech has been extensively studied in the last decades (e.g. McNeill, 1992; Kendon, 2004; Mittelberg, 2006; Zlatev & Andrén, 2009; Andrén, 2010; Cienki, 2013; Müller et al., 2014; Zlatev, 2014; Cartmill et al., 2017; Williams, forthcoming). One of the issues under discussion concerns the relationship between gestures and *image schemas*. The latter notion was introduced by Johnson (1987, p. xiv), and defined as “recurring, dynamic patterns of our perceptual interactions and motor programs that give coherence and structure to our experience.” Such structures have been suggested to be crucial for our conceptualisation of events (Williams, forthcoming). These claims imply that image schemas are prelinguistic cognitive structures. However, as argued by Zlatev (2005), image schemas may not be independent of linguistic expressions such as prepositions, but actually gain their relatively abstract nature precisely due to the semantics of such linguistic expressions (*in vs. out, up vs. down*, etc.).

Cienki (2013) and Williams (forthcoming) have proposed that image schemas such as CONTAINER and PATH may underlie the structure and meaning of many gestures, such as when one traces the motion of a moving object in space. However, the gestures that can be analysed in such terms are to a considerable extent abstract (Andrén, 2008; Cienki, 2013), while the first iconic (i.e. resemblance-based) gestures of children have been shown to be more specific (Zlatev, 2014). On this basis, Zlatev has proposed that what grounds the meaning of iconic gestures (and possibly also the first linguistic expressions) are rather *mimetic schemas*, based on covert or overt imitation of practical actions like KISS, EAT, KICK, APPLY LOTION. Andrén (2010) conducted a longitudinal study of Swedish children’s gestures between the ages 18 to 30 months and noted that iconic gestures during the period were largely based on “typified actions”, which is in accordance with the hypothesis that they emerge as mimetic schemas. Further, Andrén noted a “dip” in children’s use of iconic gestures during the last

months of the studied period, which could perhaps mark the onset of the transition to more adult-like abstract gestures. However, the notion of a dichotomy between concrete and abstract gestures may be too simplistic, and the assumed transition could turn out to be a *gradual* one – there could be more continuity than discontinuity between younger and older children’s gestures, as well as those of adults. The question then arises, *how and when does the transition from more concrete (mimetic-schematic) gestures to more abstract (image-schematic) gestures take place?* Furthermore, if such a transition does take place, one might ask whether it is due to the children’s increasing linguistic proficiency.

This thesis addresses this question by investigating children’s gestures, in particular iconic gestures, between the ages of 3 and 4 years. This period in gesture development is understudied (Andrén, 2010; Zlatev, 2014, 2015a), and yet seems to carry important information regarding a possible transition in children’s cognitive and semiotic understanding. Iconic gestures in particular are valuable in this respect due to their potential status as representations or *signs* (see Chapter 2), and thus for the emergence of sign use in ontogeny. Finally, if children during the studied period do produce image-schematic gestures, and if they do so only with corresponding image-schematic language, then this should have theoretical implications for the status of image schemas. The research questions asked are thus:

RQ1. *Is there a transition towards abstractness in children’s iconic gestures in their fourth year of life, and if so, it is relatively gradual, or abrupt?*

RQ2. *Can such a transition be characterized as one from mimetic-schematic to image-schematic gestures?*

RQ3. *What is the role of language for the transition in question?*

To address these questions, this thesis investigates video-recorded, longitudinal data from three Swedish children (the same studied by Andrén, 2010) from the age of 34 months to 47 months.

The remainder of the thesis is divided into six chapters. Chapter 2 presents the theoretical background and framework for the study, concluding in general hypotheses. Chapter 3 presents the methodology and specifies the hypotheses. The results are given in Chapter 4, and Chapter 5 presents the discussion of the results, restating the original aim of the study and puts the findings into context. Finally, Chapter 6 presents the conclusions in relation to the three research questions and offers some suggestions for further research.

Chapter 2. Theoretical background

2.1 Introduction

This chapter introduces the overarching theoretical framework used for the study, based on concepts and methods from cognitive semiotics (2.2). Then, the notions of image schemas and mimetic schemas are discussed (2.3). After that, the different types of gestures are introduced, with a detailed section focusing on iconic gestures and the important differences between the dimensions Viewpoint and Mode of Representation (2.4). Section 2.5 combines schemas and gestures, and reviews studies that are relevant for the questions of this thesis. I end with a summary of the chapter (2.6), and a presentation of the general hypotheses.

2.2 Cognitive semiotics

2.2.1 General characteristics

Cognitive semiotics is a relatively new research discipline, predicated on a combination of research methods and concepts from (above all) linguistics, semiotics and cognitive science, aiming at new insights into human (and animal) meaning-making (Zlatev, 2012, 2015a; Sonesson, 2015; Konderak, 2016). It is a transdisciplinary field, with the ambition to “mend the gap between science and the humanities” (Zlatev, 2012, p. 18). While the field of semiotics suffers from a general lack of empirical methods, cognitive science often lacks the concepts and methods to study meaning-making and experience (Sonesson, 2009). Cognitive semiotics attempts to resolve this impasse with the help of phenomenology, the philosophical tradition emanating from the work of Edmund Husserl, where one of the basic ideas is to “depart from experience itself, and to provide descriptions of the phenomena of the world, including ourselves and others, as true to experience as possible” (Zlatev, 2012, p. 15). The experiential world, or the *life world*, is given to us through experience and is in a sense “constituted” through it (Sokolowski, 2000). This can serve as the basis for the formulation of a method of analysing experience as *intentionality* (Sonesson, 2012, p. 858):

The phenomenological method is based on the fact that everything, which, in the normal course of events, is available to (at least human) consciousness is present to this consciousness as something being outside of it. [...]. This is what, in the Brentano-Husserl-tradition, is known as intentionality.

Further, since human consciousness is *reflective* (as opposed to only perceptual, as supposedly that of most animals), it can turn onto itself and analyse its processes, which is technically known as the “phenomenological reduction” (see Sokolowski, 2000).

However, this is not sufficient as a method when studying meaning in general, and therefore cognitive semiotics highlights the need for a *conceptual-empirical loop* (Zlatev, 2015b, see Figure 1). This implies that we start by investigating a particular phenomenon (such as gesture) by asking “what is it?” and attempt to answer this question on the basis of experience, with as few preconceptions as possible. But then we need to go into the field and investigate with as much empirical detail as possible how this phenomenon “manifests itself”. Finally, we return to the conceptual side with more insights than when we started. In the present study this is carried out by looking at specific Swedish children’s gestures.

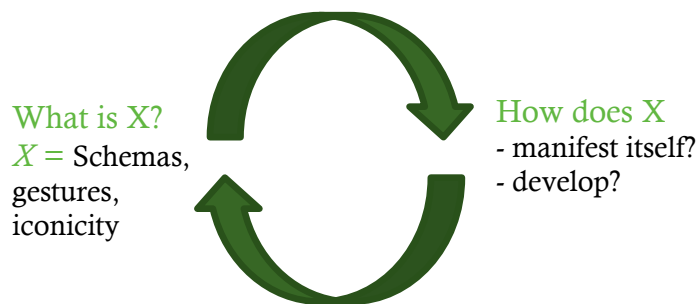


Figure 1. The conceptual-empirical loop (adapted from Zlatev 2015b, p. 1058)

Related to this is the idea of *methodological triangulation* from the cognitive semiotic perspective (Zlatev, 2012), where so-called first-person (subjective) methods, such as the phenomenological method, and third-person (quantitative) methods are combined. To do this, second-person (intersubjective) methods such as empathy are essential, since human

beings, as well as animals, are principally treated as other *subjects*, and not as objects to be studied only through “objective observation”. This approach results in a kind of methodological pluralism, which aims to:

- (a) acknowledge the validity of all methods within their respective domain of inquiry, (b) acknowledge the epistemological priority of first- and second-person methods in the study of meaning, and (c) integrate the three kinds of methods in the same project (Zlatev, 2012, p. 14).

Table 1 presents methodological triangulation as applied to the current study.

Table 1. Methodological triangulation, as applied to the present study (based on Zlatev, 2012, p. 15).

Perspective	Methods	Applied to (in this study)
First-person (“subjective”)	* Intuition-based analysis * Conceptual analysis	* Classifying gestures * Recognition of schemas in language and gesture
Second-person (“intersubjective”)	*Empathy-based understanding	* Deciding when a bodily action is communicative and/or representational
Third-person (“objective”)	*Detached observation	*Quantitative analysis

In this thesis, intuition and conceptual analysis are used in order to classify gestures based on their predominant semiotic ground (see Section 2.2.2), as well as to reach an understanding of image schemas and their relationship to iconic gestures in particular (see Section 2.5). Husserl (1970) identified *intuition* as being intersubjectively shared, and therefore not a private process (Zlatev, 2007). *Introspection*, on the other hand, is an examination of one’s own mental processes and is in principle a subjective act. It is intuition rather than

introspection that allows for conceptual analysis, which is essential for providing precise definitions of any concept.

Empathy is the ability to perceive, and in some cases, imagine the experiences of other beings because of their similarity to ourselves (Zlatev, 2009). Empathy-based understanding is similar to a concept that comes from Conversation Analysis, called *members' knowledge*. Members of society connect their present reality with their “stock of knowledge” in order to “establish what may be reasonably assumed to exist.” (Ten Have, 1990, p. 27). Thus, identification of communicative and representational acts becomes possible in the present study when conceptual analysis (first-person perspective) and the knowledge of social acts and situations (second-person perspective) are combined.

Finally, detached observation is employed to the “outputs” of the above methods in order to perform quantitative analysis and test falsifiable predictions (see Chapter 3).

2.2.2 Signs and grounds

One of the most important concepts within cognitive semiotics, as in semiotics in general, is that of the *sign*. There are numerous theories that define this notion in different ways. Some adopt broad definitions including, for example, all kinds of animal communication systems (e.g. the gestural repertoires and calls of apes, the calls and songs of marine mammals, etc.) (cf. Byrne et al., 2017; Montgomery & Radford, 2017). Others propose more narrow definitions, including only linguistic or other strictly conventional signs (e.g. Saussure, 1960). This thesis adopts the notion of sign (function) proposed by Sonesson (1989, 2010), which is intermediary. This states that a sign contains an expression and the content, which are independent from what the sign stands for (the referent). There is an asymmetry within the sign, in which the expression is more directly experienced, while the content is more in focus. Thus, there is a clear differentiation between the expression and content from the point of view of a conscious subject. Sonesson (2010) gives the example of a child using a pebble to signify candy. Another aspect of differentiation is when the expression is separated from the content in time and space, as for example in a photograph (Zlatev, 2009; Sonesson & Lenninger, 2015). The relation between expression and content has been termed the *ground*.

Following the semiotic theory of C.S. Peirce (Peirce, 1931-1958), there are three types of grounds known as *iconic*, *indexical*, and *symbolic* which are not mutually exclusive.

In the case of iconicity, the relation between expression and content is that of similarity/resemblance. This can be either more *imagistic*, as in a photograph, or more *diagrammatic*, in which the similarity between expression and content is perceived through the analogous relationship between the two, as in a map or stick figure, as shown in Figure 2 (Devyllder, 2018).



Figure 2. An imagistic icon of the Icelandic President, Gudni Th. Johannesson, and a diagrammatic icon of a human body.

An indexical ground involves either spatio-temporal contiguity, or part-whole relation between expression and content. For example, footprints in the snow are an index of the presence of another animal, smoke is an index of fire, and a pointing gesture is an index of the thing pointed to. Indexicality is only a potential sign until the subject enters the relationship and perceives the relation (Sonesson, 2000).

Finally, a symbol is a relationship built on convention, which is an agreement between two or more subjects on the existence of the relation, such as the colours of traffic lights signifying whether to stop, prepare, or go. Conventions are learned, and do not require the presence or absence of similarity or contiguity (Jakobson, 1965). Importantly, signs are never so clear-cut as to only contain one of the three types of grounds. As pointed out by Jakobson (1965, p. 26): “the difference between the three basic classes of signs is merely a difference in relative hierarchy... [based on] the predominance of one of these factors over the others.”

2.3 Schemas

The notion of the *schema* has a relatively long tradition in Western thought. In the 18th century, the philosopher Immanuel Kant proposed a structure of the imagination, which he called the *transcendental schema*, a mediator between the concept of an object, and the object itself. It is not an instantiation of the object, but a general *procedure* of serving to link the object of perception with the conception itself (Kant, 1998, [1787]). In Kant's view, imagination is an activity that orders representations in time. As pointed out by Johnson (1987, p. 153) "Kant thought of time as a pure, formal structure of consciousness to which all of our representations are subject...and since it organises all representations, it is connected to our perceptions."

The 20th century psychologist Frederic Bartlett, on the other hand, viewed schemas as an organisation of past experiences which help the individual to make sense of new and future experiences. Bartlett (1932, p. 201) argued that schemas "must always be supposed to be operating in any well-adapted organic response...[and] whenever there is any order or regularity of behaviour, a particular response is possible only because it is related to other similar responses...which operate...as a unitary mass."

Modern notions like image schemas and mimetic schemas seem to combine aspects of these views of temporality and memory, so as to provide a sort of template of experiences. But as we will see, these differ in terms of a number of characteristics, such as abstractness.

2.3.1 Image schemas

The notion of *image schema*¹ implies gestalt properties, operating as coherent, meaningful, and unified wholes that are more than the sum of their parts (Johnson, 1987). The best-known examples are notions like CONTAINER, SOURCE-PATH-GOAL and BALANCE. Johnson, who introduced these notions, regards them in a Kantian way as "bridges" between concrete and abstract thought (see also Mittelberg, 2006). More strongly, they have been regarded as the "building blocks" of cognition (Kimmel, 2005).

¹ Note that this concept is completely distinct from that of imagistic iconicity (see Section 2.2.2). The term "image" is, as well-known, an ambiguous one.

In Johnson's (1987) theory, image schemas are not so much representations as they are patterns of organism-environment interactions. They emerge as meaningful structures "chiefly at the level of our bodily movements through space, our manipulation of objects, and our perceptual interactions" (ibid, p. 29). Gibbs (2005) claims that image schemas are "emergent" properties, created on-the-fly during people's cognitive and perceptual activity. Pendlebury (1995) points out that schemas are more primitive than concepts and cannot be even remotely accessible to consciousness, and Dodge and Lakoff (2005) argue that image schemas are completely independent of both consciousness and language.

Clausner and Croft (1999), on the other hand, claim that schemas are *less* image-schematic if they are perceptual, saying that they are highly abstract structures that help our understanding of a wide variety of concepts. In their view, CONTAINER would not be an image schema due to its relation to the material world. A true image schema is central to our understanding of more elaborated concepts and does not include concepts that are connected to physical experience (ibid).

McCune (2008) is one of the few theorists who have applied the notion to cognitive development. According to her, the development of image-schematic structures such as CONTAINER and FULL-EMPTY requires the prior understanding of *reversibility* in time and space, which is a rather advanced cognitive feat.² Reversibility is the understanding that where there is an *up*, there is a *down*, or when something is *empty*, there is a reversible state of *full*. McCune (ibid) links the acquisition of such schemas with that of the corresponding words, bringing into doubt the claimed pre-verbal nature of such image-schemas.

As we can see, there is little agreement on the representational nature of image schemas. Since the notion of *representation*, at least as understood in phenomenology, "presupposes a differentiation between expression and content...implying conscious awareness of the representation as such" (Zlatev, 2005, p. 317), and image schemas are said to "typically operate beneath the level of our conscious awareness" (Johnson, 2005, p. 21), it

² Piaget (1954) introduced the concept of *reversibility*, which in his view is fully established by the developmental stage of Concrete Operational thought, around 7 years and is when children begin to use logical thought regarding physical objects. Thus, a child who has reached this stage will understand, for example, that a deflated ball has a reversible state of being inflated, and open door can be closed, etc.

would follow that they are indeed non-representational. At the same time, it is questionable whether at least more abstract image schemas, such as those that imply reversibility, are really independent of language, as pointed out by Zlatev (2005).

2.3.2 Mimetic schemas

In part due to the lack of agreement on the nature of image schemas, Zlatev (2005, p. 334) proposed the notion of *mimetic schemas*, defined as: “dynamic, concrete and preverbal representations, involving the body image, which are accessible to consciousness, and pre-reflectively shared in a community.” Examples include KICK, KISS, EAT, and DANCE. Mimetic schemas presuppose the universal human capacity for *bodily mimesis*, a concept developed by Donald (1991, p. 168): “Mimetic skills or mimesis rests on the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic.” According to Zlatev (2005), mimetic schemas utilise the conscious *body image*, rather than pre-conscious sensorimotor structures that arise from interactions with the world, as hypothesized by Johnson for image schemas. Because of their development through imitation, attention, and interaction with other people, they are interpersonal. In sum, mimetic schemas are concrete action concepts, and may arise from early understanding of daily routines and their sequences, such as eating, dressing, kissing, etc. (Mandler & Canovas, 2014). When this understanding is coupled with the production of such routines and physical actions, these mimetic schemas begin to emerge into overt enactments and gestures (Zlatev, 2014), as discussed below.

Compared to image schemas (whether these are understood as representations or not), mimetic schemas are relatively concrete, in that “each one is a generalisation of a particular bodily act...” (Zlatev, 2005, p. 318). Finally, mimetic schemas are preverbal, and may be considered a better alternative to image schemas when it comes to the grounding of language and a child’s first concepts, especially if image schemas do not exist independently of the linguistic forms used to express them. As discussed in the following sub-section, mimetic schemas may also play an important role in the emergence of gestures.

2.4 Gestures

Gestures are “naturally grounded in the concreteness of the human body as well as the social practices in which it engages” (Mittelberg, 2006, p. 14). They can thus communicate certain things better than language – specifically visual information. It is far from obvious how to define them, however. One of the leading names in the field of gesture studies, Adam Kendon, has proposed the following definition of gestures as

... 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” (Kendon, 2004, p. 15, my emphasis).

In agreement, Zlatev (2015a, p. 458) regards gestures as “expressive movements performed by the hands, the head, or any other part of the body, and perceived visually.” These definitions are very broad, however, and require further operationalisation. According to Andrén (2010), in order for a movement to be considered a gesture, it must display a high degree of *communicative explicitness* (as when a person points, and oscillates their gaze between target and audience), *semiotic complexity* (when the gesture in question is a sign, see Section 2.2.2), or both. Zlatev (2014) reformulated these criteria as three levels of *communicative intent* and *representational complexity*, respectively (see Table 2). For example, waving good-bye is not an explicit sign as it does not *stand for* a certain meaning but performs it: it *counts as* a communicative performance within the social interaction. However, because the action clearly displays communicative intent, it can be considered a gesture.

Table 2. Dimensions of gestural meaning (adapted from Zlatev, 2014, p. 7).

Level	Communicative intent (CI)	Level	Representational complexity (RC)
CI-1	No visible communicative intent	RC-1	Situation-specific: Action X contextually suggests Y
CI-2	Unclear communicative intent	RC-2	Typification: Action X <i>counts as</i> Y
CI-3	Clear communicative intent	RC-3	Explicit sign: Action or expression X <i>stands for</i> Y

Kendon (2004) identifies a gesture as containing several structural parts. The parts of the gesture together are termed the *gesture phrase*, wherein the part of the gesture usually seen as such by the observer is the *stroke*. The stroke of the gesture is “the phase of the excursion in which the movement and dynamics of ‘effort’ and ‘shape’ are manifested with greatest clarity... [it is] the phase when the ‘expression’ of the gesture is accomplished.” (ibid, p. 112). The stroke and the pause following it, termed the *post-stroke hold*, are the parts of the gesture that coincide with the semantically relevant parts of speech. Together, the stroke and the post-stroke hold are the nucleus of the gesture phrase.

Gestures can be identified as having iconic, indexical, and conventional “components”, which correspond to the three kinds of *semiotic grounds* discussed in Section 2.2.2. Just as with signs in general, in a single gesture more than one semiotic ground can be present: they are “not mutually exclusive and may coexist, in different proportions, in the same sign/gesture” (Zlatev, 2015a, p. 462). Yet, one of the grounds usually predominates, and this can serve as the basis for classification. Pointing gestures may be categorised as indexical despite having conventional components. Children’s enactments may also contain a component of conventionalisation (e.g. pretending to put on lotion), in that the actions themselves are conventional. However, the gestural form does not have the normative constraints to the degree that more conventional gestures do (Andrén, 2010), and can be regarded as primarily iconic. In the following subsections, I review the three major kinds of (representational) gestures as types, but it should be reminded that this is to be understood in terms of “prevalent ground”.

2.4.1 Indexical gestures

Indexical (or deictic) gestures bring things and people to others' attention, and "actualise target referents in the physical surroundings" (Andrén, 2010, p. 99). Pointing gestures are a typical example, but requests or showing/giving something to someone are also based on contiguity – between the intended attention of the addressee and the target. These gestures are the first to emerge in a child's repertoire (Andrén, 2010; Zlatev, 2014). Initially, so-called *imperative pointing* and reaching for something appear to function as acts of *ritualised behaviour* (Brinck, 2004), along with giving someone something and showing something to someone. Volterra et al. (2018, p. 432), summarising a long list of studies, state that these early gestures "emerge from interaction between individuals and...gain their meaning from being constituents of particular behaviour patterns." Given that early gestures such as imperative pointing do not require an understanding of the other as an intentional agent (Tomasello, 1999), they are not at level 3 of communicative intent (see Table 2), and neither can they be argued to be differentiated from their content. Thus, given the present definition, they would not qualify as a gesture.

Declarative pointing, on the other hand, does require the child to understand the intentions of the other, as well as to make an effort to have one's own intention understood. That is, a third object is brought into the communicative situation, and it is meant to be jointly attended to by both participants. Declarative pointing, possibly learned through imitation (Brinck, 2004), is both referential and communicative at level CI-3, and hence is a (proper) gesture.

Another type of declarative pointing is *abstract pointing*, which is when the pointing gesture refers to something that is not present in the communicative situation. This sort of pointing uses the gesture space to "refer to particular discourse information...potentially play[ing] a role in discourse build-up" (Gunter et al., 2015, p. 2). In other words, to the specification of actors, actions, or objects in a narrative in order to maintain cohesion. Thus, abstract pointing supplements a verbal narrative with important spatial information, allowing the interlocutors to collectively build a shared "image" within the gesture space (Han et al.,

2018). According to McNeill (1992), abstract pointing is one of the last gestures to emerge in childhood, many years into development. Given these considerations, it should be regarded as a highly abstract indexical gesture, presupposing the prior development of narrative, which itself presupposes symbolic signs.

2.4.2 Conventional gestures

Strongly conventional (or emblematic) gestures, such as the OK or NOD gestures, have “normative criteria on the proper form and meaning” (Zlatev, 2014, p. 461). In other words, there is a correct way and an incorrect way of performing the gesture. Gestures of negation and agreement are among the first conventional gestures to arise in children’s repertoire, though the headshake seems to appear earlier in development than the nod (Guidetti, 2005; Fusaro et al., 2011; Andrén, 2014). Such emblems have a semantics that is similar to that of linguistic items, as there is “a direct verbal translation or dictionary definition, usually consisting of a word or two or perhaps a phrase” (Kendon, 2004, p. 96).

Another class of gestures that have conventional qualities are so-called *recurrent gestures*, like BRUSH-ASIDE. Such gestures are conventionalised and culturally shared, and have pragmatic meaning in that they act on the discourse and are used to display or accomplish an action within the communicative situation (Cienki et.al., 2014), rather than represent something within it. According to Bressem and Müller (2014) recurrent gestures are originally motivated by practical actions and are part of a repertoire (similar to mimetic schemas, see 2.3.2). They differ from emblems in that they are not word-like (Ladewig, 2013). Furthermore, the form of the recurrent gesture contributes to its meaning, which implies that they are also iconic. The border between emblems and recurrent gestures, however, is not clear. Many emblems, such as SHRUG, could be regarded as either. As pointed out by Debras (2017, p. 3): “The fact that some gestures defined as emblems can add a conventionalised supplementary meaning to speech suggests that there is an overlap between emblems and recurrent gestures.”

2.4.3 Iconic gestures

Iconic gestures show a “resemblance between the movements of the whole body, or parts of it, and properties of intended actions, objects, or whole events” (Zlatev, 2015a, p. 461). There have been several different accounts in the literature that attempt to categorise types of iconicity in gesture (McNeill, 1992; Zlatev & Andrén, 2009; Brown et al., 2019; Cartmill et al., 2017; Müller, 2014; Frederiksen, 2017). Some of these have relied on differences in perspective (viewpoint), while others on more specific criteria, based on the nature of the expression-content resemblance, as explained below.

McNeill (1992) distinguishes between gestures that are either performed from a *character viewpoint* or an *observer viewpoint*. For example, if the gesturer performs a walking like motion with their whole body, this is of the first kind. If they rather illustrate it with moving fingers, it is of the latter. Analogously, Zlatev & Andrén (2009, p. 387) distinguish between first-person perspective gestures (1pp) and third-person perspective gestures (3pp), where the first display “explicit or implicit mapping of the whole body onto the signified, even if only a part of the body is thematic.” In the case of 3pp gestures “the articulating parts of the body figure as observed objects, isolated from the rest of the body... [and not] bear[ing] any relation to the signified.”

Similarly, Brown et al. (2019) distinguish between *egocentric* and *allocentric* gestures instead of the character and observer viewpoint distinctions. The former are performed “with reference to the position and orientation of a person’s body”, while the latter are performed “with reference to objects and their surrounding environment.” (p. 3). According to the authors, some gestures combine both, such as the CALL-ME gesture, where the hand becomes a phone. In this case, the body part is an object (allocentric), but it is performed egocentrically.

However, the latter may be a conflation of perspective and what may be called, following Müller (2014), *modes of representation*. Frederiksen (2017) questions categorising gestures based on viewpoint and proposes that the imagery of what is represented, and the mode of representation, are more important. Müller (2014) has influentially distinguished between the following modes of representation:

- In *Acting*, the hands mime or enact instrumental actions, as when you pretend to eat with an invisible spoon.
- In the case of *Representing*, the hands or body take on the qualities of that which is being represented, as in the example of the CALL-ME gesture mentioned above. Notably, on this account the gesture is not “allocentric”, and is still being performed with a character viewpoint, but uses a different Mode of Representation from when, say, you pretend to hold an invisible phone to your ear (see Figure 3).
- *Moulding* in Müller’s terms is when the speaker moulds the external attributes of objects, such as their shape or size. This gesture uses the whole hand or hands, and as such is usually done from a character viewpoint. In line with this, Hassemer and Winter (2016, p. 5) call such height and shape gestures “holding gestures”, as the hands maintain surface contact with an imagined object (see Figure 4).
- In *Tracing* the hands trace or “draw” paths of motion, or an object’s shape, in the air. These two sub-types, however, need to be distinguished. As pointed out by Brown et al. (2019), tracing can either be static or dynamic. Static tracing occurs when the tracing finger simply outlines the shape or size of an entity (Drawing), and dynamic tracing is when a path of motion is traced (Tracing) (see Figure 5).

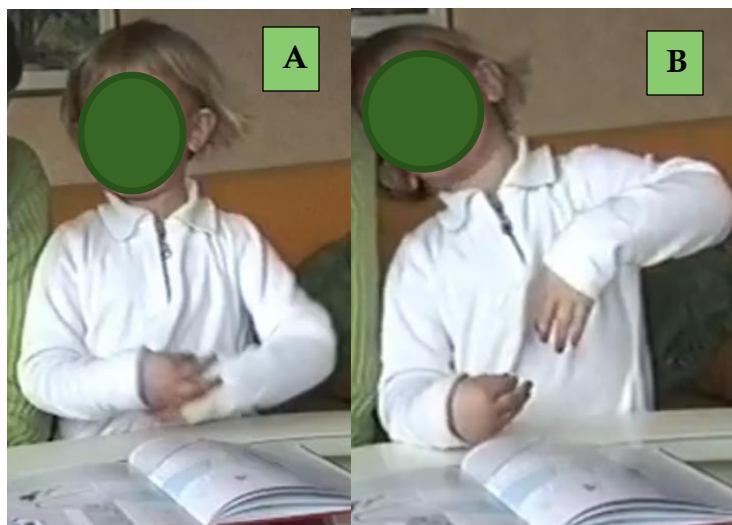


Figure 3. Representing: she *is* the glass that tips over. (A) is preparation, (B) is the stroke of the gesture. "Den ha vält ner på golvet" (see Appendix B).



Figure 4. Moulding: size (A) and distance/length (B) (see Appendix C).



Figure 5. Tracing (A): "Och i loftet!" (see Appendix D) & Drawing (B); "Leka med dom." (see Appendix E).

The scheme proposed by Cartmill et al. (2017) is similar, but less differentiated: in *hand-as-hand*, the hand acts on an imaginary object, in *hand-as-object* the hand mimes the shape of an object, and in *hand-as-neutral* the movement and location of the hand conveys a shape, location or path or motion. It thus appears that this first category maps to Acting and Moulding, the second to Representing, and the third to the static and dynamic kinds of Tracing.

As can be seen from this review, it is beneficial to hold the dimensions of Viewpoint and Mode of Representation distinct, and analyse gestures as combinations of these, as far as possible. While Character/1pp and (en)Acting gestures would seem to largely overlap, it is possible to use a representing/hand-as-object, and still use a Character Viewpoint, as illustrated. In the case of Tracing, it will most often be a matter of Observer Viewpoint, but there are occasions where it may take on a quality of mimicking an action, such as when the finger is used as if it were a marker or pen, and the Tracing takes place on an empty piece of paper (Andrén, 2010), which implies a Character Viewpoint. In regard to Mode of Representation, this example would technically be *both* Representing and Tracing, because the finger *is* the drawing tool. Any type of Tracing can in principle be regarded in this way,

but as the referent being highlighted in these cases is the movement and not the object the finger represents, they can be considered as *Tracing*, and not *Representing* in analysis.

A final point to consider is if iconic gestures are performed with objects or not, which is an independent dimension from those above. A child playing with a toy, making a car “drive”, or a doll “dance” is involved in *Representing*, but this could be either from an Observer Viewpoint, or from a Character Viewpoint if, for example, the doll is imagined to be the child’s baby.

2.5 Gestures, schemas, and language

The classifications of gestures, and especially of iconic gestures, discussed in the previous subsection, have been argued to have important developmental implications. Starting with Viewpoint, McNeill (1992) found that younger children produce Character Viewpoint gestures far more often than Observer Viewpoint gestures. He proposed that early iconic gestures are of the “inside observer type” where, for example, a story is told from the perspective of an observer who is taking part in the events of the story rather than an observer narrating from “outside” of the story. Even when children begin to use Observer Viewpoint gestures in their third year, there are still elements of Character Viewpoint that are not present in adult gestures. It is, however, not fully clear whether these features of early iconic gestures have to do with Viewpoint or with Mode of Representation.

With regard to Mode of Representation, Andrén (2010) and Zlatev (2014) found that Acting gestures are the most frequent kind between the ages of 18-26 months old. Being shaped by the community’s typical bodily actions, such gestures clearly correspond to mimetic schemas. Thus, even if bodily mimesis is implicated in the development of all gestures, Zlatev (2014, p. 24) argued that only for iconic gestures is there good evidence that they emerge as mimetic schemas. While Williams (forthcoming) accepts that action-based iconic gestures are motivated by mimetic schemas, he proposes that other iconic gestures

should rather be seen as “shaped...by image schemas in cognitive models”³ (p. 8). The proposal is summarised as follows:

The emerging picture is one in which mimetic schemas are apparent in early gestures produced in first-person perspective or character viewpoint, where the face, hands, and body are engaged in enacting the intended meaning, while image schemas at the heart of more abstract cognitive models (such as those involved in time-telling) become more apparent in gestures as children build knowledge and apply it in discourse, notably where gestures are produced from a third-person perspective or observer viewpoint, depicting with the hands in the space in front of the body. Both types of gestures continue into adulthood, with mimetic schemas evident in pantomimic gestures and image schemas evident in referential gestures (whether concrete or abstract) as well as in pragmatic gestures that act on the discourse... (Williams, forthcoming, p. 32).

Cienki (2013, p. 426) similarly observes that gestures that are based on image schemas “seem to relate to ideas on the general level of types of processes, reasoning, or behaviour, while those based on mimetic schemas...concerned more particular ideas, like negation or dismissiveness.” It thus seems reasonable to assume that mimetic schema-based gestures emerge before image schema-based gestures. When it comes to adult gestures, however, the situation is more complex. Both Cienki (2013) and Mittelberg (2018) argue that adult gestures motivated by mimetic schemas are produced consciously and have handshapes that are specifically related to instrumental actions, while those grounded by image schemas have less specific handshapes with simple motions, and are hence more abstract, and are not consciously produced. This conclusion is consistent from Johnson’s (2005) claim that image schemas operate below consciousness and Zlatev’s (2005) proposal that mimetic schemas *are*

³ Conversely, Cienki (2008, 2013) describes image-schematic gestures in terms of the mapping between two conceptual domains. In general, the argument is that the source domain is exemplified by the gesture itself, and the target domain is the concrete situation which is being talked about. For example, a straight, downward chop used in conjunction with the words ‘honest’ and ‘moral’ “reflects a cognitive model for moral behaviour based on the STRAIGHT image schema and its metaphorical projection MORAL IS STRAIGHT.” (Williams, forthcoming, p. 8).

accessible to consciousness.⁴ However, it is not clear if this reasoning can be mapped over to development, as it would imply the children are first more conscious, and then less conscious in their use of gestures, which is rather counter intuitive.

Several studies have shown that children's understanding of action-based iconicity emerges much earlier than perception-based iconicity, especially when the latter concerns size and shape. In a study conducted by Novack et al. (2018), pincer-like hand shapes were used to assess whether children between 30 months and four years old could understand iconic gestures when asked for either a small or large rubber duck. Their results showed that when performed along with an instrumental (i.e. Acting) reach towards the toy, the younger children had no trouble understanding the meaning of the gesture. However, when performed without the instrumental reach (i.e. Representing), they failed to understand the gesture information, suggesting that such gestures "may constitute a greater representational leap...than the mapping between the movement of the hands and the object on which the action was performed (p. 1810). Thus, younger children do seem to understand the iconic link between handshape and object attributes, but only when the gesture is performed embedded within an action. When the gesture solely represents information, children 2.5 years old and younger do not understand these gestures as signs.

Moulding can be said to be even more abstract than Representing, as what is being represented is not an object, but only one of its features, such as shape or size. Still, Acting, Representing, and Moulding all use the body or hands to create a direct link between the gesture and referent through size/shape or action information, while Tracing and Drawing do not. With Tracing and Drawing, the similarity lies in the movement of the hand or finger (or foot). To remind, tracing is the similarity between the path of motion being gestured and the path of motion that an actual object or person (or even more abstractly an event or idea) takes. Drawing is the similarity between the trajectory of the hand or finger, and the static

⁴ This is also consistent with the findings from an experiment conducted with a man who, in his late teens lost all haptic senses and coordination from the neck down. The point of the study was to observe his gesture performance by looking at gesture-speech synchronisation, location of hands, and handshape during narrative, while cutting off visual access to his hands. The results showed that Acting gestures (from the character viewpoint) clearly derived from practical actions required conscious control, while those gestures that are integrated with speech did not (McNeill, Quaegebeur & Duncan, 2010).

shape of an object or person it creates. The difference is, again, of the latter solely representing information, and the former incorporating action information, implying that Drawing is even more abstract than Tracing. In these gestures, the “essence [is distilled] out of cognitive or physical processes...by abstracting away from the elements or ideas undergoing the process” (Mittelberg & Waugh, 2014, p. 1753). As discussed with respect to the abstractness of image schemas earlier, and the possibility that this is not a precondition for, but consequence of language (Zlatev, 2005), the same could be the case for more abstract gestures, which do not begin to arise until the third year, which is also when language beings to “spurt” (Andrén, 2010; Zlatev, 2014).

It is important to consider language and its role for both schemas and gestures. In most cases, speech is essential for comprehending the gesture’s meaning, and its schematic nature is not understandable without it (Cartmill et al., 2012; Müller et al., 2014; Frederiksen, 2017). Even the exact shape of (iconic) gestures can depend on the presence or absence of speech. This was shown by Özcaliskan et al. (2016), who replicated a study conducted by Kita and Özyürek (2003) and found that the iconic gestures produced depended on whether they were accompanied by speech or not. English-speakers and speakers of Turkish produced different gestures from each other when using them with language, but similar ones without language. The use of objects together with gestures is also important as it might be that “the integration of speech with iconic gestures starts in contexts involving objects...which may be a step in the development of iconic gestures that are detached from objects” (Zlatev & Andrén, 2009, p. 393). There are practically no empty-handed Representing or Drawing gestures performed by children before the age of 30 months. On the other hand, there are multiple such gestures being performed with an object, i.e., doll-kiss/walk/dance, car-drive (Zlatev & Andrén, 2009; Andrén, 2010; Zlatev, 2014).

Using the concepts of different kinds of iconicity discussed in Section 2.2.2, we could say that all but Acting gestures (which may be considered imagistic) are more or less *diagrammatic*, in that their shape and movement are analogous to the referent’s. This should

also help in mapping gestures to schemas,⁵ since a gesture (and its corresponding speech) may analogously refer to an element of image-schematic meaning, such as CONTAINMENT, or SOURCE-PATH-GOAL. Schematic language consists of grammatical expressions such as prepositions (*in, on, above, under, to, from*), many of which are about space and motion (see Section 2.3.1). According to McCune (2006, p. 252), “early understanding of concrete motion-event expression could form the basis...for more abstract linguistic meaning later on.” Talmy (2015) mentions several such closed-class, schematic forms, including epistemic, causal, configurational, temporal, and spatial expressions. Young children, however, first learn core “dynamic event words”, related to path, motion, and figure/ground relations (McCune, 2008).

According to McCune (2006, p. 238), dynamic event words are defined as referring to “some aspect of one or more reversible motion events extended in time and involving spatial and temporal relationship(s) between entities.” Because these early words all form contrastive pairs, it seems likely that language itself plays a role in their differentiation (Zlatev, 2005). It may be that the language that accompanies gestures such as Moulding, Tracing, and especially Drawing, all of which are assumed to appear later in development than Acting or Representing, is precisely of this grammatical type.

2.6 Summary and general hypotheses

This chapter presented some of the concepts from cognitive semiotics relevant for the present study, including the concept of the *sign* and its different kinds of *grounds* (Section 2.2). Some of the background to the schema concept was presented in Section 2.3, along with the two kinds of schemas discussed in cognitive linguistics, namely image schemas and mimetic schemas. An overview of different kinds of gestures, ending with a thorough outline of iconic gestures and their manifestations was given in Section 2.4. The most important conclusion was that *Viewpoint* and *Mode of Representation* are two independent dimensions and should be

⁵ Most of the literature describing the connection between schemas and gestures tend to rely on the idea that metaphor is at play (McNeill, 1992; Cienki, 2013; Mittelberg, 2006, 2018; Williams, forthcoming). McNeill (1992) calls abstract iconics ‘metaphorics’, and Mittelberg (2006) claims that gestures are “outlets for metaphoric articulations of emotions and abstract thought...” (p. 14).

treated as such, even while they interact. The use of objects or not in gestures was stated to be a third dimension. Finally, a discussion of the possible relations between (iconic) gestures and schemas was offered in Section 2.5, also highlighting the need to consider the role of language.

While young children do not produce metaphoric gestures (in which image schemas could possibly be involved), it seems reasonable to expect *increased levels of abstractness* in children's gestures, corresponding to the distinction between mimetic and image schemas. This chapter has made the point that gestures from the Observer Viewpoint are more abstract than gestures from the Character Viewpoint. Furthermore, the assumption made here is that in terms of Mode of Representation there is a cline of abstractness: Acting > Representing > Moulding & Tracing > Drawing. On this basis, we can formulate the following set of possibly competing hypotheses concerning children's gestures in the fourth year of life:

H1. There will be a progression from the Character to the Observer Viewpoint.

H2. There will be a progression from Acting, to Representing, to Moulding & Tracing, and finally to Drawing, in terms of Mode of Representation.

H3. More abstract gestures (in terms of either H1 or H2) would (a) occur more with language than less abstract ones and (b) this language would be more grammatical than lexical.

Chapter 3. Methods

3.1 Introduction

In order to investigate a possible developmental transition from more concrete, mimetic-schematic gestures to more abstract, image-schematic gestures, an analysis of the gestures of three Swedish children was conducted. This chapter presents the data used for the study (3.2) and the procedure of their analysis, including transcription, identification and coding (3.3). Then, an overview of the general analysis of iconic gestures and language is presented (3.4). Finally, a summary of the chapter is provided, and the hypotheses presented in chapter 2 are operationalised as specific hypotheses (3.5).

3.2 Data

The data consisted of longitudinal video-recordings and transcriptions of the interactions between three Swedish children and their mothers. The children in this study are the same as those analysed by Zlatev and Andrén (2009) and Andrén (2010), but here examined at a later age. The children's ages ranged from 34 months to 47 months, spanning roughly the entire fourth year of life (Table 3), which is both the period when gestures are expected to change in the direction of abstractness, and the period that has been relatively less studied in the developmental literature (cf. Andrén, 2010 and Section 2.5).

For the present study, four data points were chosen for each child, depending on availability. The recordings used were collected and transcribed by Ulla Richthoff (2000) and are available online at <https://www.childes.talkbank.org> (Plunkett & Strömqvist, 1992). The transcriptions were in CHAT format, which may be processed with the CLAN software (MacWhinney, 2000). All recordings were between 21 and 34 minutes in duration.

Table 3. The data-set with 4 data points per child, with age given in months (m) and days (d)

Child	#1	#2	#3	#4
ANTON	34m 4d	38m 15d	42m 15d	47m 3d
BELLA	36m 7d	38m 19d	42m 14d	45m 12d
HARRY	37m 21d	39m 11d	40m 1d	46m 18d

3.3 Procedure

3.3.1 Transcription

As mentioned, all but two transcriptions had been prepared in CHAT format by Ulla Richthoff. The following two recordings were transcribed by myself: Bella, #3, and Bella, #4. Furthermore, some of the other recordings had unfinished transcriptions, in which case I transcribed the remainder. In these latter cases, after identifying each instance of gestures in the video-file (see below), the child's speech directly preceding, coinciding, or following the gesture was transcribed.

3.3.2 Identification and semiotic ground

As a first step all gestures (not just iconic ones) were identified based on the criteria for high representational complexity and/or communicative intent (see Section 2.4), and their dominant semiotic grounds were decided using a combination of the first- and second-person methods, summarized in Table 1 (Section 2.2.1).

Indexical (IND) gestures were identified as those gestures that bring things and people to others' attention. Usually they were pointing gestures, though they were also requests or showing/giving something to someone. IND gestures were further coded for whether they were declarative (DEC) or imperative (IMP). These two types were identified based on their level of communicative intent (Section 2.4, Table 2), the latter being on level 2, and the former on level 3. Any possible abstract pointing was similarly coded as a declarative indexical gesture. Given that imperative pointing does not qualify as a gesture according the criteria used (see Section 2.4.1), IMP gestures did not undergo further analysis.

Conventional (CON) gestures were identified as expressions that display intersubjectivity in the “the strong sense of being mutually known, rather than being (individual) ‘habits’.” (Zlatev & Andrén, 2009, p. 386). Thus, gestures such as DONE, SHRUG, and early negation, which includes waving the hands and turning the head away, were coded as CON. These gestures are deemed to be on level 3 of communicative explicitness.

Iconic (ICO) gestures were identified based on their resemblance between the movements of the body and the properties of the referent. These gestures are on level 3 of communicative explicitness, referential complexity, or both.

Gestures were coded as UNCLEAR in cases where either the classification of the gesture was impossible due to unclear speech (given that the language used is often crucial to the interpretation of gestures, see Section 2.5), or the movement was simply unclear in whether it was an intentional communicative and representational gesture or not. In this case, intuition and empathy-based understanding were not enough to determine the expression’s dominant semiotic ground, and thus these were not included in further analysis.

3.3.3 Coding

Coding was carried out by myself for every child and took place using the CLAN software (MacWhinney, 2000) using so-called coder mode. The coding scheme is shown in Figure 6.

```

%ges:
" $ICO
:UNCLEAR
" $IND
:IMP
:DEC
:UNCLEAR
" $CON
:UNCLEAR
%spa:
" $VIEW
:CHA
:OBS
" $MOR
:ACT
:REP
:MOU
:TRA
:DRA
" $IMI
:YES
:NO
" $OBJ
:YES
:NO
" $SPCH
:YES
:NO

```

Figure 6. The gesture coding scheme in CLAN's "coder mode"

For example, when Anton was playing with Playmo, he pretended to eat the toy food. This was coded as ICO on the gesture tier (%ges), and as \$VIEW:CHA \$MOR:ACT \$OBJ:YES \$IMI:NO \$SPCH:YES on the speech act tier (%spa). In other words, the viewpoint was coded as Character Viewpoint because he was eating the food. The Mode of Representation (MOR) was Acting because his hands did not represent anything other than his own hands. An object was used, in this case the toy meat, and because he said *namm namm*, speech was coded as YES (see Appendix A).

Within CLAN, the commands window allows for automatic analysis of codes using either the **FREQ** command, which calculates the frequencies of specific words in a file, or the **CHAINS** command, which is used to track sequences of interactional codes across speakers. The former was used to calculate of the individual gesture components coded using %ges. Thus, the software indicated how often the codes IND, CON, and ICO appeared in each individual file, allowing for further analysis using the excel program. The **CHAINS**

command, on the other hand, was used to analyse the distribution of the values for the different codes. For example, Figure 7 shows that on line #48 in the transcription, an original (not imitated) iconic gesture was coded, which was produced from the Character Viewpoint, through the Acting MOR, with an object and no corresponding speech.

\$imi:no	\$mor:act	\$mor:rep	\$obj:no	\$obj:yes	\$spch:no	\$spch:yes	\$view:cha	\$view:obs	line #
1	1	0	0	1	1	0	1	0	48

Figure 7. CHAINS %spa

3.3.4 Statistics

Because of the low number of participants, descriptive statistics were preferred over inferential statistics. Standard deviation (SD) was computed for the gesture tokens for both Viewpoint and MOR in order to help visualise inter-individual variation.⁶ A second coder would ideally have coded at least part of the data for consistency, but for practical reasons, this was not possible. This limitation might have an effect on the study's reliability.

3.4 General analysis

3.4.1 Iconic gestures

When the gesture was specified for ICO, classification in terms of Viewpoint and MOR was performed. Identification of Viewpoint was not based on handshape, which, as pointed in Section 2.4, is rather indicative of MOR, which should be held distinct. Instead, identification was based on both the child's gaze and the location of the gesture in the gesture space. Following Frederiksen (2017), gaze, body orientation, and the movement of the hands relative to the body were considered. When the gesture was performed in the speaker's locus, which is the "point in gesture space inhabited by the speaker's upper body" (ibid, p. 685), or if the gaze was imitating the referent's gaze direction (or both), it was considered a Character

⁶ Here, the SD is only a rough measure as only three children are studied.

Viewpoint gesture (Engberg-Pedersen, 2015). On the other hand, when the speaker's gesture was performed outside their own locus, or their gaze was on the addressee (or both), the gesture was coded as Observer Viewpoint.

A combination of first-person and second-person methods was used in order to decide (see Section 2.2.1), and likewise for MOR, which followed the description in Section 2.4.2 and was identified based on whether the gesture was of the *Acting* type, *Representing*, *Moulding*, *Drawing*, or *Tracing*. To remind, Drawing corresponded to the outlining of an object's shape or general attributes with the finger, and Tracing to the tracing of a path or trajectory (Brown et al., 2019). Moulding included both a gesture where the hands mould an imaginary object, and a measuring gesture, where the hands are held out to indicate size. This could either be Character or Observer Viewpoint, depending on whether the gesture was performed in the child's own locus or a locus outside of their own gesture space, as in measuring a distance indicated on a map.

It was also coded whether the gesture could be considered an imitation of the caregiver in the immediately preceding discourse. Following Zlatev (2014), imitation was coded only when the child repeated the caregiver's gesture within the space of two turns. Given that such imitated actions could not be attributed to level 3 on either communicative intent or representational complexity, they were not considered as gestures and therefore not subjected to further analysis.

Finally, gestures were coded with regard to whether they were performed while using an object or not. In the case of Tracing, an object was coded when it was performed and based on the depictions in a book or puzzle. Similarly, Drawing was coded with an object when shapes within a book or puzzle were outlined, but without an object when the drawing happened in a *supporting* space, which is when a table or a wall acts as an easel for the gesture (Andrén, 2010). This distinction was made based on whether the gesture was produced because of an already existing outline or whether the outline was of their own making.

3.4.2 Language

Co-gesture speech (i.e. speech aligned with the whole gesture phrase, and not necessarily the stroke, see Section 2.4) was specified (YES/NO). Then, the type of language was noted. Three groupings of linguistic expressions were common: (1) Demonstratives, (2) Lexical expressions, and (3) Grammatical expressions. Demonstratives bring attention to the referents and include expressions such as “like this/that” (*sån här, sån, såna, så och så*, etc.) or “here/there” (*här/där*). Lexical expressions are those that are open-classed and have referential content, such as nouns, adjectives, or action verbs. Finally, grammatical expressions are those that contain prepositions (see Section 2.5), sentence adverbs like “not” (*inte*), and have reversibility, such as “up/down” (*långt upp i loftet / den ha vält ner*), “all/none” (*allihopa / ingen*), and “in/out” (*gå ut / här inne*) (see Section 2.3.1). These expressions belong to closed classes, and express more abstract meanings such as causal, spatial, and temporal relations (Talmy, 2015; McCune, 2006). Speech was coded as UNCLEAR when it was not possible to determine to which group of linguistic expressions it belonged. Some linguistic expressions, similar to the case with the semiotic ground of gestures, contain one or more of the above types of expressions. Thus, the dominant category was decided using the following factors:

- 1) It was deemed grammatical (GRA), if the meaning of the gesture was the same as the meaning of a grammatical expression used, no matter where the stroke falls. For example, if an expression includes the preposition “in” (*i*), and the gesture clearly means that something goes into something else, like “in the air” (*i loftet*) where the gesture is a finger tracing a trajectory through the air like the path of an airplane (see Appendix D).
- 2) It was deemed Demonstrative (DEM) if (a) the meaning of the stroke and post-stroke hold overlapped with a determiner, and (b) the gesture did not have the meaning of a grammatical term. For example, the expression “they are not here” (*de är inte jo här*) was treated so since (a) the gesture was co-temporal with “here” (*jo här*), rather than

the word “not” (*inte*), and (b) did not express the meaning of the latter (see Appendix E).

3.5 Summary and Predictions

This chapter provided an overview of the methodology of the study. Based on the theoretical overview and the general hypotheses put forth at the end of chapter 2, the following more *specific hypotheses* (SHs) could be formulated concerning (predominantly) iconic gestures, as operationalisations of the general hypotheses:

SH1. *Character Viewpoint gestures will outnumber Observer Viewpoint at the beginning of the studied time period, and vice versa for the end of the period.*

SH2. *Acting, in terms of MOR, gestures will outnumber Representing gestures at the beginning of the time period, and vice versa at the end of the time period, and this difference will be even clearer for Moulding, Tracing, and Drawing.*

SH3. *Observer Viewpoint gestures, as well as Moulding, Tracing, and Drawing will (a) more often occur with speech than the other, less abstract gestures, and (b) be accompanied by more Grammatical expressions, with Demonstrative and Lexical expressions accompanying Character Viewpoint, Acting, and Representing gestures.*

Chapter 4. Results

4.1 Introduction

The results presented in this chapter were obtained based on the methods and hypotheses presented in Chapter 3. The first part of this chapter considers the overall proportions of gesture components throughout the period (4.2). The second part considers the iconic gestures and their production in regard to Viewpoint and Mode of Representation (MOR) (4.3, 4.4). Finally, the third part of the chapter is concerned with the children's corresponding speech (4.5). In this chapter, the children are considered together as a group due to the overall low number of ICO gestures. Individual differences are discussed in Chapter 5.

4.2 Gesture proportions

The overall proportions of types of gestures based on the predominant semiotic ground throughout the period are displayed in Table 4.

Table 4. The number of tokens for each gesture type (= dominant semiotic ground) and the corresponding percentage of gestures that belong to each type.

Child	Data point	#1 (%)	#2 (%)	#3 (%)	#4 (%)
	Gesture type				
Bella	ICO	10 (14)	6 (7)	16 (24)	3 (6)
	IND	41 (55)	51 (59)	20 (29)	18 (35)
	CON	23 (31)	29 (34)	32 (47)	31 (60)
Harry	ICO	0 (0)	8 (17)	8 (11)	7 (9)
	IND	61 (84)	22 (48)	47 (65)	65 (88)
	CON	12 (16)	16 (35)	17 (24)	2 (3)
Anton	ICO	0 (0)	16 (33)	4 (8)	16 (26)
	IND	61 (78)	27 (55)	34 (67)	37 (61)
	CON	17 (22)	6 (12)	12 (24)	8 (13)
Total	ICO	10 (4)	30 (17)	28 (15)	26 (14)
	IND	163 (72)	100 (55)	101 (53)	120 (65)
	CON	52 (23)	51 (28)	61 (32)	41 (22)

In general, the highest proportion of gestures were those with a predominantly indexical (IND) ground, with an overall number of 484 gestures performed throughout the period for all of the three children. This was followed by gestures with a predominantly conventional (CON) ground, with the overall number of 205 gestures performed throughout the period for all of the three children. Finally, gestures with a predominantly iconic (ICO) ground were 97. Figure 8 shows the relative proportions for each respective data point.

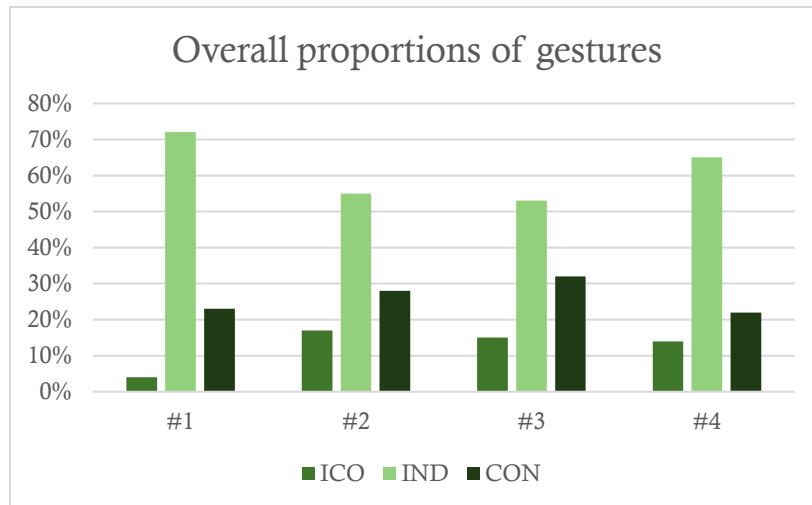


Figure 8. Overall proportions of gesture types (dominant ground) over time

On the whole, the relative proportions were rather similar, and the observed variation over time may be attributed to differences in the activities that the children partook in during the different data points. For example, in #1 neither Harry nor Anton produced any ICO gestures, possibly due to the fact that the children were engaged in reading, eating, and looking at pictures. The following sections evaluate each of the three specific hypotheses presented at the end of Chapter 3.

4.3 Hypothesis 1

Hypothesis 1 predicted that Character Viewpoint gestures would outnumber Observer Viewpoint gestures at the beginning of the time period, and vice versa for the time period. Table 5 presents the data points for the gestures performed from each Viewpoint, as well as the *standard deviations* (SD), and Figure 9 shows the relative proportions of Viewpoint throughout the time period for all three children.

Table 5. Tokens of gestures from the Character and Observer Viewpoints throughout the time period.

	Character Viewpoint	SD	Observer Viewpoint	SD
#1	8	3,77	2	0,94
#2	21	4,97	10	0,47
#3	12	4,97	16	2,19
#4	9	1,41	17	5,66
Total	50		45	

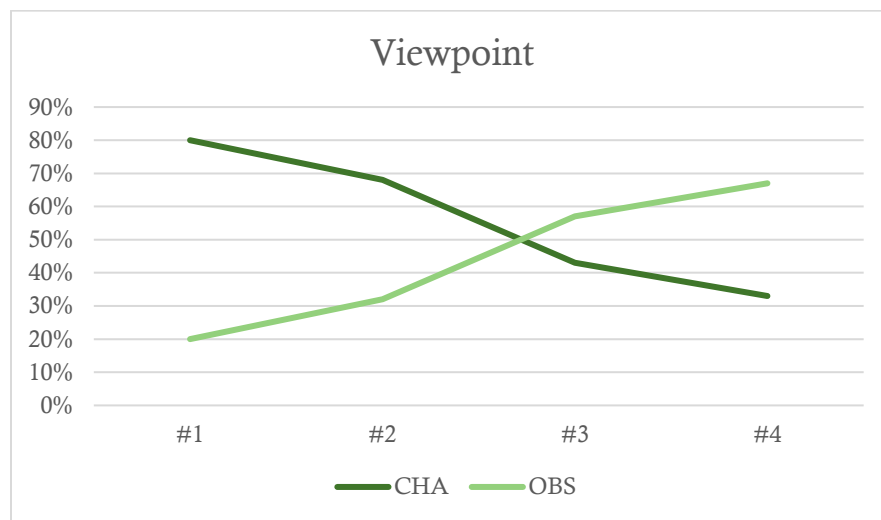


Figure 9. Character and Observer Viewpoint gestures over time.

As can be seen, while the proportion of Character Viewpoint gestures was at 80% at the beginning of the period, there was a clear decrease in their use over time as Observer Viewpoint gestures steadily rose throughout the period, going from a proportion of 20% up to 67% of all gestures performed at the end of the period. Thus Hypothesis 1 was supported. The significance of the SD is discussed in Chapter 5.

4.4 Hypothesis 2

The second hypothesis predicted that Acting gestures would outnumber more abstract gestures at the beginning of the time period, and vice versa for the end of the time period, and this difference will be even clearer for those gestures grouped under the heading “Other” (Moulding, Tracing, and Drawing). Table 6 presents the data points for the gestures performed from each MOR, as well as the SD for each age bracket. Figure 10 shows the general trend of use of the different Modes of Representation throughout the period.

Table 6. Tokens of gestures in the Acting, Representing, and Other MORs.

	Acting	SD	Representing	SD	Other	SD
#1	6	2,83	2	0,94	2	0,94
#2	16	6,18	10	1,70	4	1,23
#3	9	4,24	17	1,23	2	0,47
#4	5	1,70	6	0,0	14	5,91
Total	36		35		22	

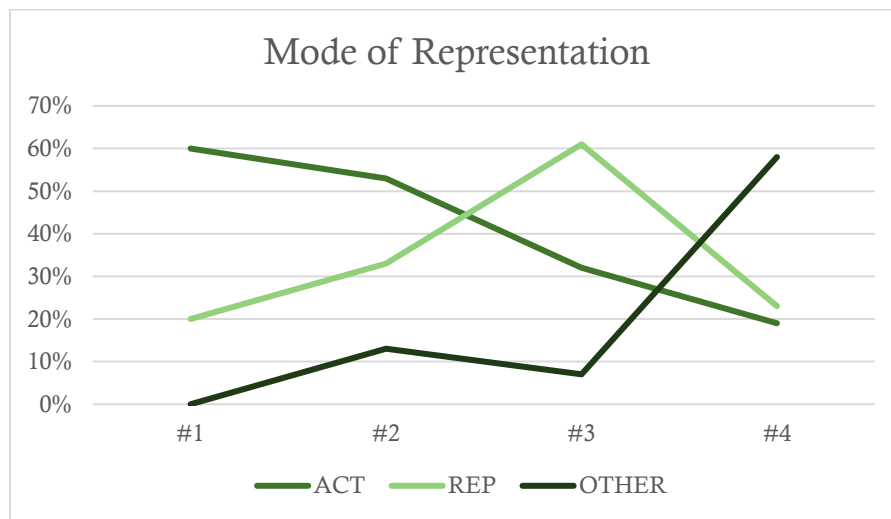


Figure 10. MOR over time.

To remind, Harry and Anton performed no iconic gestures in age bracket #1. Despite this, Acting gestures clearly dominated at the beginning of the time period, decreasing steadily with each data point. Conversely, Representing gestures increased in use throughout the time period, though drastically decreasing around age bracket #3, giving way to a steep rise in Other gestures. These results clearly support Hypothesis 2, with the SD discussed in Chapter 5.

4.5 Hypothesis 3

Hypothesis 3 stated that (a) Observer Viewpoint gestures, as well as Moulding, Tracing, and Drawing will more often occur with speech than the other, less abstract gestures, and (b) Demonstrative and Lexical expressions will accompany more concrete gestures (i.e. Character Viewpoint, Acting, and Representing), while Grammatical expressions will accompany more abstract gestures (i.e. Observer Viewpoint, Moulding, Tracing, and Drawing). Figures 11a and 11b show the proportion of gestures used with speech throughout the period, with respect to Viewpoint, and MOR.

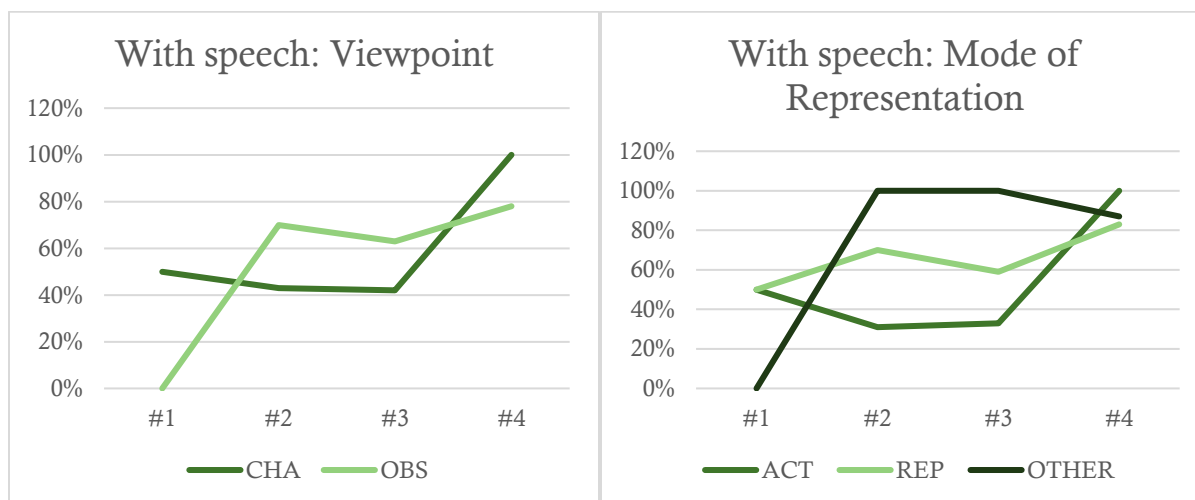


Figure 11a and Figure 11b: Viewpoints and MORs with speech.

During the first age bracket (#1), the two most abstract types of gestures (OBS and OTHER) were least likely to be used with speech, but it should be reminded that the two boys,

Anton and Harry, performed no iconic gestures in their first data points (see Section 4.2). With that in mind, one can see from age bracket #2 that both Character and Acting gestures were used less with speech during the beginning of the period than Observer, Representing, and Other gestures, in line with the hypothesis. In general, though, there was a rise in the use of speech with all gestures, including Character Viewpoint and Acting gestures, which, at the end of the studied period (#4) were just as frequently used with speech as Observer and Representing gestures. Table 7 shows the tokens of gestures used with speech for the different values of both Viewpoint and MOR.

Table 7. Tokens of speech in regard to Viewpoint and MOR.

	CHA	OBS	ACT	REP	OTHER
#1	4	0	3	1	0
#2	9	7	5	7	4
#3	5	10	3	10	2
#4	9	14	5	5	14
Total	27	31	16	23	20

Concerning the second part of the hypothesis, Table 8 shows the data points for the linguistic expressions used during gestures from the Character and Observer Viewpoints.

Table 8. Tokens of linguistic expressions in regard to Viewpoint

	Character Viewpoint			Observer Viewpoint		
	DEM	LEX	GRAM	DEM	LEX	GRAM
#1	0	3	1	0	0	0
#2	4	5	0	1	5	1
#3	0	4	0	0	9	1
#4	2	6	1	6	3	6
Total	6	18	2	7	17	8

Grammatical expressions occurred only twice together with gestures from the Character Viewpoint, while occurring far more with Observer Viewpoint gestures, especially at the end of the studied period. Most co-gesture expressions occurring with Character Viewpoint gestures were lexical in all age brackets, while almost half of the expressions in age brackets #2 were demonstrative. The lack of data from age bracket #1 is, again, due to the fact that only Bella performed ICO gestures for that period. Lexical expressions were the most frequent as co-gesture speech for both Character and Observer Viewpoints, with demonstratives being intermediary.

Table 9 shows the data points for the linguistic expressions used during the Acting, Representing, and Other (i.e. Moulding, Tracing, Drawing) Modes of Representation. As can be seen, lexical expressions are almost exclusively used during Acting and Representing gestures throughout the whole time period (#1 - #4). Grammatical expressions occurred solely with Representing and Other gestures (Moulding, Tracing, and Drawing), and demonstratives were overwhelmingly used with Other gestures (see Chapter 5 for a closer look at these individual Modes of Representation). These results can in part be said to support the hypothesis, though the lack of demonstratives with Acting and Representing gestures was unexpected.

Table 9. Tokens of linguistic expressions in regard to Acting, Representing, and Other Modes of Representation

	Acting			Representing			Other		
	DEM	LEX	GRAM	DEM	LEX	GRAM	DEM	LEX	GRAM
#1	0	3	0	0	0	1	0	0	0
#2	1	4	0	0	6	1	4	0	0
#3	0	3	0	0	9	0	0	1	1
#4	1	4	0	1	4	0	6	1	7
Total	2	14	0	1	19	2	10	2	8

4.5 Summary

The key results may be summarised as follows. At the beginning of the studied period, Character Viewpoint gestures and gestures from the Acting MOR were performed far more frequently than Observer Viewpoint gestures and gestures from the Representing and Other MOR. Towards the end of the studied period, this trend reversed. Interestingly, during age bracket #3, there was a sharp decline in the use of Representing gestures, and a sharp rise in the use of Other gestures.

Concerning co-gesture speech, gestures performed from the Observer Viewpoint and Representing MOR were used far more often with speech than gestures from the Character Viewpoint and Acting MOR, and Moulding, Tracing, and Drawing gestures (Other) were almost exclusively used with speech. Towards the end of the studied period, there was a rise in the use of co-speech gesture for Character Viewpoint and Acting, so that during age bracket #4, almost all gestures were performed with language. There were only two grammatical expressions that coincided with the Character Viewpoint, and a small number of demonstrative expressions. Observer Viewpoint gestures coincided with lexical and grammatical expressions, as well as a small number of demonstrative expressions. With regard to MOR, both Acting and Representing were almost exclusively used with lexical expressions, while grammatical and demonstrative expressions occurred most often with Other gestures.

Thus, all three hypotheses were supported. Moulding, Tracing, and Drawing were here treated homogenously for clarity, but in the next chapter, each of these Modes of Representation will be looked at separately, as well as factors like inter-individual variation, and relations to general hypotheses in Chapter 2.

Chapter 5. Discussion

5.1 Introduction

The previous chapter presented the results of the study based on the hypotheses given at the end of Chapter 3. Here, it is appropriate to take the discussion a step up and review the results with regard to the more general hypotheses put forth at the end of the theoretical discussion in Chapter 2. Up until now, image schemas and mimetic schemas have not been discussed in relation to the data presented, and thus each section will take a look at their possible role in connection to the hypotheses. To remind, the general hypotheses were as follows:

H1. There will be a progression from the Character to the Observer Viewpoint.

H2. There will be a progression from Acting, to Representing, to Moulding & Tracing, and finally to Drawing, in terms of Mode of Representation.

H3. More abstract gestures (in terms of either H1 or H2) would (a) occur more with language than less abstract ones and (b) this language would be more grammatical than lexical.

5.2 Viewpoint (H1)

The first hypothesis predicted that the children's gestures would begin as mainly Character Viewpoint gestures, but with time they would be performed mainly from an Observer Viewpoint. This prediction was based on gesture theories put forth by McNeill (1992), and on research carried out by Zlatev (2014), who argued that early iconic gestures are derived from mimetic schemas (see Sections 2.3.2 and 2.5). Williams (forthcoming) similarly proposed that, unlike image schemas, mimetic schemas underlie gestures where their meaning is "enacted" from a first-person perspective. In the present study, at the beginning of the time period, most gestures were indeed performed from a Character Viewpoint, though only Bella performed iconic gestures during age bracket #1. However, as can be seen from Table 5 (see Section 4.3), in age bracket #2 inter-individual variation (see SD) is higher for Character Viewpoint gestures than for Observer Viewpoint gestures, with the majority of Character

Viewpoint gestures performed by one child, in this case Anton. Observer Viewpoint gestures, on the other hand, were more evenly spread out among the children, suggested by the low SD. This trend is similar for age bracket #3, with the majority of Character Viewpoint gestures performed by Bella. No such gestures were performed by Harry, and only one was performed by Anton. This upsurge in Character Viewpoint gestures for Bella involved pretend play.

In age bracket #4 there was a trend reversal. Inter-individual variation is higher for the Observer Viewpoint, suggesting that the majority of those gestures were performed by one child, in this case again, Anton, while both Harry and Bella performed only two each. It can be noted that Anton was 47.1 months old in age bracket #4, and this may have contributed to the upsurge in Observer Viewpoint gestures (see 5.3 for more discussion). Despite this, the low inter-individual variation in age-brackets #2 and #3 for Observer Viewpoint suggests that the transition from Character to Observer period may have started earlier than the studied period, in line with the findings of Zlatev's (2014) study, where all gestures performed from Observer Viewpoint were performed with objects, rather than being empty-handed (see Section 5.5).

A final point to make regarding Viewpoint is that of “dual viewpoint” (Engberg-Pedersen, 2015). Anton, in age bracket #2, performed a gesture that consisted of him holding Playmo “meat”, and allowing his Playmo doll to “taste” it. This was analysed as Character Viewpoint, since his eyes were fixed on the doll as if within an interaction. However, he uttered *namnam* – indicating “eating”, and thus an action from Observer Viewpoint. In age bracket #4, Bella and her mother played with Lego blocks, and as they discussed where to put a particular block on the Lego house, Bella displayed a gesture from the Character Viewpoint. She held the block away from her own locus and looked up at it as if she were a small Lego character looking at the space which the block should occupy on the house. She said, simultaneously, “here?” (*sån?*), and then looked at her mother, continuing to hold the gesture. She thus removed herself from the action (Character) and placed herself outside of it (Observer). These two “dual viewpoint” examples were treated as single gestures and were

thus not marked twice in Table 4 (Section 4.2). They were, however, marked twice in Table 5 (Section 4.3), due to the plurality of viewpoint.

5.3 Mode of Representation (H2)

The second hypothesis predicted that Acting would give way to Representing, which would then give way to the more abstract gestures Moulding, Tracing, and Drawing. This prediction was based on several studies claiming that those gestures that are motivated by mimetic schemas are similar to instrumental actions (Williams, forthcoming; Andrén, 2010; Zlatev, 2014; Cienki, 2013; Mittelberg, 2018; Novack et al., 2018). In support, Acting gestures were more frequently produced than other gestures at the beginning of the time period (see Table 6). However, the inter-individual variation was quite large for Acting gestures in both age bracket #2 and #3. In age bracket #2, the majority of Acting gestures were performed by Anton, while none were performed by Harry, and only 2 by Bella. In age bracket #3, all Acting gestures were performed by Bella, and in age bracket #4, the majority were performed by Harry, while only one was performed by Bella, and none by Anton. This variation may be due to the activities that the children took part in with Acting, i.e. mimetic-schematic gestures depending more on context than on age.

Representing gestures, on the other hand, were more evenly spaced between the children, as can be seen from the overall low SD in Table 6 (Section 4.4). During age bracket #2, Anton performed one Representing gesture, while Harry and Bella performed a similar amount (5 and 4, respectively). During age bracket #3, the children all performed a similar amount of Representing gestures, and during age bracket #4, they all performed the same amount of Representing gestures. The difference between the inter-individual variation seen between Acting and Representing gestures is indicative of a steady movement towards the more abstract Representing gestures, which may have begun before the studied period.

When it comes to “Other” gestures, there was quite a lot of variation. In age bracket #2, Harry performed three Other gestures, Anton performed one, and Bella none at all. In age bracket #3, Bella and Harry each performed one Other gesture, and Anton performed none. The inter-individual variation in age bracket #4 is extreme. During that time, Anton

performed the majority of Other gestures (13), Harry performed only one, and Bella performed none at all. This could perhaps be due to the fact that Anton is older than the other two children during #4.

Drawing strong conclusions about the data would be far-fetched, though it is interesting to note that Anton produced by far the most Other gestures in age bracket #4, leading one to wonder whether this trend would continue for the other children after the studied period. This is something to examine in future research.

5.4 Language (H3)

The final hypothesis had two parts. The first stated that language would accompany abstract gestures more often than concrete gestures, and the second stated that the expressions that accompany those abstract gestures would be predominantly grammatical. These predictions were based on the idea that abstract gestures are more likely to be motivated by image schemas, which have also been argued to be grammatical (Zlatev, 2005) and reversible (McCune, 2008) expressions in language.

Figures 11a and 11b (Section 4.5) show that the first part of the hypothesis was supported, though only during age brackets #2 and #3. The Moulding, Tracing, and Drawing (“Other”) gestures were almost exclusively used with speech throughout, though during age bracket #1, Bella performed two Tracing gestures without speech, tracing a path within a book. Age bracket #4 showed an increase in co-gesture speech for both Acting and Character Viewpoint gestures, to the point where all gestures, and not only the more abstract ones, were almost entirely used with speech. It would be interesting to know if this trend would continue after the studied period. If so, this would indicate that the semiotic systems of language and gesture solidify their link in cognition around the end of the fourth year, working together to relay ideas in a similar way to the polysemiotic communication of adults.

As for the second part of the hypothesis, grammatical expressions accompanied mostly Other gestures, though also occurred (though much less) with Representing gestures. As can be seen in Table 11, demonstrative, but crucially not lexical expressions occurred just as frequently as grammatical expressions.

Table 10. Tokens of linguistic expressions used with Other gestures

	Moulding	Tracing	Drawing
Demonstrative	5	2	3
Lexical	0	1	1
Grammatical	1	5	2

The majority of Moulding gestures with demonstrative expressions indicates that Moulding is inherently a gesture meant for other people, which calls to question the role of image schemas in the production of such gestures. To remind, Cienki (2013) and Mittelberg (2018) both claim that gestures motivated by image schemas should be less specific and less consciously produced (see Section 2.5). This is the exact opposite of the role that Moulding gestures seem to fulfil, and gives credence to the idea that Acting, Representing, and Moulding gestures belong together in the way that they are produced, while Tracing and Drawing are different, and more image schema-like.

Tracing and Drawing gestures were produced with a similar number of demonstrative and grammatical expressions. *Appendices A-E* display the linguistic expressions used in conjunction with all Modes of Representation, including an explanation of the gesture, the reasoning behind some of the more complex cases, and any image-schematic meanings. All grammatical expressions and/or corresponding gestures in the Tracing and Drawing MOR can be argued to correspond to image schemas. On the other hand, only one grammatical expression and its corresponding gesture in the Representing MOR had an image-schematic meaning. The image schemas that occurred are limited, seemingly restricted to the SOURCE-PATH-GOAL and CONTAINER schemas.

All of the image-schemas that occurred with Drawing gestures were exclusively of the CONTAINER type. Two image-schematic Drawing gestures were accompanied by the demonstrative expressions “play with them” (*leka med dom*) and “it is not here, though” (*de är inte jo här*). These gestures group objects and colours together in a CONTAINER using a

drawing motion with the finger. It is not strange that Drawing exclusively accompanies the CONTAINER schema, since Drawing is static (see Section 2.4.3), while Tracing, which often accompanies the SOURCE-PATH-GOAL schema, is dynamic.

The image schemas that occurred within the Tracing MOR were both CONTAINER and the SOURCE-PATH-GOAL image schema. In some cases, within the same expression both image schemas occurred: one that corresponded with the gesture and one part of the phrase, and one that corresponded to another part of the phrase but not the gesture, i.e. the expression “far up into the air” (*långt upp i loftet*). The expression was accompanied by a gesture where the finger traces a path into the air. Thus, *upp* and the gesture both express SOURCE-PATH-GOAL meaning, while the preposition (*i*) indicates the CONTAINER image schema, with the air being the container into which the airplane flies.

The fact that image-schematic meaning occurred in abundance in Tracing and Drawing MOR indicates their abstractness in relation to the other Modes of Representation. It is also striking how often they co-occurred with grammatical expressions that corresponded exactly to these same image schemas. On the one hand, this supports an image schema analysis of such gestures. On the other, it argues against their pre-linguistic nature. If not directly derived from language, they appear at least to be closely connected to it.

5.5 Other issues

This section reviews the aspects of the data that does not pertain to the hypotheses specifically, including some of the UNCLEAR gestures and the use of objects. Before the age of 30 months, empty-handed Representing or Other gestures were almost non-existent (Andrén, 2010; Zlatev, 2014) (see Section 2.5). The data in Zlatev’s (2014) study showed that gestures from the third-person perspective, or Observer Viewpoint, were all performed with objects. Furthermore, Zlatev and Andrén (2009) suggested that speech and iconic gestures begin their integration by using objects as a bridge. Thus, the pattern of emergence of more abstract gestures might come to light when the simultaneous use of objects is examined. Table 12 shows the number of gestures used with objects with the Character and Observer Viewpoints, and the proportions of all such gestures.

Table 11. Tokens and percentages of object-use during gestures with different Viewpoint

	Character Viewpoint (%)	Observer Viewpoint (%)
#1	5 (63%)	2 (100%)
#2	14 (67%)	10 (100%)
#3	9 (75%)	10 (63%)
#4	3 (38%)	13 (68%)

As can be seen, objects tended to be used with Observer Viewpoints to a greater extent, which is to be expected if object use contributes to the greater abstractness of gestures.⁷ This tentative hypothesis, however, does not seem to be supported when MOR is considered together with Viewpoint (Table 13).

Table 12. Tokens of object-use during Viewpoint and MOR combination gestures

	CHA;ACT (%)	CHA;REP	OBS;REP (%)
#1	5/6 (83%)	0/2	0/0
#2	14/15 (93%)	0/2	6/7 (86%)
#3	12/12 (100%)	0/3	10/15 (67%)
#4	0/0	0/1	3/6 (50%)

Interestingly, Character Viewpoint Acting gestures with object-use increased over time rather than decreased. On the other hand, the trend regarding object-use and Observer Viewpoint Representing gestures goes steadily down, starting at 86% during age bracket #2 and ending at 50% during age bracket #4. The fact that no objects were used during Character Viewpoint Representing gestures is expected, since the entire body represents something other than itself. Again, the situations and activities that the children took part in during the sessions may have

⁷ It would be interesting to know how these numbers look when the activities during the sessions are restricted to an environment of chatting only. If Observer Viewpoint gestures would increase in number over time, this would indicate the important role of objects, especially if, with objects, those gestures far outnumber gestures without objects at the beginning of the studied period.

an effect on the data, and it would be interesting to see the results of a study which looks at sessions between children and others in which they are only chatting. These results also shine a light on an important issue regarding the overlap between Viewpoint and MOR. In general, the Character Viewpoint could occur with all Modes of Representation except for Drawing and Tracing, which were exclusively from the Observer Viewpoint, unless, again, they took place in a way which suggests the finger as a representation of a writing implement (see Section 2.4.3). Furthermore, the Acting MOR never occurred with the Observer Viewpoint, except in the cases of dual viewpoint, where there was a movement or co-occurrence of one viewpoint to the other (Section 5.2).

Looking at abstract pointing, which, according to McNeill (1992) may not occur until around the age of 12 years old, the two sole (possible) cases were marked as UNCLEAR, due to unclear speech. Again, speech is important for understanding a gesture's referential meaning (see Section 2.5), and this inevitably has an effect on coding (Cartmill et al., 2012; Müller et al., 2014; Frederiksen, 2017). In the data analysed, there were a few other cases of UNCLEAR gestures, though not as many as occurred in Anton's data, specifically age-bracket #4. Most of these gestures seemed like Drawing gestures, but it was impossible to tell due to the fact that, for the most part, his mouth was full of food, and his speech was simply unclear. During the analysis of Anton's linguistic expressions, one expression was labelled as UNCLEAR, even though the gesture itself was not (Character Viewpoint + Representing).

Finally, there were no truly metaphoric gestures in the data, as both metaphoric phrases and gestures appear much later in development than non-metaphoric phrases and gestures (Andrén, 2010; Winner, 1988; Waggoner & Palermo, 1989). However, there was one gesture in particular that was misunderstood by the parent to be a case of metaphor. The gesture was performed by Harry, who had remarked something about the time of an activity, making a gesture where he waved his right arm/hand seemingly over his right shoulder. His mother did not understand the word that came with the gesture, interpreting it as an adult metaphoric gesture meaning "yesterday." However, when she tried to confirm that this was what he said, he refused, saying more clearly "today". Thus, the gesture received the label UNCLEAR.

5.6 Summary

This chapter looked at the hypotheses put forth at the end of Chapter 2 in view of the results in Chapter 4. The first hypothesis regarded Viewpoint (5.2), and we saw that as a group, Character Viewpoint gestures were in the majority at the beginning of the time period. However, the extreme variability in the inter-individual production of Character Viewpoint gestures suggests that *Observer Viewpoint gestures are a more reliable indicator of the transition towards abstractness in the data*. A larger data-set might shed more light on this issue.

The second hypothesis regarded Mode of Representation (5.3), and we saw that as a group, Acting gestures far outnumbered Representing gestures at the beginning of the time period. However, similarly to the case with Viewpoint, there was large inter-individual variability in the production of Acting gestures, which is natural, given that the vast majority of Character Viewpoint gestures are also Acting gestures.

The production of Moulding, Tracing, and Drawing gestures were more evenly spaced between the children during age brackets #2 and #3, while their production doubled during age bracket #4. This might have been due to the fact that Anton, who performed all but one of these gestures during that time, was older than the two other children. A data set that extends beyond the fourth year might shed more light on this.

The third hypothesis regarded the production of language (5.4). We saw that Character Viewpoint and Acting gestures, though used less with speech than the other gestures at the beginning of the time period, were used as much with speech as the other more abstract gestures during the final age bracket, indicating that language and gesture become more closely integrated towards the end of the fourth year. Furthermore, we saw that Tracing and Drawing were the most closely associated with image-schematic meaning, both in the type of their meaning and the language used. This is an important finding, as it both supports the hypothesis that image schemas underlie the meaning of more abstract gestures (Cienki, 2013; Williams, forthcoming; Mittelberg, 2018), as well as the idea of such image schemas being based, at least in part, on language (Zlatev, 2005).

Finally, regarding objects and other gestures, we saw that more abstract gestures with Observer Viewpoint are more often produced with objects like toys. However, this topic was not systematically examined, and no conclusions can be drawn from these preliminary findings so far.

Chapter 6. Conclusions

This thesis explored the gestures of three Swedish children in their fourth year of life, with a focus on their iconic (e.g. meaning-resembling) gestures, as these most clearly correspond to two kinds of schemas that have been argued to organize meaning-making: *mimetic schemas* (Zlatev, 2005; Zlatev, 2014; Andrén, 2010; Mandler & Canovas, 2014) and *image schemas* (Cienki, 2013; Williams, forthcoming; Mittelberg, 2018). Further, it investigated the role of language in the development of these gestures towards abstractness. Image schemas and gestures are a topic that has been extensively studied, though usually with regard to metaphor in adult gesture (Mittelberg, 2006, 2008, 2018; Cienki, 2008; Williams, forthcoming). The novelty of the approach adopted in this thesis is that it explored the possibility of a developmental transition from mimetic schemas underlying younger children's gestures towards image schemas possibly underlying more abstract gestures. Four data points per child, covering natural interactions with their caregivers in their fourth year of life, were examined.

The first research question (RQ1) asked whether there is a transition towards abstractness in these children's gestures, and whether it is gradual or abrupt. First, it was established on the basis of theoretical reasoning that mimetic-schematic gestures should be predominantly Character Viewpoint-based and from the Acting Mode of Representation. Image schema gestures should rather be mainly Observer Viewpoint-based, and Tracing/Drawing in terms of Mode of Representation. The analysis showed that so-called Representing and Moulding gestures are intermediate. Mimetic-schematic gestures varied extensively between the children and the data points, while image schema gestures did follow the hypothesis: fewer at the beginning at the period, and much more toward the end. This may indeed reflect a trend towards abstractness, and further, one that is more gradual than abrupt. More data would however be needed to support this conjecture.

The Drawing/Tracing gestures, which were presumed to be the most image schematic and abstract, took a leap in age bracket #4, though when looking closer, this was due to only one child. During age brackets #2 and #3, the number of image schematic gestures were

evenly distributed among the children, whose ages during that time were almost the same. Further studies, with a greater number of children observed monthly between the ages of 30 to 50 months, would be necessary to answer the question of graduality or abruptness.

The second research question (RQ2) asked whether the aforementioned transition could be characterised as one from mimetic-schematic to image-schematic gestures. Following the discussion from above, the results of the study did not show a clear transition from one to the other. Rather, mimetic schema gestures continued to be performed from the beginning to the end of period, though indeed, their relative proportions decreased, while image-schematic Tracing and Drawing gestures increased in number and proportion. The reason for these developments remains to be determined, but given the present data, it does not seem to be a matter of “abstraction” from the more concrete, mimetic schemas. Importantly, clear image-schematic gestures do not appear until language has been firmly established and the link between gestures and language begins to cement, which leads naturally to the third research question (RQ3).

What is the role of language in the transition towards abstractness in children’s gestures and to image schemas? This is a large and complex question, and the study cannot answer it definitely. Still, it was significant that abstract gestures did not begin to emerge until language had been firmly established and abstract linguistic expressions only occurred alongside the most abstract gestures. Whether this is a consequence of language development itself cannot be decided, for it could be that (a) there is a separate factor that underlies both in cognitive development or (b) these are coincidental developments of gestures and language. Still, the latter would appear odd, given the tight connection between the semiotic systems of gesture and language. Thus, the results support the proposal that while mimetic schemas may be regarded as pre-/non-linguistic, image schemas may very well be dependent on language.

Empirically, the study may at least be regarded as a “pilot”, as it creates a good foundation for future studies to build on. What would be needed are data from more children, and a longer, more steady time period of observation. Ideally, the ages of the children should be from 30 months old, aiming for the end of Andrén’s (2010) and Zlatev’s (2014) studies, and reaching to at least 50 months old with monthly intervals. This allows for a margin of

error and may shed more light on the peculiarities of Anton's abstract gesture development at 47 months. Furthermore, during data collection, the activities of the children should be controlled for, with less variation, and perhaps more chatting. Reading books and playing with puzzles is not ideal when conducting a study on children's gestures. Finally, following Zlatev (2014), it would be interesting to have a data set of children from two separate cultures.

Despite its limitations, this cognitive semiotic study of children's development has cast light on important issues concerning changes in gesture, and to some extent language, that happen in the fourth year of life. With the help of the *conceptual-empirical loop* (see Chapter 2), key concepts were explicated, including different kinds of iconic gestures, the relationship between Viewpoint and Mode of Representation, and different kinds of schemas. The principle of *methodological triangulation* helped to connect the loop, using the conceptual side of gestures and schemas to explain the data, and conversely, used the data to add to the knowledge of gestures and schemas. Mimetic schemas were shown to be less dependent on language, and the gestures motivated by them largely based on context. Image schemas were proposed to be post-linguistic, given that the gestures motivated by them were more abstract and produced from an Observer Viewpoint.

This study is the first to combine variables such as Viewpoint and Mode of Representation with the theory of image schemas in the analysis of children's gestures and semiotic development. Importantly, it was found that the ages between three and five years should be studied more extensively, for it is precisely this time period that appears to tell us the most about the transition between concrete and abstract meaning making.

References

- Andrén, M. (2008). Iconicity, object manipulation, and schematicity in the development of children's bodily communication. Talk presented at the third conference of the German Cognitive Linguistics Association, Leipzig, Germany, September 2008.
- Andrén, M. (2010). *Children's Gestures from 18 to 30 months*. Department of Linguistics and Phonetics, Centre for Languages and Literature, Lund University.
- Andrén, M. (2014). Multimodal constructions in children: Is the headshake part of language? *Gesture*, 14(2), 141-170.
- Bartlett, F. C. (1932). *Remembering*. London: Cambridge University Press.
- Bressem, J., & Müller, C. (2014). The family of away gestures: Negation, refusal, and negative assessment. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Jana Bressem (eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. (Handbooks of Linguistics and Communication Science 38.2.), 1592-1604.
- Brinck, I. (2004). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly*, 3(4), 429-446.
- Brown, S., Mittermaier, E., Kher, T., & Arnold, P. (2019). How pantomime works: implications for theories of language origin. *Front. Commun*, 4(9).
- Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal cognition*, 20(4), 755-769.
- Cartmill, E. A., Demir, Ö. E., & Goldin-Meadow, S. (2012). Studying Gesture. *Research Methods in Child*, 208.
- Cartmill, E. A., Rissman, L., Novack, M. A., & Goldin-Meadow, S. (2017). The development of iconicity in children's co-speech gesture and homesign. *Language, Interaction and Acquisition*, 8(1), 42-68.
- Cienki, A. (2008). Why study metaphor and gesture? In: Cienki, A., and Müller, C. (eds.), *Metaphor and gesture*, 5-25.

- Cienki, A. (2013). Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics. Published Under the Auspices of the Spanish Cognitive Linguistics Association*, 11(2), 417-432.
- Cienki, A. J., Müller, C., Fricke, E., Ladewig, S., McNeill, D., & Bressemer, J. (2014). Ways of viewing metaphor in gesture. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Jana Bressemer (eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. (Handbooks of Linguistics and Communication Science 38.2.), 1766-1781.
- Clausner, T. C., & Croft, W. (1999). Domains and image schemas. *Cognitive linguistics*, 10, 1-32.
- de Saussure, F. (1960). *Course in General Linguistics*, tr. Wade Baskin. London: Peter Owen Limited.
- Debras, C. (2017). The shrug. *Gesture*, 16(1), 1-34.
- Devyllder, S. (2018). Diagrammatic iconicity explains asymmetries in Paamese possessive constructions. *Cognitive Linguistics*, 29(2), 313-348.
- Dodge, E., & Lakoff, G. (2005). Image schemas: from linguistic analysis to neural grounding. In: Hampe, B., & Grady, J. E. (eds.), *From perception to meaning: Image schemas in cognitive linguistics*, 57-91.
- Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
- Engberg-Pedersen, E. (2015). Perspective in signed discourse: the privileged status of the signer's locus and gaze. *Open Linguistics*, 1(1).
- Frederiksen, A. T. (2017). Separating viewpoint from mode of representation in iconic co-speech gestures: insights from Danish narratives. *Language and Cognition*, 9(4), 677-708.
- Fusaro, M., Harris, P. L., & Pan, B. A. (2012). Head nodding and head shaking gestures in children's early communication. *First Language*, 32(4), 439-458.
- Gibbs, R. W. (2005). The psychological status of image schemas. In: Hampe, B., & Grady, J. E. (eds.), *From perception to meaning: Image schemas in cognitive linguistics*, 113-136.

- 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.
- Gunter, T. C., Weinbrenner, J. E., & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in psychology*, 6, 80.
- Han, T., Kennington, C., & Schlangen, D. (2018). Placing Objects in Gesture Space: Toward Incremental Interpretation of Multimodal Spatial Descriptions. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Hassemer, J., & Winter, B. (2016). Producing and perceiving gestures conveying height or shape. *Gesture*, 15(3), 404-424.
- Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy*. Translated, with an Introduction, by David Carr. Evanston: Northwestern University Press.
- Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13(51), 21-37.
- Johnson, M. (1987). *The Body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.
- Johnson, M. (2005). The philosophical significance of image schemas. In: Hampe, B., & Grady, J. E. (eds.), *From perception to meaning: Image schemas in cognitive linguistics*, 15-33.
- Kant, I. (1998). *Critique of Pure Reason*. Translated and edited by Paul Guyer & Allen W. Wood.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kimmel, M. (2005). Culture regained: Situated and compound image schemas. In: Hampe, B., & Grady, J. E. (eds.), *From perception to meaning: Image schemas in cognitive linguistics*, 285-311.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, 48(1), 16-32.

- Konderak, P. (2016). On evolution of thinking about semiosis: semiotics meets cognitive science. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, (2), 82-103.
- Ladewig, Silva H. (2013). The cyclic gesture. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Jana Bressem (eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. (Handbooks of Linguistics and Communication Science 38.2.), 1605–1618. Berlin/Boston: De Gruyter
- Louhema, et al., under review. Translating from unisemiotic to polysemiotic narratives: A study of language and gesture in Finnish speakers. *Sign System Studies*.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mandler, J. M., & Cánovas, C. P. (2014). On defining image schemas. *Language and Cognition*, 6(4), 510-532.
- McCune, L. (2006). Dynamic event words: From common cognition to varied linguistic expression. *First Language*, 26(2), 233-255.
- McCune, L. (2008). *How children learn to learn language*. Oxford University Press.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
- McNeill, D., Quaegebeur, L., & Duncan, S. (2010). IW-“The man who lost his body”. In *Handbook of phenomenology and cognitive science* (pp. 519-543). Springer, Dordrecht.
- Mittelberg, I. (2006). *Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar*. Cornell University.
- Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In: Cienki, A., and Müller, C. (eds.), *Metaphor and gesture*, 115-154.
- Mittelberg, I. (2018). Gestures as image schemas and force gestalts: A dynamic systems approach augmented with motion-capture data analyses. *Cognitive Semiotics*, 11(1).
- Mittelberg, I., & Waugh, L. R. (2014). Gestures and metonymy. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Jana Bressem (eds.), *Body*

- *Language – Communication. An International Handbook on Multimodality in Human Interaction*. (Handbooks of Linguistics and Communication Science 38.2.), 1747-1766.
- Montgomery, J. C., & Radford, C. A. (2017). Marine bioacoustics. *Current Biology*, 27(11), R502-R507.
- Müller, C. (2014). Gestural modes of representation as techniques of depiction. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Jana Bressemer (eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. (Handbooks of Linguistics and Communication Science 38.2.), 1687-1702.
- Müller, C., Fricke, E., Cienki, A., & McNeill, D. (2014). *Body-Language-Communication*. de Gruyter.
- Novack, M. A., Filippi, C. A., Goldin-Meadow, S., & Woodward, A. L. (2018). Actions speak louder than gestures when you are 2 years old. *Developmental psychology*, 54(10), 1809.
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition*, 148, 10-18.
- Peirce, Charles S. 1931–1958. *Collected Papers I-VIII*. [Hartshorne, C.; Weiss, P.; Burks, A. (eds.)] Cambridge: Harvard University Press.
- Pendlebury, M. (1995). Making sense of Kant's schematism. *Philosophy and Phenomenological Research*, 55(4), 777-797.
- Piaget, J. (1954). The development of object concept (M. Cook, Trans.). In J. Piaget & M. Cook (Trans.), *The construction of reality in the child*, 3-96. New York, NY: Basic Books.
- Plunkett, K., & Strömquist, S. 1992 The acquisition of Scandinavian languages. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition: Volume 3*, 457-556. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Richthoff, U. (2000). *En svensk barnspråkskorpus: Uppbyggnad och analyser*. PhD thesis, Department of Linguistics, University of Göteborg.
- Sokolowski, R. (2000). *Introduction to phenomenology*. Cambridge university press.

- Sonesson, G. (1989). *Pictorial concepts: inquiries into the semiotic heritage and its relevance to the interpretation of the visual world*. Sweden: Lund University Press.
- Sonesson, G. (2000). Iconicity in the ecology of semiosis. *Iconicity-A Fundamental Problem in Semiotics*, 59-80.
- Sonesson, G. (2009). View from Husserl's Lectern. *Cybernetics & Human Knowing*, 16(3-4), 107-148
- Sonesson, G. (2010). From mimicry to mime by way of mimesis: Reflections on a general theory of iconicity. *Σημειωτική-Sign Systems Studies*, 38(1-4), 18-66.
- Sonesson, G. (2012). The phenomenological road to cognitive semiotics.
- Sonesson, G. (2015). From remembering to memory by way of culture. A study in cognitive semiotics. *Southern Journal of Semiotics*, 5(1), 25-52.
- Sonesson, G., & Lenninger, S. (2015). The psychological development of semiotic competence: From the window to the movie by way of the mirror. *Cognitive Development*, 36, 191-201.
- Talmy, L. (2015). Relating language to other cognitive systems: An overview. *Cognitive Semantics*, 1(1), 1-44.
- Ten Have, P. (1990). Methodological issues in conversation analysis1. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 27(1), 23-51.
- Tomasello, M. (1999). The human adaptation for culture. *Annual review of anthropology*, 28(1), 509-529.
- Volterra, V., Capirci, O., Rinaldi, P., & Sparaci, L. (2018). From action to spoken and signed language through gesture. *Interaction Studies*, 19(1-2), 216-238.
- 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.
- Williams, R. (forthcoming). The Source-Path-Goal Image Schema in Gestures for Thinking and Teaching.
- Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. Cambridge, MA: Harvard University Press.

- Zlatev, J. (2005). What's in a schema? Bodily mimesis and the grounding of language. *Cognitive Linguistics Research*, 29, 313.
- Zlatev, J. (2007). Spatial semantics. In: Geeraerts, D., & Cuyckens, H. (eds.), *The Oxford handbook of cognitive linguistics*, 318-350.
- Zlatev, J. (2009). The semiotic hierarchy: Life, consciousness, signs and language. *Cognitive Semiotics*, 4(Supplement), 169-200.
- Zlatev, J. (2012). Cognitive semiotics: An emerging field for the transdisciplinary study of meaning. *Public Journal of Semiotics*, 4(1), 2-24.
- Zlatev, J. (2014). Image schemas, mimetic schemas and children's gestures. *Cognitive Semiotics*, 7(1), 3-29.
- Zlatev, J. (2015a). The Emergence of Gestures. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence*, 458-477. West Sussex: Wiley Blackwell.
- Zlatev, J. (2015b). Cognitive semiotics. In: Trifonas, P. P. (ed.), *International handbook of semiotics*, 1043-1067. Springer, Dordrecht.
- Zlatev J. (2019) Mimesis Theory, Learning, and Polysemiotic Communication. In: Peters M. (ed.), *Encyclopedia of Educational Philosophy and Theory*. Springer, Singapore.
- Zlatev, J., & Andrén, M. (2009). Stages and transitions in children's semiotic development. *Studies in language and cognition*, 380-401.

Appendix A: Language aligned with *Acting* gestures

Words that semantically coincide with the gesture phrase are underlined and in bold, and, where appropriate, vice versa within the explanation of the gesture. No special indication is given in cases where the whole gesture phrase coincides with the corresponding linguistic phrase and vice versa.

	Language	Gesture	Notes
DEM	Så! (There!)	Uses Playmo bucket to <u>pour</u> “water” on a Playmo tree after caregiver’s instruction.	
	Sån? (Like that?)	Asks the opinion of caregiver about the placing of a Lego block.	
LEX	Ja ska <u>banka</u> på dina <u>känningar</u> (I will <u>knock on your feels</u>)	Makes a <u>knocking</u> gesture in the air with a toy in her hand, <u>patting</u> the känningar with her other hand.	
	Känningar (feels)	Knocks on caregiver’s <i>känningar</i> with a toy to illustrate.	
	Dunkdunk (x2) (heart beating)	Uses a toy stethoscope to “listen” to caregiver’s heart.	
	Såna här som han ska <u>hålla i</u> (Like that <u>which he should hold on to</u>)	Puts arms out in front of body as if <u>grabbing two poles, bouncing up and down.</u>	There is a component of DEM (<i>Såna här</i>). However, it was placed as LEX because the gesture phrase semantically corresponded with <i>hålla i</i> . It was not placed with GRA despite the preposition “in” (<i>i</i>), because the point of the phrase is to hold (<i>hålla</i>).
	Namnam	Is feeding Playmo toy and speaks from the OBS viewpoint.	
	CocaCola nu (now CocaCola)	Holds a Playmo drink container up to his lips, says the phrase, and then proceeds to “drink” the contents.	

Då måste jag <u>skära</u> på <u>köttet</u> (Then I <u>must cut the meat</u>)	Picks up a Playmo knife and proceeds to “ cut ” Playmo meat.	
Klippklipp (cut cut)	Holds a toy medical instrument up to the caregiver’s skin.	
Nu kommer blodet, hjälp! (The blood is coming now, help!)	Places a cloth representation of blood on caregiver’s hand and holds it there.	
(x) på lite olja. ([x] a little oil)	Makes a puzzle piece that looks like a person put “oil” on the depicted bus central to the puzzle.	
Dom ha spillt där. (They have spilled there).	“Dries” an area on a puzzle where a bus is central.	This has a component of DEM (<i>där</i>). However, it was placed as LEX because he is “drying” the area, and not showing where “they” spilled.
Dom ha gjort sönder glas i hans bus (They broke a glas in his bus).	“Vacuums” the “broken glass” on the puzzle’s depicted bus.	
Torkatorka (dry dry)	“Cleans” the puzzle’s depicted bus.	

Appendix B: Language aligned with *Representing* gestures

	Language	Gesture	Notes
DEM	Sån här <u>pinne</u> (This kind of stick)	Draws attention to finger as it takes on the form of an upright stick.	Even though the gesture corresponds with the Lexical portion of the sentence, the aim of the interaction was to draw attention to the gesture (2pp was used to decide).
LEX	Nu åker han kring om bordet (Now he drives around the table)	Makes a Playmo character push a Playmo wheelbarrow around the table.	
	Mamma!	The doll is “calling out” for its mother.	
	Å, de var <u>en som</u> hade ingen cykel, å <u>sen hade</u> ingen hjälm dom åkte motorcykel, å sen hade ingen hjälm när dom åkte cykel (Yes, there was one that had no bicycle, and then had no helmet when they drove a motorcycle, and then had no helmet when they rode a bicycle).	She is counting on her fingers the first two of these three instances, using her fingers for holding ideas rather than to actually count (thus an ICO Representing gesture, and not a CON gesture).	Even though there is a sentence adverb that coincides with the gesture, the gesture does not mean “then”. Rather, it represents a whole idea.
	Hjälp! Ett djur! (Help! An animal!)	Playacting with toys.	
	Voomvoom!	Holds a toy car in hand and moves it fast back and forth in front of body to illustrate how fast it goes.	

Karusell!	The child is the carousel, showing how it moves.	
Sen kommer vatten ut och spruter så (<u>Then the water comes out and sprays, like that</u>).	He is showing how water comes out of a fire truck.	Despite the GRA element (kommer <i>ut</i>), the gesture semantically corresponds with <i>spruter så</i> . Had the gesture corresponded with <i>ut</i> , it would have been a TRA gesture, the <i>ut</i> displaying the trajectory, and not simply representing the spraying water itself.
Å, åka (Yes, drive).	Makes a small train of Lego carts “drive” around the table.	
Kopplakoppla (x2) (Trot, trot)	Makes a toy horse trot.	
Namnam	Makes a toy horse eat.	
Hej, muah (Hey, [kiss]).	Makes the toys kiss each other.	
Å, det har hon (Yes, she has it).	Moves hand back and forth across mouth to indicate that, yes, she did have a piece of tape across her mouth.	
Då krockade han (<u>Then he crashed</u>).	Flings outspread hands into the air in front of face outside of own locus, palms facing the caregiver.	
Och sen går (And then goes)	Makes a toy plough move.	This has a component of GRA (<i>och sen</i>), but the gesture semantically corresponds with <i>går</i> .
Plattplatt (Flat flat)	Narrates the “flattening” of the “earth” with a figurine of an old farming tool.	
Det var jätte stort! (It was really big!)	Rubs hands over body, face, and hair, to indicate <i>jätte stort</i> .	

	Pumpapumpa (Pump pump).	Pumps “water” on a toy fire truck as if someone else would do on a real fire truck.	
	Den <u>hoppa</u> r (It jumps).	Finger demonstrates to the caregiver how it jumps.	
GRA	Den ha vält <u>ner</u> på golvet (It fell down to the floor).	While sitting, puts arms in the air next to face to resemble a glass and <u>pretends to fall sideways</u> like a glass would fall off a countertop.	The gesture corresponds with the GRA expression, and they mean the same thing.
	Allihopa (Everyone).	Places hand in the air above a group of toys.	This is the image schema CONTAINER, as the gesture places an “umbrella” around a group of objects. Phrase contains reversibility.

Appendix C: Language aligned with *Moulding* gestures

	Language	Gesture	Notes
DEM	Sån här liten (Small, like this)	Measuring with the hands	
	Sån här stor (Big, like this)	Measuring with the hands	
	Sån liten (This small)	Measuring with the hands	
	Sån här (x2) (Like this)	Measuring with the hands	
GRA	Mellan pinna såna (Between sticks like this)	Holds palms out and measures a distance from which two <i>pinna</i> s stand from each other.	Both the word <i>mellan</i> (preposition) and the gesture correspond to GRA.

Appendix D: Language aligned with *Tracing* gestures

	Language	Gesture	Notes
DEM	Här de börjar, så och så (They start here, like this and this).	Tracing a path where the length of a puzzle worm starts and how it curves into an S.	This gesture contains the image schema SOURCE-PATH-GOAL, due to the beginning phrase, <i>här de börjar</i> (they start here), indicating the SOURCE in the schema (1pp was used to decide).
	Här åker trollet (The troll travels here).	Demonstrates the path of the troll across the tabletop.	
LEX	Dom hoppa r och hoppa r (They jump and jump).	Uses finger to indicate a trajectory by which a cat jumps.	This is an instance of TRA rather than REP or ACT because his hand is not the cat (OBS;REP), and his hand is not “holding” a cat (CHA;REP).
GRA	Långt upp i loftet (Far up into the air).	Finger goes up in the air as if it is an airplane flying high into the sky.	The stroke of the gesture corresponds with the GRA expression (<i>upp</i> and <i>i</i>). Here is the image schema SOURCE-PATH-GOAL (the gesture and the adverb [up]), as well as the image-schema CONTAINER which corresponds to the preposition (<i>i</i>).
	Gå ut på den andra sidan (Go out on the other side).	Had been tracing a path around a picture when finger rushes up diagonally from the bottom of the page to the top.	The stroke of the gesture corresponds with the GRA expression (<i>ut</i>). The gesture and its semantically corresponding expression (<i>ut</i>) indicates the CONTAINER schema. <i>Ut</i> contains reversibility.
	Och i loftet (And in the air). Och går ner (And goes down).	Finger is flung into the air in a sort of arch, copying the trajectory of an airplane.	The stroke of the gesture corresponds with the GRA expression (<i>i loftet</i>). The first linguistic expression indicates the CONTAINER

		The finger comes down and “lands” on the book.	image schema, while the second linguistic expression and the gesture contain the SOURCE-PATH-GOAL schema.
	Här åker trollet, <u>här inne</u> (The troll travels here, inside here).	Traces the troll’s path along the tabletop and <u>enters</u> into a crack between the table and the wall.	The gesture corresponds with the GRA expression (<i>här inne</i>). Here, only the language (<i>inne</i>) corresponds to the image schema CONTAINER, as the gesture neither contains any referents nor does it create a boundary.

Appendix E: Language aligned with *Drawing* gestures

	Language	Gesture	Notes
DEM	De är inte jo <u>här</u> (It is not here, though).	Draws a circle with finger around an area on a completed puzzle, indicating where the colour red is not situated.	This has a component of GRA (<i>inte</i>), but the gesture semantically corresponds with <i>här</i> . The gesture indicates the CONTAINER image schema, containing the colour within a boundary.
	En sån här (One like this)	<i>Sån här</i> is referring to a circle. The child draws with the finger around the empty spot of puzzle where a circular puzzle piece should go.	
	Leka med <u>dom</u> (Play with them).	<u>Draws a circle</u> around a depicted group of characters in a book.	Both the gesture and the corresponding word <i>dom</i> are used demonstratively. The gesture indicates the image schema CONTAINER, in that by putting the characters in the circle, they are contained within a boundary.
LEX	Träden, träden (The trees, the trees).	Finger Draws a treeline on the wall next to him.	
GRA	Här i jungeln ha <u>dom (x)</u> (They have [x] here in the jungle).	Draws a circle around an area within an environment depicted in a book in which “they” have something.	Even though the gesture corresponds temporally with <i>dom (x)</i> , the circle drawn corresponds semantically with <i>här i jungeln</i> . The gesture indicates the image schema CONTAINER, creating a boundary within which “they” have (x).

	Såna <u>inne</u> (Like this inside).	Draws a circle with finger around an area depicted in a book.	<p>There is an element of DEM (såna), but the GRA word (<i>inne</i>) corresponds with the gesture.</p> <p>The gesture and the language indicate the CONTAINER image schema.</p>
--	--------------------------------------	---	---