Jordan Zlatev

Image schemas, mimetic schemas and children's gestures

Abstract: Mimetic schemas, unlike the popular cognitive linguistic notion of image schemas, have been characterized in earlier work as explicitly representational, bodily structures arising from imitation of culture-specific practical actions (Zlatev 2005, 2007a, 2007b). We performed an analysis of the gestures of three Swedish and three Thai children at the age of 18, 22 and 26 months in episodes of natural interaction with caregivers and siblings in order to analyze the hypothesis that iconic gestures emerge as mimetic schemas. In accordance with this hypothesis, we predicted that the children's first iconic gestures would be (a) intermediately specific, (b) culture-typical, (c) falling in a set of recurrent types, (d) predominantly enacted from a first-person perspective (1pp) rather than performed from a third-person perspective (3pp), with (e) 3pp gestures being more dependent on direct imitation than 1pp gestures and (f) more often co-occurring with speech. All specific predictions but the last were confirmed, and differences were found between the children's iconic gestures on the one side and their deictic and emblematic gestures on the other. Thus, the study both confirms earlier conjectures that mimetic schemas "ground" both gesture and speech and implies the need to qualify these proposals, limiting the link between mimetic schemas and gestures to the iconic category.

Keywords: imitation, mimesis, convention, communicative intent, typification, iconic, representation, language

DOI 10.1515/cogsem-2014-0002

1 Introduction

The concept *image schema* was initially defined as "a recurring dynamic pattern of our perceptual interactions and motor programs that gives rise to coherence and structure to our experience" (Johnson 1987: xiv). It emerged in discussions of how linguistic meaning and abstract thought can possibly be "grounded"

Jordan Zlatev: Lund University. E-mail: jordan.zlatev@ling.lu.se

in perception and action. Frequently cited image schemas are notions such as PATH and CONTAINER. While agreeing on their presumed foundational role, theorists have expressed rather different views on their nature. Are image schemas *conscious* or at least accessible to consciousness (Gibbs 2005; Langacker 2006) or part of the "cognitive unconscious" (Lakoff and Johnson 1999)? Are they *representational* (Mandler 2005) or non-representational, interactional structures (Lakoff and Johnson 1999)? Are they tied to *individual* bodies and brains (Dodge and Lakoff 2005) or intersubjectively shared (Johnson 1987)? Are they based on *concrete bodily actions* (Johnson 1987), or on basic process of consciousness such as "perceptual meaning analysis" (Mandler 2005) or "mental scanning" (Langacker 2006)? Are they essentially *identical with linguistic meanings*, especially closed-class morphemes such as spatial prepositions (Johnson and Rohrer 2007), or rather reflect pre-linguistic experiential structures/processes that motivate but do not determine linguistic meaning (Zlatev 1997, 2011)? Are they *universal*, as most often assumed, or to a large extent culture-specific (Kimmel 2005)? Issues such as these became obvious with the publication of a volume of edited papers on image schemas (Hampe 2005; Zlatev 2005) and remain controversial. Still, as mentioned above, the majority of researchers using the concept image schema employ it to refer to relatively abstract structures, albeit experientially grounded ones, such as PATH, CONTAINER and VERTICALITY, with a strong aspiration for universality.

I introduced the concept of *mimetic schema* in a series of publications (Zlatev 2005, 2007a, 2007b), combining ideas from evolutionary theory, in particular the notion of *mimesis* as "conscious, self-initiated, representational acts that are intentional but not linguistic" (Donald 1991: 168) and ideas from developmental psychology, in particular Piaget's theory that the first cognitive representations in childhood arise through acts of *sensory-motor imitation* which are eventually internalized. This process of internalization then makes possible *representational imitation* where "the interior image precedes the exterior gesture, which is thus a copy of an 'internal model' that guarantees the connection between the real, but absent model, and the imitative reproduction of it" (Piaget 1962: 279). The foremost theoretical role of mimetic schemas is similar to that of image schemas: to provide a basis for explaining the evolution and development of language. With the intention of avoiding the ambiguities surrounding image schemas pointed out above, mimetic schemas were defined as "dynamic, concrete and preverbal representations, involving the body image, accessible to consciousness and pre-reflectively shared in a community" (Zlatev 2005: 334), or alternatively as "fairly specific, cross-modal, consciously accessible representations based on imitation, and largely shared within a (sub)culture" (Zlatev 2007b: 131). For present purposes, the following (hypothetical) properties of mimetic schemas are

central: *preverbal*, *body-based*, *representational, intermediately specific* and *culturally shared*.[1]

It is not hard to see that together these properties constitute the necessary ingredients for a pre-linguistic semiotic system for *communication*, which was not the focus of Piaget, interested as he was in the development of thought and "the child's construction of reality." The main theoretical advantage of mimetic schemas compared to image schemas is that they can help explain, almost literally, the "grounding" of *both* communication and thought through action and imitation, in both evolution and development. It is possible that, as Piaget considered, they function "internally," or *covertly*, consistently with phenomenological accounts of imagery (Thompson 2007) and neuroscientific findings, usually expressed in terms of "simulation" (cf. Gallagher 2005; Zlatev 2007b for the problematic overuse of this term). Most important for present concerns, however, is my earlier proposal that mimetic schemas are crucially implicated in gestures: "[. . .] mimetic schemas underlie both speech and gesture, thereby accounting for the close synchronization of the two modes of expression" (Zlatev 2005: 335). However, this was stated somewhat in passing without adequate substantiation.

The aim with the present article is to elaborate the proposal of a close connection between mimetic schemas and gestures, in particular with respect to ontogenetic development. In Section 2, I suggest that the properties of *gestures*, as defined by Andrén (2010), correspond in a fairly direct way to those of (overt) mimetic schemas. On this basis, I formulate the general hypothesis that children's iconic gestures emerge as preverbal, overt mimetic schemas. Furthermore, to the extent that imitation plays a major role in the acquisition of children's deictic and emblematic gestures, bodily mimesis can be expected to be crucially implicated in their development as well.

To test these hypotheses, section 3 describes an empirical study of the early gestures of three Swedish and three Thai children based on longitudinal naturalistic data, focusing on the ages 18, 22, and 26 months. To anticipate, the analysis shows that not only *emblematic* gestures, such as NOD-HEAD and WAVE-BYE, and *deictic* gestures, as INDEXFINGER-POINT, constitute socially shared types, realized by recurrent instances in the children's data, but also that the children's *iconic* gestures to a large extent do likewise (see Section 3, and Appendix A for definitions of theoretical concepts). It is such iconic gestures, especially when enacted from a first-person perspective (1pp), that most clearly correspond to overt, com-

1 See the cited publications for explications of these properties. For example, the feature "intermediately specific" means that mimetic schemas are intermediary in specificity compared to relatively abstract image schemas such as UP on the one side, and action schemas in use (Newton 2003) controlling individual actions on the other.

municatively used mimetic schemas. Section 4 discusses the implications of the findings, including the need to revise, to some extent, the concept of mimetic schemas, as previously characterized. Finally, Section 5 sums up the argument that mimetic schemas, much more so than image schemas, as most often defined, provide a productive concept for understanding the ontogeny of children's gestures.

2 The upper and lower limits of mimetic schemas and gestures

Gesture is not a technical but an everyday term commonly referring to expressions using the hands or the whole body. This is both an advantage and disadvantage, since "the confusion in the literature and the arguments about "what is gesture" derive from the attempts of various investigators to co-opt this everyday word and make it a technical term" (Adam Kendon, private communication). Indeed, there is a good deal of controversy on how exactly to define gestures. Are they *nonverbal* or not (McNeill 1985)? It has been well established that they are closely associated with speech (McNeill 1992), but claiming that "gesticulations are obligatorily accompanied by speech," and conflating gesticulations with gestures in general can give the impression that non-speech accompanied gestures or pantomimes are not gestures at all.[2] Need gestures be *empty-handed*, or can they also involve objects in their articulation (Andrén 2010)? Is their function primarily communicative or cognitive (i.e. for self-expression and thought)? Are they performed and understood consciously or unconsciously? Are all gestures representational, and for those that are: are they understood on the basis of their iconicity or through their conventionality (Streek 2009)? These questions are, of course, not so much empirical as conceptual: what is meant by "verbal," "representational" etc. One may note that some of these questions are reminiscent of those concerning image schemas, stated at the onset.

For the purpose of a study of Swedish children's gestures in the second and third year of life, Andrén (2010) combined semiotic analysis and empirical

2 For example, in the following recent formulation of (what was earlier called) *Kendon's Continuum*: "Gesticulations are obligatorily accompanied by speech but have properties unlike language [. . .] Speech-linked gestures are also obligatorily performed with speech [. . .] Signs are obligatorily not performed with speech and have the essential properties of language." (McNeill and Sowa 2011: 43). It is not that the existence of other, non-speech linked gestures is denied, but rather that they often fail to be mentioned and somehow disappear from view in McNeill's and many other gesture scholars' writings.

observation, defining gestures as actions that qualify either as explicitly communicative or as explicit signs, or both. This definition rests on the identification of two separate dimensions of gestural meaning, which I will refer to, somewhat differently from Andrén, as *communicative intent* and *representational complexity*. Within each of these two dimensions, three different levels can be distinguished, as shown in Table 1. Following Andrén (2010), gesture may be defined as a bodily act that fulfills level 3 on at least one, and possibly both, of these dimensions.

According to this definition, a gesture is either an act that displays features of communicative intentionality, such as mutual gaze and persistence (CI-3), or has the character of an explicit sign, clearly differentiated from what is stands for by the subject (Sonesson 2007, 2009; Zlatev 2009; Andrén 2010) (RC-3), or possibly, but not necessarily, both. For example, a performance of WAVE-BYE may be used as the appropriate move at the end of a social interaction. This is meaning on level RC-2, similar to that of performative speech acts (cf. Searle 1995; Sinha 2010), rather than RC-3, since it *counts as* a type of action which is part of the social interaction rather than a *representation* of it.[3] However, if the act is directed to a certain recipient, having visible features of communicative intentionality (CI-3), it would qualify as a gesture. On the other hand, an act of "symbolic play" performed in solitude would also qualify as a gesture, since (full) symbolic play displays understanding on the part of the child of the stands-for (RC-3) relation (McCune 2008). Many gestures, e.g. a pantomime performed for the benefit of another, would involve both CI-3 and RC-3. Furthermore, most gestures may be distinguished from the "signs" of signed languages on the basis of the properties

Table 1: Defining the concept of gesture, following (Andrén 2010: 68): Level 3 must be reached on at least one of the two dimensions of Communicative intent and Representational complexity.

Level	Communicative intent (CI)	Level	Representational complexity (RC)
CI-3	Explicitly other-oriented action: Clear communicative intentionality	RC-3	Explicit signs: Expression X stands for meaning Y
CI-2	Action framed by mutual attunement: Unclear communicative intentionality	RC-2	Typified acts: Performance X counts as doing action-type Y
CI-1	Side effect of co-presence: No visible communicative intentionality	RC-1	Situation-specific acts: Performance X contextually suggests Y

3 A WAVE-BYE gesture may also be used to quote a gesture that another person performed in a previous situation and, in this case, it would rather *stand for* what this other person did (RC-3). The semiotic properties of a specific action are always relative to how it is used in a given situation.

normativity and *systematicity*: signed-language signs, like words of spoken languages have meanings which are not just pre-reflectively shared, but commonly known, and combined according to commonly known rules/norms (Itkonen 2003). Emblematic gestures, like the OK-gesture, take an intermediary position: they have normative criteria on both expression and meaning, but lack the (extended) systematicity of both spoken and signed languages.

Returning to mimetic schemas, their most succinct characterization was that of "categories of acts of overt or covert bodily mimesis" (Zlatev 2007b: 133), where *bodily mimesis* was defined as follows:

> **Def:** A particular bodily act of cognition or communication is an act of **bodily mimesis** if and only if:
> a) It involves a cross-modal mapping between *exteroception* (i.e. perception of the environment, normally dominated by vision) and *proprioception* (perception of one's own body, normally through kinesthetic sense);
> b) It is under conscious control and corresponds to – either iconically or indexically – to some action, object or event, while at the same time being *differentiated* from it by the subject;
> c) The subject *intends* the act *to stand for* some action, object or event for an addressee (and for the addressee to recognize this intention);
> d) Without the act being conventional-normative, and
> e) Without the act dividing (semi)compositionally into meaningful sub-acts that systematically relate to each other and other similar acts. (Zlatev 2007b: 132)

It is important to note that features (a) and (b) alone are sufficient for *pre-reflective sharing* on the level of types and not only specific instances.

> Mimetic schemas will be overwhelmingly shared among the members of a community who engage in close face-to-face or rather body-to-body interaction. A particularly important form of such interaction consists of (overt or covert) *imitation*. This yields what Arbib (2003, 2005) calls *representational parity* and the content of a mimetic schema, e.g. JUMPING, will be similar for the one who performs the act and for the one who observes it, imitates it, and internalizes it. [. . .] There will be a limited set of mimetic schemas with a specific cultural community, and their parity will be not only on the level of individual actions, but of *types*. (Zlatev 2007b: 142)

Thus, mimetic schemas correspond to what Andrén, following Schutz (1953), calls *typifications*, in which a particular act "counts as" an instance of a socially acknowledged type, in most cases having a name in the community language. When such an act is performed demonstrably for the sake of someone else (e.g. "Look at me: I am *dancing*!"), this performance would qualify both as gesture and as an instance of a mimetic schema. On the other hand, property (a) alone in the definition of mimesis above, and the lowest levels of Representational complexity

and Communicative intent (Table 1), are clearly not sufficient for either mimetic schemas or gestures. Hence, the lower limits of both concepts are quite similar. As can be seen from the negative criteria (d) and (e), the upper limit of bodily mimesis, and hence of mimetic schemas, also corresponds to that of gestures, with the exception of emblems qualifying as gestures but not as mimetic schemas. When children pass to using semiotic norms, in gesture as well as in language, the role of overt mimetic schemas may be transcended, but not their covert role: "The use of collective signs (words) [. . .] presupposes that the child learns them. This he does through imitation, by means of which he has become capable of representative thought. Moreover, the interior imitative images continue to serve [. . .] even when language comes to be used" (Piaget 1962: 279).

In sum, there appears to be considerable overlap between the concepts of mimetic schema and gesture. In terms of representational format, mimetic schemas can be *overt* (performed) or *covert* (imagined), and, at least in the first case, there is a close affinity with gestures. In terms of granularity, mimetic schemas are on the level of typified actions such as DANCE and HIT rather than on the level of specific movements (e.g. dancing can be done in different ways) or more abstract schemas such as PATH and VERTICALITY. Iconic gestures may constitute mimetic schemas in this respect in at least three different ways: as communicatively used realizations of such actions, as enactments of such actions, performed from a first-person perspective (1pp) and as relatively detached third-person perspective (3pp) representations of such actions.[4] There are similarities in terms of development as well. Mimetic schemas are assumed to emerge through imitation, including interactive processes such as role-reversal (Andrén 2010). But such processes are also instrumental for acquiring deictic gestures such as pointing (Tomasello 1999) and emblematic gestures like the OK-gesture, the latter being cultural conventions by definition. Thus, one could expect the processes of schematization and social typification to be rather similar in the case of children's acquisition of iconic gestures on the one hand, and of pointing and emblems on the other. Still, there is an important difference. What children imitate in the case of mimetic schemas and early iconic gestures are practical actions like EAT and KICK. In the case of deictic and emblematic gestures, like INDEXFINGER-POINT and WAVE-HELLO, it is actions that are communicative from the start.

Thus, we can formulate the general hypothesis that *children's iconic gestures emerge as overt mimetic schemas*: as imitations of practical, culture-typical actions. With respect to other types of gestures, in particular deictic and emblem-

4 Of course, this is not to say that iconic gestures are limited to expressing such meanings. Especially in adults, they can signify more abstract concepts or properties than concrete actions.

atic ones, the applicability of a mimetic-schema analysis is more tentative. The following section describes an empirical study, performed together with Mats Andrén, intended to test the hypothesis.

3 A study of Swedish and Thai children's gestures

3.1 Data

To investigate the relationship between mimetic schemas and children's gestures, we performed a new analysis on data that have been used in past studies (Zlatev and Andrén 2009; Andrén 2010): a multimodal corpus of longitudinal data from three Swedish and three Thai children engaged in naturalistic caretaker-child interactions, with each utterance linked to the video-files using CLAN (MacWhinney 2000). The Thai data was recorded in Bangkok during the first two years of the millennium. The Swedish data were originally recorded by Ulla Richthoff in Gothenburg in the 1990s (Richthoff 2000). For the present study, we focused on three data points per child: at the beginning (18 months), middle (22 months) and end (26 months) of the corpus. 10 minutes in the beginning of each session (6 children x 3 sessions x 10 minutes = 3 hours) during which the children were clearly visible, were coded according to the following categories and procedure.

3.2 Procedure

3.2.1 Identification

All instances of children's gestures in the data were identified following the definition of gesture given in Section 2 after viewing the material repeatedly. This was performed by myself for the Thai data and by Mats Andrén for the Swedish data. A single gesture could occur together with zero, one, or more spoken utterances. It was also possible to have more than one gesture overlapping with a single utterance.

3.2.2 General analysis

A first step of analysis was performed, also individually by the two coders for the Thai and Swedish data respectively, on the basis of the following predefined *general semiotic categories*. Each gesture was coded as being predominantly deictic (DEI), iconic (ICO) or emblematic (EMB). While we have earlier emphasized that

these semiotic categories, originating from Peirce and recurring under different terms in practically all gesture analyses (cf. Kendon 2004), are not mutually exclusive and should be viewed as co-present "components" (Zlatev and Andrén 2009) or "aspects" (Andrén 2010), this was not part of the present analysis for two reasons. The first was methodological: it was easier to introduce a greater conformity of analysis by making the categories mutually exclusive. The second was theoretical: while having components of several different semiotic categories, a particular gesture appears to be dominated by one particular category (cf. Jakobson 1965), either on the basis of its proximity/directionality (DEI) with respect to the referent, its similarity to the intended meaning (ICO), or its conventionally/normatively defined form and meaning (EMB).

Thus, DEI gestures were defined to be communicative actions that indicate or individuate an external target (a physical object, a person, location, direction, sound or even a whole event). This includes not only different types of pointing, but also acts that bring an object to the attention of the addressee (e.g. showing, grasping, giving). The criterion for EMB gestures was the existence of intersubjective criteria for the gesture's proper form and meaning.[5] ICO gestures were defined to be those where there is *resemblance* between the movements of the whole body, or parts of it, and properties of intended actions, objects, or whole events. As explained in Section 2, apart from cases of explicit signs (e.g. KICK performed in the air as a pantomime of actual kicking), typified acts that were used communicatively with or without objects (e.g. demonstrably KISSING a teddybear as part of the communicative interaction with the parent) were also coded as ICO gestures. Furthermore, all ICO gestures were analyzed as being performed from either a *first person perspective* (1pp) or *third person perspective* (3pp).[6] 1pp ICO gestures were "reduced" versions of practical actions, either performed in the as-if mode (KISS, DANCE, HIT etc.) or else being explicitly communicative (cf. Table 1). 3pp ICO gestures were enacted with the hand(s), possibly together with an artifact. e.g. DRIVE-CAR, performed by pushing a toy car back and forth.

3.2.3 Typification

In the next step of the analysis, the gestures were coded (still individually by the two coders) into types. In the case of EMB gestures, this was fairly straightforward

5 These appear to correspond to what McNeill and Sowa (2011: 27) call morphs, having "standards of good form and syntagmatic values."

6 The distinction is similar to that of "character-viewpoint" vs. "observer-viewpoint" (McNeill 1992, 2005), but is less concerned with where exactly in gesture space the gesture is performed.

as they correspond to familiar norms (HEAD-SHAKE, WAVE-BYE etc.). For DEI gestures, two types were defined in advance: INDEXFINGER-POINT (prototypical pointing with index finger), and GRAB (reaching out and demonstrably grasping an object).[7] Others types were allowed to be defined by the individual coder in the following manner. On the basis of similarity in expression and meaning, ICO were grouped in types by each coder and, hence, for each cultural group. These types were given labels using English glosses that attempted to capture their typified meaning, often translations of Swedish or Thai phrases that the parents or the children used in conjunction with the gestures: KICK, DANCE, APPLY-LOTION. On all levels of the analysis (DEI/ICO/EMB, 1pp/3pp for ICO, and the specific gesture types) it was also possible to choose the generic code UNCLEAR, so that hard-to-classify instances would not need to be coerced into types.

3.2.4 Type calibration

After each coder had performed several passes through the data and converged on an analysis that was stable, we jointly analyzed approximately 50% of the coded gestures: 6 sessions of the Thai data, and 3 sessions of the Swedish data. Only the following changes to the individual codes were made: First, different labels that referred to the same gesture types were made identical. Second, if a type was present in one language and absent in the other, and a gesture in the second language was found to instantiate this type better than the type it was originally assigned to, the gesture was re-assigned to the new type. This step was performed in only a few cases and with utmost care, in order not to make the gestures in the two groups appear more similar than warranted. Third, gestures classed as belonging to "types" that were found not to be consistent across the different occurrences were demoted to the UNCLEAR class.

3.2.5 Coding of imitation and co-ordination with speech

A final aspect of the analysis was to add for each gesture codes reflecting whether it was produced in imitation to the gesture of an adult or not (IMI), and whether it was produced in conjunction with speech or not (SPCH). Imitation was opera-

7 The act of grabbing something usually requires coordination with a spoken utterance to stand out as a communicative act. Hence, the type GRAB does not refer to any act of "grabbing" an object but to a category of communicative acts that can be described as "Look, I am taking this."

tionalized as the presence of a gesture of the same type by an adult interlocutor in at most two preceding (parent) turns, and co-speech was not just fully articulated speech but any vocalization that seemed "speech like" or "sound-symbolic" (e.g. a crash-like sound when two cars smash together), rather than, for example, crying. The values of the IMI and SPCH codes were also jointly checked and agreed upon by the two coders for the same 50% of the data.

3.3 Predictions

On the basis of the theoretical background and hypothesis given at the end of Section 2, we formulated the following predictions (expectations):

First, in accordance with the general hypothesis that children's first iconic gestures are instantiations of mimetic schemas, it was expected that the majority of ICO gestures would correspond to socially acknowledged types. The same was expected to be the case for DEI and EMB gestures, where the latter are typified per definition. Hence, we predicted that the proportions of typified gestures for all three categories would be approximately the same. For evaluating the prediction, "typified" was operationalized as (a) having at least two instances per type and (b) not coded with the code UNCLEAR.[8]

Second, it was predicted that many, though not all, of the types found would be specific to each cultural group, since mimetic schemas arise "locally" through imitation processes. This prediction stands in contrast to the (most often) assumed universality of image schemas.

Third, it was expected that the majority of ICO types would be intermediately specific in granularity, corresponding to practical actions. Again, this prediction contrasts with what would be expected from an image-schema analysis: more abstract and universal types such related to schemas such as PATH, CONTAINER and VERTICALITY.

Fourth, we predicted that 1pp ICO gestures would both outnumber and developmentally precede 3pp ICO gestures. Once more, this prediction contrasts with an image-schema perspective, where the distinction is (usually) not considered relevant. The precedence 1pp > 3pp ICO gestures follows naturally from the

8 On the one hand, these exclusions were motivated by an ambition to provide a conservative measure of gesture-typification: gestures that occurred only once were regarded as "non-typified," even if they were known to conform to types in reality. On the other hand, the measure was quite liberal, since we did not require any particular temporal distance between the tokens, allowing e.g. repetition of an action several times in a row.

assumption that they emerge as mimetic schemas, i.e. as typifications of bodily actions.[9]

Fifth, it was expected that 3pp gestures display a greater dependence on imitation of the parent's gestures than 1pp gestures, which emerge both earlier and more gradually. A more tentative prediction was that a similar relationship would also hold between INDEXFINGER-POINT and other DEI gestures. The motivation for this latter prediction was the higher degree of conventionality of INDEXFINGER-POINT and the assumption that the more conventional a gesture is, the more likely it is that it would be produced in a context in which an adult interactive partner has produced it earlier.

Sixth, and finally, since the SPCH code served as an indication of the extent to which various kinds of gestures are coordinated with speech and since the degree of such co-ordination could be said to reflect how "verbal" gestures are (McNeill 1985), it could be expected that gestures that display property (d) in the definition of bodily mimesis, i.e. emblematic gestures, and those that are acquired later (3pp after 1pp), would have higher ratios for the SPCH code.

3.4 Results and discussion

The 180 minutes of data altogether contained 973 gestures produced by the six children, divided by the major semiotic categories and the two culture/language groups as shown in Table 2, yielding an average of 8 gestures per minute for the Swedish children and 2.8 for the Thai children. This difference is above all due to a substantial difference between the two data sets regarding the overall number of utterances produced by the children. If gesture rate is instead calculated as gestures per utterance (an utterance includes speech and/or gesture), the rate is

Table 2: Total number of gestures in the 180 min of data (2 × 9 sessions), by language/culture group and major semiotic category

	Swedish	Thai	Total
DEI	470	146	616
EMB	118	31	149
ICO	133	75	208
Total:	721	252	973

9 The prediction is also in line with earlier gesture theories (Werner and Kaplan 1963; McNeill 1992), though their reasons for making such a prediction (e.g. increasing "symbolic distance") are not as transparent as those that follow from mimetic schema theory.

instead 0.47 gestures per utterance for the Swedish children and 0.34 gestures per utterance for the Thai children, which is a fairer measure of the children's actual gesture production.

Table 3 below shows all the gestures that met the criterion for typification of at least two instances, presented by semiotic category and language. Altogether, 68 gesture types were attested. 10 of these were found in both languages, thus yielding 58 separate types of gestures, accounting for 889 of the 973 gestures. As can be seen by comparing the number of tokens in Table 2 and 3, and more graphically in Figure 1, a strong majority of the gestures for both groups proved to be typified.

As shown by Figure 1, the typification levels for the three categories DEI, EMB and ICO were similar, thus confirming the first expectation. The levels for the EMB and ICO categories in the two cultural groups are nearly identical. The difference of 13% for the DEI category can be explained by considering the character of the pointing gestures (i.e. excluding other DEI gestures) in the two cultures. 96.7%

Table 3: Gesture types (with at least two tokens) for each language, with corresponding tokens and number of data sessions (max 9), in which they occurred. Types that were attested in both languages are given in boldface. The total number of non-overlapping types is thus 58. 3pp indicates ICO gestures performed from a third-person perspective (Zlatev and Andrén 2009).

Semiotic category	Swedish children's gestures			Thai children's gestures		
	Types	# Sessions	# Tokens	Types	# Sessions	# Tokens
DEI	**Indexfinger-point**	9	296	**Indexfinger-point**	9	81
	Put	8	39	**Give**	4	11
	Give	7	26	**Show**	4	9
	Grab	6	28	Reach-to-person	3	6
	Reach-to-thing	6	11	**Grab**	3	4
	Remove	5	19	**Reach-to-thing**	3	4
	Show	5	16	**Surface-point**	3	4
	Surface-point	4	5	Throw-to-person	1	2
	Beg	2	9			
DEI all:	9 types	9 (> 1)	449	8 types	7 (> 1)	121
EMB	**Nod-head**	6	55	**Shake-head**	5	9
	Shake-head	5	37	**Nod-head**	4	7
	Done-clap	2	9	Wave-hand-no	3	5
	Wave-bye	2	5	Wai	3	5
	Wave-hello	2	2	Clap-hands	2	4
	Gone	1	3			
	Thanks-bow	1	2			
EMB all:	7 types	5 (> 1)	113	5 types	5 (> 1)	30

Table 3 (cont.)

Semiotic category	Swedish children's gestures			Thai children's gestures		
	Types	**# Sessions**	**# Tokens**	**Types**	**# Sessions**	**# Tokens**
ICO	Cheek-cuddle	3	4	Kick	3	5
	Doll-walk (3pp)	2	13	Dance	2	5
	Car-drive (3pp)	2	12	Hug	2	5
	Doll-hello (3pp)	2	6	Smell-kiss	2	5
	Kiss	2	6	Cat-scratch	2	2
	Pour	2	5	**Kiss**	2	2
	Feed	2	4	Wave-away	2	2
	Pat	2	4	Hit-person	1	12
	Stir	2	3	Apply-lotion	1	6
	Throw	2	2	Doll-dance (3pp)	1	6
	Doll-kiss (3pp)	2	2	Bite-kiss	1	4
	Shiver	1	10	**Feed**	1	3
	Phone	1	7	Hi-there!	1	2
	Doll-jump-down (3pp)	1	6	Scare-dog	1	2
				Bang-door	1	2
	Sit-in-car	1	5			
	Comb	1	4			
	Feed-drink	1	4			
	Wipe-mouth	1	3			
	Sing-sway	1	3			
	Turn-knob	1	2			
	Doll-drive (3pp)	1	2			
	Search	1	2			
	Put-lid-on	1	2			
	Eat	1	2			
ICO all:	24	11 (> 1)	113	15	7 (> 1)	63
Total:	40 types	25	675	28 types	19	214

and 77.9% were INDEXFINGER-POINT for the Swedish and the Thai children, respectively. This may be related to the fact that INDEXFINGER-POINT is considered "rude" to a greater extent in Thailand than in Sweden (cf. Zlatev and Andrén 2009).

The high typification levels in all three categories, and especially the ICO category, are even more impressive given that the results stem from relatively scarce data: 3 × 10 minutes for each child. One may therefore expect that only gestures that recur often will be captured by the sampling process. On the other hand, a possible objection could be that the estimates rest on a relatively liberal opera-

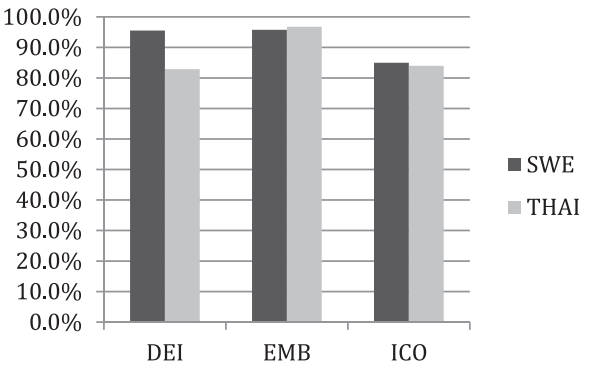

Fig. 1: Typification levels for the three semiotic categories of gestures, for the two cultural groups

tionalization of typification, possibly including repetitions of a gesture type in a particular situation, attributed to "priming" in a recent analysis of gesture recurrence (McNeill and Sowa 2011: 28). However, as shown in Table 3, the majority of types consisted of instances from more than one recording session: 25 out of 40 for the Swedish children, and 19 out of 28 for the Thai children. Joined, this means that 44 of all 68 types (or 64.7%) consisted of instances that could not be of the "priming" kind.[10] Still, one can notice a difference between the ICO category on the one hand, where a (slight) majority of the types come from a single session: 13 of 24 for Swedish, 8 of 15 for Thai; and DEI and EMB on the other hand, where there were altogether only three such types for both language groups. Yet another estimate supports the conclusion that the results are not due to the local (priming) recurrence: 89% of all gesture instances belong to a type that occurs across sessions (between any two, or more, sessions of the 18 sessions in the data).

Comparing the types between the language groups for the purpose of evaluating the second expectation concerning a high degree of culture-specificity of mimetic schemas and gestures, there are several ways to perform this. Considering only the gestures presented in Table 3, we can note that only two of the large number ICO types were attested in both data sets: FEED and KISS. It is further notable that there were two other culture-specific types of "kissing" in the Thai data, labeled BITE-KISS (where the lower jaw is extended in an exaggerated way) and SMELL-KISS (a kind of sniffing the adored one). They also have separate names in

10 Note that this estimate is based on counting types that occurred in both groups twice, and excluding all instances with the code UNCLEAR, or which occurred only once, i.e. those that were excluded from the "typification level" estimates in Figure 1.

Thai: *man-khîaw* and *hɔ̌ɔm*.[11] For the much smaller number of EMB types, two (HEAD-NOD and SHAKE-HEAD) were attested in both groups. Of the DEI gestures, not only the predefined two types, but also GIVE, SHOW, REACH-TO-THING, and SURFACE-POINT were attested in both the Swedish and Thai data. This manner of comparison gives the impression of a relatively higher degree of culture-specificity of ICO gestures, compared to DEI and even EMB. However, this may give a somewhat skewed impression, since ICO gestures are, by their nature, of a more "open-class" kind, in the manner of nouns and verbs, while DEI gestures may be compared to "closed class" words like prepositions and determiners, and EMB to a relatively fixed set of "idioms." Thus, there are simply many more ICO types than those of the other semiotic categories, and they are likely to be under-sampled in the data. That is why we performed a second form of comparison. We reversed the question, and asked how many of the types that occurred across children (and not just across sessions) had instances from only one of the language/culture groups, respectively from both. Of the 27 such types found, 15 had instances from both groups,[12] while 12 had instances only from one of the groups. Of the 15 "shared" types, 7 were DEI, 3 were EMB, and 5 were ICO. Conversely, of the 12 "non-shared" gestures, 3 were DEI, 4 EMB and 5 ICO. In sum, the second expectation predicting a degree of culture-specificity, in particular for ICO types, can be seen as supported, but with the reservation that a surprisingly high degree of co-occurrence in the two groups was found.

Concerning the third prediction that ICO gestures would be "intermediately specific" in the manner of mimetic schemas rather than image schemas, we can consider the list of ICO types in Table 3. While in some cases the labels are not self-explanatory, I hope that they are sufficiently so to allow the reader to appreciate that the expectation was overwhelmingly supported. In fact all ICO gestures, with the possible exceptions of WAVE-AWAY in the Thai data, were either demonstrably performed typified actions (RC-2 + CI-3) such as DANCE, or re-enactments (RC-3) like APPLY-LOTION, in which the Thai child engaged in "applying" imaginary skin lotion to various parts of her body. Concerning a possible analysis in terms of image schemas, we can note that several of the attested gesture types can perhaps be said to express abstract notions such as AWAY FROM EGO (e.g. KICK, HIT-PERSON, WAVE-AWAY) vs. TOWARD EGO (e.g. KISS, APPLY-LOTION), but what exactly would such generalizations explain? I return to this in Section 4.1.

11 Since the children performed these in "the air" or with dolls, they were analyzed as ICO gestures (RC-3). If performed with others, they would have been analyzed as EMB instead.

12 The higher figure here is due to the fact that types with single instances in one of the languages were considered too, while these were excluded from earlier estimates.

The fourth expectation, that among ICO gestures, those performed from a first-person perspective (1pp) would both outnumber and developmentally precede 3pp gestures, was confirmed somewhat more ambiguously than the preceding three. On the one hand, looking again at Table 3, it can be seen that there were only five types in the Swedish data and one for the Thai data that were 3pp. Furthermore, all of these involved the manipulation of a toy, such as a doll or a car, rather than a body part to stand for something else, as typical for adult "observer viewpoint" gestures (cf. Andrén 2010). Even if two of the five 3pp types among the Swedish children (CAR-DRIVE, DOLL-WALK) were the most numerous ones in terms of instances of a single type, counting the number of instances for 1pp (74) and 3pp (39) gestures separately, shows a clear dominance for the first category.

Figure 2 shows the average number 1pp and 3pp ICO gestures per session, also including ICO "types" attested with a single instance, or gestures coded as UNCLEAR but still allowing 1pp vs. 3pp discrimination. The results further confirm the prediction that 1pp gestures outnumber 3pp gestures – for all three ages. However, there is no obvious trend for the *emergence* of 3pp gestures since there are roughly as many 3pp gestures at 18 months as at 26 months. As it turns out, the Swedish child Tea at 18 months is a clear outlier in this respect, since all of the 3pp instances at 18 months come from this session and none of the other children perform 3pp gestures at this age.[13] All of the 3pp gestures Tea produces are tokens of three distinct types and all these three types are first performed by the mother,

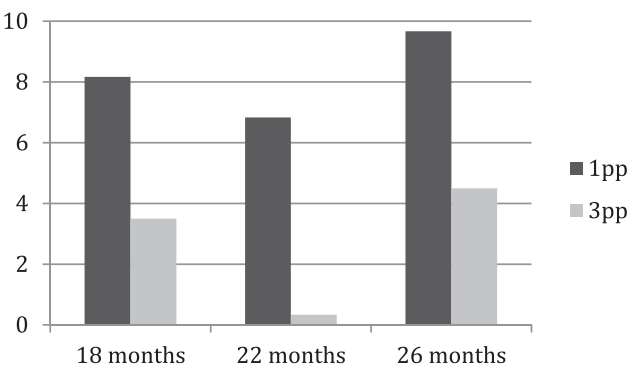

Fig. 2: Average number of first-person perspective (1pp) and third-person perspective (3pp) ICO gestures per session for the three age periods, in Thai and Swedish data combined

13 Tea was also part of the data analyzed by Andrén (2010), and in this study there were many more sessions included for each child. Therefore it is possible to say that even for Tea there are only a few 3pp gestures after this 18 month session until around 26 months.

i.e. there is a strong effect of imitation involved. If this outlier were to be removed, the data would indeed confirm the expectation that 3pp gestures emerge during the period studied here, and that 1pp gestures appear earlier in development.

However, the removal of outliers is always problematic from a methodological point of view, especially in cases with relatively few subjects such as the present. An alternative way to look at the question of developmental precedence of 1pp/3pp gestures is consider each child separately. This provides some support for the expectation of a later emergence of 3pp gestures, since only one child performed 3pp gestures at 18 months (Tea) and one child at 22 months (Harry), whereas at 26 months three children did so: Tea, Harry (Swedish), and Jam (Thai). This is also in accordance with the findings of Andrén (2010) regarding a strong and relatively sudden upsurge in the number of 3pp gestures around 24 months in six Swedish children (including a larger number of recordings of the three Swedish children studied in the present study).

As pointed out, within the ICO category, 1pp gestures can be naturally seen as deriving from mimetic schemas, but what about 3pp gestures? And what can be said about the role of bodily mimesis for DEI and EMB gestures? To remind, the fifth prediction concerned a more obvious role for imitation in the case of 3pp than 1pp ICO gestures, and for INDEXFINGER-POINT compared to other DEI gestures, with EMB gestures more or less on the level of INDEXFINGER-POINT. The assumption was that in the case of 3pp gestures, INDEXFINGER-POINT, and EMB gestures, the children learn and encounter these as explicitly communicative gestures directly, rather than first passing through a stage in which they are learned as practical actions. The IMI coding, as explained in 3.2.5, was used to test this assumption. We limited the analysis to the first token of each type in each session – whether this was imitated or not – because this shows who is responsible for the introduction of the gesture type in the interaction: the child or the parent. Figure 3 shows the tendencies for various gesture types to be imitated as operationalized. The results confirm both parts of the prediction: 3pp gestures were imitated considerably more often than 1pp gestures (37% vs. 22% of the time), while EMB and INDEXFINGER-POINT were more often imitated than other DEI gestures. In other words, for each part of the prediction, the gestures that were more conventional-normative, showed a higher dependence on direct imitation than those that where less so.

While falling outside of the expectation, we may ask why ICO gestures as a whole (1pp and 3pp) were found to be imitated more often than the other kinds. A possible answer lies in what was suggested earlier: ICO gestures involve a greater number of distinct types, and thus constitute a more "open class" category than EMB and DEI gestures. Since most ICO gestures (as here argued) are learned as sociocultural types by imitating parents' spontaneous or stylized actions, rather

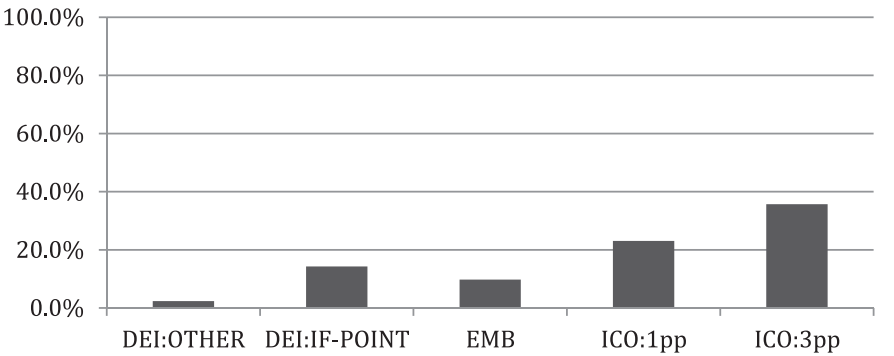

Fig. 3: Levels of imitation for classes of gestures, showing lower degrees for DEI gestures other than INDEXFINGER-POINTING (on the level of EMB) and for ICO gestures performed from a first person perspective (1pp) compared to a third-person perspective (3pp)

than being the children's spontaneous creations, their much greater number would require more "scaffolding" from parents through imitative processes than the more limited inventory of EMB and DEI gestures.

The sixth and final prediction concerned relative comparisons between the same types and categories as the previous expectation, but in this case regarding the extent to which they would tend to be coordinated with spoken utterances or not. To the extent that EMB gestures and INDEXFINGER-POINT are more "post-mimetic" (cf. Section 2), they could be expected to be more often combined with speech than other DEI gestures, and likewise for 3pp compared to 1pp ICO gestures. As Figure 4 shows, the prediction was not supported. The first thing to notice is that most of the gestures of all analyzed classes were performed together with speech. 3pp ICO gestures were found to co-occur with speech just as often as 1pp ICO gestures, and there is no clear tendency for EMB gestures and INDEXFINGER-POINT to co-occur with speech more often than other DEI gestures. Thus, the degree of coordination between speech and children's gestures could not be explained by how "mimetic" a particular gesture is.

Instead, the results point to a difference in the degree to which gestures of the three general semiotic categories are coordinated with speech: DEI gestures tend to do so most often, EMB gestures less so, and ICO gestures the least of the three categories.[14] This runs counter to the common idea that EMB gestures are not "true" co-speech gestures, or at least less so than ICO gestures (McNeill 2005). Just because EMB gestures can be performed without speech, and because they

14 Again, this is in accordance with findings presented by Andrén (2010: 148 and 176).

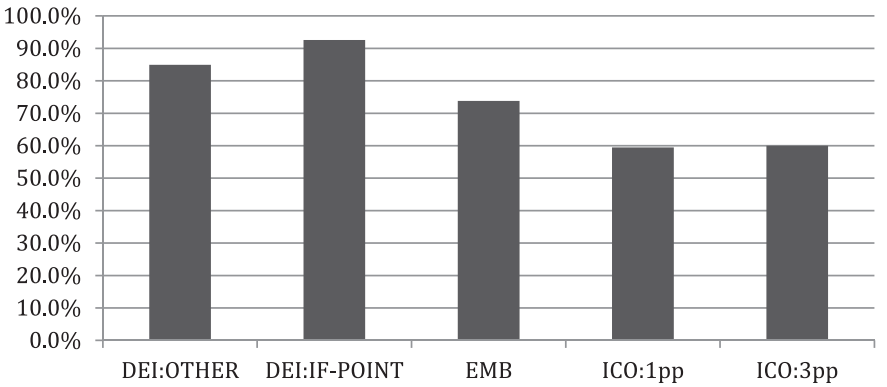

Fig. 4: Levels of coordination with speech for different kinds of gestures (note differences compared to Figure 3).

can be easier to interpret without the support of speech, does not mean that they are actually used less often with speech.

4 General discussion: mimetic schemas or image schemas in the emergence of gestures?

The results from the case study of Thai and Swedish children's gestures described in the previous section should not be over-interpreted, given that their empirical basis was fairly restricted: six children, with 30 minutes of data each, from three age points: 18, 22 and 26 months. Still, they were intriguing and relevant for the major issues of this article. On the one hand, the findings have implications for the relative merits of image schemas and mimetic schemas for the "grounding" of gesture. On the other hand, there are implications for clarifying and further developing the theory of bodily mimesis and mimetic schemas. The discussion below interlaces the two kinds of implications.

To remind, both image schemas and mimetic schemas have been conceived of as *bodily* (a more adequate term than the popular term "embodied") *gestalts*: experiential structures not derivable from their composite parts, and arising through bodily action and imitation. Their main difference lies in the much higher level of abstractness of image schemas, which coalesces with the view that they constitute a restricted and presumably universal set. The results of the study quite clearly support the view that children's iconic gestures emerge as mimetic schemas rather than as image schemas. The evidence for this are the

findings that the children's first iconic gestures (in the data) fall into socially acknowledged types – almost to the same extent as deictic and emblematic gestures (cf. Figure 1) – and that these types are on the level of specific actions, such as kicking, kissing, applying-lotion, dog-scaring etc. (cf. Table 3). Some of these were common to both language/culture groups, but some were culture-specific (or at least culture-typical). Importantly, the study provided clear evidence that children's iconic gestures are (in general) initially enacted (1pp) and only later performed from a third-person perspective (3pp). In gesture studies, a distinction between the two sub-categories is sometimes made by calling the first "symbolic" and only the second "iconic," terminology intended to emphasize the view that they have "different sources," as expressed by Nicoladis (2002: 245):

> [. . .] iconic and symbolic gestures seem to have different sources. Iconic gestures are thought to be created on-line (McNeill, 1992) while children's symbolic gestures may have been learned from parents (Caselli, 1990). To the extent that children's symbolic gestures are learned from parents, they might be better classified as conventional gestures. Iconic gestures appear in children's productions as early as two years of age and are usually accompanied by similar-meaning speech.

While the present study in part concurs with the view that 1pp and 3pp iconic gestures have important developmental differences, it contradicts the interpretation given above that only "symbolic" (1pp) gestures are "learned from parents": 3pp iconic gestures also appeared in types, and the level to which they were imitated was actually considerably higher than for 1pp gestures (Figure 3). There were no differences between the two sub-classes with respect to coordination with speech (Figure 4).

At the same time, by considering different categories and sub-categories of children's gestures, the study may serve to nuance the rather sweeping statement that "[. . .] mimetic schemas underlie both speech and gesture" (Zlatev 2005: 355), quoted in Section 1. It also raises questions about earlier definitions of mimetic schemas. In particular, it appears that the qualifiers "preverbal" and "representational" were assigned rather hastily to the concept in earlier treatments (e.g. Zlatev 2007b). The most important revision that needs to be acknowledged is that while imitation (processes) indeed appears to be crucial for the development of all three categories of gestures, *what is imitated is not of the same kind*. This would indeed imply that different kinds of gestures do not follow the same developmental trajectory, and if claimed to "correspond to" or even "emerge from" mimetic schemas, then these schemas cannot be of the same kind either.

The empirical findings supporting the need for revising mimetic schema theory are those which showed that children's first iconic gestures are much closer to specific, practical actions, and fall in a more open-ended repertoire of

types than their deictic and emblematic gestures. Furthermore, it appears that it is above all iconic gestures that develop into *explicit signs* (RC-3, Table 1). In effect, the developmental route of iconic gestures can be said to primarily follow the right-hand side (Representational complexity) of Table 1: from practical action, through typification, differentiation, to explicit signs. With their presumed representational character, mimetic schemas correspond most closely to the upper levels of this developmental route.

On the other hand, deictic gestures and, to a large degree, even emblematic gestures are not so much representations, or explicit signs (RS-3), but explicit, performative communicative actions (CI-3): a pointing gesture or a hand-shake do not represent an act of directing attention and a greeting, respectively; they count as such (RC-2). Like 1pp iconic gestures, they are also imitated and typified, but not initially as practical but directly as communicative actions. This can account for their divergent properties compared to iconic gestures found in the study: (a) fewer types, (b) greater likelihood to overlap across languages, and (c) less reliance on direct imitation (at least during the period of study).

All this means that it is problematic for the concept of mimetic schemas to be extended to deictic and emblematic gestures, such as those listed in Table 3.[15] The most consistent solution, without undue twisting of the tangible notion "representation," is simply not to attempt such an extension. Thus, the logic of the study leads to the conclusion that while a broad notion of bodily mimesis is certainly relevant for the development of all gestures, *only ICO gestures emerge as mimetic schemas*. Furthermore, since 3pp ICO gestures also appear to be learned directly as communicative acts, and not via the "dyadic mimesis" of practical acts, the degree to which they are connected to mimetic schemas and their bodily imagery remains an open question.

Another aspect of mimetic schemas in relation to children's gestures that needs revision is the qualifier "preverbal," listed as one of their definitional properties (cf. Section 1). This qualifier gives the impression of mimetic schemas – possibly realized in gestures of the kind shown in Table 2 – as first fully formed and only then linked with linguistic expressions. Considering that most of the types given in Table 3 have verbal labels that are used both by the parents and the children, often in association with producing the gestures, such a proposal would be as problematic as similar ones concerning image schemas (cf. Lakoff 1987: 447). Thus, what can be regarded as "preverbal" are the typified (practi-

15 Concerning deictic gestures, the part of the definition of bodily mimesis concerning "indexical correspondence" in (b) would need to be modified. On the other hand, emblematic gestures are already regarded as "post-mimetic," in a manner similar to lexical items, and thus do not correspond to "categories of acts of bodily mimesis," i.e. mimetic schemas.

cal) actions (RC-2) and possibly also covert schemas of the kind hypothesized by Piaget (1962). However, *overt* mimetic schemas are shaped both by the bodily actions and by the linguistic practices of the community in which the child is embedded from the start.

The final, and only unsupported, prediction in the study was therefore based on a rather simplistic assumption of correspondence between: mimetic/post-mimetic = preverbal/verbal = less/more coordination with speech. Still, the finding that the children's iconic gestures as a whole co-occurred less often with speech than those of the other categories (Figure 4) lends some additional support to what was suggested above: that (1pp) iconic gestures have the closest affinity to mimetic schemas. We may also note that the correlation between gesture category and speech predicted by this analysis (ICO < EMB) is the converse of the one most commonly claimed in gesture studies (McNeill 1985; Nicoladis 2002).

What about image schemas in children's gestures? While their value as an analytical concept for the study of adult gestures has been clearly shown (Cienki 2005; Mittelberg 2010), it is much less clear what part they play, if any, in the emergence of gestures in ontogeny. As suggested section 3.5, it is possible to seek (and find) generalizations such as the one that certain gestures express the "schema" AWAY FROM EGO VS. TOWARD EGO, which could be generalized to PATH, or perhaps PATH in combination with CONTAINER. But as in some image schematic analyses of linguistic meaning (e.g. Johnson and Rohrer 2007), it is not clear what such generalizations amount to: phenomenological descriptions of the child's experiences, analytical generalizations, or discoveries of conceptual primitives on an invisible, "sub-personal" level. Image-schema analysts have oscillated between such interpretations and have often ignored the fact that they are quite different (cf. Zlatev 2010).

But as Cienki (2013) has recently proposed, mimetic schemas and image schemas may have rather complementary roles for gesture studies: more pantomimic gestures in adults correspond to mimetic schemas while more abstract ones to image schemas. A possible theoretical synthesis between the mimetic and image schema approaches could further be enhanced by closer empirical studies of children's gestures, in particular during the period of the suggested "gesture explosion" around 3–4 years of age (McNeill 2005), where indeed there seems to be a transition in the character of children's gesturing in the direction of greater abstractness. How this transition takes place and what the acquisition of language and, in particular, of grammatical morphemes with more abstract semantics contributes to the transition is yet to be determined.

Finally, both mimetic and image schema analyses appear troublesome in accounting for the first kind of children's gestures: pointing and other deictics at the end of the first year (Bates et al. 1979) as well as for emblematic gestures.

The reason for this could be called the "representational bias" of both constructs (even though image schemas are sometimes called non-representational), while deictics and emblems are, as here argued, primarily interpersonal communicative acts.

5 Summary and conclusions

This article theoretically and empirically considered the applicability of the concepts of image schemas and mimetic schemas to children's early gestures. By and large, the analysis confirmed earlier proposals that mimetic schemas, rather than image schemas "ground" gestures in development (Zlatev 2005). However, this was shown to apply only to iconic gestures and above all to those pantomime-like gestures where the subject (child) enacts a particular practical action.

An attractive characteristic of mimetic schemas is that they are thought to have covert (imagistic) as well as overt (bodily) manifestations. But in considering the latter, it is problematic to regard them as "preverbal." Even if the learning and typification of practical, culture-typical actions through observation and participation does to some extent precede language, in social interaction overt mimetic schemas inevitably become entangled with speech and communication in general. Thus it is hard to argue that their preverbal character remains intact, rather than being gradually "sedimented" with communicative, linguistic meaning. Iconic gestures should therefore not be regarded as direct realizations of mimetic schemas, but as complex orchestrations of visible and audible communicative actions – as well as imagery that may have its basis in bodily mimesis and mimetic schemas.

Furthermore, our empirical study showed that iconic gestures were, in at least one respect, similar to children's deictic and emblematic gestures: nearly all could be accounted for by a given number of types, apparently acquired through imitative processes. Thus, a more general capacity for bodily mimesis, for which human beings seem to be uniquely endowed (Donald 1991, 2001; Zlatev 2008), can still be regarded as an indispensable cradle, if not ultimate ground for gesture and language.

Acknowledgments: The empirical study described in Section 3 was performed together with Mats Andrén, who also provided careful and critical comments to several previous versions of the text. The article would have been impossible without him. In addition, I am most grateful for the constructive and critical comments on earlier versions of this text from (in alphabetic order) Paul Bouissac,

Alan Cienki, Adam Kendon, Michael Kimmel, Brian MacWhinney and Göran Sonesson, as well as from an anonymous reviewer and comments from the participants of the Cognitive Semiotics seminar at Lund University.

References

Andrén, Mats. 2010. *Children's gestures between 18 and 30 months*. Lund: Lund University PhD thesis.

Arbib, Michael A. 2003. The evolving mirror system: A neural basis for language readiness. In Morten Christiansen & Simon Kirby (eds.), *Language evolution*, 182–200. Oxford: Oxford University Press.

Arbib, Michael A. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences 28*. 105–167.

Bates, Elisabeth, Bretherton, I., Camaioni, L., & Volterra, Virginia 1979. *Emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

Cienki, Alan. 2005. Image schemas and gesture. In Beate Hampe (ed.), *From perception to meaning: Image schemas in cognitive linguistics,* 421–441. Berlin: Mouton de Gruyter.

Cienki, Alan. 2013. Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics* 11/2: 417–432.

Dodge, Elan & Lakoff, George. 2005. Image schemas: From linguistic analysis to neural grounding. In Beate Hampe (ed.), *From perception to meaning: Image schemas in cognitive linguistics*, 57–91. Berlin: Mouton de Gruyter.

Donald, Merlin. 1991. *Origins of the modern mind. Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Donald, Merlin. 2001. *A mind so rare: The evolution of human consciousness.* New York: Norton.

Gallagher, Shaun. 2005. *How the body shapes the mind.* Oxford: Oxford University Press.

Gibbs, Raymond. 2005. The psychological status of image schemas. In Beate Hampe (ed.), *From perception to meaning: Image schemas in cognitive linguistics,* 113–135. Berlin: Mouton de Gruyter.

Hampe, Beate (ed.). 2005. *From perception to meaning: Image schemas in cognitive linguistics*. Berlin: Mouton de Gruyter.

Itkonen, Esa. 2003. *What is language? A study in the philosophy of linguistics.* Turku: Turku University Press.

Jakobson, Roman. 1965. Quest for the essence of language. *Diogenes*, 13, 21–38.

Johnson, Mark & Rohrer, Tim. 2007. We are live creatures. Embodiment, American pragmatism, and the cognitive organism. In Tim Ziemke, Jordan Zlatev, & Roz Frank (eds.), *Body, Language and Mind. Vol 1. Embodiment*, 17–54. Berlin: Mouton de Gruyter.

Johnson, Mark. 1987. *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Johnson, Mark. 2005. The philosophical significance of image schemas. In Beate Hampe (ed.), *From perception to meaning:* Image schemas in cognitive linguistics, 15–33. Berlin: Mouton de Gruyter.

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

Kimmel, Michael. 2005. Culture regained: Situated and compound image schemas. In Beate Hampe (ed.), *From perception to meaning: Image schemas in cognitive linguistics,* 285–312. Berlin: Mouton de Gruyter.

Lakoff, George & Johnson, Mark. 1999. *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Langacker, Ronald. 2006. Subjectification, grammaticaliztion, and conceptual archetypes. In Angeliki Athanasiadou, Costas Canakis, & Bert Cornillie (eds.), *Subjectification: Various paths to subjectivity*, 17–40. Berlin: Mouton de Gruyter.

MacWhinney, Brian. 2000. *The CHILDES project (3rd edition). Volume 1: Tools for analyzing talk: Transcription format and programs*. Mahwah, NJ: Lawrence Erlbaum.

Mandler, Jean. 2005. How to build a baby: III. Image schemas and the transition to verbal thought. In Beate Hampe & Joseph E. Grady (eds.), *From perception to meaning: Image schemas in cognitive linguistics,* 137–163. Berlin: Mouton de Gruyter.

McCune, Lorraine. 2008. *How children learn to learn language.* Oxford: Oxford University Press.

McNeill, David & Sowa, C. 2011. Birth of a Morph. In Gale Stam & Mika Ishino (eds.) *Integrating Gestures: The Interdisciplinary Nature of Gesture*. Amsterdam & Philadelphia: John Benjamins.

McNeill, David. 1985. So you think gestures are nonverbal? *Psychological Review, 92*, 350–371.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: Chicago University Press.

McNeill, David. 2005. *Gesture and thought*. Chicago: University of Chicago Press.

Mittelberg, Irene. 2010. Geometric and image-schematic patterns in gesture space. In Vyvyan Evans & Paul Chilton (eds.), *Language, cognition and space: The state of the art and new directions*, 351–385. London: Equinox.

Newton, Natika. 2003. Representation in theories of embodied cognition. *Theoria et Historia Scientarum*, 7(1), 181–193.

Nicoladis, Elena. 2002. Some gestures develop in conjunction with spoken language and others don't: evidence from bilingual prescholars. *Journal of Nonverbal Behavior* 26(4), 241–265.

Piaget, Jean. 1962. *Play, dreams and imagination in childhood*. London: Routledge.

Richthoff, Ulla. 2000. *En svensk barnspråkskorpus: Uppbyggnad och analyser*. PhD thesis, Department of Linguistics, University of Gothenburg.

Schutz, Alfred. 1953. Common-sense and scientific interpretation of human action. *Philosophy and Phenomenological Research*, 14(1). 1–38.

Searle, John. 1995. *The construction of social reality.* London: Penguin.

Sinha, Chris. 2010. Language as a biocultural niche and social institution. In Vyvyan Evans and Stephanie Pourcel (eds.), *New Directions in Cognitive Linguistics.* 289–309. Amsterdam & Philadelphia: John Benjamins.

Sonesson, Göran (2007). From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In Tim Ziemke, Jordan Zlatev, & Roz Frank, *Body, Language and Mind. Vol 1: Embodiment,* 85–128. Berlin: Mouton de Gruyter.

Sonesson, Göran (2009). The view from Husserl's lectern: Considerations on the role of phenomenology in cognitive semiotics, *Cybernetics and Human Knowing,* 16(3–4), 107–148.

Streeck, J. 2009. *Gesturecraft: The manu-facture of meaning.* Amsterdam & Philadelphia: John Benjamins.

Thompson, Evan (2007). *Mind in Life: Biology, phenomenology and the sciences of mind.* London: Belkarp Press.
Tomasello, Michael. 1999. *Cultural origins of human cognition.* Cambridge: Harvard University Press.
Werner, H. & Kaplan, B. 1963. *Symbol formation*. New York: Wiley.
Zlatev, Jordan (1997). *Situated embodiment: Studies in the emergence of spatial meaning*. Stockholm: Gotab.
Zlatev, Jordan. 2005. What's in a schema? Bodily mimesis and the grounding of language. In Beate Hampe (ed.), *From perception to meaning: Image schemas in cognitive linguistics*, 313–343. Berlin: Mouton de Gruyter.
Zlatev, Jordan. 2007a. Language, embodiment and mimesis. In Tim Ziemke, Jordan Zlatev, & Roz Frank (eds.), *Body, Language and Mind. Vol 1. Embodiment*, 297–337. Berlin: Mouton de Gruyter.
Zlatev, Jordan. 2007b. Intersubjectivity, mimetic schemas and the emergence of language. *Intellectica,* 46/47, 123–152.
Zlatev, Jordan. 2008. The co-evolution of intersubjectivty and bodily mimesis. In Jordan Zlatev, Tim Racine, Chris Sinha, & Esa Itkonen (eds.), *The shared mind: Perspectives on Intersubjectivity,* 215–244. Amsterdam & Philadelphia: John Benjamins.
Zlatev, Jordan. 2009. The Semiotic Hierarchy: Life, Consciousness, Signs and Language, *Cognitive Semiotics*, 4, 169–200
Zlatev, Jordan. 2010. Cognitive linguistics and phenomenology. In D. Schmicking & Shaun Gallagher (eds.), *Handbook of Phenomenology and Cognitive Sciences*, 415–443. Dordrecht & New York: Springer.
Zlatev, Jordan. 2011. From cognitive to integral linguistics and back again. *Intellectica,* 56, 125–147.
Zlatev, Jordan & Mats Andrén. 2009. Stages and transitions in children's semiotic development. In Jordan Zlatev, Mats Andrén, Carita Lundmark, & Marlene Johansson Falck (eds.), *Studies in Language and Cognition*, 380–401. Newcastle: Cambridge Scholars.

Bionote

Jordan Zlatev is Professor of General Linguistics at Lund University, where he is one of the research directors of the transdisciplinary Centre for Cognitive Semiotics (CCS) established in 2009. His research focuses on language as a predominantly conventional-normative semiotic system for communication and thought, and especially on its bio-cultural evolution and development. He is the author of *Situated Embodiment: Studies in the Emergence of Spatial Meaning* (1997), co-editor of *The Shared Mind* (2008) and *Moving Ourselves, Moving Others* (2012), and has published over 50 articles in peer-reviewed journals and books.