

Pantomime as the original human-specific communicative system

Jordan Zlatev^{†,‡}, Przemysław Żywiczyński^{*,‡} and Sławomir Wacewicz [‡]

[†]Division for Cognitive Semiotics, Centre for Languages and Literature, Lund University, Lund, Sweden and [‡]Centre for Language Evolution Studies, Nicolaus Copernicus University, Torun, Kujawsko-Pomorskie, Poland

*Corresponding author: przemek@umk.pl

Abstract

We propose reframing one of the key questions in the field of language evolution as *what was the original human-specific communicative system?* With the help of cognitive semiotics, first we clarify the difference between signals, which characterize animal communication, and signs, which do not replace but complement signals in human communication. We claim that the evolution of bodily mimesis allowed for the use of signs, and the social-cognitive skills needed to support them to emerge in hominin evolution. Neither signs nor signals operate single-handedly, but as part of *semiotic systems*. Communicative systems can be either *monosemiotic* or *polysemiotic*—the former consisting of a single semiotic system and the latter, of several. Our proposal is that pantomime, as the original human-specific communicative system, should be characterized as polysemiotic: dominated by gesture but also including vocalization, facial expression, and possibly the rudiments of depiction. Given that pantomimic gestures must have been maximally similar to bodily actions, we characterize them as typically (1) dominated by iconicity, (2) of the primary kind, (3) involving the whole body, (4) performed from a first-person perspective, (5) concerning peripersonal space, and (6) using the Enacting mode of representation.

Key words: evolution of communication; iconicity; gesture; polysemiotic; sign and signal; semiotic system

1. Introduction: reformulating the question

Debates in the field of language evolution often focus on the nature of human *protolanguage*, understanding this notion along the well-known conception proposed by Bickerton (1990: 128) as ‘a more primitive variety of language’ that serves as a stepping stone in language evolution. In particular, theorists differ on whether such protolanguage was ‘musical’, ‘gestural’, ‘lexical’, or otherwise (Fitch 2010; Żywiczyński, Gontier and Wacewicz 2017). This starting point, however, assumes

that the first step toward human communicative specificity was in fact a *language*. Even if ‘primitive’, we are still left with the question of how it could evolve in the first place. This is a non-trivial question since, as argued by Donald (1998: 140): ‘there are important fundamentals missing from the primate mind, without which protolanguages could not emerge; I shall call these the ‘cognitive prerequisites’ of protolanguage’. The challenge is to spell out these prerequisites.

Research in the evolution of language has for a long time attempted to establish prerequisites for language

that are absent in non-human animals in general, and in the extant non-human apes in particular (e.g. Johansson 2005; Fitch 2010). Early research primarily focused on the anatomical differences between the vocal tract of modern humans and non-human apes or other hominin species (Lieberman and Crelin 1971). While this line of research has generalized to a quest for *sensorimotor* differences (Boë et al. 1999, 2002; Corballis 2002; d'Errico et al. 2003), there is a degree of consensus that such differences are not fundamental and thus unable to explain the qualitative differences between animal and human communication (Fitch 2010). The focus has thus shifted to a quest for *cognitive* prerequisites, with proposals such as human-specific forms of 'theory of mind' (e.g. Bräuer et al. 2007), meta-representation (e.g. Suddendorf and Whiten 2001; Dunbar 2007; Horton and Brennan 2016), and memory (Coolidge and Wynn 2005; Hurford 2011; Tallerman 2011; Corballis 2013). Others have focused on the *social* (or rather, socio-cognitive) aspects such as pro-sociality (Tomasello 2008), complex action imitation (Arbib 2012), and trust (Knight 1998; Waciewicz and Żywicznyński 2018).

As a complement to these approaches, we here focus on *semiotic* (i.e. meaning-making, and in particular communicative) prerequisites for the evolution of language (Donald 1991; Deacon 1997; Hurford 2011). Further, by adopting concepts from the new discipline of *cognitive semiotics*, which integrates theories and methods from the traditional fields of semiotics, linguistics, and cognitive science (Zlatev 2009, 2015, 2018; Sonesson 2010; Konderak 2018), our aim is to integrate evolutionary considerations regarding semiotic, cognitive, and socio-environmental factors into a unified approach. This should be applicable not only to the evolution of language, but to the evolution of *polysemiotic communication* (Green 2014; Zlatev 2019): the combination of a number of different semiotic systems within an integrated communicative system, as we explain in the following sections.

To summarize our proposal, we begin by proposing that the first semiotic threshold that our ancestors needed to pass was to evolve the ability to use *signs*, as opposed to *signals* (Sonesson 2010; Zlatev 2009, 2015, 2018). Both signs and signals form *semiotic systems*, which can be sign systems or signal systems. Language constitutes a paradigmatic *sign system*, and, arguably, so do gesture and depiction (Zlatev 2019). The well-known vervet monkey alarm calls (Seyfarth and Cheney 1990), on the other hand, constitute a *signal system*. When two or more semiotic systems (of either kind) are

combined in an integrated communicative system, the latter is *polysemiotic*; else it is *monosemiotic*.

The question substituting the one about the original 'protolanguage' can thus be reformulated as: *What was the original human-specific communicative system?* As reflected in the title, the answer we propose is that it was *pantomime*, understood as a communicative system, with *gesture* as the sign system at its core, but also containing vocalizations, and at least some aspects of depiction. The latter systems would with time evolve into full-fledged speech (a sub-system of language) and drawing (a sub-system of depiction), respectively. We thus propose that pantomime was from its onset *polysemiotic*, combining different semiotic systems, as well as *multimodal*, involving different sensory channels. Importantly, these two notions are not synonymous (Stampoulidis, Bolognesi and Zlatev 2019).

In Sections 4 and 5, we provide empirical support for this theoretical proposal. First, we outline a scenario for the evolution of pantomime on the basis of mimesis theory (Donald 1998, 2012, 2013; Zlatev 2008b, 2014a), and then ask: *what kind of gestures would have characterized the core of pantomime?* Since, following mimesis theory, they should have been of the kind that are representational, but most similar to practical actions, we use distinctions made in semiotics and gesture studies to describe such 'pantomimic gesture'. We conclude in Section 6 by summing up the implications of our proposal for pantomime as the original human-specific communicative system.

2. From signals to signs

A crucial, though theoretically underdeveloped distinction is that between two different kinds of semiotic units: *signals* and *signs*. Although there have been attempts to blur this distinction (Hauser and Konishi 2003; Dennett 2017), it is affirmed by a majority of researchers in language evolution and animal communication (Deacon 1997; Hurford 2007; Tomasello 2008; Zlatev 2003). It is also pivotal for Vygotsky's (1978) developmental semiotics.

The assumption that there is a fundamental difference between the meaning-making involved in language versus various natural forms of communication is reflected in the well-known methodological statement of Burling (2005: 16; cf. Hurford 2007): 'We will understand more about the origins of language by considering the ways in which language differs from the cries and gestures of human and non-human primates than by looking for ways in which they are alike'. The point is

that from a semiotic perspective, most of the ‘cries and gestures’ of animals, as well as those of human beings (like laughter or yawning), are qualitatively different from signs such as (1) a word like *dog*, (2) an iconic (i.e. resemblance-based) gesture that resembles a dog, or (3) a picture of a dog. The latter are examples of *signs*, while the former are *signals*. By adapting and extending the definition of what distinguishes language from signal systems made by Zlatev, Persson and Gärdenfors (2005), we can characterize the differences between signals and signs in terms of the criteria given in Table 1.

In terms of (#1), animal signals like alarm calls or bird song (Marler 1991) clearly involve a degree of learning, and are not completely innate (i.e. genetically determined). However, such learning is limited in terms of both mechanisms and in scope, making signals typically common for the species as a whole (Hauser 1996). Signs (like words and most human gestures), on the other hand, are learned socially, with imitation playing a key role (Piaget 1962; Richerson and Boyd 2005). This is key for them to form sign systems that are *open* (see Section 3), leading to almost limitless variation across cultural groups.

With respect to (#2), typical signals like chimpanzee food cries are produced with limited volitional control (Deacon 1997). Of course, there are well-known *audience effects*, showing for example that ‘a visible audience increase[s] the rate of food calling for a large, sharable quantity of food, yet decrease[s] the rate of a small, non-sharable quantity’ (Brosnan and de Waal 2002: 211, cited in Hurford 2007: 233). Thus chimpanzee calls, and even more so chimpanzee gestures (Tomasello 2008), cannot be regarded as fully out of conscious control. Likewise, it could be doubted how voluntary the production of some human signs, like showing another driver the middle finger, actually is. Still, the two types clearly differ in the degree of voluntary control involved in their production, a typical kind of signal use leaning towards the automatic side of the continuum, and sign use—towards the pre-meditated side.

Vygotsky’s ideas concerning the difference between signs and signals in the context of child development are very much in line with this claim (Vygotsky 1978). On his account, it is through signs that children can break away from stimulus-response behaviour, understood as an automatic reaction of an organism to a change in the environment. The sign provides a mediating link between Stimulus and Response, and thus allows greater control of one’s behaviour. That non-human animals are to some degree able to acquire this was shown in a study based on a reversed-reward contingency task (Boysen et al. 1996), where chimpanzees were taught to point at lesser and larger amounts of food, receiving what they did *not* point to. It was found that they could not stop pointing at the larger amount, despite the fact that this gave them the smaller reward. However, when they were given ‘tokens’ standing for the two different kinds of portions, they could learn the task, pointing to the token (sign) for the lesser quantity, so as to receive the larger one. The sign had mediated between S and R, in Vygotsky’s terms, making the chimpanzees less controlled by their direct perceptions and wants.

What (#3) emphasizes is that despite ‘audience effects’, and some other limited forms of context sensitivity, signals are provoked by the current situation: be this external conditions such the presence of food or danger, or internal conditions like hunger. This makes them tightly linked to the here or now; in semiotic terms, they are *indexical* to one or more elements of the present situation. Note, however, that this does not make them indexical signs, which requires clear *differentiation* from the referent (e.g. Sonesson 2007; see below). Signs are largely independent of the physical context in which they are produced, and denote *objects* (in a general sense of the term, including properties or events) that may or may not be currently present, thereby displaying *displacement* (Hockett 1960).

From the perspective of the recipient (#4), signals are likewise inflexible, leading to more or less fixed patterns

Table 1. Six key features distinguishing signals and signs.

Feature	Signals	Signs
(#1) Learning	Mostly innate	Mostly learned socially
(#2) Conscious control	Low	High
(#3) Production	Tied to a particular context	Flexible, relatively independent of specific context
(#4) Interpretation	Relatively fixed response	Flexible and ‘negotiable’
(#5) Communicative relations	Mostly dyadic: a. Communicator–Recipient b. Object–Communicator	Mostly triadic: • Communicator–Object–Addressee
(#6) Untruthful use	Hard	Easy

of response; for example alarm calls tend to be promptly followed by the appropriate flight responses. A sign can, on other hand, always be interpreted in more or less various ways: ‘*Beware of the DOG*. Hmm, really? Do they really have a dog, and if so, is it really as dangerous as the sign says?’ This brings us to (#5): communicative signs are always *triadic*, in the sense that they involve a communicator, a denoted object, and an audience (that could be the communicator himself, once the sign has been internalized, and used for thought). A sign could be a pointing gesture, where the communicator intends to bring the attention of the addressee to a relevant object, and for the addressee to recognize this, rather than just to look in a given direction (Tomasello 2008; Zlatev, Brinck and Andrén 2008). In contrast, signals are *dyadic*, even when they involve two dyads: (1) communicator and audience, and (2) object and communicator. If only (1) is the case, signals are clearly dyadic, as in mating calls. If both (1) and (2) are the case, we have so-called ‘functionally referential’ signals like the alarm calls mentioned above. This, however, is qualitatively different from the true referential triangle of intentional *denotation* involved in sign use. This difference is consistent with properties (#2–4): a signal is typically produced involuntarily, as a response to something in the environment, leading to an equally direct recipient reaction; a sign is produced voluntarily, denoting something that is either present or absent in the situation, and the effect on the audience cannot in general be predicted.

Finally, we come to criterion (#6), which follows from what was said so far, but deserves to be spelled out on its own. Signal-based communication tends to be ‘honest’, i.e. non-manipulative, to the degree that the interests of the producers and receivers are aligned—one example being closely related individuals, between which cooperative communication can evolve through ‘inclusive fitness’ (Hamilton 1964). Any conflict of interests creates motivation for ‘deception’, so that to remain ‘honest’, signals must usually be stabilized by the cost of a signal, as in the case of handicaps such as the famous peacock’s tail (Zahavi 1975; Smith and Harper 2003). Signs, on the other hand, can easily be used untruthfully with anyone. Indeed they often are, in all human cultures, and yet this does not lead to a breakdown in communication (Dor 2017). This is so because sign-based communication more often than not follows Gricean principles (Grice 1957, 1968, 1975; cf. Scott-Phillips 2008; Waciewicz and Żywiczyński 2018), so that as a default, the communicator is expected to provide honest communication, and actually delivers on that expectation. This, on the face of it, is a puzzle, as honesty is neither guaranteed by the cost of the sign nor by the

alignment of interests between the communicator and recipient. A candidate solution of this apparent paradox is given by the recent ‘social turn’ in evolutionary linguistics, which states that an advanced form of intersubjectivity and a high degree of trust were preconditions, not only for language, but also for the evolution of sign use in general (Dor, Knight and Lewis 2014; Zlatev et al. 2008).

Therefore, we may agree with Eco (1976), who stated that a sign is ‘everything that can be used in order to lie’ (p. 7). However, Eco had a rather broad notion of the sign, including the manipulative signals of animals, mentioned earlier. Consider the red deer dynamically lowering its larynx to increase the functional length of its vocal tract so as to produce a more impressive roar, leading the audience to overestimate its body size (Fitch and Reby 2001). Does this classify as lying? We think not, as lying presupposes the ability to strategically and flexibly choose between being truthful or not, which does not apply to the roaring deer. In terms of the well-known typology of deception (Mitchell 1986), only human beings, and possibly some non-human great apes, are capable of the highest form of *intentional deception*. Chimpanzees are capable of this, but characteristically not through their (vocal) signals, but through bodily actions (De Waal 2007). Thus, we could revise Eco’s statement: *You can lie with signs, but not with signals*.

In sum, the discussion in this section, based on the six criteria of Table 1, shows that signs and signals differ qualitatively rather than quantitatively. While we allow for the gradation between signals and signs on the individual properties in Table 1, taken together they constitute a *qualitative difference*. While it is not impossible for non-human animals to learn signs under special conditions, their communication in the wild overwhelmingly takes place through signals. This makes good evolutionary sense, as sign use implies a great degree of *freedom*: in learning (#1), control on whether to communicate or not (#2), choice in topic (#3 and #5), interpretation (#4), and to be truthful or not (#6). To manage this requires cognitive and social prerequisites that are absent in non-human animals. To rephrase the conclusion from Zlatev et al. (2005: 3), where the focus was on language and not on sign use in general, ‘no animal in the wild has anything approaching [a] socially transmitted, voluntarily controlled, contextually flexible, triadic semiotic system’, which is a wider notion than language.

The discussion so far has defined signs in contrast to signals in terms of a set of particular properties, resembling Hockett’s (1960) ‘design features’. But is there not

a deeper theoretical basis for these differences? Following proposals from phenomenology (Sokolowski 2000) and cognitive semiotics (Zlatev 2009, 2018), we may define sign use as follows:

DEF. A sign $\langle E, O \rangle$ is used (produced or understood) by a subject S , if and only if:

(a) S is made aware of an intentional object O by means of expression E , which can be perceived by the senses.

(b) S is (or at least can be) aware of (a).

First, this definition implies that sign use requires a conscious subject, in both production and comprehension. It does *not* imply that signs are only used in intra-personal communication as a doctor observing a given symptom E (like coughing), could conclude that it is a sign of a given intentional object O (like Covid-19). However, the definition implies that E is not just associated with O , as a stimulus is to a response, but signifies it. Condition (b) guarantees the latter. Signals, like vervet monkeys' alarm calls, may fulfill condition (a), but not (b). Definitions of sign use based on 'double asymmetry' between E and O , with E being more directly perceivable, and O more in focus (Sonesson 2007, 2010) are compatible with our definition, but we submit that ours is more general. It is hard to see why there would be no asymmetry and differentiation between, for example, alarm calls and the corresponding predators. On our account, it is the reflective awareness of the directed relation between E and O in (b) that distinguishes signals from signs.

The link between E and O constitutes the *semiotic ground* of the sign (Ahlner and Zlatev 2010). Following an influential interpretation of Peircean semiotics (Sonesson 2010), this ground can be of three different kinds: *iconic* (based on the expression-object resemblance, E - O), *indexical* (contiguity in space/time between E and O) and *symbolic* (based on convention, that is: common knowledge that E denotes O). The three coexist in a single act of sign use, but typically one predominates (Jakobson 1965). For example, a pointing gesture typically involves deixis (a special kind of indexicality) with respect to its object, resemblance (i.e. iconicity) with the intended gaze alternation or motion of the addressee, and conventionality (i.e. symbolicity), since there are different norms for pointing in different cultures. Yet, it is generally considered an indexical sign, as the first kind of ground is arguably most essential for establishing the E - O relation.

But what allowed the evolution of a communicative system based on signs? Once again, we are back to the

question of prerequisites. Section 4 will spell out several of the social-cognitive adaptations necessary for learning signs, including enhanced (i) learning of bodily skills, (ii) intersubjectivity and trust, and (iii) imagination, all of which were provided by *bodily mimesis*. Jointly, they paved the way for the evolution of signs, and further, of *sign systems*, a notion that we explain in the following section.

3. Semiotic and communicative systems

While sign use constitutes an important semiotic threshold, individual signs are of little use on their own. Rather, the power of sign use comes from combining signs to form more or less complex messages such as stories. Further, depending on their properties and interrelations, signs form systems. The best-known and arguably most fundamental ones are the sign systems of *language* (with speech and writing as sub-systems), *gesture* (where the expression of the sign is always produced by the body of the communicator)¹ and *depiction* (where the expression consists of marks made on a two-dimensional surface).² Table 2 shows some of the properties of these systems. Excluding for the time being signed languages, language is either spoken or written, and in the latter case it is characterized by a high degree of *permanence* of its expressions, at least in prototypical written media. In either case, it has so-called 'double articulation': phonemes or graphemes (or other elements that are meaningless in themselves) combine systematically to form meaningful 'morphemes'. Its semiotic ground is predominantly conventional, even if iconicity and indexicality are also present (Jakobson 1965). Its 'syntagmatic' (sequential) relations are characterized by a high degree of compositionality, where the meaning of a composite sign is built up from the meanings of its constituent signs, and the 'rules' for combining these.

In the case of gesture and depiction, on the other hand, the predominant grounds are iconic and indexical, even if conventionality is also important (see Section 5). Gestures with a dominant iconic ground denote their objects on the basis of resemblance more than on shared conventions, and the same goes for depiction. It is usually possible to map specific elements of the expression to specific ones on the object. For example, a gesture of two fists bumping into one another to denote a car crash corresponds to the two cars, and a picture of a forest will typically have individual trees. But these elements will combine in a holistic way as in perception, and not as in language, where the words/morphemes in 'car crash' and 'many trees' combine as heads and modifiers. Further, there is nothing corresponding to phonemes in

Table 2. The sign systems of language, gesture and depiction, with some of their properties.

Properties	Sign systems			
	Language		Gesture	Depiction
	Speech	Writing		
Production	Vocal	Material	Body	Material
Perception	Auditory (+ Visual)	Visual	Visual (+Auditory + Tactile)	Visual (+ Tactile)
Permanence	Very low	High	Low	Intermediate
Double articulation	Yes		No	No
Semiotic grounds	Conventional > Iconic + Indexical		Iconic + Indexical > Conventional	Iconic + Indexical > Conventional
Syntagmatic relations	Compositional		Linear	Possibly linear

gesture and depiction. Their signs can be analysed into phrases and units (Kendon 2004; Green 2014), but these are not made up of minimal distinctive elements.

Further, gesture and depiction also have much less systematic manners in arranging sequences of signs, making it more difficult, though not impossible (e.g. Sibierska 2017), to express complex messages such as narratives. In terms of permanence, both are intermediate to speech and writing. The permanence of gesture is similar to but not identical with that of speech, as one can ‘hold’ gestural expressions in mid-air, so to speak, when the need for emphasis or audience attention requires it. The permanence of depiction, on the other hand, is similar to that of writing.³

Given the many similarities between gesture and depiction as semiotic systems, what are the main differences between them? In terms of production, in depiction the producer’s physical action results in a static expression, with shape and color as the most important features for establishing its meaning. In the case of gesture, the dynamic bodily action itself constitutes the expression, and this can either denote intentional objects, or contribute to the social interaction more generally in so-called *pragmatic gestures*. This makes gestures particularly suitable for face-to-face communication (Goffman 1963), where it is necessary that interactants may easily change the roles of producer and recipient. Turn-taking is traditionally understood in terms of how speech is distributed in conversation (e.g. Schegloff 1996, 1998), but research has shown the importance of gestures in the process (Kendon 1967; Duncan 1972; Ho et al. 2015; Torreira et al. 2015). This is not to say that depiction cannot be used interactively, as in some traditional forms of narration, where speaking, gesturing and drawing in the sand fluently combine (Green 2014). There is also a large body of experimental-semiotic research on the interactive use of drawings, which leads to the emergence of communicative conventions (Galantucci 2017).

Yet, given that gesture is part and parcel of the living body and can be used in any context, while depiction requires specific material resources, gesture appears to show advantages over depiction for the purpose of managing communicative interactions.

Sign systems such as language, gesture and depiction are clearly semiotic systems (Zlatev 2019), but the terms *semiotic* has a broader application than signs, and concerns other kinds of meaning such as *affordances* (Gibson 1979; Sonesson 2007), and indeed signals, as defined in Section 2. Hence, we propose to use the term *semiotic system* as a superordinate concept, comprising both sign systems as those in Table 2, as well as *signal systems* such a bee dance and various species’ alarm calls. The differences between signs and signals discussed in Section 2 naturally carry over to their corresponding systems. This implies that sign systems are *open*, with no limit on what they can express by either inventing new signs, or novel combinations of existing signs. Signal systems, on the other hand, are closed, with strict limits on what can be used to communicate with them. For example, putty-nosed monkeys are known to have two basic alarm calls: ‘pyows’ that act as a general warning call, and ‘hacks’ that indicate an approaching eagle (Arnold and Zuberbühler 2012). They have no dedicated call for, say, humans or any other special kind of predator. Occasionally, they may combine calls into ‘pyow-hack’ sequences, but these are not compositional, as they form a functionally different type of signal, used for initiating the movement of a group (Schlenker et al. 2017).

Communication in *only* language, gesture, or depiction is thus *monosemiotic*. But if we consider actual human communication, especially in face-to-face contexts, interactants typically use both spoken language and gestures (McNeill 1992; Kendon 2004), and when the opportunity arises, also depiction. Such polysemiotic communication (Zlatev 2019)⁴ often takes place in

highly integrated ways (Green 2014), and thus we can speak of polysemiotic *communicative systems*. Their composing semiotic systems can be either sign systems, signal systems, or the combination of sign and signal systems, for example, speaking while using spontaneous facial expressions. In the following section, we propose that pantomime can be characterized as such a polysemiotic communicative system, evolving from bodily mimesis.

4. Mimesis and pantomime

4.1 Mimesis theory

The essential ‘cognitive prerequisites’ for the evolution of any kind of sign system can be arguably summarized through a single notion: *mimesis* (Donald 1998, 2012, 2013). Stemming from the Greek *mīmeisthai* (‘imitate’), within current evolutionary research the concept of (bodily) mimesis is used to encompass ‘...pantomime, imitation, gesturing, shared attention, ritualized behaviors, and many games. It is also the basis of skill rehearsal, in which a previous act is mimed, over and over, to improve it’ (Donald 2001: 240). What is common to these various capacities is according to Donald (2012: 180) ‘an embodied, analogue, and primordial mode of representation’.

According to Donald’s general theory of human cognitive-semiotic evolution, it was the ‘unified neuro-cognitive adaptation’ (Donald 2012: 181) of mimesis, emerging ca. 2 MYA in a particular species of hominins that served as the watershed between human beings and other animals, prior to the evolution of language (Zlatev 2014b). While the original function of mimesis was likely tool production, it eventually extended to cover much else. First, as noted repeatedly by Donald, mimesis allowed the development of increasingly complex bodily skills, based on the ability to ‘compare, in imagination, the performed act with the intended one’ (Donald 2012: 182). This would have been essential in, for example, the learning of projectile throwing, jumps, dance, and many kinds of rituals. Second, and related to this, mimesis allowed human-specific forms of social learning, such as complex imitation (Arbib 2005), ‘over-imitation’ (Horner and Whiten, 2005), and more generally, pedagogy (Gergely and Csibra 2006). In other words, bodily mimesis was not only a purely motor-cognitive adaptation, but also as a *social-cognitive* one, co-evolving with aspects of intersubjectivity such as trust and empathy (Hutto 2008; Zlatev 2008a). Third, as a consequence of the previous two points, mimesis propelled the evolution of *imagination* (a form of intentionality that is not directed to what is present but to what is

absent, see Sokolowski 2000) to unprecedented levels (Zlatev 2014b, 2019).

However, none of these capacities necessarily involve sign use as defined in Section 2, since they do not imply clear awareness of the link between expressions and objects, and denotational relations between them. To help distinguish different ‘kinds’ or functions of mimesis, as well as to distinguish it more clearly from language, a more formal definition of *bodily mimesis* such as the following could be used:

... an act of cognition or communication is an act of bodily mimesis if: (1) it involves a cross-modal mapping between exteroception (e.g. vision) and proprioception (e.g. kinesthesia); (2) it is under conscious control and is perceived by the subject to be similar to some other action, object or event, (3) the subject intends the act to stand for some action, object or event for an addressee, and for the addressee to recognize this intention; (4) it is not fully conventional and normative, and (5) it does not divide (semi)compositionally into meaningful sub-acts that systematically relate to other similar acts, as in grammar (Zlatev 2014b: 206).

Starting from the end, the last two points are meant to distinguish acts of mimesis from any kind of language. The various kinds of capacities discussed above, such as skill rehearsal, imitation and teaching, involve points (1) and (2), and their combination may be referred to as *dyadic mimesis*. While this *may* involve sign use, as in private pretend play, it does not have to, as in acts of *demonstration*, as discussed below. Adding (3) brings an explicitly communicative, ‘Gricean’, element, leading to *triadic mimesis* (Zlatev 2008b; Zlatev et al. 2013) and thus a case of sign use.

But what type of communicative system would evolve from this? Given the focus on the physical body, the core of the system should be constituted by the sign system of *gesture*. Further, given that such gestures should ‘be similar to some other action, object or event’ (see the citation above), we can expect these gestures to be of the kind that *maximally resemble* their intentional objects. In Section 5, we explore properties of such ‘pantomimic gesture’ based on a number of distinctions that have been made in the literature. But triadic mimetic acts could also leave traces on a surface, and thus include a preliminary form of depiction (Zlatev 2019). And it may involve vocalizations, as either signals or signs. It is this polysemiotic communicative system emerging from bodily mimesis

that corresponds to the notion of *pantomime*, previously defined as:

a non-verbal, mimetic and non-conventionalised means of communication, which is executed primarily in the visual channel by coordinated movements of the whole body, but which may incorporate other semiotic resources, most importantly non-linguistic vocalisations. [This may] holistically refer to a potentially unlimited repertoire of events, or sequences of events, displaced from the here and now (Żywiecziński et al. 2018).

This definition needs to be elaborated further by (a) underlining the *polysemiotic* character of pantomime, (b) the fact that the dominant sign system in pantomime is *gesture*, and (c) pointing out that such ‘pantomimic gesture’ is a gradient phenomenon, also with respect to conventionality. We turn to (b) and (c) in Section 5, but first it remains to clarify our concept of pantomime with respect to demonstration, vocalizations and (proto)drawing.

4.3 Demonstration, vocalizations, and drawing

Gärdenfors (2017) highlights the similarities between pantomime and ‘demonstration’, by which he means a teacher performing an action for the benefit of a student. He defines it as follows:

(D1) *The demonstrator actually performs the actions involved in the task.*

(D2) *The demonstrator makes sure that the learner attends to the series of actions.*

(D3) *The demonstrator intends that the learner perceives the right actions in the correct sequence.*

(D4) *The demonstrator exaggerates and slows down some of the actions in order to facilitate for the learner to perceive important features.*

On the one hand, demonstration differs from practical action, since its main goal is not a practical end, but for the student to understand how to perform such actions. Further, Gärdenfors underlines the mimetic nature of demonstration, involving conscious (volitional) use, other-directedness, and ‘mind-reading’. Demonstration bears a strong similarity with pantomime, since criteria D2-D4 can function almost verbatim as criteria for defining pantomime, substituting ‘addressee’ for ‘learner’. However, D1 is clearly absent in pantomime, given that in the latter the communicator at best ‘pretends’ to perform the actions, without

actually doing so. This leads Gärdenfors to conclude that pantomime is a form of pretence (Leslie 1987). If so, it is a very special form, as the intention of the communicator is to have this ‘pretence’ understood as such, in line with (3) in the definition of mimesis.

Returning to the definitions from above: pantomime, but not demonstration, implies triadic mimesis, and consequently sign use, in the semiotic system of gesture. Considering the definition of sign use given in Section 2, in pantomime the bodily action (i.e. the gestural expression E) is meant to bring into focus what the bodily action denotes, rather than the action itself. In demonstration, on the other hand, the student must attend to the bodily action itself to be able to learn it as best as he or she can. Thus, despite the similarity between the two behaviours, and their basis in bodily mimesis, they are importantly different. Evolutionarily, demonstration can at best be considered a precursor of pantomime.

What about the status of vocalizations and drawing in pantomime, given that we defined it as a communicative system that is dominated by gesture, but is nevertheless polysemiotic? Using the distinction between signs and signals, as defined in Section 2, it is reasonable to assume that at an initial stage of pantomime, vocalizations served as signals, with features such as little volitional control and dyadic nature, similarly to the vocal calls of non-human animals, such as alarm and food calls. The basis for this conjecture is evidence for greater volitional control of the body than of the vocal apparatus in the great apes, and presumably in the ‘last common ancestor’ (Tomasello 2008). This has been challenged (e.g., See 2014), but the exact degree of voluntary control of vocalization is not essential. More important is to acknowledge that with evolutionary time, vocalizations would inevitably have come under *increased* volitional control, and assumed the status of expressions in sign-relation. This is also implied by Donald’s notion of ‘voco-mimesis’ (Donald 2013).

For example, consider the Bayaka nomads of the Western Congo Basin, who typically incorporate the vocalizations of the hunted game into their hunting narratives (Lewis 2014). This is clearly based on the basic mimetic ability to produce self-initiated and representational acts, with the vocalization denoting the corresponding animal on the basis of an iconic ground, i.e. triadic mimesis.⁵ In the further evolution of polysemiotic communication, pantomime would inevitably have comprised such vocalizations, which would have contributed to the overall message.

In fact, this can be seen as the process towards the evolution of speech, given that recent research shows

that an iconic ground is to a considerable extent present along with a symbolic one across many languages (Ahlner and Zlatev 2010; Dingemanse 2012) and that it can contribute to successful communication, in particular with respect to movement-related concepts (Imai and Kita 2014). On the other hand, gesture has a greater potential for ‘bootstrapping’ a communication system than non-linguistic vocalization (e.g. Fay et al. 2013; Schouwstra et al. 2019). Further, when gesture is combined with non-linguistic vocalizations, this does not necessarily increase the accuracy in identifying communicated meanings over the purely gestural communication (Fay et al. 2014; Zlatev et al. 2017). Such findings support our claim that gesture played a dominant role compared to vocalization in pantomime, at least in its early stages.

As for depiction, it should be noted that the definition of triadic mimesis applies to bodily acts, and not to their products. Thus, an already completed painting would not qualify as triadic mimesis. Indeed such ‘exograms’ were considered by Donald to evolve at a much later stage, following rather than preceding language. The cross-cultural phenomenon of sand drawing, however, has helped focus attention on *depiction as a process*, produced more or less spontaneously in an interactive context (Green 2014). Thus, a gesture that leaves a trace on a surface could indeed be triadic-mimetic, and the relatively greater permanence of the trace (see Table 2, Section 3), would also be able to contribute to the communication of a complex message, such as a story. Hence, analogously to the case with vocalizations, we can presume that pantomime (i.e. triadic mimesis) would also have comprised at least a form of proto-depiction, which given time and appropriate context could have evolved into depiction proper.

5. Characteristics of pantomimic gesture

We have so far argued that pantomime, evolving from bodily mimesis, was the original human-specific communicative system, and that it was a polysemiotic system, with the sign system of gesture at its core. How can we characterize the nature of such gesture, and what can we say about its further evolution, analogously to the way vocalizations evolved into speech, and tracing into depiction? We introduce the notion of *pantomimic gesture* not as a specific gesture type, as used in various classifications of modern gestures (McNeill 1992, 2005; Goldin-Meadow 1999; Marentette et al. 2016), but as a label for the gestures that characterized the original form of pantomime, which as already stated, must have been maximally similar to the actions and objects they

represented. In the terms of Werner and Kaplan (1963) they are the kind that must have had the least ‘symbolic distance’.

To characterize these, we here adopt a number of dimensions discussed in cognitive semiotics and gesture studies, focusing on (a) those that have *empirical implications*, (b) on *complementary* dimensions, and not on terminological differences of essentially the same phenomena and (c) on these dimensions *as continuous rather than discrete*. That is, the dimensions may be used as criteria to classify a gesture as ‘more-or-less pantomimic’. This is appropriate given our evolutionary perspective, as we conclude in Section 5.7. The first three dimensions are more general and derived from basic semiotic properties. The latter three focus on its more specific characteristics and refer to distinctions made in the gesture studies literature. While we cite some of the authors involved in developing these notions, we deviate from some of their interpretation and terms, for the sake of the points (a–c) above.

5.1 Ground: predominantly iconic

As pointed out repeatedly in previous sections, considered as a sign system, gesture is in general dominated by iconic (resemblance-based) and indexical (proximity-based) semiotic ground (see Table 1, Section 3). However, this is a macroscopic generalization, and we need to make space for qualifications for a number of cases where conventionality (i.e. symbolicity as a semiotic ground) becomes an essential aspect of gestural meaning.

First, and most obviously, there are well-known gestural *emblems* (Ekman and Friesen 1969; or *quotable gestures* (Kendon 1984) like the OK sign, which need to be learned more or less as words. Here, condition (4) in the definition of bodily mimesis in Section 3 is obviously not fulfilled. Second, there are *recurrent* gestures that are common in a culture, without being as conventionalized as emblems. Many iconic gestures are not just spontaneous productions, but precisely such recurrent, *typified* representations, which, interestingly, can also be seen in the gestures of 2- to 3-year-old children (Andrén 2010; Zlatev 2014c).⁶ In relation to this, what is being represented, the object, also undergoes typification; for example, the way eating will be gestured in Sweden and India will differ given that the action of eating takes different forms in different cultures. Finally, we should also mention gestures that do not represent, and thus do not function as signs, so much as help coordinate social interaction, so-called *pragmatic gestures* (Poggi and Zomparelli 1987). While these may have iconic origins (Müller 2016), they are highly

conventionalized, more or less dependent on language, and hence should be classified as ‘post-mimetic’.

We can then conclude that pantomimic gesture is predominantly iconic, but also at least to some degree combining a symbolic (conventional) ground, as in typification. Conventionality may be expected to increase in repeated transmission.

5.2 Iconicity type: mostly primary

Sonesson (1997) made the distinction between *primary* and *secondary* iconicity concerning the sign system of depiction, but it is more general, and can for example be applied to ‘sound symbolism’ (Ahlner and Zlatev 2010). In primary iconicity, the similarity between expression and object is sufficient for understanding that the former represents the latter. In secondary iconicity, it is the reverse: knowing that a given expression represents a given object is a necessary condition for the similarity to be perceived. Given this absolutely converse relation, it is natural to interpret the opposition in actual cases in terms of *proportions*: the understanding of a given iconic sign will be based on a ratio between primary and secondary iconicity, and not either-or (Giraldo 2019).

In the language evolution literature, the ‘self-sufficiency’ of pantomimic gesture to communicate various objects (events, properties, etc.) is often emphasized by stating that it should be comprehensible in the absence of any ‘verbally established’ context (e.g. Arbib 2012; Żywicznyński et al. 2018). But this does not imply that no context is needed—a pantomimic gesture that is as iconic as possible, say one representing kissing, will be performed and interpreted differently given different conventions concerning how kissing is performed (Zlatev 2014c). In general, the gestures on the ‘left side’ of the dimensions described below (Table 3) can be expected to be more primarily iconic than those on the ‘right side’.

5.3 Body expression: typically the whole body

Many everyday actions, such as walking, pushing, jumping etc., involve coordinated muscular activity across the entire body, and to represent these as iconically as possible would require a similar use of the whole body. For this reason, Żywicznyński et al. (2018) proposed ‘whole-bodiness’ as an important part of the definition of pantomime. It should be stressed, however, that this requirement concerns *pantomime as a communicative system* and does not need to hold for every individual instance of pantomime, so it is compatible with regarding gestures made with only the hands and arms, or even only with the head as pantomimes (e.g. Gärdenfors

2017; Brown et al. 2019). Rather than framing this as a terminological issue, we can again see this dimension as a gradual one. In other words, the greater the involvement of the entire body in a gestural representation, the more pantomimic it is. But even just opening one’s mouth to enact eating is to some degree pantomimic.

5.4 Viewpoint: mostly first-person

McNeill (1992) famously distinguished between gestures performed from a *character* and from an *observer viewpoint*. In the first case, the gesture ‘incorporates the speaker’s body into the gesture space, and the speaker’s hands represent the hands . . . of a character’ (McNeill 1992: 119). In the second, the gesture ‘excludes the speaker’s body from the gesture space and his hands play the part of the character as a whole’ (1992: 119). Analogously, Zlatev and Andr n (2009) distinguished between first-person perspective (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’. Note that the last qualification implies that even a ‘part-body’ gesture can be 1pp. For example, a pantomime of hammering (Figure 1) primarily relies on the movements performed with the dominant hand.

In the case of 3pp gestures, ‘the articulating parts of the body figure as observed objects, isolated from the rest of the body’ (McNeill 1992: 387). To revert to the example of hammering, the use of 3pp could consist in a clenched fist of the dominant hand representing the hammer, performing the hammering movement against the palm of the non-dominant hand (Figure 2). Finally, Brown et al. (2019)’s distinction between ‘egocentric’ and ‘allocentric’ gestures amounts to the same difference in viewpoint: the former are performed ‘with reference to the position and orientation of a person’s body’, while the latter: ‘with reference to objects and their surrounding environment’ (p. 3).

Here we use the terms ‘first-person’ (1pp) and ‘third-person’ (3pp) as more neutral and less loaded, as the other pairs have unwanted associations. Note that this does not say anything about the number of people whose perspective is taken. For example, when representing a transitive event, e.g. pushing somebody, the gesturer can adopt a two-person strategy (Adornetti et al. 2019), whereby first Agent, and then Patient is represented, but in each case adopting an appropriate 1pp gesture. Or consider gesturing the transitive event of being strangled by somebody, by applying your hands onto your neck; here, your body, with the neck as the focal part, represents the Patient while the hands

Table 3. The six dimensions, as characteristics of pantomimic gestures.

Dimension	More pantomimic	Less pantomimic
Dominant semiotic ground	Iconic ground	Symbolic ground
Type of iconicity	More primary iconicity	More secondary iconicity
Body	Using more of the body	Using less of the body
Viewpoint	Mostly first-person	Mostly third-person
Space	Action in peripersonal space	Action in extrapersonal space
Modes of representation	More Enacting	More Embodying and Tracing

represent the action performed by some distinct Agent (Figure 3).

In general, pantomimic gesture can be expected to be more 1pp rather than 3pp, and for a transition from the first to the latter to take place in evolution, with the increase of ‘symbolic distance’ and conventionalization. Not so much as an argument for this claim, but as an illustration of this transition on a parallel level, it has been shown that children’s first iconic gestures are 1pp (McNeill 1992; Zlatev 2014c), while 3pp gestures dominate the co-speech gestures of adults. Skarphedinsdottir (2019) observed a transition from 1pp to 3pp in the 4th year of life of three Swedish children.

5.5 Space: mostly peripersonal

Brown et al. (2019) have recently proposed a complex classification scheme for pantomime that partly overlaps with ours, but it would take us too far aside to make a detailed comparison. We simply borrow one dimension that is distinct from the others we discuss. As mentioned in Section 5.4, their distinction of ‘egocentric’ versus ‘allocentric’ gestures corresponds to the distinction 1pp versus 3pp. In addition, they propose that this distinction corresponds to another distinction: gestures representing actions in *peripersonal space* (i.e. the space immediately surrounding one’s own body) versus gestures representing actions in *extrapersonal space* (i.e. the space far from one’s body, see Cléry et al. 2015). But the two pairs (1pp/3pp and peri/extrapersonal space) do not overlap. For example, it is possible to have 1pp gestures in the case of ‘personification gestures’ (see below), where one pretends to be a bird, a tree, a house, etc.—objects that are clearly in the extrapersonal space. Conversely, it is possible to include 3pp gestures (e.g. the hand representing a mirror) in a gesture of an action that takes place in peripersonal space: observing oneself in a mirror. Alternatively, we could decide that this is primarily 1pp (and peripersonal), but that it combines a different mode of representation (see Section 5.6). Once more, we can say that the peripersonal/extrapersonal space dimension constitutes a cline, and that



Figure 1. Hammering represented by a first-person perspective gesture (1pp).

pantomimic gestures will tend to be of the first kind (overlapping in part with 1pp), but not exclusively so.

5.6 Mode of representation: mostly enacting

Over a number of publications, and with various terminologies, Cornelia Müller (e.g. 2013, 2016) has distinguished between ‘modes of representation’ that are not based on viewpoint or space, but on how properties of gestures map onto properties of the objects that are represented. At least five such modes can be distinguished, and we separate ‘tracing’ and ‘drawing’, which are conflated by both Müller (see above) and Brown et al. (2019):

- In the *Enacting* mode, the body of the gesturer strictly maps onto the (human) body of the object. This mode is thus appropriate for representing ‘actual manual activities, such as grasping, holding, giving, receiving, opening a window, turning off a radiator, or pulling an old-fashion gear shift’ (Müller 2013: 128).
- *Moulding* consists in using the hands to shape a 3-dimensional ‘transient sculpture’ of an object (e.g. a bowl).



Figure 2. Hammering represented by a third-person perspective gesture (3pp).



Figure 3. A 1pp gesture of a transitive event employing a 2-person strategy and incorporating both Agent and Patient.

- In *Embodying*, the hand or hands are used to stand for an object as a whole.
- In the case of *Tracing*, the gesture follows the path of a moving object.
- In *Drawing*, it makes a 2-dimensional outline of the shape of an object.

Again, it is important to point out the principled independence of this dimension compared to these above, even if there may be (strong) tendencies for co-occurrence. Enacting gestures can be 1pp, as in the case where there is an almost one-to-one mapping between the gesture and the corresponding action (Figure 1), but enacting gestures can also be 3pp, as when one is pretending to be strangled by someone else (Figure 3). Embodying gestures are usually 3pp as in Figure 2, but it is also possible to use 1pp gesture with the Embodying mode, as when one pretends to be an inanimate entity, such as a house (Figure 4).

A case in point is a special kind of enacting gesture known as *personification*, which maps ‘the body of a non-human entity onto the human body, using the human head to represent parallel locations on a non-human head, the human body to represent a non-human body, and human appendages to represent non-human appendages’ (Hwang et al. 2017: 4; Ortega and Ozyurek 2019). For example, the gesturer’s head would be mapped onto the bird’s head, the gesturer’s body onto the bird’s body and the gesturer’s hands onto the bird’s wings (Figure 5). This gesture is performed with the whole body (Section 5.3), with a 1pp viewpoint (Section 5.4), but represents something in the extrapersonal space (Section 5.5), since flying is not something that the human body affords.

Moulding also does not correspond neatly to one of the previous dimensions. It has a 1pp aspect in that the hands are shaped as-if ‘holding’ the object, but also a 3pp aspect: the invisible object. An Embodying gesture, on the other hand, can be performed from 1pp viewpoint representing something in extrapersonal space (Figure 4) or from a 3pp viewpoint representing something in peripersonal space, as in a ‘phone me’ gesture, when the hand takes the shape of a telephone (Figure 6).

Finally, Drawing and Tracing modes would normally correspond to 3pp viewpoint and to extrapersonal space, but if one is pretending to draw on a surface, one could argue that at least the latter takes on a 1pp viewpoint, and the whole gesture concerns an action in peripersonal space.

5.7 Summary

The six dimensions discussed in this section were described as clines, and shown to be at least in part independent of one another. We can thus use them, as shown in Table 3, to characterize *the degree to which an individual gesture can be described as pantomimic*: the more features it has on the left side, the more pantomimic it will be. At the same time, as the first dimension



Figure 4. A 1pp embodying gesture ('being a house').

(semiotic ground) is not absolute but proportional, so are the others: the more of the body as part of the expression, the more 1pp dominates, the more peripersonal space, the more Enacting mode of representation—the more pantomimic the gesture will be. This gives us a set of empirically definable criteria through which we can characterize the nature of pantomimic gesture. As for the evolution from pantomimic to modern gestures, this can be envisaged as a progression from the left to the right side in Table 3, not in the sense of the first kind 'disappearing', but rather as the latter appearing as a layer 'on top' of the former. In a very general sense, such a progression is what we observe in studies on the conventionalization of bodily-manual communication through repeated use, such as in emerging signed languages (e.g. Mineiro et al. 2017) and laboratory experiments in the 'silent gestures' tradition (e.g. Motamedi et al. 2018).

6. Summary and conclusions

In this article, we have proposed to reframe one of the key questions in the field of language evolution: from *What was the original protolanguage?* to *What was the original human-specific communicative system?* The difference is not only terminological. While the notion of 'protolanguage' has remained controversial over the past decades, that of a communicative system should be less so. We defined it as consisting of either a single semiotic system, making it monosemiotic, or of multiple such systems, and thus polysemiotic. Further, we defined semiotic systems as either consisting of signs or of signals. Language and gesture are examples of sign systems, while alarm calls and spontaneous facial expressions exemplify signal systems.



Figure 5. Personification: 1pp, extrapersonal space, enacting gesture ('flying as a bird').



Figure 6. A 1pp embodying gesture, which brings extrapersonal space into peripersonal space.

With the help of cognitive semiotics, we clarified the difference between signals, which characterize animal communication, and signs, which do not replace but complement signals in human communication, allowing a number of distinct semiotic features, and in particular *denotation*, implying awareness of the Expression–Object directed link. This, on its side, underlies the ability to *lie*, as opposed to simply deceive. Further, we argued that the evolution of (bodily) mimesis was what allowed for signs to emerge in hominin evolution, as well as the social-cognitive skills needed to support them, including advanced imitation, skill rehearsal, and cultural learning. We described how pantomime should be seen as naturally evolving from mimesis, as a polysemiotic communicative system, dominated by gesture but also including vocalizations, facial expressions, and possibly the rudiments of depiction.

We conclude that pantomime was the most likely original human-specific communicative system, both in terms of temporal primacy, and in the sense that it was the relatively undifferentiated whole from which

modern human systems emerged. This common origin could possibly be considered as one of the factors that keep different sign and signal systems aligned into complex polysemiotic communication, including the well-known alignment between language and gesture.

Concerning language evolution, our main thesis of pantomime as the original human communicative system is consistent with pantomimic accounts of language origin (Arbib 2012; Levinson and Holler 2014; Zlatev et al. 2017; Żywiczyński et al. 2018), which emphasize that language arose primarily out of visually perceived communicative action, with iconic gesture playing the dominant role in transmitting referential information. But it extends these by both characterizing more explicitly the notion of pantomime, and by emphasizing that the ‘target’ of the evolutionary process is not simply language, but modern polysemiotic communication – and above all the interaction of language, gesture and depiction (Zlatev 2019).

The implications of our ‘pantomime-first’ proposal go against claims that the original ‘protolanguage’ was monosemiotic: either vocal (e.g. Dunbar 1996), or gestural (e.g. Corballis 2002). However, it also contradicts so-called ‘multimodal accounts’ of language emergence, which stress a tight integration and equipotentiality of sign systems from the very beginning, as in the case of McNeill’s ‘growth point’ theory (McNeill 2012) or Kendon’s ‘speech-kinesic ensemble’ (Kendon 2014), given that we claim that gesture was the fundamental component of pantomime. We spelled out six different dimensions that can help characterize *prototypical* pantomimic gesture as: (1) mostly iconic, (2) with primary iconicity dominating, (3) using the whole body, (4) performed from a first-person perspective, (5) concerning actions or objects in peripersonal space, and (6) using the enacting mode of representation. The conventionalization of gesture and its establishment as its own semiotic system can be characterized as changes along these six dimensions, in the direction away from prototypical pantomime.

To conclude, this article has offered a number of novel ideas to the field of language origins. Conceptually, we have endeavored (1) to clearly define the distinction between signals and signs, with an explicit theoretical definition of the latter, (2) to distinguish two kinds of semiotic systems: signal and sign systems, and describe how they can combine to form polysemiotic communicative systems, (3) to define pantomime as a polysemiotic communicative system, and (4) specify features of its core system gesture, on the basis of specific dimensions. Empirically, we have brought these concepts, along with evidence supporting the pivotal role of bodily mimesis in human evolution, to make the claim that

pantomime, as here described, was the first communicative system that was based on signs rather than signals, and thus ushered humanity into new, and still uncharted territory, for the better or for the worse of the planet.

Notes

1. One could argue that the everyday use of the term “gesture” is much narrower, focusing on manual communicative movements that are perceived visually. In earlier work (Zlatev et al.) we have therefore contrasted gesture and pantomime, and distinguished between “gesture-first” and “pantomime-first” theories of language origins. Here, however, we refer to the broader notion of gesture as a type of semiotic system.
2. Other candidates that have been proposed are *ritual* (Knight 1998), *dance* (Mithen 2005; Savage 2015; Maraz et al. 2015), and *music* (Monelle 1991) to that extent that these are composed of signs. However, they differ from the definition of sign use provided in Section 2, by lacking clear intentional objects, and thus being non-referential. Music, for example, involves a complex network of relations between pitch, rhythm and dynamics (Nettl 2000). It appears to resemble language in some respects, given that all musical cultures employ a small set of ‘notes’, usually consisting of 5 or 7 frequency values and their integer multiples, to build a potentially infinite number of musical phrases (Fitch 2010; Arom 2000; Nettl 2000). However, these phrases do not denote, with the relatively rare exception of programmatic music (Giraldo 2019). There have even been attempts to link specific emotions to certain musical phrases: descending melodic contours with sadness, rising contours with happiness and so on (Nikolsky 2015) but this does not mean that descending and rising contours denote these emotions, unlike the linguistic signs “sad” and “happy”.
3. It is the *medium* (e.g. sand vs. stone) in which depiction is made that will largely determine its degree of permanence, and the same could be said of writing. Thus, the difference in permanence is not so much between the semiotic systems of writing and depiction themselves, but between their prototypical medial expressions. We thank Linea Brink Andersen for helping make this clarification.
4. Many authors refer to such communicative acts as ‘multimodal’, understanding language and gesture as ‘communicative modalities’ (Vigliocco et al. 2014). However, this terminology is problematic

because it confuses sign systems, or ‘modes’ in social semiotics (Kress 2009), and perceptual modalities, and there is no one-to-one correspondence between these, as also indicated in Table 1.

5. Interestingly, they use very similar vocalizations during their hunt, to lure the corresponding animals. Here, their signs are obviously not intended to be understood as such by the animals, but as signals, and are hence a clear case of deception.
6. This appears to correspond to the notion of *autonomous gestures* proposed by Marentette et al. (2016). There is much variability in the terminology used in the literature.

Acknowledgements

We are grateful to Linea Brink Andersen, Göran Sonesson, Simon Devylder and two anonymous reviewers for insightful comments to an earlier version of this paper. We wish to thank Aleksandra Poniewierska, Aleksandra Szczepańska and Michał Szczepański for their help in preparing the figures.

Funding

The work of J.Z. on this article was supported by the Swedish Research Council project *Polysemiotic Communication: Integrating Language, Gesture and Depiction in Three Cultural Practices*, grant 2018-01281. P.Z. was supported by the Polish National Agency for Academic Exchange (NAWA) under grant agreement PPN/BEK/2018/1/00289/U. S.W. was supported by the Polish National Science Centre (NCN) under grant agreement UMO-2017/27/B/HS2/00642.

Author contributions

J.Z. led the writing of all sections but Section 5, and was responsible for general organization and argumentation. P.Z. collaborated with J.Z. on all the sections of the article, and was specifically responsible for writing Section 5. Together with S.W., he edited the text. S.W. contributed ideas to Section 2, provided feedback on many intermediary versions of the text, and produced the figures. Together with P.Z., he edited the text.

Conflict of interest statement The authors declare that there is no conflict of interest regarding the publication of this article.

References

Adornetti, I., Ferretti, F., Chiera, A., Waciewicz, S., Żywicznyński, P., Deriu, V. and Valeri, G. (2019) ‘Do children with Autism Spectrum Disorders understand pantomimic events?’ *Frontiers in Psychology*, 10: 1382.

- Ahlner, F., and Zlatev, J. (2010) ‘Cross-Modal Iconicity: A Cognitive Semiotic Approach to Sound Symbolism’, *Sign System Studies*, 38: 298–348.
- Andrén, M. (2010) *Children’s Gestures between 18 and 30 Months*. Lund: Media Tryck.
- Arbib, M. (2005) ‘From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics’, *Behavioral and Brain Sciences*, 28: 105–68.
- Arbib, M. (2012) *How the Brain Got Language*. Oxford: Oxford University Press.
- Arnold, K., and Zuberbühler, K. (2012) ‘Call Combinations in Monkeys: Compositional or Idiomatic Expressions?’, *Brain and Language*, 120/3: 303–9.
- Arom, S. (2000) ‘Prolegomena to a Biomusicology’. In: Wallin, N. L., Merker, B. and Brown, S. (eds) *The Origins of Music*, pp. 27–29. London: MIT Press.
- Bickerton, D. (1990) *Language and Species*. Chicago: University of Chicago Press.
- Boysen, S. T. et al. (1996) ‘Quantity-Based Interference and Symbolic Representations in Chimpanzees (*Pan troglodytes*)’, *Journal of Experimental Psychology: Animal Behavior Processes*, 22: 76–86.
- Boë, L. J. et al. (2002) ‘The Potential Neandertal Vowel Space Was as Large as That of Modern Humans’, *Journal of Phonetics*, 30/3: 465–84.
- Boë, L. J., Maeda, S., and Heim, J. L. (1999) ‘Neandertal Man Was Not Morphologically Handicapped for Speech’, *Evolution of Communication*, 3/1: 49–77.
- Bräuer, J., Call, J., and Tomasello, M. (2007) ‘Chimpanzees Really Know What Others Can See in a Competitive Situation’, *Animal Cognition*, 10/4: 439–48.
- Brown, S. et al. (2019) ‘How Pantomime Works: Implications for Theories of Language Origin’, *Frontiers in Communication*, 4: 9.[10.3389/fcomm.2019.00009]
- Burling, R. (2005) *The Talking Ape*. Oxford: Oxford University Press.
- Cléry, J. et al. (2015) ‘Neuronal Bases of Peripersonal and Extrapersonal Spaces, Their Plasticity and Their Dynamics: Knowns and Unknowns’, *Neuropsychologia*, 70: 313–26.
- Coolidge, F. L., and Wynn, T. (2005) ‘Working Memory, its Executive Functions, and the Emergence of Modern Thinking’, *Cambridge Archaeological Journal*, 15/01: 5–26.
- Corballis, M. C. (2002) *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Corballis, M. C. (2013) ‘Mental Time Travel: A Case for Evolutionary Continuity’, *Trends in Cognitive Science*, 17: 5–6.
- De Waal, F. (2007) *Chimpanzee Politics: Power and Sex among Apes*. Baltimore: JHU Press.
- Deacon, T. (1997) *The Symbolic Species*. New York: Norton.
- Dennett, D. (2017) *From Bacteria to Bach and Back. The Evolution of Minds*. New York: Norton & Company.
- Dingemanse, M. (2012) ‘Advances in the Cross-Linguistic Study of Ideophones’, *Language and Linguistics Compass*, 6: 654–72.

- Donald, M. (1991) *Origins of the Modern Mind: Three Stages in the Evolution of Human Culture*. Cambridge, MA: Harvard University Press.
- Donald, M. (1998) 'Mimesis and the Executive Suite: Missing Links in Language Evolution'. In: Hurford, J. R., Studdert-Kennedy, M. and Knight, C. (eds) *Approaches to the Evolution of Language: Social and Cognitive Biases*, pp. 44–67. Cambridge: Cambridge University Press.
- Donald, M. (2001) *A Mind so Rare. The Evolution of Human Consciousness*. New York: Norton. Emery, NJ.
- Donald, M. (2012) 'The Mimetic Origins of Language'. In: Tallerman, M. and Gibson, K. R. (eds) *The Oxford Handbook of Language Evolution*, pp. 180–183. Oxford: OUP.
- Donald, M. (2013) 'Mimesis Theory Re-Examined, Twenty Years after the Fact'. In: Hatfield, G. and Pittman, H. (eds) *Evolution of Mind, Brain and Culture*, pp. 169–192. Philadelphia, PA: University of Pennsylvania.
- Dor, D. (2017) 'The Role of the Lie in the Evolution of Human Language', *Language Sciences*, 63: 44–59.
- Dor, D., Knight, C., and Lewis, J. (2014) *The Social Origins of Language* (Vol. 19). Oxford: Oxford University Press.
- Dunbar, R. (1996) Grooming, Gossip and the Evolution of Language. London: Faber & Faber
- Dunbar, R. (2007) 'The Social Brain Hypothesis and Its Relevance to Social Psychology'. In: Forgas, J., Haselton, M. and von Hippel, W. (eds) *Evolution and the Social Mind: Evolutionary Psychology and Social Cognition*, pp. 21–31. Psychology Press: New York.
- Duncan, S. D. (1972) 'Some Signals and Rules for Taking Speaking Turns in Conversations', *Journal of Personality and Social Psychology*, 23: 283–92.
- d'Errico, F. et al. (2003) 'Archaeological Evidence for the Emergence of Language, Symbolism, and Music – an Alternative Multidisciplinary Perspective', *Journal of World Prehistory*, 17: 1–70.
- Eco, U. (1976) *A Theory of Semiotics*. Bloomington: Indiana University Press.
- Ekman, P., and Friesen, W. V. (1969) 'The Repertoire of Nonverbal Behavior: Categories, Origins, Usage and Coding', *Semiotica*, 1: 49–98.
- Fay, N., Arbib, M., and Garrod, D. (2013) 'How to Bootstrap a Human Communication System', *Cognitive Science*, 37: 1356–67.
- Fay, N. et al. (2014) 'Creating a Communication System from Scratch: Gesture Beats Vocalization Hands Down', *Frontiers in Psychology*, 5: 354 [10.3389/fpsyg.2014.00354]
- Fitch, W. T. (2010) *The Evolution of Language*. Cambridge: Cambridge University Press.
- Fitch, W. T., and Reby, D. (2001) 'The Descended Larynx is not Uniquely Human', *Proceedings of the Royal Society, Biological Sciences*, 268/1477: 1669–75.
- Galantucci, B. (2017) *Experimental Semiotics. Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
- Gärdenfors, P. (2017) 'Demonstration and Pantomime in the Evolution of Teaching', *Frontiers in Psychology*, 8: 415.
- Gergely, G., and Csibra, G. (2006) 'Sylvia's Recipe: The Role of Imitation and Pedagogy in the Transmission of Cultural Knowledge'. In: Enfield, N. J. and Levinson, S. C. (eds) *Roots of Human Sociality: Culture, Cognition, and Human Interaction*, pp. 229–255. Oxford: Berg Publishers.
- Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
- Giraldo, V. (2019) 'Referential Iconicity in Music and Speech', *Public Journal of Semiotics*, 9/1: 41–77
- Goffman, E. (1963) *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: The Free Press.
- Goldin-Meadow, S. (1999) 'The Role of Gesture in Communication and Thinking', *Trends in Cognitive Sciences*, 3/11: 419–29.
- Green, J. (2014) *Drawn from the Ground: Sound, Sign and Inscription in Central Australian Sand Stories*. Cambridge: Cambridge University Press.
- Grice, H. P. (1957) 'Meaning', *Philosophical Review*, 66: 377–88.
- Grice, H. P. (1968) 'Utterer's Meaning, Sentence-Meaning, and Word-Meaning'. In: Kulas, J., Fetzer, J. and Rankin, T. (eds) *Philosophy, Language, and Artificial Intelligence. Studies in Cognitive Systems*, vol 2, pp. 225–42. Dordrecht: Springer. 225.
- Grice, H. P. (1975) 'Logic and Conversation'. In: Cole, P. and Morgan, J. L. (eds) *Speech Acts*, pp. 41–58. New York: Academic Press.
- Hamilton, W. D. (1964) 'The Genetical Evolution of Social Behaviour', *Journal of Theoretical Biology*, 7: 1–16.
- Hauser, M. D. (1996) *The Evolution of Communication*. Cambridge, MA: MIT press.
- Hauser, M. D., and Konishi, M. (eds). (2003) *The Design of Animal Communication*. Cambridge, MA: MIT press.
- Ho, S., Foulsham, T., and Kingstone, A. (2015) 'Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions', *PLoS ONE*, 10/8: e0136905.
- Hockett, C. F. (1960) 'The Origin of Speech', *Scientific American*, 203: 89–97.
- Horner, V., and Whiten, A. (2005) 'Causal Knowledge and Imitation/emulation Switching in Chimpanzees (*Pan Troglodytes*) and Children (*Homo Sapiens*)', *Animal Cognition*, 8/3: 164–81.
- Horton, W. S., and Brennan, S. E. (2016) 'The Role of Metarepresentation in the Production and Resolution of Referring Expressions', *Frontiers in Psychology*, 7: 1111.
- Hurford, J. R. (2007) *The Origins of Meaning; Language in the Light of Evolution*. Oxford: OUP.
- Hurford, J. (2011) 'The Origins of Meaning'. In: Tallerman, M., Gibson, K. R. (eds) *The Oxford Handbook of Language Evolution*, pp. 370–81. Oxford: Oxford University Press.
- Hutto, D. D. (2008) 'First Communion: Mimetic Sharing without Theory of Mind'. In: Zlatev, J., Racine, T., Sinha, C. and Itkonen, E. (eds) *The Shared Mind: Perspectives on*

- Intersubjectivity*, pp. 246–76. Amsterdam: John Benjamins Publishing.
- Hwang, S. et al. (2017) ‘Of the Body and the Hands: Patterned Iconicity for Semantic Categories’, *Language and Cognition*, 9/4: 573–602.
- Imai, M., and Kita, S. (2014) ‘The Sound Symbolism Bootstrapping Hypothesis for Language Acquisition and Language Evolution’, *Philosophical Transactions of the Royal Society B*, 369:20130298.
- Jakobson, R. (1965) ‘Quest for the Essence of Language’, *Diogenes*, 13/51: 21–37.
- Johansson, S. (2005) *Origins of Language: Constraints on Hypotheses*. Amsterdam: John Benjamins.
- Kendon, A. (1967) ‘Some Functions of Gaze Direction in Social Interaction’, *Acta Psychologica*, 26: 22–63.
- Kendon, A. (1984) ‘Did Gesture Have the Happiness to Escape the Curse at the Confusion of Babel?’ In: Wolfgang, A. (ed.) *Nonverbal Behavior: Perspectives, Applications, Intercultural Insights*, pp. 75–114. Lewiston, NY: C.J. Hogrefe.
- Kendon, A. (1985) ‘Some Uses of Gesture’. In: Tannen, D. and Saville Troike, M. (eds) *Perspectives on Silence*, pp. 215–34. Norwood, NJ: Ablex.
- Kendon, A. (2004) *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, A. (2014) ‘Semiotic Diversity in Utterance Production and the Concept of “Language”’, *Philosophical Transactions of Royal Society B*, 369/1651: 20130293.
- Knight, C. (1998) ‘Ritual/Speech Coevolution: A Solution to the Problem of Deception’. In: Hurford, J., Studdert-Kennedy, M. and Knight, C. (eds) *Approaches to the Evolution of Language. Social and Cognitive Bases*, pp. 68–91. Cambridge: Cambridge University Press.
- Konderak, P. (2018) *Mind, Meaning, Semiosis: Ways to Cognitive Semiotics*. Lublin: UMCS Publishing.
- Kress, G. (2009) *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Leslie, A. M. (1987) ‘Pretense and Representation: The Origins of “Theory of Mind”’, *Psychological Review*, 94/4: 412–26.
- Levinson, S. C., and Holler, J. (2014) ‘The Origin of Human Multi-Modal Communication’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369/1651: 20130302.
- Lewis, J. (2014) ‘BaYaka Pygmy Multi-Modal and Mimetic Communication Traditions’. In: Dor, D., Knight, C. and Lewis, J. (eds) *The Social Origins of Language*, pp. 77–91. Oxford: Oxford University Press.
- Lieberman, P., and Crelin, E. S. (1971) ‘On the Speech of Neanderthal Man’, *Linguistic Inquiry*, 2/2: 203–22.
- Maraz, A. et al. (2015) ‘Why Do You Dance? Development of the Dance Motivation Inventory (DMI)’, *PLoS One*, 10/3: e0122866.
- Marentette, P. et al. (2016) ‘Gesture and Symbolic Representation in Italian and English-Speaking Canadian 2-Year-Olds’, *Child Development*, 87: 944–61.
- Marler, P. (1991) ‘Differences in behavioural development in closely related’. In: Bateson, P. (ed) *The development and integration of behaviour: Essays in honour of Robert Hinde*, pp. 41–70. Cambridge: Cambridge University Press.
- McNeill, D. (1992) *What Gestures Reveal about Thought*. Chicago: The University of Chicago Press.
- McNeill, D. (2005) *Gesture and Thought*. Chicago: The University of Chicago Press.
- McNeill, D. (2012) *How Language Began: Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press.
- Mineiro, A. et al. (2017) ‘Emerging Linguistic Features of Sao Tome And Principe Sign Language’, *Sign Language & Linguistics*, 20/1: 109–28.
- Mitchell, R. W. (1986) ‘A Framework for Discussing Deception’. In: Mitchell, R. W. and Thompson, N. S. (eds) *Deception: Perspectives on Human and Nonhuman Deceit*, pp. 3–40. New York: State University of New York Press.
- Mithen, S. (2005) *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld and Nicholson.
- Monelle, R. (1991) *Linguistics and Semiotics in Music*. London: Routledge.
- Motamedi, Y. et al. (2018, April 14). The emergence of spatial modulation in artificial sign languages <<https://psyarxiv.com/p6zy4/>> (last accessed 21 June 2020)
- Müller, C., (2013) ‘Gestural Modes of Representation as Techniques of Depiction’. In: Müller, C. et al. (eds) *Body–Language Communication: An International Handbook on Multimodality in Human Interaction*, pp. 1687–1701. Berlin, Germany: De Gruyter Mouton.
- Müller, C. (2016) From Mimesis to Meaning: A Systematics of Gestural Mimesis for Concrete and Abstract Referential Gestures. In: Zlatev, J., Sonesson, G. and Konderak, P. (eds) *Meaning, Mind and Communication: Explorations in Cognitive Semiotics*, pp. 211–26. Frankfurt am Main, Germany: Peter Lang.
- Nettl, B. (2000) ‘An Ethnomusicologist Contemplates Universals in Musical Sound and Musical Culture’. In: Wallin, N. L., Merker, B., and Brown, S. (eds) *The Origins of Music*, pp. 463–72. Cambridge: The MIT Press.
- Nikolsky, A. (2015) ‘Evolution of Tonal Organization in Music Mirrors Symbolic Representation of Perceptual Reality. Part-1: Prehistoric’, *Frontiers in Psychology*, 6: 1405.
- Ortega, G., and Özyürek, A. (2019) ‘Systematic Mappings between Semantic Categories and Types of Iconic Representations in the Manual Modality: A Normed Database of Silent Gesture’, *Behavior Research Methods*, 52: 1–17.
- Piaget, J. (1962) *Play, Dreams, and Imitation in Childhood*. New York: Norton.
- Poggi, I., and Zomparelli, M. (1987) ‘Lessico e Grammatica Nei Gestii e Nelle Parole’. In: Poggi, I. (ed.) *Le Parole Nella Testa. Per Una Educazione Linguistica Cognitivista*, pp. 291–327. Bologn: Mulino.

- Richerson, P. J., and Boyd, R. (2005) *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
- Savage, P. E. et al. (2015) 'Statistical Universals Reveal the Structures and Functions of Human Music', *Proceedings of the National Academy of Sciences*, 112/29: 8987–92
- Schegloff, E. A. (1996) 'Turn Organization: One Intersection of Grammar and Interaction', *Studies in Interactional Sociolinguistics*, 13: 52–133
- Schegloff, E. A. (1998) 'Body Torque', *Social Research*, 535–96
- Schlenker, P. et al. (2017) 'Titi Semantics: Context and Meaning in Titi Monkey Call Sequences', *Natural Language & Linguistic Theory*, 35/1: 271–98.
- Schouwstra, M., de Swart, H., and Thompson, B. (2019) 'Interpreting Silent Gesture: Cognitive Biases and Rational Inference in Emerging Language Systems', *Cognitive Science*, 43/7: e12732.
- Scott-Phillips, T. C. (2008) 'On the Correct Application of Animal Signalling Theory to Human Communication'. In: Smith, A., Smith, K. and Ferrer i Cancho, R. (eds) *Proceedings of the 7th International Conference on the Evolution of Language*, pp. 275–82. Singapore: World Scientific.
- See, A. (2014) 'Reevaluating Chimpanzee Vocal Signals: Toward a Multimodal account of the Origins of Human Communication'. In: Pina, M. and Gontier, N. (eds) *The Evolution of Social Communication in Primates*, pp. 195–215. New York: Springer.
- Seyfarth, R., and Cheney, D. (1990) 'The Assessment by Vervet Monkeys of Their Own and Another Species' Alarm Calls', *Animal Behaviour*, 40/4: 754–64.
- Sibierska, M. (2017) 'Storytelling without Telling: The Non-Linguistic Nature of Narratives from Evolutionary and Narratological Perspectives', *Language & Communication*, 54: 47–55.
- Smith, M. J., and Harper, D. (2003) *Animal Signals*. Oxford: Oxford University Press.
- Skarphedinsdottir, C. S. (2019) Image schemas, mimetic schemas, and the development of Swedish children's gestures. Unpublished M.A. Lund University.
- Sokolowski, R. (2000) *Introduction to Phenomenology*. New York: Cambridge University Press.
- Sonesson, G. (1997) 'The Ecological Foundations of Iconicity'. In: Rauch, I. and Carr, G. F. (eds) *Semiotics around the World: Synthesis in Diversity*, pp. 739–42. Berlin & New York: Mouton de Gruyter.
- Sonesson, G. (2007) 'From the Meaning of Embodiment to the Embodiment of Meaning: A Study in Phenomenological Semiotics'. In: Ziemke, T., Zlatev, J. and Frank, R. (eds) *Body, Language and Mind*. Vol 1: Embodiment, pp. 85–128. Berlin: Mouton de Gruyter.
- Sonesson, G. (2010) 'From Mimicry to mime by Way of Mimesis: Reflections on a General Theory of Iconicity', *Sign Systems Studies*, 38/1/4: 18–65.
- Stampoulidis, G., Bolognesi, M., and Zlatev, J. (2019) 'A Cognitive Semiotic Exploration of Metaphors in Greek Street Art', *Cognitive Semiotics*, 12/1: 1–20.
- Suddendorf, T., and Whiten, A. (2001) 'Mental Evolution and Development: Evidence for Secondary Representation in Children, Great Apes, and Other Animals', *Psychological Bulletin*, 127/5: 629.
- Tallerman, M. (2011) Protolanguage. Gibson, K. R., and Tallerman, M. (eds) *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press, 479–491.
- Tomasello, M. (2008) *The Origins of Human Communication*. Cambridge, MA: MIT Press.
- Torreira, F., Bögels, S., and Levinson, S. (2015) 'Breathing for Answering: The Time Course of Response Planning in Conversation', *Frontiers in Psychology*, 6: 284.
- Vigliocco, G., Perniss, P., and Vinson, D. (2014) 'Language as a Multimodal Phenomenon: Implications for Language Learning, Processing and Evolution', *Philosophical Transactions of the Royal Society of London B*, 369: 20130292. [10.1098/rstb.2013.0292]
- Vygotsky, L. S. (1978) *Mind in Society*. Cambridge, MA: Harvard University Press.
- Waciewicz, S., and Żywiczyński, P. (2018) 'Language Origins: Fitness Consequences, Platform of Trust, Cooperation, and Turn-Taking', *Interaction Studies*, 19/1–2: 167–82.
- Werner, H., and Kaplan, B. (1963) *Symbol Formation*. New York, NY: Wiley.
- Zahavi, A. (1975) 'Mate Selection – a Selection for a Handicap', *Journal of Theoretical Biology*, 53/1: 205–14.
- Zlatev, J., and Andrén, M. (2009). Stages and transitions in children's semiotic development. In Zlatev, J., Andrén, M., Johansson-Falck, M. and Lundmark, C. (eds) *Studies in Language and Cognition*, pp. 380–401. Newcastle: Cambridge University Press.
- Zlatev, J., Waciewicz, S., Żywiczyński, P. and van de Weijer, J. (2017). Multimodal-first or pantomime-first? Communicating events through pantomime with and without vocalization. *Interaction Studies*, 18, 455–79.
- Zlatev, J. (2003) 'Meaning = Life (+ Culture): an Outline of a Unified Biocultural Theory of Meaning', *Evolution of Communication*, 4/2: 253–96.
- Zlatev, J. (2008a) 'The Coevolution of Intersubjectivity and Bodily Mimesis'. In: Zlatev, J., Racine, T. P., Sinha, C. and Itkonen, E. (eds) *The Shared Mind: Perspectives on Intersubjectivity*, pp. 215–244. Amsterdam: John Benjamins.
- Zlatev, J. (2008b) 'From Proto-Mimesis to Language: Evidence from Primatology and Social Neuroscience', *Journal of Physiology – Paris*, 102: 137–52.
- Zlatev, J. (2009) 'The Semiotic Hierarchy: Life, Consciousness, Signs and Language', *Cognitive Semiotics*, 4/Suppl.: 169–200.
- Zlatev, J. (2014a) 'Bodily Mimesis and the Transition to Speech'. In: Pina, M. and Gontier, N. (eds), *The Evolution of Social Communication in Primates*, pp. 165–78. New York: Springer.
- Zlatev, J. (2014b) 'Human Uniqueness, Bodily Mimesis and the Evolution of Language', *Humana Mente. Journal of Philosophical Studies*, 7/27: 197–219.

- Zlatev, J. (2014c) 'Image Schemas, Mimetic Schemas, and Children's Gestures', *Cognitive Semiotics*, 7: 3–30.
- Zlatev, J. (2015) 'Cognitive Semiotics'. In: Trifonas P. (ed.) *International Handbook of Semiotics*, pp. 1043–67. Springer: Dordrecht.
- Zlatev, J. (2018) 'Meaning Making from Life to Language: The Semiotic Hierarchy and Phenomenology', *Cognitive Semiotics*, 11/1:
- Zlatev, J. (2019) 'Mimesis Theory, Learning and Polysemiotic Communication'. In: Peters, M. (ed.) *Encyclopedia of Educational Philosophy and Theory*. Dordrecht: Springer.
- Zlatev, J., Persson, T., and Gärdenfors, P. (2005) 'Bodily Mimesis as "the Missing Link" in Human Cognitive Evolution', *Lund University Cognitive Studies*, 121: 1–45.
- Zlatev, J., Brinck, I., and Andrén, M. (2008) 'Stages in the Development of Perceptual Intersubjectivity', *Enacting Intersubjectivity*, 10: 117–32.
- Zlatev, J. et al. (2013) 'Understanding Communicative Intentions and Semiotic Vehicles by Children and Chimpanzees', *Cognitive Development*, 28: 312–29.
- Zlatev, J. et al. (2008) *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins.
- Żywiczyński, P., Gontier, N., and Wacewicz, S. (2017) 'The Evolution of (Proto-) Language: Focus on Mechanisms', *Language Sciences*, 63: 1–11.
- Żywiczyński, P., Wacewicz, S., and Sibierska, M. (2018) 'Defining Pantomime for Language Evolution Research', *Topoi*, 37/2: 307–18.