Journal of Language Evolution, 2018, 1–15
doi: 10.1093/jole/lzx021
Target article

The role of pantomime in gestural language evolution, its cognitive bases and an alternative

Ekaterina Abramova*

Faculty of Philosophy, Theology and Religious Studies, Radboud University Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

*Corresponding author: e.abramova@ftr.ru.nl

Abstract

This article examines a popular trend of postulating that gestures have played a crucial role in the emergence of human language. Language evolution is frequently understood as a transition from a system, in which signals (whether vocal or manual) have fixed meanings and are used asymmetrically by senders and receivers, through specific cognitive and neurological changes, to a system, in which signals are (1) flexibly referential, i.e., can stand for a variety of ideas and (2) intersubjective, i.e., can be used equally in production and comprehension with any member of the community. The function assigned to gestures in gesture-first theories is to provide a first version of the more advanced open-ended communication in the form of spontaneous pantomimes that initiates a subsequent expansion of this system, its conventionalization and eventually a switch to the vocal modality. In the present article, I examine a particular theory that claims that pantomime was enabled by changes within the system of complex action recognition, and imitation. I argue that while the theory is promising, the notion of a pantomime it employs, presupposes two sophisticated abilities that themselves are left unexplained: symbolization and intentional communication. I point out two ways to remedy the situation, namely, constructing a leaner understanding of pantomime or supplementing the theory with an explanation for the emergence of these abilities. In this article I pursue a third option: identifying an alternative mechanism that can lead to a suitably complex language precursor while avoiding pantomime and its problematic cognitive bases altogether. This mechanism is ontogenetic ritualization, a well-known process responsible for the development of gestures in non-human primates. I outline the possibility that when placed in appropriate sociocultural circumstances, in which complementary actions around objects are required, this process can lead to signals that are modestly referential and intersubjective.

Key words: gestural protolanguage; communicative intentions; ontogenetic ritualization; symbolization; pantomime

1. Introduction

Theories that propose that human language emerged from gestures, rather than directly from vocalizations, have been around for a long while (Hewes 1976). What often fuels this idea is a common intuition that expressing and understanding our thoughts in gestures is somehow cognitively easier or more natural because there is a possibility for signals to resemble the meaning to be conveyed (Another common motivation for gestural theories is the claim that in non-human primates (and our last common ancestor) gestures are more voluntary and flexible than vocalization and therefore could have provided a communicative scaffolding long before we acquired full vocal control. This view has been recently challenged (Slocombe et al. 2011)).

There is a great variety of gestural theories and not all subscribe to the same claims. Some maintain that gestures really came first in evolution (Stokoe 2001; Corballis 2002; Arbib 2005), while others insist on a multimodal beginning (Kendon 1973; Goldin-Meadow 2011; McNeill 2012). Some see the primary role of gestures in their semantic potential (Corballis 2002; Tomasello 2008; Arbib 2012), while others focus on their proto-syntactic qualities (Stokoe 2001; Armstrong and Wilcox 2007). There is a difference in whether the first gestural expressions were more like words (Hewes 1973) or more like complete propositions (Arbib 2012). And finally, certain theories are pitched at a more personal level of explanation (Zlatev 2007), while others pursue a more sub-personal, mechanistic route (Corballis 2002; Arbib 2012). In this article, I will restrict myself to one particular instantiation of a gestural theory to examine one particular notion that seems central to many of them, namely pantomime. I consider the critique to be made in the paper applicable to any gestural theory that relies on this concept.

Pantomime is a spontaneous bodily mode of communication, in which meaning is conveyed through resemblance. For example, molding your hand into a round shape and moving it close to your lips might signify *cup* or *drinking*. This communicative trick is often presented as more powerful than what is available to non-human primates and our last common ancestor (LCA) but less powerful than conventional and compositional language proper. At the same time, because it is bodily and can rely on similarity between form and meaning, it is seen as not requiring much sophisticated cognitive machinery and therefore potentially within easy evolutionary reach for our LCA.

The suggestion that pantomime is necessarily cognitively easy has been challenged from the perspective of cognitive development and language acquisition (Irvine 2016). What I wish to do in the present article is contribute to this challenge by situating the discussion of pantomime in a larger theoretical context. Specifically, reliance on pantomime in gestural accounts betrays several assumptions about what human language is and what was required for its emergence. (By far the most important underlying assumption is that language is seen as primarily a business of getting a message across, transmitting ideas between individuals. Since this ‘externalization view’ underlies most of current language evolution theories and is extensively discussed elsewhere (Hawkey 2008; Smit 2016), I will not address it in this article.) The two key properties ascribed to our communicative system are symbolization and communicative intentions. The first is defined along Piagetian (1962) lines as an ability to differentiate between a signal and its referent and understand that the former *stands in* for the latter. The second is frequently unpacked as the speaker demonstrating his intention to communicate and the hearer having to infer that intention. Neither of the two is uncontroversially easy and if pantomime requires them, then any theory that relies on pantomime needs an explicit account of how they emerged.

In mounting the challenge to the role of pantomime in language evolution, the article proceeds in three stages. In the first stage (Section 2), to make the discussion more concrete, I describe one specific recent gestural theory: the so-called Mirror System Hypothesis (MSH) of Arbib (2012). The choice is motivated by the fact that the communicative breakthrough that pantomime is supposed to provide, cries for an explicit mechanistic account of what exactly changed in the minds and brains of our ancestors and how it happened. The MSH is a theory that proposes that it is changes to the mirror neuron system (MNS) that enabled the key transition.

In the second stage (Section 3) I start by explaining the features of symbolization and communicative intentions and highlight their controversial status. I also show that pantomime does indeed rely on them both and consider whether MSH provides a clear account of how such sophisticated skills appear in evolution. I argue that in fact the account cannot yet be considered complete.

In the third stage (Sections 4 and 5) I suggest that the shortcomings of a pantomime-based account do not warrant abandoning a gestural theory. I propose considering a different mechanism for an expansion of a gestural repertoire: ontogenetic ritualization transformed through changing sociocultural settings. Ontogenetic ritualization is a well-known simple process in which communicative gestures emerge from repeated social interactions. I outline a possibility for this process leading to gestures, which are sufficiently complex to constitute viable precursors to language, but which at the same time do not face the problems inherent in reliance on pantomime. Circling back to the example of MSH, I conclude by discussing the effect of abandoning the concept of a pantomime for a theory that employs it.

2. The MSH

The discovery of mirror neurons (neurons that fire when executing and observing an action) and specifically their location in the primate homologue of the Broca's area has fueled the proposal that the system for grasping and imitation could have served as a basis for the emergence of gestural intentional protolanguage (Rizzolatti and Arbib 1998). This preliminary suggestion has been elaborated into the MSH by Arbib (2012) as a neurological implementation of the gestural language evolution scenario and has been since adopted by many of its other proponents (Zlatev 2008; Corballis 2010).

It is important to note from the start that MSH does not state that the possession of a MNS automatically leads to either action understanding (as some have interpreted Gallese and Goldman 1998) or language, only that it provided the basis for mechanisms that transformed a non-linguistic brain into a language-ready brain that could support manual protolanguage. Obviously, considering the changes to social and technological life of our ancestors would still be needed to fully spell out the details of such an account but that is not the focus of MSH. The focus is rather on describing the changes to the MNS (and systems 'beyond the mirror') that would be required to result in a gestural protolanguage sufficiently different from a simple gestural communication system, such as that currently present in non-human primates, that it could jump-start a series of transitions that would eventually lead to human language.

Specifically, Arbib (2005, 2012) proposes seven stages of language evolution. In the first three stages the capacities are shared with the LCA of humans and other primates. What we have there is an action system for grasping (stage 1), a MNS which responds to the perception of grasping (stage 2) and allows both current primates and our ancestors to respond to and imitate simple object-directed actions (stage 3). The discussions about the difference between imitation in non-human primates and humans are still ongoing but there seems to be an overall consensus that several kinds of this skill can be distinguished. For example, *stimulus enhancement* is apparent imitation resulting from one's attention being directed to a particular part of the environment and executing appropriate action on that part. *Emulation* is observing an action and attempting to reproduce its result without copying the manner in which it has been achieved. Arbib calls both of these mechanisms *simple imitation* and distinguishes it from *true* or *complex imitation*, in which a novel action, which is outside the imitator's own repertoire is replicated in detail by recognizing that the overall action is composed of familiar sub-actions. (Note that this process may still require fine-tuning and involve flexibility of execution, what matters is that what is delivered at the end is a replication of means, not just ends of the overarching action.)

Only humans seem to be capable of complex imitation and hence the emergence of this capacity is assigned to the truly novel stage 4. The (extended) MNS that subserves this type of imitation is able to recognize not only single actions but also compound sequences, not only transitive actions but also intransitive sub-components, i.e. movements not explicitly and directly related to a visible goal-related object. No claim is made on whether such imitation evolved because of its usefulness in acquiring language or for other reasons, such as, for instance, tool-use. What is claimed, however, is that the new capacity is recruited for communication purposes (This is necessary because the theory holds that gestures already present in apes are a result of simple imitation and therefore something else is required for gestural protolanguage, see a diagram in (Arbib 2012: 215)) and begins a chain of changes that eventually lead via pantomime (stage 5a) to the emergence of an open-ended range of conventionalized, progressively abstract gestures called protosigns (stage 5 b), protospeech (stage 6) and then multimodal syntactically structured human language (stage 7). As the changes related to complex imitation and resulting pantomime seem to be the most crucial, I focus here on stages 4 and 5a.

Now, complex imitation is said to be the key novelty because it provided a foundation for pantomime. As such, pantomime in MSH is said to rest on three abilities (to be explained below):

- the recognition that a partial action A serves to achieve the overall goal G of the behaviour of which it is part;
- using the recognition that a partial action A that serves to achieve the goal G for assisting the other in the achievement of G;
- the reversal of that recognition to *consciously create actions that will stand in metonymic(part-whole) relationship to some overall goal, whether praxic or communicative* (Arbib 2012: 218–19).

All three abilities are normally used for imitation of complex hierarchically structured actions. In the praxic context, the overarching goal G of an action is an achievement of some praxic aim and the means for achieving it are intransitive sub-components $A_1 \ldots A_n$ of an overall behaviour. The first property means simply that recognizing one of the component As might lead to the recognition of G. The second property means that if one can recognize the purpose of the sub-components (their sub-goals in relation to G), one can assist the other in achieving G (perhaps given some cooperative motivation). Finally, one acquires a reverse insight (the third property above) that performing appropriate sub-components will achieve G, constituting a case of complex imitation. At the same time, however, this praxic mechanism gets exapted for communicative purposes, namely performing a sub-action *A* with the hope that the other will recognize *G* and assist in completing the action. In this case, the movement becomes communicative because the goal is now to get the other to think of *G* and thereby elicit a desired response.

Why do the mechanisms underlying complex imitation get transformed in this way? It is suggested that our ancestors learned that producing imitative movements in instrumentally inappropriate contexts leads to desired effects and this got reinforced as a communicative strategy. And what is more, what starts out as a stock of pantomimes about manual actions is soon extended to 'the ability to in some sense project the degrees of freedom of movements involving other effectors, [other animals, such as a bird flying,] and even, say, of the passage of wind through the trees' (pp. 214–15). As a result, representing non-human actions or objects via gestures becomes available.

Arbib argues that crucial to this latter transition is the class of so-called *quasi-mirror neurons*, which are to be distinguished from *potential mirror neurons*. The latter are the neurons that as a result of experience can acquire mirroring properties. Quasi-MNs, in contrast, are those that connect observed actions not to exactly the same own actions (such as when I try to imitate my conspecific using a tool), but to 'somewhat related' movements. For example, they could be active when one tries to flap one's arms to imitate a bird flying for the purpose of miming *bird* or *flying*. (One should note that both the existence of such quasi-MNs and their supposed role in pantomime has yet to be demonstrated empirically, although see Aziz-Zadeh et al. (2012) for systems level evidence of understanding conspecifics with substantially different bodies.)

In the end, what emerges is a flexible ability and social practice of producing and understanding pantomimes for a variety of meanings, which lays the basis for future, more conventional, gestural protolanguage. It is said to exhibit certain properties definitional of language and therefore constitute an adequate language precursor. These are as follows:

- *parity of meaning*—the understanding of the signal is shared between its sender and receiver, so that the same individual can both produce and understand the same signal.
- *symbolization*—an individual can associate an open set of communicative forms with an open class of events.
- *intended communication*—the signal is meant by the sender to have a particular effect on the receiver.

Parity of meaning follows relatively straightforwardly from the properties of the MNS: my firing of the mirror neurons helps me understand instrumental actions but also communicative manual actions by extension. (This is not as straightforward because mirror neurons are mere pattern detectors and to furnish 'understanding' their firing needs to be seen in the context of the whole extended mirror neuron system. But let us grant Arbib that the basic insight holds.) Symbolization is ensured by 'projecting the degrees of freedom' from a variety of actions, objects, and events. And finally, intended communication is based on the goal-directed hierarchical nature of both instrumental and communicative actions: the fact that gestures are driven by an overarching goal *G* and require recognition of that goal on the part of the receiver. The responses are thus not elicited directly because of some reinforced connection.

3. Symbols and communicative intentions

As advertised in Section 1 and should be clear from the exposition of MSH above, Arbib explicitly relies on pantomime defined in terms of symbolization and communicative intentions. It should also be clear that it is these two features that are held up as crucial to initiating a transition away from limited communicative repertoires of our LCA and towards the richness of human language. This is unproblematic if one of the two holds: (1) both features can be shown to be cognitively lean enough to smoothly emerge from LCA's communicative precursors without e.g. themselves requiring language or (2) both features are not cognitively lean but MSH shows how they developed from other LCA's capacities. I will now tackle both options in turn.

3.1 Definitions and controversies

Symbolization, as the name suggests, is an ability to use symbols. A symbol can be understood in different ways. One, rather non-standard way to define it, is to say that a symbol is a sign which derives its meaning partially from its relationship to other signs in the system (Deacon 1996). Pantomime is not a symbol in that sense since it arises when there is no yet a (protolanguage) system to speak of. A more frequent way to define a symbol, following Peirce (1931–1958), is as a sign in which relationship between its form and meaning is arbitrary, based on convention. For example a word 'dog' in no way resembles a dog. This is to be contrasted with an index, in which relationship is based on some causal connection, e.g., smoke means fire; and with an icon, in which the connection lies in similarity, e.g. a drawing of a dog. Pantomime is clearly not symbolic in this sense, it is rather an instance of an icon. (Gestural theories typically hold that pantomimes eventually turn into Peircean symbols, usually through the process of cultural transmission, in which the forms are progressively abbreviated and conventionalized and the initial similarity link is lost (see Garrod et al. 2007, for experimental evidence of such a process)).

When somebody says pantomime displays a feature of symbolization they follow a definition initiated by Piaget (1962). According to Piaget one of the developmental milestones reached by the child is an attainment of symbolic function, i.e. a differentiation, from the subject's own point of view, between the signifier and the signified. For example, a child understands that a picture of a dog and a word 'dog' are not the same as the actual animal but rather represent it.

The notion of intentionality is yet another polysemous term in language evolution debates. On one reading, communicative act is intentional when it is goal-directed and under voluntary control, rather than reflexive. According to a set of criteria proposed in primate studies, a gesture is taken to be intentional in this sense when it is directed at an audience, persistent and flexible, that is, can be changed when the desired response is not obtained (Leavens et al. 2005). The discussion about whether primate gestures are so intentional is ongoing but there seems to be a general agreement that even if the answer is yes, it is not this kind of intentionality that is distinctive about human language and that makes it more powerful (Liebal et al. 2014: Ch. 8). Rather, intentionality implied in pantomime is specifically the notion that figures in 'communicative intentions'.

Communicative intentions (CIs) come from a theory of meaning proposed by Grice (1957). On this account, a speaker is said to mean *p* when an utterance that expresses *p* was produced with the intention of inducing a belief in the receiver via receiver recognizing the underlying intention of the speaker. This definition allows Grice to make a distinction between natural meaning where, e.g. an animal produces a cry of pain involuntarily and receives a reflex-like response, and more complex non-natural meaning, where the same animal A would produce a cry voluntarily (Communicative intentions thus depend on communication being intentional in a goal-directed sense but also require something more) with the intention of getting the receiver B to believe that A is in pain to evoke a specific response. The latter happens, furthermore, not just because B recognizes that A is in pain but rather because he recognizes that A has produced the signal with a particular speaker intention.

Grice's theory has been subsequently developed and the most popular current model is that of ostensive-inferential communication (Scott-Phillips 2015). Here speakers are analysed as having two intentions: (1) an informative intention to make a certain actual or desired state manifest to the hearer (2) a communicative intention to achieve the informative intention by making it mutually manifest to the hearer that she has this informative intention (Sperber and Wilson 1995; Origgi and Sperber 2000). An informative intention is most frequently understood as the speaker trying to influence the mind of the hearer, aiming at the hearer forming a particular belief, intention, goal, etc. A communicative intention is then the speaker signalling that he is in fact trying to communicate, i.e. it is the intention on the part of the speaker that the hearer recognizes that the speaker has an informative intention. The task of the hearer is to infer both intentions and thereby arrive at the meaning intended by the speaker. It is widely believed that CIs understood in this way is what makes human language so open-ended because communication can then be based on flexible reading of each other's intentions, rather than laborious formation of associations (see e.g. Scott-Phillips (2015), for a recent evolutionary account of this type).

Now that the definitions are in place, to see that both features figure in pantomime consider the following example. Suppose you are walking with a friend in a forest and you see a deer in the bushes nearby. Since you love animals, you are thrilled and you want to share your excitement with the friend. In order not to scare the dear off, you tug on your friend's sleeve, put a spread out hand to the top of your head and point in the direction of the bushes. The friend sees the deer and smiles.

What happened in this pantomime (skipping the tugging and the pointing for simplicity) is that you spontaneously mapped deer's antlers onto your hand, while at the same time realizing that the two are not the same but rather the hands represents the antlers (and the deer by extension). You then attempted to convey the deer idea to your friend wishing for him to recognize that you are in fact trying to communicate something and not merely scratching your head in an awkward way. Moreover, you wish the friend to recognize that what you wanted to communicate is that there is a deer and it is exciting and you want to share the excitement, rather than asking your friend if it is in fact a deer or requesting him to go and kill it so you can hang the antlers above your fireplace. The fact that your friend saw the deer and smiled indicates that he was successful in inferring your intentions.

Now, if pantomime requires both symbolization and CIs, can it be viewed as cognitively lean? Presenting a thorough analysis of these two features would go beyond the scope of this article. However, I do wish to point out that in wider debates on mind and language both are not uncontroversial.

Given the terminological confusion about the notion of a symbol and symbolic reference (despite the fact that 'more philosophic ink has been spilt over attempts to explain the basis for symbolic reference than over any other problem' (Deacon 1996: 43)), it is hard to pinpoint what specific capacity constitutes evidence for an ability to differentiate between signal and referent. It has been suggested, for example, that a chimpanzee using a particular gesture to request some response from a conspecific must be differentiating between that gesture and the response (Zlatev et al. 2005). This would place symbolization before the pantomime stage postulated by MSH and would deny that symbolization is the source of flexibility of specifically human language. At the same time, however, developmental evidence (De Loache 2004; Zlatev et al. 2013) suggests that an ability to appreciate the representational function of various semiotic vehicles (e.g. points, pictures) is acquired relatively late by children, which would mean that perhaps some experience with symbolic culture is necessary for symbolic skill to emerge. This of course would imply a reverse dependency with respect to that suggested in MSH. Treading these dangerous waters means that any theory of language evolution must be clear on what symbolization specifically means and how it appears on the evolutionary scene.

The issue of CIs is a subject of more heated debates. Despite some similarity with intentionality in the sense of goal-directedness, it is generally not assumed that communicative acts based on CIs fall out of goal-directed communicative acts for free. Rather, some further cognitive advance is thought to be required, mostly unpacked in terms of mindreading abilities (Scott-Phillips 2015). That is, having and inferring CIs requires some understanding of the other's mental states and possibly an understanding of the other's understanding of one's own mental states (since I want *you* to recognize *my* intention directed at *your* mental state, which is second-order intentionality). However, postulating that such capacities precede language just pushes the question back. We are now required to first answer how our ancestors became proficient mindreaders pre-linguistically, and given that mindreading itself could be dependent on language (Astington and Baird 2005), the whole theory might just become circular (Bar-On 2013).

3.2 The MSH solutions

As presented in Section 2, pantomime as it is currently conceived within MSH explicitly invites an analysis in terms of symbolization and standard CIs. Therefore, it needs an explicit account of an emergence of these cognitive capacities, a demonstration of why they are unproblematic or a redescription of pantomime in a way that does not rely on them. The latter could involve, for example, adopting a more minimal account of CIs or restricting pantomime in some way.

To recall, Arbib maintains that complex action recognition and imitation played a key role in language evolution. Equipped with these skills, an observer is able to recognize some novel action as composed of familiar sub-actions or their variants, which are directed at subgoals of the overall action goal. This allows for approximating that novel action by an assembly of familiar sub-actions or trying to reach sub-goals by trial and error. Both allow an observer to recognize and imitate actions outside their repertoire. At this point, although we might quarrel with this particular understanding of complex imitation (see e.g. Catmur et al. (2007) and Cook et al. (2014) for a view that does not require recognizing goals), no issue with symbolization or mindreading mechanisms arises. However, it does become problematic once we try to envision the functioning of this process in communicative settings.

We are told, namely, that pantomime emerges when our ancestors are able to 'consciously create actions that will stand in metonymic relationship [to X] . . . with the intention of getting the observer to think of a specific action or event' (Arbib 2012: 218–19). I now have to perform an action with the intention that *the hearer* recognizes *my goal* and as a result of that recognition performs compatible actions. (The definition of a pantomime has recently been updated by Arbib (2016, personal communication) to (1) X performs an intransitive action A that resembles an action B which might occur within a context C to achieve goal G with the intention that some observer Y will 'get the message' concerning some aspect of C or G; (2) Y recognizes that A does indeed resemble B and, knowing that action B might occur within a context C' with goal G' infers that the message is some aspect of C' or G' (which might not be completely equivalent to the C or G intended by A). However, while this specification lowers the requirements on recognizing the goals and meanings in precise manner and drops reference to conscious intending, it still requires production and perception of messages with a communicative intention.) Perhaps we could still say this happens somehow automatically (forgetting the emphasis on conscious intentions), just by virtue of MN firing and some cross-talk with other brain systems. This is not Arbib's aim, though, because the same mechanism should hold for more complex scenarios, i.e. miming a shape of the object or other animals. Clearly, if I am to start flapping my arms to signify *bird*, I will not be able to move the hearer to respond appropriately directly. An automatic MN response would lead her to perhaps imitate my flapping, which is not the response I want. Instead, I need to have an intention to instill the image of a bird in the hearer, together with an understanding that I am communicating and of the reason for why I am communicating *bird* and hope she will recognize my intention.

Perhaps, quasi-MNs that allow for 'projecting the degrees of freedom' from birds to arms explain how I am able to accomplish the symbolic mapping. Alternatively, mirroring has nothing to do with the process and other brain systems are the solution. After all, there must have been a way for our ancestors to recognize animals and their characteristic motions. One could then tell a story of how the brain regions responsible for this recognition got connected to brain regions that guide manual actions. However, this could not be a mere association because otherwise any associative link would automatically be referential. What we need is an explicit account of what makes the association an instance of a *standing in* relationship. For example, it is unclear at present why a sub-action of some larger action sequence should stand in for the overall goal, just as it is unclear how recognizing a bird flight could lead to representing that bird or flight with one's hands. Saying, for example, that the required mechanism 'involves not merely changes internal to the mirror system but its integration with a wide range of brain regions' (Arbib 2012: 215) amounts to saying that we need to explain the changes to the whole brain. In this case, however, the crucial explanatory force is yet to come from an account of this integration in communicative contexts.

In response to the problematic nature of CIs, current efforts in the field of language evolution go in a direction of providing a more minimal interpretation of this construct. For example, Zlatev's theory (Zlatev et al. 2005; Zlatev 2008) relies on a notion of *bodily* mimesis, rather than pantomime. This is defined as a bodily act which (1) involves a cross-modal mapping between exteroception and proprioception; (2) is under conscious control and corresponds to some action, object or event, while at the same time being differentiated from it by the subject; (3) is intended by the subject to stand for some action, object or event for an addressee (and for the addressee to recognize this intention); (4) is not conventional and not compositional. Pantomime is a prime example of bodily mimesis. The second property above corresponds to symbolization in MSH and the third is a restatement of CIs.

One could argue, although no explicit attempt has been yet made by Zlatev, that on this reading CIs are not at all problematic because bodily mimesis does not rely on higher-order propositional attitudes, only on simulation. (A version of such an argument has been provided by Hutto (2008). However, he so far has not explicated how specifically mimesis-based CIs are to be understood within a larger non-representational view on cognition that he advocates.) CIs realized in simulation would mean that the speaker simulates the hearer's simulation of the speaker to convey his communicative intention while the hearer simulates the speaker's intentions to understand the utterance. Since mirror neuron activity has been frequently linked to simulation, this reply could salvage MSH.

However, it is not at all clear that simulation version of CIs would be cognitively cheaper than operating with something like higher-order propositional attitudes, typically invoked in discussions of CIs. Mathematical analyses show that at least a particular formalization of intentional communication (as Bayesian inference) is computationally intractable (Rooij et al. 2011). In general, if a certain problem is found to be intractable, this is regardless of the algorithm that implements it (That is, to make the claim that simulation-based CIs are unproblematic, one would need to formalize them and show that communication based on such CIs is tractable) (Van Rooij 2008, 2012). This is likely because the source of complexity is not the type of representations that implement the mechanism but the layered structure of CIs as described in a post-Gricean tradition. That is, if CIs are defined as a speaker's capacity to intentionally affect mental states of the addressee and appreciate that they need to recognize the speaker's intentions directed at their mental states, the key feature of this capacity is the layered, second-order structure. Whether this structure is realized in propositional attitudes or bodily perspective-taking skills is irrelevant to the computational complexity of the skill.

Another solution that is currently on the market is reconceiving CIs in terms of speakers trying to influence the hearer's behaviour, rather than their states of mind (Moore 2015, 2016). The idea here is that once overly mentalistic description of what is going on in a communicative act is abandoned, the troubling features of CIs can disappear as well. The speaker now does not need to hold higher-order beliefs about mental states of the hearer, all he needs is intending to change the hearer's behaviour. For example, when I am pointing out a deer to you, I do not intend that you *believe* there is a deer but that you *look* at the deer. The MSH could perhaps take this reanalysis on board but, again, it has to be made specifically clear how such a redescription fits into a network of mechanisms that implement complex imitation and other brain systems. For example, instead of saying that pantomime is aimed at 'getting the observer to think' of something, one could say that by performing a sub-action of a larger action, the mimer aims at getting the audience to complete that action. And perhaps, there is not even a desire to get one's CI recognized, but only an appreciation that pantomime needs to be seen to be effective. However, it then should be spelled out how the same mechanisms can be applied to sophisticated pantomimes such as miming a bird flying, or, if they cannot be so applied, what needs to be added.

To sum up this section, neither symbolization nor communicative intentions are yet explicitly addressed within MSH framework. In the meantime, then, I wish to consider an alternative response to the pantomime conundrum. The strategy here will be to see if one can get a gestural protolanguage that goes a little beyond what is available to non-human primates but does not require flexible and open-ended pantomime as its central building block. It should be stressed that an alternative account is not necessarily at odds with other solutions discussed above and could, for example, be combined with a leaner reanalysis of CIs.

4. Communication reenvisioned

In a recent critique of a Gricean view on language evolution, Bar-On (2013) suggests that rather than searching for the origins of CIs, perhaps a more straightforward approach would be to investigate how non-Gricean signals that are characteristic of non-human primates (and presumably our LCA) could have changed gradually so as to take on more linguistic character. That is, to fill the role of a language precursor, a communicative system has to exhibit proto-semantic and proto-pragmatic features that go beyond what researchers find insufficient in contemporary non-human primate communication systems but at the same time avoids presupposing symbolization or CIs. Bar-On postulates that *expressive communication* could fit this job description.

Expressive communication is often discussed in the context of natural meaning, such as a cry of pain described above. What is often said of such communication is that it is rigid, emotional, merely imperative and merely dyadic. That is, expressive signals are produced involuntarily as part of a general state of an animal and they move the receiver in a reflexive kind of way. They are not used to communicate intentionally and certainly not to talk about states of affairs. Clearly, if we think language proper is exactly the opposite, it is hard to see how to build a bridge from one to another. However, perhaps we were simply too quick in characterizing expressive signals as being so inadequate.

Bar-On argues that acts of expressive communication are in fact flexible (individually variable, context-sensitive, learnable) and while they do not refer to the world, they can nevertheless be world-involving. At least some expressive signals express not only producer's internal state but also the external cause or object of that state and thereby direct the receiver's attention to that object. They can show how the producer is disposed to act with respect to some part of the world, giving the receiver an opportunity to respond appropriately. At the same time, there is no need to postulate a desire to inform or a (conscious) intention to affect the other's state of mind and a response can be based on some type of simple resonance, rather than reading producer's intentions.

How does all this relate to the gestural language evolution theory? If one wishes to focus on explaining an emergence of a *gestural* protolanguage that goes beyond what could be available to our LCA with non-human primates, one needs to focus on a particular kind of expressive communication, namely, one realized in manual modality. Gestural theories invoke such communication as one of the reasons for even considering a gestural protolanguage (e.g. due to their flexibility, see ft. 1). However, it is also often noted that they are limited in precisely some of the ways that make expressive communication inadequate. That is, even though primate gestures are seen as relatively flexible, they are also said to be invariantly dyadic (not referring to the world), imperative rather than declarative and insufficiently intersubjective ('Intersubjectivity' is a term that normally means a lot more in different fields of philosophy and psychology. Here we will merely take it to mean that a sign is understood in roughly the same way when produced and comprehended and, moreover, understood in a similar way by all members of the community), i.e. they have meaning which holds for particular dyads that employ them and that meaning is often specific to particular individuals within a dyad, rather than having a status of a shared form with meaning that holds for both parties concerned or for the whole group.

The reason for inadequacy of gestures is often sought in the mechanism through which they emerge.

The consensus seems to be that while some gestures seem to have evolved specifically for communicative purposes, most originate from instrumental movements through phylogenetic or ontogenetic ritualization (OR). The latter involves a transformation of instrumental social actions which are used to affect the behaviour of a conspecific into communicative actions by repeated interaction between individuals. Specifically,

> *In ontogenetic ritualization two organisms essentially shape one another's behavior in repeated instances of a social interaction. The general form of this type of learning is:*
> *Individual A performs behavior X;*
> *Individual B reacts consistently with behavior Y;*
> *Subsequently B anticipates A's performance of X, on the basis of its initial step, by performing Y; and*
> *Subsequently, A anticipates B's anticipation and produces the initial step in a ritualized form (waiting for a response) in order to elicit Y.*
> *The main point is that a behavior that was not at first a communicative signal becomes one by virtue of the anticipations of the interactants over time (Tomasello and Zuberbühler 2002: 205).*

For example, if an individual A wants to embrace another individual B, they might start out by pulling B closer and embracing them. Over time, B will begin to anticipate the desire of A in just the beginning of a pull and A will learn that already an abbreviated pull (which would be motorically ineffective), will elicit the correct response from B. As a result, the pull will be abbreviated even further and a communicative 'embrace me' gesture will emerge (Liebal and Call 2012).

The evidence for OR as a mechanism for the emergence of gestures is a high degree of variability in individual repertoires which are specific to particular couples of individuals. (Of course, even in this case certain degree of uniformity is to be expected, given that social interactions in which individuals participate (and from which gestures derive) are relatively similar across the group.) For example, Halina et al. (2013) conducted a study trying to trace the emergence of gestures through OR in bonobos. They focused on a mother–infant carry action and gesture asking (1) whether there is indeed substantial variation among gestural repertoires within the group and (2) if gestures are structurally similar to their originating actions. They found the predicted variability between dyads but also that the gesture form depends on the role of the individual in the shared activity. That is, mothers use gestures that stem from their carrying the infant and infants use gestures that are a result of their being carried.

Now, a gesture acquired through ontogenetic ritualization is a routinized expressive gesture. It is used to express a particular motivational state of the animal and elicits a response directly because it has been acquired in tandem with the emergence of the signal itself. While one could describe OR gestures in terms of symbolization and CIs, nobody seems willing to do so. For example, Tomasello (2008: 296) himself states that:

> *the meaning or communicative significance of intention-movements is inherent in them, in the sense that they are one part of a pre-existent meaningful social interaction . . . individuals do not need to learn . . .to connect the signal with its 'meaning'—the 'meaning' comes built in.*

That is, the meaning of an OR gesture does not consist in an explicit signifier-signified relation and does not need to be conveyed or recovered with the help of communicative and informative intentions. For the same reason, however, OR and gestures it leads to are viewed as insufficient to scaffold the emergence of human language as they can only mean whatever they developed to mean in a particular context for a particular dyad. As a consequence, they seem to lack the properties that create an open-ended system in which infinitely many signals can be created and understood almost on the spot. Thus, Tomasello abandons OR in favour of recursive mind-reading, shared intentionality and cooperative motivations. Arbib deems it necessary to propose an intermediate stage of the emergence of complex imitation before primate-like gestures can be turned into more proper gestural protolanguage.

Hurford (2007), another language evolution researcher, discusses OR in the context of possible learning mechanisms that could form the seed of human language. This evolutionary target is defined in terms of 'signals with an arbitrary, non-iconic, non-indexical, and not physically causal, relationship to their function' (p. 200), possessing a quality of reciprocity, i.e. can be used from the sender or receiver side indicating that the signal is understood intersubjectively. Given these criteria he also finds OR insufficient because it occurs predominantly in asymmetric interactions and the resulting gestures are often tied to a particular ontogenetic stage, not persisting as signals when they are no longer needed. Finally, since OR requires a history of interaction and gradual mutual shaping of the gestures, 'it would not be possible, in a large social group, for each individual to participate in such a history of interaction with all the others, so ontogenetic ritualization as a direct source of group-wide arbitrary learned illocutionary signals . . . is unlikely' (pp. 200–1).

Interestingly, Hurford also notes that things could be different if the social situation were changed, e.g. family ties being more extended. He dismisses this quickly because there are no reports of this, i.e. of more complex forms of OR-derived gestures. We need to consider, however, if this is a relevant objection. After all, we are trying to tell an evolutionary story about the stages of communication which cannot be observed and which fit into the gap between what we think our ancestors (and their brains) were capable of and current language. Just because there are no reports of something at present does not mean that it is a logical impossibility.

Let us then redefine a strategy for a plausible gestural language evolution account. Our explanatory target will be a gestural protolanguage (For the purpose of this article, I assume that some type of gestural scenario is correct and hence we inherit all the theoretical issues associated with it, e.g. how gestural protolanguage could have switched to vocal modality. I believe that there is in fact a need for a more multimodal account but I merely focus here on the potential changes to gestural modality), which does not require symbolization or CIs, while at the same time possessing more linguistic qualities than gestures currently observed in non-human primates. It is a communicative system in which gestures are as follows:

- *triadic*, i.e. relate to objects in some way (this replaces *symbolization*).
- *reciprocal*, i.e. the signal can be used equally by senders and receivers.
- *intersubjective* in a wider sense, i.e. the same signal can be used with multiple partners (the last two properties capture parity of meaning).

Our new explanation will be a simple OR mechanism, which in non-human primates produces dyadic, imperative, non-reciprocal gestures. What I want to probe is whether it is conceivable that the processes of ritualization could under certain conditions lead to a gestural protolanguage as defined above. I will suggest that indeed it is, given appropriate sociocultural changes.

5. Joint actions and a gestural alternative

The route to gestural protolanguage I will investigate is through an increased involvement of joint action and use of objects. Most attention in the study of language evolution has been devoted to the role of particular kinds of objects in our LCA's life, namely Oldowan and Acheulean stone tools. For example, in theories that focus on the emergence of syntax, the ability to process complex hierarchical action sequences is seen as promoting the emergence of hierarchical structure in grammar (Stout and Chaminade 2012). In relation to MSH, Arbib (2011) makes a link between language evolution and tool use through the MNS: evolving abilities to monitor and plan one's own manual actions, including tool use, led to the capacity for complex imitation, which then enabled pantomime.

Theories of this type undoubtedly have a lot of merit and it is not my intention to oppose them. However, I want to emphasize that our data on technological skills available to our ancestors is limited by archaeological record. In this case, the conclusions that are drawn from tools to language and cognition are based on objects that are preservable due to their material (stone). This is limited in several ways. First, according to the recent estimates, the split between humans and chimpanzees has been placed at around 7 mya (White et al. 2009; Young et al. 2015) while the earliest pre-Oldowan stone tools date to about 3.3 mya (Harmand et al. 2015). This creates plenty of time in which tools made of organic materials (stones, sticks, bones, shells) might have been used but were not preserved. Second, the archaeologically preserved tools can tell us something about the process of their production but they do not preserve the context in which they were made and used. Most importantly for my purposes, we cannot easily infer the *social* context of interaction with objects. Finally, tools are just one type of object that could have played a role in language evolution, everyday items like food might have been equally important.

Despite archaeological uncertainties, in discussions of tools and language it is often assumed that object-based interactions were solitary. In contrast to this view, Reynolds (1993) emphasizes that one of the main differences in tool use between humans and other primates is its social dimension:

> *Tool making is usually presented as an artisan working alone. In reality artifacts are usually constructed together, in a chain of complementary actions guided largely by anticipation of what the other participant will do. The basic principles of a manufacturing system [is] task specialization, symbolic coordination, social cooperation, role complementarity, collective goals, logical sequencing of operations, assembly of separately manufactured parts .... The essence of human technological activity is anticipation of the action of the other person and performance of an action complementary to it, such that the two people together produce physical results that could not be produced by the two actions done in series by one person (p. 412).*

Apparently, across the globe, even if tasks are 'simple' enough to perform on one's own, they are always performed together, with complementary roles and division of labour. In contrast, when primates use objects, they do so individualistically and hence also without role complementarity. Such an activity as using sticks to dig holes in termite nests and fish termites out of them is not performed jointly by chimpanzees.

If Reynolds is right in emphasizing sociality, there must have been a point in time where activity around objects did become more social. One could argue that language enabled this transition but it could also be that such increased sociality actually enabled language. Furthermore, if he is right that 'the essence of human technological activity is anticipation of the actions of the other', we could conceive of a language evolution stage where *triadic ontogenetic ritualization* (TOR) is a route through which gestures that involve objects can be created.

Let us suppose now, that due to some changes in hominid evolution there is an increased pressure for tool use and joint action more broadly. Such a change is acknowledged by all language evolution researchers mentioned above. Let us also suppose, however, that rather than these two pressures acting independently, there arises a need to make or use tools and other objects together, cooperatively, leading to an increase in triadic interactions in the community (i.e. interactions between two individuals around an object, rather than merely between two individuals). If such interactions are sufficiently regular, it is plausible that similar process of ritualization can occur, just by virtue of motor simplification and anticipation, resulting in gestures. By analogy to dyadic OR, we could envision the following sequence of events:

1. Individual A performs behaviour X with respect to object T;
2. Individual B reacts consistently with behaviour Y towards the object T;
3. Subsequently B anticipates A's performance of X, on the basis of its initial step, by performing Y; and
4. Subsequently, A anticipates B's anticipation and produces the initial step in a ritualized form (waiting for a response) to elicit Y towards the object T (e.g. making the joint action more efficient or initiating it).

For example, in a hunting scenario, one could conceive of the following. A is chasing a wild pig (or some other small animal) and trying to hit it by throwing stones at it but the pig keeps running away. B joins the chase and tries to manoeuvre the pig closer to A's position by making loud noise and waving his arms. Both succeed, kill the pig and a whole tribe feasts on it. The hunting trick gets repeated until A can anticipate that making stone throwing movements in the presence of a pig is enough to request B to execute his maneuvering routine.

If the scenario above is too contrived or complex, consider another one. A and B walk through the forest and encounter a tree with a large amount of tasty fruit on it. Since they have evolved a more upright posture and become larger, and the fruit hangs on rather thin branches, they cannot reach it by climbing the tree. By trial and error, they discover that if one of them pulls on the branch and brings it down, the other one can pick the fruit and both can feast on it. The interaction is repeated until A can make a reaching motion towards the branch to request B to help pick the fruit.

What would such a view on gestural language evolution entail? Our target explanandum has been a gestural protolanguage that possesses proto-semantic and proto-pragmatic qualities (being triadic, reciprocal and intersubjective) without requiring complex abilities for symbolization and CIs. We have seen that OR-produced gestures do not seem to be appropriately described in terms of such abilities. The gestures that emerge in the context of previous history come with a meaning built in and so do not rely on an ability to connect some linguistic form with an internally generated meaning that the form stands in for. Being an instance of expressive communication, an OR gesture also does not rely on production and comprehension of CIs. Does it mean that both symbolization and CI are dispensed with altogether? And does communication based on TOR gestures have sufficient complexity to play a role of a stepping stone to language?

I believe the general strategy for answering these questions would be to emphasize that we should not expect a miracle solution in which symbolization and CIs emerge as general cognitive skills that enable flexible communication about arbitrary actions, objects, and events. Rather, we should look for a slow build-up of a variety of extensions to simpler communicative systems, where only at the end of this process something that can be described as symbolization and CIs are in place. That is, they are not *enabling conditions* for flexible communication but *an emergent result* of such communication. The extension that TOR provides is just one achievement in this lengthy process.

Since TOR gestures emerge in triadic interactions, they are effectively imperative gestures for acting on an object in a specific way and in that sense are weakly referential. They are referential in the same way that widely discussed vervet monkey alarm calls are 'functionally referential'. That is, these monkeys are famously said to emit particular calls in response to particular natural predators that lead to particular responses on the part of their receivers. The calls are referential because they are in some sense 'about' the predators but they are weakly so because they are also innate, relatively inflexible and apparently not produced on the basis of symbolization and CIs (Wheeler and Fischer 2012, for a recent critical overview). In addition, there does not seem to be a way to decide whether e.g. an eagle alarm call means 'Eagle!' or 'Hide in the bushes!', that is, whether the call is a declarative sign for the object or an imperative sign for the action and whether the signal is providing information or manipulating the receiver (Rendall et al. 2009; Seyfarth et al. 2010). In the same way, a TOR gesture employed in a fruit picking scenario could mean 'Lower the branch', 'Let us feast on this tasty fruit', 'pull down', 'branch' or 'fruit'. There is no fact of the matter which of these meanings is the correct interpretation and therefore whether a TOR gesture is merely imperative or already declarative and therefore stands for some object (cf. Wittgenstein's (2009) analysis of language games).

At this point, a defender of a pantomime-based account might insist that we still need an explanation for how a TOR gesture might get transformed into one that *really* refers to objects. After all, pantomime based on a general capacity for symbolization, has the advantage of giving our ancestors a way to express a variety of ideas about actions, objects, and events. What is more, within MSH, it is the ability to entertain such diverse thoughts and a need to distinguish pantomimes 'for action' from pantomimes 'for objects' that lead to the practice of introducing small modifications into them, thus fueling the transition to conventionalized protosigns and paving the road to language.

I would counter this insistence by noting that such a proposal is based on a view in which speaking is an expression of fully formed thoughts with determinate contents and therefore mental richness precedes linguistic sophistication and is its evolutionary source. A different view is possible, however (along the lines of Vygotsky (1987)). Namely, that a distinction between linguistic forms that express actions or objects emerges from a growing set of signals as they come to function in differing contexts, including the context of other signals. For example, the same TOR gesture could come to be accompanied by different sounds and thereby acquire a function of requesting an action whereas the sounds come to stand for the different objects on which the action is requested (Hutchins and Johnson 2009). It is this development in turn that would enable thoughts with determinate contents, i.e. thoughts that refer to objects vs thoughts that request actions.

Before such a stage, what matters is that a TOR-gesture is functionally about objects and flexibly acquired in development. In this way, it is a step beyond the communicative systems of non-human primates, whether they are monkey alarm calls or dyadic OR gestures.

Moving on to reciprocity and intersubjectivity, the TOR alternative holds that they can be a result of new types of interactions. That is, primate OR gestures are in some sense private and asymmetric because they emerge from interactions that do not go beyond the dyad and in which the bodily roles are complementary but the background is only of one's own know-how. As a result, a single individual is typically only on the producing or a receiving end of a certain gesture. However, if more complex interactions enter the life of the group, part of these interactions could be such that they occur between grown-up individuals who are similar in their background practical knowledge and action possibilities, allowing for their roles to be in principle interchangeable. That is, I can pull the tree branch so that you collect the fruit or vice versa. After ritualization of such a scenario, the same individual would have an opportunity to both produce a certain gesture and respond to it, creating a reciprocal gesture in Hurford's terminology. (In fact, such a situation, in which an individual is in some sense aware of the meaning of one's own gesture because they have experienced it from both sides, as a producer and an addressee can be seen as a source of symbolization on sociocultural accounts of cognition such as that of Vygotsky (1987) or Mead (1962). Developing this argument goes beyond the scope of this article.) Finally, if such cooperative interactions are important enough for the whole group and scaffolded by the use of the same types of objects, there is no *in principle* reason why a single individual cannot enter the same interactions with multiple others and therefore the gestures that emerge cannot go beyond a single dyad.

In sum, the novel features of TOR gestures with respect to the types of gestures in contemporary non-human primates would not be a result of an entirely new cognitive mechanism. Instead, they would be a result of the same mechanism employed in the context of a new set of social practices.

6. TOR in relation to MSH

Having started from MSH, a particular pantomime-based account, a few more words are in order on how it compares to the ideas presented here. The precise comparison would of course require two things: clarity on what constitutes the central ingredients of MSH and a fully developed TOR-based account. In particular, if it is absolutely essential to MSH that there is a stage of pantomime, which in turn requires symbolization and CIs, while a fully developed TOR account somehow precludes such a pantomime, then the two would be incompatible. However, if what is central to MSH is an involvement of MNS together with complex imitation (stages 1–4 in MSH), some form of a gestural protolanguage stage (5) and a subsequent emergence of language proper (stages 6 and 7), and if a TOR idea can accommodate all these, then the two could be merged modifying thereby the original MSH. While a TOR-based account is not yet fully developed, I can offer some broad strokes of where it is headed, which suggest adopting the latter option.

More specifically, I see a TOR-based gestural communication positioned just before the emergence of complex imitation. This possibility is particularly plausible given the developments that occurred within the MSH framework in the past few years, namely explicit modelling of ontogenetic ritualization as it occurs between interacting individuals.

Arbib et al. (2014) present a computational model of the emergence of a 'beckoning' gesture through a history of 'changes in the brains of two agents during interactive [mutual] behavioral shaping'. The underlying neural architecture employed by the authors is a model of the MNS, which implements a planning of sub-actions that serve a particular goal. In the simulation, the model starts from a child's goal to bond socially with the mother, which is initially accomplished by tugging. The mother's MNS enables her recognize the bonding goal of the child from haptic information (of tugging) and she responds by moving closer and embracing it. Over time the mother's MNS learns to recognize the goal earlier in the trajectory from increasingly more abbreviated haptic and then visual information until it is sufficient for the child to reach towards the mother to accomplish the goal. On the other hand, the child learns to associate proprioceptive reaching state with the distal goal (bonding), leading to the formation of a new beckoning gesture.

This model has not been explicitly postulated to fit into the broader MSH evolutionary trajectory and it is not yet clear whether the activity of the MNS indeed plays a role in the formation of primate gestures. What is widely acknowledged, however, is its involvement in recognition of object-directed actions, especially if such actions are within the observer's own repertoire. It has also been found to be active when a complementary action is required (Newman-Norlund et al. 2007). If two individuals frequently perform object-directed actions jointly, especially if their roles are interchangeable, it is very plausible that the MNS would facilitate action recognition in such contexts. It is also not difficult to imagine that just as in the OR model described above, the MNS of one individual would learn to recognize the collaborator's goals based on visual information of their manual movements until these movements are abbreviated into gestures, and still respond to them appropriately.

Now, at the end of such a process one would have a MNS responsive to increasingly intransitive manual movements (ritualized gestures) just as in complex imitation. This responsiveness would, however, be a result of the gesturing experience itself, not an exaptation of mechanisms originally employed in instrumental actions and put to use in communicative settings. Such a view is consistent with an associative view on the MNS and its evolution, according to which mirroring properties to a certain type of input are acquired through experience of that particular type (Cook et al. 2014).

What has additionally been argued from the perspective of an associative view (e.g. Heyes (2010, 2013)) is that the evolutionary changes to the MNS and other brain systems do not need to be seen as an acquisition of specific novel functions (such as a capacity to imitate or produce pantomimes). These changes can be more subtle—increased attention to certain type of input, faster learning and processing, improved motor control. If this is so, then as the process of ritualization became more important, it could also lead to an increased speed with which such gestures can emerge, improved execution and perhaps an ability to learn them by imitation.

In sum, the TOR alternative would not be an argument against the role of the MNS or complex imitation in the emergence of gestural protolanguage. Rather, it sets up a scenario in which a simple associative mechanism can lead to gestures that occur in triadic contexts and then mirroring properties enhance this process in various ways leading to an even greater repertoire of such gestures. As the repertoire increases, cultural transmission can then result in the emergence of structure and convention. Whether this process still requires a stage of pantomimes and strictly gestural protosigns, or rather it brings us straight to a multimodal language is up for debate. I believe the latter is more parsimonious and conducive to a multimodal view. However, one could see the growing TOR-repertoire as laying a groundwork for pantomime-ready brain and a certain convention of triadic communication.

The proposal presented in this article is obviously incomplete and in need of further development. However, I hope to have shown that it is a viable option to be explored.

Acknowledgements

E.A. gratefully acknowledges the funding provided by the Netherlands Organisation for Scientific Research (NWO) within the programme PhDs in the Humanities, research project PGW-12-32. E.A. is especially thankful to Marc Slors, Bart Geurts, and Michael Arbib for helpful comments and extensive discussion on previous versions of this article.

References

Arbib, M. A. (2005) 'From Monkey-Like Action Recognition to Human Language: an Evolutionary Framework for Neurolinguistics', *The Behavioral and Brain Sciences*, 28/2: 105–24.

——. (2011) 'From Mirror Neurons to Complex Imitation in the Evolution of Language and Tool Use', *Annual Review of Anthropology*, 40/1: 257–73.

——. (2012) *How the Brain got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press.

Arbib, M., Ganesh, V., and Gasser, B. (2014) 'Dyadic Brain Modelling, Mirror Systems and the Ontogenetic Ritualization of Ape Gesture', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369/1644: 4–14.

Armstrong, D. F., and Wilcox, S. E. (2007) *The Gestural Origin of Language*. Oxford: Oxford University Press.

Astington, J. W., and Baird, J. A. (2005) *Why Language Matters for Theory of Mind*. Oxford: Oxford University Press.

Aziz-Zadeh, L. et al. (2012) 'Understanding Otherness: the Neural Bases of Action Comprehension and Pain Empathy in a Congenital Amputee', *Cerebral Cortex*, 22/4: 811–19.

Bar-On, D. (2013) 'Origins of Meaning: Must we "go Gricean"?', *Mind & Language*, 28/3: 342–75.

Catmur, C., Walsh, V., and Heyes, C. (2007) 'Sensorimotor Learning Configures the Human Mirror System', *Current Biology*, 17/17: 1527–31.

Cook, R. et al. (2014) 'Mirror Neurons: From Origin to Function', *The Behavioral and Brain Sciences*, 37/2: 177–92.

Corballis, M. C. (2002) *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

——. (2010) 'Mirror Neurons and the Evolution of Language', *Brain and Language*, 112/1: 25–35.

Deacon, T. W. (1996) *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: Norton.

DeLoache, J. S. (2004) 'Becoming symbol-minded', *Trends in Cognitive Sciences*, 8/2: 66–70.

Gallese, V., and Goldman, A. (1998) 'Mirror Neurons and the Simulation Theory of Mind-reading', *Trends in Cognitive Sciences*, 2/12: 493–501.

Garrod, S. et al. (2007) 'Foundations of Representation: Where Might Graphical Symbol Systems Come From?', *Cognitive Science*, 31/6: 961–87.

Goldin-Meadow, S. (2011) 'What Modern-Day Gesture Can Tell Us About Language Evolution'. In: Tallerman M. and Gibson K. R. (eds) *The Oxford Handbook of Language Evolution*, pp. 545–57. Oxford: Oxford University Press.

Grice, P. (1957) 'Meaning', *The Philosophical Review*, 66/3: 377–88.

Halina, M., Rossano, F., and Tomasello, M. (2013) 'The Ontogenetic Ritualization of Bonobo Gestures', *Animal Cognition*, 16/4: 653–66.

Harmand, S. et al. (2015) '3.3-million-year-old Stone Tools from Lomekwi 3, West Turkana, Kenya', *Nature*, 521/7552: 310–5.

Hawkey, D. J. C. (2008) 'Beyond the Individual in the Evolution of Language', Unpublished doctoral dissertation, University of Edinburgh.

Hewes, G. (1976) 'The Current Status of the Gestural Theory of Language Origin', in Harnad S. R., Steklis H. D., and Lancaster J. (eds) *Origins and evolution of language and speech*, Vol. 280, pp. 482–504. New York: New York Academy of Sciences.

Hewes, G. W. (1973) 'Primate Communication and the Gestural Origin of Language', *Current Anthropology*, 14/1/2: 5–24.

Heyes, C. (2010) 'Mesmerising Mirror Neurons', *NeuroImage*, 51/2: 789–91

——. (2013) 'What Can Imitation do for Cooperation?' In Stereiny K., Joyce R., Calcott B., and Fraser B. (eds) *Cooperation and its Evolution*, pp. 313–32. MIT Press. (in press).

Hurford, J. (2007) *The Origins of Meaning: Language in the Light of Evolution*. Oxford: Oxford University Press.

Hutchins, E., and Johnson, C. M. (2009) 'Modeling the Emergence of Language as an Embodied Collective Cognitive Activity', *Topics in Cognitive Science*, 1/3: 523–46.

Hutto, D. D. (2008) 'First Communions: Mimetic Sharing without Theory of Mind'. In Zlatev J., Racine T. P., Sinha C., and Itkonen E. (eds) *The Shared Mind: Perspectives on Intersubjectivity*, pp. 245–76. Amsterdam: John Benjamins.

Irvine, E. (2016) 'Method and Evidence: Gesture and Iconicity in the Evolution of Language', *Mind & Language*, 31/2: 221–47.

Kendon, A. (1973) 'Some Considerations for a Theory of Language Origins', *Man*, 26: 199–221.

Leavens, D. A., Russell, J. L., and Hopkins, W. D. (2005) 'Intentionality as Measured in the Persistence and Elaboration of Communication by Chimpanzees (pan troglodytes)', *Child Development*, 76/1: 291–306.

Liebal, K., and Call, J. (2012) 'The Origins of Non-Human Primates' Manual Gestures', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367/1585: 118–28.

—— et al. (2014) *Primate Communication: A Multimodal Approach*. Cambridge: Cambridge University Press.

McNeill, D. (2012) *How Language Began: Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press.

Mead, G. H. (1962) *Mind, Self, and Society*. Chicago: University of Chicago Press.

Moore, R. (2015) 'A Common Intentional Framework for Ape and Human Communication', *Current Anthropology*, 56/1: 70–71.

——. (2016) Gricean communication, joint action, and the evolution of cooperation. Topoi, 1–13 <https://doi.org/10.1007/s11245-016-9372-5> accessed 21 December 2017.

Newman-Norlund, R. D. et al. (2007) 'The Mirror Neuron System is More Active During Complementary Compared with Imitative Action', *Nature Neuroscience*, 10/7: 817–18.

Origgi, G., and Sperber, D. (2000) 'Evolution, Communication and the Proper Function of Language'. In Carruthers P. and Chamberlain A. (eds) *Evolution and the Human Mind: Language, Modularity and Social Cognition*, pp. 140–69. Cambridge: Cambridge University Press.

Peirce, C. S. (1931–1958) *Collected Papers of Charles Sanders Peirce*, Vols. 1–8. Cambridge, MA: Harvard University Press.

Piaget, J. (1962) *Play, Dreams and Imitation in Childhood*. New York: Norton.

Rendall, D., Owren, M. J., and Ryan, M. J. (2009) 'What Do Animal Signals Mean?', *Animal Behaviour*, 78/2: 233–40.

Reynolds, P. C. (1993) 'The Complementation Theory of Language and Tool Use'. In Gibson K. and Ingold T. (eds) *Cognition, Tool Use, and Human Evolution*, pp. 407–28. Cambridge: Cambridge University Press.

Rizzolatti, G., and Arbib, M. A. (1998) 'Language Within Our Grasp', *Trends in Neurosciences*, 21/5: 188–94.

Rooij, I. v. et al. (2011) 'Intentional Communication: Computationally Easy or Difficult?', *Frontiers in Human Neuroscience*, 5: 52.

Scott-Phillips, T. (2015) *Speaking our Minds*. Houndmills: Palgrave Macmillan.

Scott-Phillips, T. C. (2015) 'Nonhuman Primate Communication, Pragmatics, and the Origins of Language', *Current Anthropology*, 56: 56–80.

Seyfarth, R. M. et al. (2010) 'The Central Importance of Information in Studies of Animal Communication', *Animal Behaviour*, 80/1: 3–8.

Slocombe, K. E., Waller, B. M., and Liebal, K. (2011) 'The Language Void: the Need for Multimodality in Primate Communication Research', *Animal Behaviour*, 81/5: 919–24.

Smit, H. (2016) 'The Transition from Animal to Linguistic Communication', *Biological Theory*, 11/3: 158–72.

Sperber, D., and Wilson, D. (1995) *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stokoe, W. C. (2001). *Language in Hand: Why Sign Came Before Speech*. Washington, DC: Gallaudet University Press.

Stout, D., and Chaminade, T. (2012) 'Stone Tools, Language and the Brain in Human Evolution', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367/1585: 75–87.

Tomasello, M. (2008) *Origins of Human Communication*. Cambridge: MIT Press.

——, and Zuberbühler, K. (2002) 'Primate Vocal and Gestural Communication'. In *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, pp. 293–29. Cambridge: MIT Press.

Van Rooij, I. (2008) 'The Tractable Cognition Thesis', *Cognitive Science*, 32/6: 939–84.

—— (2012) 'Self-Organization Takes Time Too', *Topics in Cognitive Science*, 4/1: 63–71.

Vygotsky, L. S. (1987) *The Collected Works of LS vygotsky, Volume 1: Problems of General Psychology*. New York: Plenum Press.

Wheeler, B. C., and Fischer, J. (2012) 'Functionally Referential Signals: a Promising Paradigm Whose Time Has Passed', *Evolutionary Anthropology*, 21/5: 195–205.

White, T. D. et al. (2009) 'Ardipithecus Ramidus and the Paleobiology of Early Hominids', *Science*, 326/5949: 75–86.

Wittgenstein, L. (2009) *Philosophical Investigations*. Oxford: Wiley-Blackwell.

Young, N. M. et al. (2015) 'Fossil Hominin Shoulders Support an African Ape-Like Last Common Ancestor of Humans and Chimpanzees', *Proceedings of the National Academy of Sciences of the United States of America*, 112/38: 11829–34.

Zlatev, J. (2007). 'Embodiment, Language and Mimesis'. In Ziemke T., Zlatev J., and Frank R. M. (eds) *Body, Language and Mind, vol.1: Embodiment*, pp. 297–337. Berlin: Mouton.

—— (2008) 'From Proto-mimesis to Language: Evidence from Primatology and Social Neuroscience', *Journal of Physiology, Paris*, 102/1–3: 137–51.

——, Persson, T., and Gärdenfors, P. (2005) 'Bodily Mimesis as "the missing link" in Human Cognitive Evolution', *Lund University Cognitive Studies*, 121: 1–45.

—— et al. (2013) 'Understanding Communicative Intentions and Semiotic Vehicles by Children and Chimpanzees', *Cognitive Development*, 28/3: 312–29.