ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/70855/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Irvine, Elizabeth 2016. Method and evidence: Gesture and iconicity in the evolution of language. Mind and Language 31 (2) , pp. 221-247. 10.1111/mila.12102

Publishers page: http://dx.doi.org/10.1111/mila.12102

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Method and evidence: Gesture and iconicity in the evolution of language?

Abstract: The aim of this paper is to mount a challenge to gesture-first hypotheses about the evolution of language by identifying constraints on the emergence of symbol use. Current debates focus on a range of pre-conditions for the emergence of language, including co-operation and related mentalising capacities, imitation and tool use, episodic memory, and vocal physiology, but little specifically on the ability to learn and understand symbols. It is argued here that such a focus raises new questions about the plausibility of gesture-first hypotheses, and so about the evolution of language in general. After a brief review of the methodology used in the paper, it is argued that existing uses of gesture in hominid communities may have prohibited the emergence of symbol use, rather than 'bootstrapped' symbolic capacities as is usually assumed, and that the vocal channel offers other advantages in both learning and using language. In this case, the vocal channel offers a more promising platform for the evolution of language than is often assumed.

Word count: 9992

1. Introduction

In recent years there has been an increasingly wide acceptance of the idea that

---

Many thanks to Kim Sterelny and the philosophy of biology group at ANU for their help throughout this project, to numerous audiences at EvoLang, the AAP and elsewhere for their valuable feedback, to Matteo Colombo and anonymous referees for their comments, and to Sean Roberts for ongoing discussions.
Address for correspondence: John Percival Building, University of Cardiff, Cardiff CF10 3EU, UK
Email: irvinee@cardiff.ac.uk

gestures played a significant role in the evolution of symbolic language (Arbib et al. 2008; Arbib 2005; Fay et al. 2013; Tomasello 2008; Corballis 2003, 2009; Sterelny 2012a). In particular, it has been hypothesized that the first (proto-)linguistic systems, with fairly small vocabularies and little to no grammar, were either exclusively, or primarily, gestural in nature. The use of vocalisations is argued to have come late on the scene, and taken over from gestures because vocalisations 'free up the hands', allow for communication in the dark or in dense forests, and so on.

Perhaps the most convincing evidence for gesture-first hypotheses comes from research on primate communication and physiology[1]. It is claimed that primate vocalisations are automatic expressions of emotion, with no intentional vocal control (e.g. Tomasello 2008; Sterelny 2012a). The primate vocal tract is also very different to that of modern humans. Non-human primates can produce a much smaller range of sounds, so have limited expressive capacity. It is claimed that they cannot control their breathing which is essential to speech, and they also have little to no capacity for vocal learning (Fitch et al. 2013; Fitch, 2000). Both in terms of the apparent lack of top-down control of vocalisations, and in terms of existing physiology, the vocal channel seems to provide a poor evolutionary platform for linguistic communication.

In contrast primates have intentional control of hand and body movements, and already use gestures in reasonably flexible ways to communicate (Genty et al., 2009; Hobaiter & Byrne, 2011; Liebal et al., 2004; Pika et al., 2005; Roberts et al., 2013; Tomasello, 2008). Further, since imitation learning was likely important in

[1] This research makes it possible to make a case for gesture-first hypotheses without relying on controversial mirror-neurons (Cook et al. 2014; Heyes, 2010a, 2010b), and is perhaps best developed by Tomasello and Sterelny.

early hominid populations, memory for sequences of hand movements could have already been the target of selection[2]. So, based purely on an assessment of the available platforms for language, gestures looks like the best bet (Arbib et al., 2008; Arbib, 2005; Corballis, 2003, 2009; Fay et al., 2013; Sterelny, 2012a; Tomasello, 2008).

In particular, it has been suggested that the physiological changes required for speech are so great (and costly), and that speech requires such fine-grained motor control, that the relevant evolutionary pressures could only have come from an existing, and potentially fairly complex, gestural system of communication. As Sterelny (2012a) notes: 'We evolved speech as a result of living in a world in which communication was already important' (p. 2143). If one makes the further plausible assumption that only a reasonably complex language-like system could have provided strong selective pressures for high levels of expressivity and control, then one has an argument in favour of the existence of an early gestural *language*, again, potentially fairly complex, which was only later followed by vocal language[3].

Another factor is that gesture provides a much better channel than speech for generating iconic signs, where the form of the sign 'resembles' its meaning. For example, a sign can be iconic if it mimics an action that is referred to, if the shape of the sign resembles the shape of the referent, or the meaning is somehow otherwise obvious (e.g. hands moving outwards for 'bigger'). By using iconic signs, adults can communicate fairly easily in the gestural channel in the absence of shared language

[2] Thanks to Kim Sterelny for this point.
[3] Though one of course still has to say something about why gesture was largely abandoned as the primary modality for delivering semantic content.

(Fay et al., 2013). Spatial and temporal features, motion, actions, and objects associated with actions can all be represented with gestures, and understood apparently without much cognitive effort.

Gesture and iconicity therefore go together in gesture-first hypotheses about language evolution; early communication, and then linguistic systems, were made up of gestures that took the form of pantomime and iconic signs (Donald, 2001; Sterelny, 2012a; Tomasello, 2008). Further, it seems plausible that iconic gestures would have made it possible for individuals with no symbolic capacities to 'bootstrap' themselves up into successful symbol and language users; shifting naturally from using iconic pantomimic signs to the kind of arbitrary symbols now found in many languages.

There are potential problems with these hypotheses though. First, recent research has shown that primate vocalisations are not always automatic and unintentional. Primates can inhibit calls, change their timing and duration, and alter them according to social contexts and so appear to satisfy criteria for first order intentionality (Genty et al., 2009; Pollick et al. 2005; Salmi et al. 2013; Schel et al. 2013; Seyfarth & Cheney, 2010; Slocombe et al., 2010), and at least gorillas seem to have some (limited) control over their breathing and vocal apparatus (Clark & Perlman, 2014). In this case, the primate vocal channel may not provide such a poor platform for the evolution of language as is often made out.

There is also further room to question the assumption that the physiological changes required for speech could only have been selected for in the context of an existing gestural linguistic system. Vocal sounds are already used to communicate in non-human primates, and may have played other roles in hominid societies too. For example, it has been suggested that music, or music-like communication, also played various roles in social bonding or displays in hominid groups (Dunbar 1998; Falk, 2009; Fitch, 2005; Mithen, 2005; and see Ackermann et al., 2014 for neurophysiological support). While not endorsing any of these hypotheses in particular, they at least show that there are no conclusive reasons to think that the physiological and other changes required for speech could *only* have occurred in the presence of an existing gestural *language*, that for some reason was no longer sufficient for communicative needs, rather than resulting from some other set of cumulative effects.

Second, and the topic of this paper, there is research that suggests that the gestural channel may be at a disadvantage compared to the vocal channel when it comes to enabling individuals to understand and learn symbols. Symbolic capacities are rarely discussed in debates over language evolution, despite being a prerequisite for language use and language learning (Deacon, 1998 is a rare and controversial exception). Presumably, this is because a range of animals can be taught to use symbols, but don't have much motivation to communicate, hence the focus on social aspects of communication and co-operation. However, there are factors that make symbol use and symbol learning easier or harder, and which may serve to place significant constraints on the contexts or modalities in which symbol use can emerge, even for groups where co-operative communication is the norm.

So, for instance, it is often suggested that since gestures were likely already so widely used by hominids (e.g. for communication, teaching skills), they could also be used as (iconic) symbols. However, the findings discussed below strongly suggest that it is easier to treat something as a symbol when it is not already the target of fairly specific, important, or automatic processing. That is, the preponderance of gestures and their many important roles in hominid lives may have made it *harder*, not easier, to treat them as symbolic. There is also evidence that learning mappings between symbols and objects is easier when visual attention can be fixed on the referent only, and not, for example, split across both the referent and a gesture. So, while gestural communication was (and still is) important, using and learning symbols may be far easier in the vocal channel. In this case, there is room to question whether iconic gestures played a positive role in the transition from non-symbolic communication to symbolic language, and more generally, whether linguistic symbols first emerged in the gestural channel.

However, while the aim here is to challenge the intertwined roles of iconicity and gesture in the evolution of language as found in gesture-first hypotheses, it is not to rule these hypotheses out entirely. The arguments made here suggest at least an earlier transition from gestural to vocal language than usually assumed, and highlights the need for more fine-grained analyses of the relationship between symbolic capacities and other features of language and communication. With other evidence and arguments this work could contribute to an argument for a vocal-first trajectory, but it is not attempted here.

The paper is organized as follows. Gesture-first hypotheses are described more fully in Section 2. Section 3 will outline some methodological constraints on using developmental and other evidence to evaluate evolutionary hypotheses. Research on human infants' use and understanding of iconicity is briefly reviewed in Section 4, and then used as a foil in a longer discussion of infants' interpretation of gestures in Section 5. Section 6 presents a general set of constraints on successful symbol use, along with a discussion of how these interact with different communicative modalities. In Section 7 a brief overview of interactions between modality and symbol learning is presented, and linked to a general advantage for vocal over gestural languages. These discussions are summarized in Section 8, and Section 9 concludes.

2. Signs to Symbols with Gesture?

The target question that many gesture-first theorists are interested in is how language, as a symbolic communication system, got going in the first place. Many animals communicate with signs, where the signs co-vary and are physically associated with the presence of their referent, and only refer in the here and now. However, humans do something different; they use symbols. Symbols are 'stimulus independent' so denote rather than co-vary with their referents, and can be used to refer to the elsewhere and elsewhen[4].

In the Cultural Learning approach to language evolution (Tomasello, 2008), the evolution of basic theory of mind (understanding of others' mental states), co-operation, and the kind of cognitive 'de-coupling' that comes with teaching others by rehearsing and exhibiting actions offline (Sterelny, 2012a), create the cognitive

[4] There is clearly a vast amount more to say about what linguistic symbols are, and how they differ from the signs found in animal communication systems, but the present brief formulation will do for now. It will become more obvious throughout the paper what being a symbol user amounts to, in particular in terms of the cognitive capacities and learning mechanisms involved in symbol use.

platform for symbolic communication. Hurford (2007) has also argued for the crucial role of conceptual knowledge and episodic memory in making it possible to share meanings at all. Once you have hominids who share a lot of common ground (shared background) and are set up to understand communicative intentions, you have hominids who seem set up to easily comprehend iconic (gestural) signs, and later iconic symbols. Adults, after all, communicate with iconic signs and symbols and pantomime fairly easily, providing there is enough context to make it obvious what kind of things one might want to communicate about.

So, the standard picture, both implicit and explicit in gesture-first theories, goes something like this. First, there is chimp-like gestural communication with signs, then once the cognitive platform outlined above has evolved sufficiently (for other purposes), pointing, pantomime and perhaps other iconic gestural signs start to be used, transitioning to iconic symbols over time. Again, iconicity plays a crucial role here, since '[e]ven for minds adapted to language and to modern human life—even for minds that can use arbitrary, purely conventional symbols— iconicity is advantageous. Presumably, it is easier to remember or to recognize iconic signs' (Sterelny 2012a, p. 2144). Via processes of conventionalization, these iconic symbols change to arbitrary (non-iconic) symbols over time (on how this shift works and why it is beneficial for large vocabularies, see Gasser, 2004; Monaghan et al., 2014). Given the significant adaptive advantages of using language, and the apparent advantages of the vocal channel over the gestural channel, strong selective pressures then act on the vocal channel to turn it into the main channel for carrying semantic content.

However, while enculturated and some captive chimps point, the use of pantomime and iconic gestures to communicate is a missing link. Non-human primates rarely, if ever, pantomime communicatively, and what look like iconic gestures instead seem to be short-hand versions of actions. These result from the process of ontogenetic ritualization, where over time sequences of actions become shorter and can be used to prompt a conspecific to complete a standard interaction. These gestures can include begging gestures, raised arms to indicate a desire to be picked up, grabbing but with little force to initiate play, and so on (Halina et al., 2013; Liebal & Call, 2012). These gestures can look iconic to human observers, but do not serve as iconic gestures for their users. That is, the gestures are not understood via recognition of a similarity relation between the gesture and its meaning, but function instead as ritualized, short-hand ways of triggering an action, and can be fairly unique to interacting dyads (e.g. mothers and offspring). The cognitive capacities driving successful communication in the chimp case are very different to those involved in communication via iconicity.

There are also likely to be limitations on the kind of pantomimes that 'make sense' to early hominids. Pantomimes of human actions seem more likely to be understood than pantomimes of an animal's behavior that is based on a different body plan. As noted in Tomasello et al. (1999), to understand a pantomime of a bird flapping its wings, one 'must effect some kind of iconic mapping of wings to arms' (p. 581). Pantomimes of birds and perhaps other prey animals require one extra cognitive step in order to be understood, and so may not be among the first set of gestural symbols. However, despite these problems, the gestural channel does seem like a serious contender because of its potential ability to bootstrap communication from fairly simple signs, relying on existing domain-specific capacities (e.g. action understanding) as well as basic theory of mind (understanding of communicative intentions), to more general, complex and abstract symbolic communication.

3. Methodology

Questions about language evolution are now often regarded as requiring multi-disciplinary approaches. The evidence used in this paper to challenge gesture-first hypotheses comes mainly from developmental and comparative studies of symbolic capacities, but as this is often deemed to be particularly problematic, this section outlines the ways that the evidence is used to try to allay some of these worries.

The major concern in using developmental evidence is in avoiding the ontology/phylogeny fallacy. Here, this stems from the fact that contemporary language change and language learning tracks features of fully language-ready contemporary humans, and that this may have little in common with the evolution of language and the cognitive systems of pre-linguistic hominids. That is, development trajectories may have very little in common with evolutionary trajectories. There are however several responsible ways to use developmental evidence, sketched below.

One way that developmental psychology can contribute to evolutionary hypotheses is by essentially being embedded in psychological research more generally. That is, developmental psychology offers a source of evidence that, with others (particularly comparative psychology), can be used to isolate and differentiate cognitive capacities that might otherwise be thought to run together, describe how cognitive capacities can be graded, and can be used to illustrate how different cognitive capacities interact. In particular, as used in this paper, this can be used to challenge common assumptions about what kind of tasks are cognitively 'easy' or more basic in both developmental and evolutionary terms.

For example, debates about the development of theory of mind serve to pinpoint which specific tasks really do require mentalising (attributing and reasoning about mental states) and which do not, and what other processes can generate complex social behaviours that seem to (but do not) rely on mentalising (Baillargeon et al., 2010; Heyes, 2014a, 2014b; Senju et al., 2011). Along with comparative research on non-human primates, this can be used to inform hypotheses about the evolution of theory of mind such that they are sensitive to subtle differences in cognitive capacities and the kinds of cues they rely on, and that identify precisely what kinds of tasks, situated in a particular ecological or social scenario, that mentalising is really necessary for.

Relatedly, developmental psychology, along with comparative studies of animal cognition, can also be used to make reasonable guesses about some of the cognitive capacities of early hominids, such as what kind of domain general learning capacities they may have had. For example, developmental and comparative research provides good reason to think that early hominids had the capacity to fast-map associated items (discussed in more detail below).

With background information on the distinctions between cognitive capacities and their interrelations, as well as reasonable guesses on the domain general capacities that early hominids had, it is possible to place constraints on evolutionary hypotheses. So for example, research on symbolic capacities can be used to identify constraints on the ways that symbol use is likely to have emerged; in particular here on the modality that is most likely to support the emergence of symbols. It is fully acknowledged that these constraints are one set among many, but these constraints may be reasonably strong; while symbol learning is now accompanied by dedicated cultural and social scaffolding which can help overcome cognitive limitations, hominid language learners would not have had access to this. In this case, their cognitive features, perhaps particularly those related to symbol use, could have had a significant impact on the form of early linguistic systems.

4. Comprehension of Iconic Symbols by Pre-Linguistic Individuals

The argument starts with the weakest evidence. The general finding briefly reviewed below is that iconicity (where the form of a sign or symbol resembles its meaning, such as pantomimed actions) plays very little or no role early language acquisition or comprehension in human infants (both hearing and deaf). This is inconsistent with at least a simplistic prediction from gesture-first hypotheses; that if the recognition of iconicity is claimed to be cognitively ‘easier’ than the use and comprehension of symbols, then one might expect this to be true of developmental sequences. This is liable to be dismissed according to the ontogeny/phylogeny fallacy noted above, but the evidence reviewed below provides a useful foil to subsequent discussions about the relationship between iconicity and symbolic capacities in an evolutionary setting.

4.1 Hearing Infants

Children point a lot to communicate with others, but iconic gestures appear to come late on the scene in language acquisition. For example, children's use of pointing predicts their rate of acquisition of nouns, but their use of an iconic gesture to represent a verb comes six months after they have successfully learned the word for the verb (Ozcaliskan et al., 2013). Children are also no better at matching quantity to iconic signs (numbers of fingers) than to arbitrary signs (number words), when asked to give someone a number of objects, or to indicate how many objects are present (Nicoladis et al., 2010). More generally, Namy (2001) found that children could map arbitrary words, gestures, non-verbal sounds and pictograms to referents equally well, and in (Namy et al., 2004) infants at the early stages of symbolic development (18 month olds), mapped iconic symbols just as well as arbitrary symbols. It seems as though any symbol, embedded in an obvious naming routine can be learned just as easily; iconic symbols show no advantage.

Finally, Namy (2008) investigated the ability of infants to recognize iconicity, that is, to understand the meaning of iconic symbols based on the recognition of resemblance relations between the symbol and its meaning (again, these relations could be ones of mimicry of actions, or similarities between perceptual features of the symbol and its referent). She found that recognition of iconicity only emerges robustly around age 2 (Tolar et al., 2008 find that recognition of iconicity only stabilizes at 3 years old). So, recognition of iconicity comes online *after*, not before, the emergence of symbolic development, so it seems that at least for contemporary human infants 'symbolic insight does not originate in or rely upon iconicity' (Namy, 2008, p. 845).

4.2 Deaf Infants

A similar picture arises from work on the acquisition of sign language. Unlike most vocal languages, iconicity is fairly prevalent in sign languages (Perniss et al., 2010; Taub, 2000; Vermeerbergen, 2006), so deaf infants get a significant amount of input in the form of meaningful iconic gestures. Yet again here iconicity plays very little role in language acquisition. The degree to which a sign resembles its referent does not predict its age of acquisition (Orlansky & Bonvillian, 1984). Signs that are highly iconic yet also highly morphologically complex (e.g. 'to give', where both hands move away from the body and open) are learned late, not early (Meier, 1987), and phonological complexity and motor constraints seem to trump any effects of iconicity in gesture production (Meier et al., 2008).

So, even for languages where iconicity is a major feature, it does not seem to play a positive role in language acquisition. Accordingly, in their review of gesture's role in language, Goldin-Meadow and Alibali (2013) state that 'the iconicity found in a sign language does not appear to play a significant role in the way the language is processed or learned' (p. 270).

5. Default Interpretations of Gesture

From the evidence briefly reviewed above, recognition of iconicity does not appear to precede the development of symbolic capacities in modern infants. However, as noted earlier, this may well just illustrate an ontology/phylogeny difference. Modern humans have undergone a large amount of biological evolution, and language learning is now heavily socially scaffolded, so what is easy for human infants (e.g.

symbol learning) may not be what was easier for hominids (e.g. perhaps recognition of iconicity).

However, related work on infants' understanding of symbols poses more serious worries about iconic gestures and their interaction with symbolic and other capacities. The work reviewed below also challenges some of claims made by Namy about the range of signals that infants readily accept as symbolic, as it seems that iconic gestures are not usually treated as symbols, and perhaps not as iconic, after all. Importantly, the ability to treat iconic gestures (particularly pantomimed actions) as symbols may be fairly hard in cognitive terms, and in ways that are relevant to the cognitive capacities and communicative contexts of early hominids. If this is the case, then it is less plausible that iconic gestures bootstrapped symbolic capacities in these individuals.

5.1 Iconic Gestures and Activity Schemas

The original hypothesis that infants treat iconic gestures differently to other potential symbols was developed in Tomasello et al. (1999). In one of the experimental manipulations used in this study, infants were asked to select an object by an experimenter, where the required object was represented by a gesture. The gestures were often pantomimes of actions that one could perform with or on the object, or were otherwise highly iconic, and so assumed to be easy to comprehend. Sometimes the infant had previously observed the gesture being modeled with the appropriate object (e.g. hammering action with a hammer), but sometimes not. The authors found that older infants (over 35 months) could pick out the right object both when they had seen the iconic gesture being used with the object before, and when it was a novel iconic gesture. In contrast, younger infants (18 and 26 months old) could pick out the right object only when they had previously seen the gesture paired with the object. That is, younger infants could not spontaneously pick out an appropriate referent for a novel gesture based on its iconic properties.[5]

This suggests an initial deflationary account of what is going on when young infants respond to iconic gestures (here, mainly pantomimes of actions). The idea is that infants are not learning symbol-referent pairings via iconicity (resemblance relations between symbol and meaning), but are learning about 'activity schemas' or 'action schemas' that include both an action and an object (e.g. hammering actions go with hammers). When a schema is activated by a gesture, such as pantomimed hammering, the infant is able to associate it with the right object because they have observed this pairing before. Iconicity is playing no role here though; there is no interpretation of the gesture as 'resembling' an object, but just an association in the form of an action or activity schema.

5.2 Fast Mapping and Activity Schemas

Marentette and Nicoladis (2011) tested this idea further to see if children use iconic gestures as symbolic 'labels', similar to how they use words or signs in sign language, or if they understand them via activity schemas as outlined above. They did this by seeing how well 2-4 year olds could 'fast-map' objects to words, arbitrary gestures and iconic gestures. Fast-mapping is an exclusion-based form of associative learning hypothesized to be used in language learning. Roughly, the idea of exclusion learning

[5] Subsequent findings from Striano et al. (2003) show that 26 month olds could also do this, but this seemed to be because infants were already very familiar with the objects and actions used in the study.

is that learners assume that novel symbols always represent novel objects (and vice versa), so if faced with a novel object one can exclude all known symbols as possibly referring to it (and vice versa). Making this simple assumption means that during the course of learning the space of possible mappings between symbols and referents can be drastically reduced over time, since the more mappings you know, the more mappings you can exclude (for alternative mechanisms see Smith & Yu, 2008; Trueswell et al. 2013). Children appear to use fast-mapping to learn symbols in natural languages, so the test is to see if they also use fast-mapping to learn gesture-referent pairings, and so treat (iconic) gestures as symbols.

In Marentette and Nicoladis's study, children are first introduced to objects and the symbols that refer to them, either novel words, arbitrary gestures, or iconic gestures, in an explicit naming ceremony ('this is an X', etc). At the testing phase, infants then have to choose which of two objects a symbol represents. Since the experiment includes very few object-symbol exposures, the only way infants can make object-symbol mappings, and so succeed on the test task, is using exclusion-based learning (fast-mapping).

The authors found that arbitrary words were fast-mapped over all ages, unsurprisingly, since infants seem to be able to use fast-mapping (or something like it) to learn words in natural languages from about 12-13 months onwards. They found that fast-mapping of iconic gestures (here, mostly pantomimes of actions) was not as good, but got better over age. Again, this may be unsurprising, given that apparent recognition of iconicity increases from age 2 years old onwards. The interesting result is that fast-mapping of arbitrary gestures was generally poor for children over all ranges, even when the exclusion based strategy was an obvious one (e.g. when they already knew the symbol for one of the two objects they had to choose between).

Marentette and Nicoladis's explanation of these results tracks the previous one. They argue that children over all age ranges apply a 'default interpretation' to actions, in terms of forming activity schemas that link actions to objects. This default interpretation is applied to both the iconic and non-iconic gestures used in the study. This works as follows. Children faced with iconic gestures try to link the gesture with the object to form an activity schema. This isn't too hard for iconic gestures in the form of pantomimes of actions that one can do with/on the object; the mapping is a fairly natural one given the affordances of the object. When children are faced with an arbitrary gesture, they also try to form an activity schema. However, now the gestures now don't 'make sense' as actions that one could do with/on the object; they have nothing to do with the affordances of the object. In this case, they fail to construct an activity schema, so fail to make any gesture-object mappings.

It is important to note the two things that are *not* going on here, which should also make the hypothesis more clear. First, for younger infants who cannot recognize iconicity in gestures (cannot understand the meaning of novel iconic symbols via recognition of resemblance relations between gestural symbols and meanings), iconicity is clearly not playing a role in linking apparently iconic gestures to objects. Instead, they are making simple associations between particular actions and particular objects.

Second, none of the infants in the age range studied (2-4 year olds) treated either iconic or arbitrary gestures on a par with other potential symbols, since they failed to apply the method of fast-mapping, even when the gestures were presented in standard naming routines. This was true even for fairly competent symbol users (4 year olds). Instead of treating pantomimed actions as symbols, Marentette and Nicoladis (2011) suggest that 'children prefer to interpret gestures as descriptors…as providing information about what one can do with an object' (p. 394). The cognitive process of learning about objects and their function through gestures and activity schemas is very different to the process of tagging objects with symbolic gestural labels.

Indeed, the existence of this default interpretation of gestures via the construction of activity schemas should not be surprising when viewed in developmental and evolutionary terms. First, it helps on a basic level of object categorization, in terms of segregating items into different kinds of tools, raw materials, and the like: 'Attending to the action of objects is an important means by which children understand objects and their categorization…' (Marentette and Nicoladis 2011, p. 395).

Second, much has been made of increasing manual skill and tool use as a driver for hominid cognitive evolution (for particular relevance to language see e.g. Iriki & Taoka, 2012; Stout & Chaminade, 2012; Stout, 2011). The ability to make and use tools in particular sequences, and importantly to teach and learn these skills, requires a complex set of motor, social and planning abilities. The ability to learn from gestures via the construction of activity schemas would seem to play a crucial role in these activities, and so be of significant evolutionary importance. Indeed, it is these existing uses of gesture that are supposed to drive the bootstrapping process from iconic gestures to symbolic understanding.

However, the problem for gesture-first hypotheses, as discussed below, is that given just how default this processing is, it might actually provide a significant developmental and evolutionary hurdle to get over, rather than play a positive role in the process of bootstrapping symbolic capacities. That is, existing ways of interpreting gestures (e.g. via activity schemas) seem likely to prohibit the interpretation of gestures as symbols, rather than make it easier.

6. Symbols

The research outlined here comes from DeLoach's work on symbolic understanding in children and her notion of dual-representation (e.g. Deloach 2002, 2004), and related work on the use of symbols in the reverse contingency task in primates and children (Apperly & Carroll, 2009; Boysen & Berntson 1995; Boysen et al., 1996; Carlson et al. 2005). This research suggests that the easiest kinds of symbols to use are those that are not obviously related to the referents in particular ways. In particular, it suggests that the kind of action-based iconic gestures supposed to play a crucial role in language evolution are particularly hard to treat as symbols.

6.1 Dual-Representation

Deloach's work focuses on identifying the factors that make it easier or harder for children to treat something as a representation or symbol. Experimentally, the situation is usually a variant on the following. A child is shown a scale model of a room, and familiarised with the similarities between the scale model and the room it represents. Something is then hidden in the scale model and the children have to search for it in the room. Despite the apparent simplicity of this task, it is only reliably accomplished by children around 3 years old. Children younger than this have sufficient spatial skills, and the ability to identify correspondences between models and objects (Troseth et al., 2007), but fail at transferring the knowledge of the scale model to the room it represents. Some variants aid performance, such as increasing the similarity between the model and the room, giving exhaustive instructions, practice on easier tasks, while taking these away make the task almost impossible (for summary see e.g. Deloach 2002, 2004). Two variants are particularly interesting; hiding the model behind a pane of glass and using only a 2-D picture of the model also aid performance.

DeLoach's theory as to what explains this variety in task performance, and the relatively late-emerging ability to perform the task, is that children have to achieve 'dual representation'. Children are clearly able to represent the scale models as interesting objects in their own right, as things that they want to interact and play with. But in order to perform well on the task they must also represent the scale model as a symbol; as standing in for something else. Children therefore have to represent the same thing in two different ways, and this is what they find challenging.

In particular, when the salience of the scale model is increased, making it a really interesting object in own right, this seems to make it more difficult to treat it as anything else (e.g. a symbol). However, when the salience of the scale model is decreased, by putting it behind a pane of glass or showing only a 2D photograph of it, or otherwise making it impossible for children to interact with it, this makes dual representation easier, and so makes the task possible. As Deloach (2002) states: 'To achieve dual representation in the first place, a child has to inhibit responding to a symbolic artifact exclusively or primarily as an object' (p. 330).

6.2 Distancing

Relatedly, experimental work using the reverse-contingency task suggests that some kinds of similarity between a sign/symbol and its referent can make it harder to treat it as a symbol. This builds on the concept of 'psychological distancing' (Sigel, 1970)[6]. This refers to the ability of an individual can detach themselves or their behaviour from the immediate context, for example when they can inhibit routine behaviours that are elicited by a particular stimulus. Symbols provide one way of generating psychological distance between an individual and a stimulus by essentially replacing the stimulus with something behaviourally neutral. The factors relevant to generating psychological distance, and so functional symbols, are reviewed below.

The reverse-contingency task is used to test the ability of an individual to inhibit impulsive behavior, and sometimes extended to test how symbol use can help with this. The task is fairly simple; participants are presented with two arrays, one with a small number of valued items (e.g. sweets), and one with a larger number of the same items. The participant has to point to the array that they don't want in order to get the one they do want, so if they want the larger group of sweets, they must point to the smaller group.

[6] There are several related ideas in the literature, e.g. Bickerton's (2009) notion of displacement, Sterelny's (2012a, 2012b) discussion of the importance of offline processing, the relationship between symbol use and mental time travel, and so on.

The initial experiments that are relevant here were carried out by Sally Boysen and colleagues on chimps (Boysen et al., 1996; Boysen & Berntson, 1995). Unsurprisingly, given chimps' general lack of self-control, they failed at this task over hundreds of trials. However, when the arrays were replaced by Arabic number symbols (which the chimps had previously been trained with), chimps could suddenly successfully complete the task, apparently by creating 'psychological distance' that allowed the chimps to inhibit their impulsive response:

'Upon introduction of Arabic symbols, performance increased immediately to more optimal levels. This suggests that the animals had in fact acquired implicit knowledge of the rule structure of the task, despite the fact that they were unable to implement this knowledge with candy arrays as stimuli. Indeed, over counterbalanced sessions with symbols and candy arrays, performance shifted immediately from significantly above chance with Arabic symbols to significantly below chance with candy arrays.' (Boysen et al. 1996, p. 84)

Similar results have since been found across other primates, including human children (Addessi & Rossi, 2011; Albiach-Serrano et al., 2007; Apperly & Carroll, 2009; Carlson et al., 2005; Evans et al., 2012; Kralik et al., 2002; Murray et al., 2005; Vlamings et al., 2006). Again, participants tend to perform badly on the task unless some modification of the task is made (e.g. only using a single set of quantities, like 1 vs. 4 items), or a symbol is used to represent the desired items (sometimes numerals, sometimes tokens etc.), and then performance greatly improves. In particular, more nuanced versions of this paradigm are useful in establishing just what kinds of similarity can obtain between the ‘symbols’ and the valued items they represent, such that participants can use them flexibly and successfully as symbols, and so succeed at the task.

For example, Boysen and Yocom (2012) found that chimps still performed badly at the task when the sweets were replaced with equal numbers of small rocks. For chimps at least, the similarity in quantity prevented them from inhibiting their standard behavioural response. This effect was replicated by Addessi and Rossi (2011), where capuchins failed the reverse-contingency task using ‘low-symbolic distance’ tokens (tokens that corresponded one-to-one with food rewards), but succeeded using ‘high-symbolic distance’ tokens (where each token represented a different number).

Two of the most detailed studies of the factors relevant to achieving ‘psychological distance’ with symbols are Apperly and Carroll (2009) and Carlson et al. (2005), with 3-4 year old children. Carlson et al. (2005) again found that performance on the reverse-contingency task differed with different symbols. Rocks (as above) were least effective, followed by dots, and the only significant improvement in performance was found when pictures of animals were used as symbols (a mouse for lower number of rewards, and an elephant for the larger number). Their explanation for these results tracks that offered above: ‘We propose that abstract symbols produc[e] psychological distance that enables [individuals] to withhold a dominant response’ (Carlson et al. 2005, p. 610).

Apperly and Carroll (2009) added to these conditions. Here, the rewards consisted of stickers, and the 'symbols' were numerals, number words (which many children could not read), dots, photographs of the stickers, and sweets (themselves obviously desirable). These were used to test whether the psychological distance offered by the symbols depended on one of more of the following factors: intrinsic vs. goal-specific desirability, and (roughly) childrens' familiarity of the relationship between the symbol and what it represents.

They found that performance was significantly improved in all 'symbol' conditions, compared to the sticker-only trials, and (differently to the original chimp case) that performance improved over trials. Interesting, this occurred for both the sweets-as-symbols trials, and for the photograph trials. This suggests that what works well as a symbol may depend on the goal of the task, and not on intrinsic desirability (e.g. sweets), and that high visual similarity between referent and symbol (here via photographs) does not necessarily have an effect.[7] Familiarity with the symbol also failed to be a significant factor, as both numerals (high familiarity) and number words (low familiarity – many could not read them) both aided performance.

In addition, when some symbol conditions, particularly number words, were immediately followed by the original, non-symbolic stickers-only condition, (the one that non-human primates and children regularly fail at), high levels of performance transferred across. That is, having formed and practiced the right strategy in the symbol condition (point to the symbol that represents the lower quantity of stickers),

[7] It is important to note that the children in the study recognized photographs as representational items. In this case, it may be that visual similarity is not a relevant factor in determining how useable a symbol is, but only if participants are already aware that it is delivered in an intrinsically representational medium.

children were able to immediately continue this strategy when presented with the real stickers (point to the smaller group of stickers to get the larger group). The authors note that this effect may however be rather short-lived.

Apperly and Carroll's explanation of these results again follows the idea of psychological distancing above; symbols help to block an impulsive response, and so give cognitive 'space' to formulate and follow another strategy. The immediate improvements in performance found in the symbol condition, (i.e. performance is significantly better from the first trial onwards), suggests that the formation of a new strategy may be reasonably easy to do once an impulsive or routine behaviour can be inhibited. The 'distancing' that symbols can provide can therefore have a significant effect on behaviour even in the absence of advanced cognitive capacities[8].

6.3 Gestural Symbols?

These results and associated concepts stand to pose a general constraint on the emergence of symbolic language. First, achieving dual representation, so treating something as both an object/thing in its own right and as a symbol, is easier with things that lack physical salience and do not invite interactions. Second, 'psychological distance' between a symbol and a referent is easier to achieve when the symbol does not elicit an impulsive behaviour. What both of these accounts have in common then is that something can be used as a symbol more easily when it does not in itself elicit an impulsive, typical, or otherwise regularised cognitive response.

[8] Symbol use also aids performance on delayed-gratification tasks for children, chimpanzees and capuchins, but again seemingly only when symbols do not track quantity of the reward in a one-to-one manner (Evans et al., 2012).

According to these accounts then, the kinds of things that are *least* likely to form easy-to-use symbols are therefore things that already played important roles in the lives of early hominids. These obviously include things that trigger cognitively entrenched processes that are crucial for an individual’s development and a community’s survival. In particular, iconic gestures that pantomime intentional actions are unlikely to fare well as symbols. This is not to deny that action-based gestures cannot be used communicatively – clearly they can. It is also likely that action-based gestures and pointing, along with elicited responses, can take you fairly far in communicative contexts set in the here and now. The difficulty arises in situations where one needs to refer to things out of sight, or in the past or future; here one needs symbols. And this is where the idea of being able to inhibit routine reactions becomes particularly important.

To illustrate: there are a few uses of gestures deemed to be crucial in the evolution of cognition and language, but all depend on eliciting a typical (automatic or routinised) response in the here and now. For example, the existing gestural communication of non-human primates consists of attempts to elicit a reactive behaviour from a conspecific in the here and now in order to complete a standard interaction (e.g. a play movement). Second, teaching manual skills using gesture also depends on eliciting reactive behaviours, such as imitation, in the here and now. Imitation plays a central role in social learning (for review see e.g. Rendell et al., 2011), and it appears to be a relatively automatic response for human children. Third, action-based (iconic) gestures are crucial in learning about intentional actions, about objects and what one can do with them, and about manual skills and tool use, mainly via the default formation of activity schemas.

Symbol use is necessary to go beyond these automatic or regularised ways of communicating set in the here and now. In particular, symbols demand a kind of 'distancing' that enables more flexible responses to be generated. Gestures that already elicit a range of cognitively default responses are precisely the sort of thing that hominid novice symbol users are likely to have found difficult to treat as symbols; they allow no 'space' for generating alternative or flexible responses. In this case, instead of existing cognitive capacities like action understanding and imitation providing a way of bootstrapping symbolic understanding via iconic gestures, they may in fact make it harder for symbol use to emerge in the gestural channel.

It is important to note here that this is a relative claim though; it is not a claim that utilising action-based iconic gestures as symbols would have been cognitively impossible. Chimps and 3-4 year old children have less executive control and inhibitory capacity than adult hominids living in co-operative and reasonably technological advanced communities. The claim is instead that turning action-based iconic gestures, that already played important roles in hominid communities, into symbols, is a great deal (cognitively) harder than usually assumed. If one is looking for the easiest path towards developing symbol use, then co-opting action-based iconic gestures is not it.

7. Gestures, Symbols and Associative Learning

This next section identifies a very different set of general constraints on language acquisition, by looking to exclusion-based associative learning (fast-mapping) again.

Before doing so, it is important to note that exclusion-based associative learning is not unique to contemporary humans, and so is relevant in this context. Exclusion-based learning and reasoning has been found in other primates (Beran, 2010; Call, 2006; Lyn & Savage-Rumbaugh, 2000) as well as dogs (Aust et al., 2008; Kaminski et al., 2004; Pilley & Reid, 2011), so can be treated as a domain general learning mechanism available to early hominids. The evidence reviewed below suggests that learning mappings between symbols and referents is likely to be easier when the symbols are in the vocal/auditory channel, rather than gestural/visual channel. This, along with a related advantage of vocal communication, provides corroborating evidence to the claim above.

7.1 Single vs. Multi-Modal Learning

Puccini and Liszkowski (2012) used a similar set-up to Marentette and Nicoladis (2011) above to investigate how well 15 month old infants fast-map symbols in the vocal and gestural channel separately, and both together. The infants watched films of explicit naming events, where objects were either labeled with a word, a gesture, or a word plus gesture. Eye-trackers were used to tell whether the infants had successfully learned the labels, by seeing which object they attended to when addressed with 'Hello, where is the [label]?'. The only relationships that were fast-mapped above chance were those between words and objects. In this study, infants failed to fast-map gestures and objects, and also failed in conditions where both words and gestures together were used to label an object.

The authors note that infants this young are able to learn multi-modally using words plus deictic gestures (pointing), but suggest that it is representational gestures (such as pantomime) that they have a problem with. The explanation here is a simple one. When presented only with sounds and objects, infants can direct all their visual attention to the object, which facilitates mapping the object to its vocal/auditory label. When pointing is used as well, the points direct visual attention to the relevant object, again facilitating fast-mapping. However, '[r]epresentational gestures, by the very nature of the visual modality, require that infants divide their visual attention between the referent and the gesture' (Puccini and Liszkowski, 2012, p. 6). Not only are infants in the multi-model presentation required to make a three-way association (word+gesture+object), but their visual attention is split in two. Even in just the gestural presentation, visual attention remains split across the gesture and the object, in a way quite unlike the word-only presentation. Word learning may therefore be easier in the vocal channel than the gestural channel, purely in terms of lessening the demands on visual attention and facilitating associative learning.

7.2 Associative Learning and Language Acquisition

There is another more general body of work that supports the idea that sustained visual attention is crucial to word learning. Smith and colleagues have used head-mounted cameras to investigate visual attention and its relation to successful world learning in infants (for review see Smith, 2013). Over many studies they have found that despite the fact that toddler visual experiences are far more dynamic and changing than adult worldviews, there are periods where single objects are attended, and these are the periods in which successful word learning tends to occur. When objects are visually larger, centered in the visual field, and gazed at for a longer period, word learning is far more likely (Pereira et al., 2014; Yu & Smith, 2012).

These periods of sustained selective visual attention on objects drastically reduce referential ambiguity in naming events, essentially by making sure that only one object is a contender as a possible referent. Importantly, these processes generate ideal periods of learning ‘for free’. As Periera et al. (2014) note, this work shows that when visual attention is directed towards a single object:

‘…there are very clean sensory moments when no additional cognitive processes would seem to be needed to determine the relevant object… [Therefore] the dynamic visual properties of naming events associated with learning versus not learning the object name also suggest that there are visual limits on object name learning.’ (Periera et al., 2014, pp. 183-184)

When symbols for visually presented objects are presented in the vocal/auditory channel, there is no need for mechanisms of attention switching, symbols and objects can be presented simultaneously, and so there are periods of time when associative learning can proceed fairly easily with no need to invoke complex cognitive processes. In contrast, when learning visually presented symbols for visually presented objects, learners must switch visual attention between symbol and object, which demands a greater amount of sustained selective attention.

It is important to note here that contemporary deaf signers clearly manage to fast-map gestures to objects, perhaps by developing greater selective attention or making use of socially scaffolded ways of learning, and that their developmental trajectory is the same as that of hearing infants. Again, the claim is a relative one; it is harder, but not impossible to learn symbol-object mappings when symbols are presented in the gestural/visual channel. But the fact that symbols delivered in the vocal/auditory domain come with easy learning episodes for free is important. It means that there is a basic modality advantage to vocal/auditory symbols over gestural/visual symbols, just in terms of providing a simple way of resolving ambiguity in input in order for associative learning to proceed.

7.3 More on Vision

Finally, a brief note on the advantages and disadvantages of the gestural and vocal modalities for communication. One question that faces gesture-first hypotheses is how and why languages shifted from the gestural to the vocal modality. Presumably, given the assumed poor evolutionary platform that non-human primate vocalisations provide, and the costs associated with changes in vocal physiology, there must have been massive selective pressures on expanding the vocal platform, given the apparent disadvantages of gestural language. However, the explanations for this are not always very satisfactory. Languages are used by individuals in a range of environments, and different modalities are better or worse for different uses. Gesture seems to win out in cases where it pays to be quiet (say in hunting). The usual ways in which vocal language is said to offer advantages is in situations where you can’t see (either in the dark, or in dense forests etc.), and for freeing up your hands (e.g. using tools, in food preparation). So, it is unclear whether switching the delivery of semantic content almost entirely from the gestural to the vocal channel can be sufficiently motivated, especially if the vocal platform is as poor as is often claimed.

Yet the vocal channel may have another advantage that is not usually raised, but is consistent with the research on visual attention above. Further, this advantage may be strong enough that it contributes to the claim that the transition from gestural to vocal (linguistic) communication happened at an early, rather than late, stage of language evolution. This is that vocal/auditory languages also free up the eyes.

Speakers of vocal/auditory languages often do look at conversational partners, but do not always do this, and do not have to in order to communicate. However, observing and following signed group conversations is surprisingly exhausting, at least for non-native speakers[9]. Unlike in vocal conversations, all visual attention must be focused on the speaker/signer at all times. It is very difficult to do anything else that requires vision at the same time.

As evidence for this, eye-tracking studies show that sign language users tend to fixate on the face of speakers/signers, both because facial movements convey a variety of types of information (emotional, lexical) and because it forms a central base for fixation. Peripheral vision is then used to process hand and body movements (Agrafiotis et al., 2003; Emmorey et al., 2009; Muir & Richardson, 2005). Muir and Richardson (2005) found that when distracting objects were present behind a speaker/signer, the percentage of time signers spent looking at them varied between 0-2%, with most signers never looking at them. Across several videos, the total proportion of time participants looked away from the speaker/signer to an uninteresting background was 0%. The implication is that in natural signed conversations, the total time that signers spend looking away from speakers/signers' faces and bodies is vanishingly small.

[9] Even fairly fluent non-native signers report that their eyes get tired well before their hands do.

This is hardly surprising since all linguistic information in sign languages is present in the visual modality, in a fairly small region of visual space, and (at least in contemporary sign languages) presented very quickly. Yet the basic fact that the direction of gaze during a signed conversation has to remain fairly static places strong constraints on the kinds of situations that it is easy to communicate in while using gesture. Vocal languages make it easy to communicate while on the move, while tracking or looking for things in the environment, while making or manipulating things, and anything else where vision is used. In this case, the constraints that gestural languages place on what you can look at while communicating (not very much) may be another significant reason to utilise the vocal/auditory modality fairly early on.

8. Adding it All Up

A number of constraints have been identified above based on research on symbolic capacities in particular, which suggest that the vocal, rather than that the gestural channel, is more likely to support the emergence of symbolic language. These are summarised below.

First, developmental research suggests that there is a cognitively default interpretation of gestures. This leads to the construction of activity schemas, where objects are associated with actions. These activity schemas can be used in gestural communication, and seem likely to have played an important role in hominid evolution, particularly in the teaching and learning of manual skill and tool use. Yet these gestures are treated neither as symbolic, nor (for younger children) as obviously iconic, even by reasonably competent symbol users (4 year old children). Importantly, this default interpretation of gestures applies not only to iconic gestures (e.g. pantomimes of actions), but to gestures more generally, so may be a cognitive response to any sign made in this modality.

Second, the concepts of dual representation and psychological distance explain further why gestures may be difficult to treat as symbolic. According to Deloach's theory of dual representation, treating something as a symbol requires representing it in two ways; first, as an object in its own right, and second, as a symbol or stand in for something else. Achieving dual representation is made far easier when the salience of the object is decreased, thus enabling an individual to inhibit their typical responses to it as an object. Relatedly, the concept of psychological distance can be used to explore the features of a successful symbol. Carlson et al. (2005) and Apperly and Carroll (2009) found that stimuli that do not strongly elicit a dominant response towards a desired object are easier to use as symbols. These can include stimuli that are not relevant to the current goal (but perhaps still desirable), images of the desired object (so long as images are treated as representational), existing symbols (numerals) and new or unfamiliar stimuli (images of other stimuli, number words). These kind of stimuli do not strongly elicit a (goal relevant) response, and allow individuals to respond to them in more flexible ways; here as symbols for desired objects.

In general then, a constraint on successful symbol use in both human and non-human primates, is for the symbol to be a stimulus that does not itself elicit an impulsive, routine, or otherwise dominant response. Yet the evidence above on the default interpretations of gestures as part of activity schemas, research on imitation and its role in learning, and the nature of primate gestural communication, strongly suggests that gestures were already being interpreted and responded to in default and routinized ways by hominids. From the point of view of symbolic understanding then, gesture does not present the best modality for symbol use to emerge in.

It is also worth reiterating how inhibition is related to symbol use. This is that non-symbolic communication, including gestural communication, is likely to be sufficient for a wide range of communicative needs in the here and how. And, as above, much of this communication may take the form of routinized behavioural interactions, or rely on dominant or automatic types of cognitive processing. Symbols are needed to expand the range of communication from these routines to something more flexible. The specific cognitive processes that serve non-symbolic gestural communication do not therefore offer a natural platform for the emergence of symbolic language; they are instead the kinds of processes that require a significant amount of inhibition in order for symbol use to emerge.

Finally, there are specific properties of the vocal modality that illustrate further (though perhaps smaller) advantages over the gestural modality. First, using associative learning to map symbols to objects is cognitively easier when symbols are presented in the vocal channel, rather than the gestural channel. The reasoning is simple: mapping a sound to a visually presented object requires much less control of selective visual attention than mapping two visually presented stimuli to each other. This is potentially important as symbol mappings were (presumably) initially learned slowly and laboriously, and in the near absence of explicit teaching or social scaffolding. In this case, any slight cognitive advantage in the learning process, here afforded by the vocal/auditory modality, may make a big difference. Relatedly, vocal languages not only free up the hands, but they also free up the eyes. Given that visual resources would have been essential in many hominid activities, there may be a significant cost associated with decreasing their availability, as is the case when using gestural language.

9. Conclusion

The aim here has been to challenge gesture-first theories of language evolution by identifying a new set of constraints on the emergence of symbolic language, based on the properties of symbolic processing. These constraints are obviously one set of many that need to be taken into account in evaluating evolutionary hypotheses, but the evidence reviewed above hopefully shows that they are worthy of being included in discussion. In particular, these constraints challenge the assumption often found in gesture-first hypotheses that existing capacities for generating and interpreting gestures could have been easily co-opted for symbolic use; the opposite has been argued to be true. In addition to this problem, the advantages of the vocal channel for learning and communication lend support to the idea that the vocal modality is not such a poor platform for the evolution of symbolic language afterall.

School of Philosophy

Australian National University

Department of Philosophy

University of Cardiff

References

Ackermann, H., Hage, S. R., & Ziegler, W. 2014: Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behavioral and Brain Sciences*, 1–84.

Addessi, E., & Rossi, S. 2011: Tokens improve capuchin performance in the reverse–reward contingency task. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1707), 849–854.

Agrafiotis, D., Canagarajah, N., Bull, D. R., & Dye, M. 2003: Perceptually optimised sign language video coding based on eye tracking analysis. *Electronics Letters*, *39*(24), 1703–1705.

Albiach-Serrano, A., Guillén-Salazar, F., & Call, J. 2007: Mangabeys (*Cercocebus torquatus lunulatus*) solve the reverse contingency task without a modified procedure. *Animal Cognition*, *10*(4), 387–396.

Apperly, I. A., & Carroll, D. J. 2009: How do symbols affect 3-to 4-year-olds' executive function? Evidence from a reverse-contingency task. *Developmental Science*, *12*(6), 1070–1082.

Arbib, M. A. 2005: From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*(02), 105–124.

Arbib, M. A., Liebal, K., & Pika, S. 2008: Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076.

Aust, U., Range, F., Steurer, M., & Huber, L. 2008: Inferential reasoning by exclusion in pigeons, dogs, and humans. *Animal Cognition*, *11*(4), 587–597.

Baillargeon, R., Scott, R. M., & He, Z. 2010: False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Beran, M. J. 2010: Use of exclusion by a chimpanzee (*Pan troglodytes*) during speech perception and auditory–visual matching-to-sample. *Behavioural Processes*, *83*(3), 287–291.

Bickerton, D. 2009: *Adam's tongue: How humans made language, how language made humans*. Macmillan.

Boysen, S. T., & Berntson, G. G. 1995: Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *21*(1), 82-86.

Boysen, S. T., Berntson, G. G., Hannan, M. B., & Cacioppo, J. T. 1996: Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *22*(1), 76-86.

Boysen, S. T., & Yocom, A. M. 2012: Sensitivity to Quantity: What Counts across Species? *The Complex Mind: An Interdisciplinary Approach*, 80-96.

Call, J. 2006: Inferences by exclusion in the great apes: the effect of age and species. *Animal Cognition*, *9*(4), 393–403.

Carlson, S. M., Davis, A. C., & Leach, J. G. 2005: Less Is More Executive Function and Symbolic Representation in Preschool Children. *Psychological Science*, *16*(8), 609–616.

Clark, N., & Perlman, M. 2014: Breath, vocal, and supralaryngeal flexibility in a human-reared gorilla. In *The Evolution of Language: Proceedings of the 10th International Conference*. Singapore: World Scientific.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. 2014: Mirror neurons: from origin to function. *Behavioral and Brain Sciences*, *37*(02), 177–192.

Corballis, M. C. 2003: From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, *26*(02), 199–208.

Corballis, M. C. 2009: The evolution of language. *Annals of the New York Academy of Sciences*, *1156*(1), 19–43.

Deacon, T. W. 1998: *The symbolic species: The co-evolution of language and the brain*. WW Norton & Company.

DeLoache, J. S. (2002). Early development of the understanding and use of symbolic artifacts. In U. Goswani (ed) *Blackwell Handbook of Childhood Cognitive Development*, Malden: Blackwell Publishing, pp. 206-226.

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

Donald, M. 2001: *A mind so rare: The evolution of human consciousness*. WW Norton & Company.

Dunbar, R. 1998: *Grooming, gossip, and the evolution of language*. Harvard University

Emmorey, K., Thompson, R., & Colvin, R. 2009: Eye gaze during comprehension of American Sign Language by native and beginning signers. *Journal of Deaf Studies and Deaf Education*, *14*(2), 237–243.

Evans, T. A., Beran, M. J., Paglieri, F., & Addessi, E. 2012: Delaying gratification for food and tokens in capuchin monkeys (Cebus apella) and chimpanzees (Pan troglodytes): When quantity is salient, symbolic stimuli do not improve performance. *Animal Cognition*, *15*(4), 539–548.

Falk, D. 2009: *Finding our tongues: Mothers, infants and the origins of language*. Basic Books.

Fay, N., Arbib, M., & Garrod, S. 2013: How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356–1367.

Fitch, W. 2000: The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267.

Fitch, W. 2005: The evolution of language: a comparative review. *Biology and Philosophy*, *20*(2-3), 193–203.

Fitch, W., Zuberbühler, K., Zimmerman, E., Schmidt, S., & Altenmüller, E. 2013: Primate precursors to human language: beyond discontinuity. *The Evolution of Emotional Communication*, 26–48.

Gasser, M. 2004: The origins of arbitrariness in language. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 434–439).

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. 2009: Gestural communication of the gorilla (Gorilla gorilla): repertoire, intentionality and possible origins. *Animal Cognition*, *12*(3), 527–546.

Goldin-Meadow, S., & Alibali, M. W. 2013: Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257-283.

Halina, M., Rossano, F., & Tomasello, M. 2013: The ontogenetic ritualization of bonobo gestures. *Animal Cognition*, *16*(4), 653–666.

Heyes, C. 2010a: Mesmerising mirror neurons. *Neuroimage*, *51*(2), 789–791.

Heyes, C. 2010b: Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, *34*(4), 575–583.

Heyes, C. 2014a: False belief in infancy: a fresh look. *Developmental Science, 17*(5), 647-659.

Heyes, C. 2014b: Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*(2), 131–143.

Hobaiter, C., & Byrne, R. W. 2011: The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*(5), 745–767.

Hurford, J. R. 2007: *The origins of meaning: Language in the light of evolution*. Oxford University Press.

Iriki, A., & Taoka, M. 2012: Triadic (ecological, neural, cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 10–23.

Kaminski, J., Call, J., & Fischer, J. 2004: Word learning in a domestic dog: evidence for "fast mapping". *Science*, *304*(5677), 1682–1683.

Kralik, J. D., Hauser, M. D., & Zimlicki, R. 2002: The relationship between problem solving and inhibitory control: Cotton-top tamarins (*Saguinus oedipus*) performance on a reversed contingency task. *Journal of Comparative Psychology*, *116*(1), 39-50.

Liebal, K., & Call, J. 2012: The origins of non-human primates' manual gestures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 118–128.

Liebal, K., Call, J., & Tomasello, M. 2004: Use of gesture sequences in chimpanzees. *American Journal of Primatology*, *64*(4), 377–396.

Lyn, H., & Savage-Rumbaugh, E. S. 2000: Observational word learning in two bonobos (*Pan paniscus*): ostensive and non-ostensive contexts. *Language & Communication*, *20*(3), 255–273.

Marentette, P., & Nicoladis, E. 2011: Preschoolers' interpretations of gesture: Label or action associate? *Cognition*, *121*(3), 386–399.

Meier, R. P. 1987: Elicited imitation of verb agreement in American Sign Language: Iconically or morphologically determined? *Journal of Memory and Language*, *26*(3), 362–376.

Meier, R. P., Mauk, C. E., Cheek, A., & Moreland, C. J. 2008: The form of children's early signs: Iconic or motoric determinants? *Language Learning and Development*, *4*(1), 63–98.

Mithen, S. J. 2005: *The singing Neanderthals: The origins of music, language, mind, and body*. Harvard University Press.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. 2014: How arbitrary is language. *Philosophical Transactions of the Royal Society B 19*(369), DOI: 10.1098/rstb.2013.0299.

Muir, L. J., & Richardson, I. E. 2005: Perception of sign language and its application to visual communications for deaf people. *Journal of Deaf Studies and Deaf Education*, *10*(4), 390–401.

Murray, E. A., Kralik, J. D., & Wise, S. P. 2005: Learning to inhibit prepotent responses: successful performance by rhesus macaques, *Macaca mulatta*, on the reversed-contingency task. *Animal Behaviour*, *69*(4), 991–998.

Namy, L. L. 2001: What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy*, *2*(1), 73–86.

Namy, L. L. 2008: Recognition of iconicity doesn't come for free. *Developmental Science*, *11*(6), 841–846.

Namy, L. L., Campbell, A. L., & Tomasello, M. 2004: The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, *5*(1), 37–57.

Nicoladis, E., Pika, S., & Marentette, P. 2010: Are number gestures easier than number words for preschoolers? *Cognitive Development*, *25*(3), 247–261.

Orlansky, M. D., & Bonvillian, J. D. 1984: The role of iconicity in early sign language acquisition. *Journal of Speech and Hearing Disorders*, *49*(3), 287–292.

Ozcaliskan, S., Gentner, D., & Goldin-Meadow, S. 2013: Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics, 35*(6), 1143-1162,

Pereira, A. F., Smith, L. B., & Yu, C. 2014: A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185.

Perniss, P., Thompson, R., & Vigliocco, G. 2010: Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227, doi: 10.3389/fpsyg.2010.00227.

Pika, S., Liebal, K., Call, J., & Tomasello, M. 2005: Gestural communication of apes. *Gesture*, *5*(1), 41-56.

Pilley, J. W., & Reid, A. K. 2011: Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*(2), 184–195.

Pollick, A. S., Gouzoules, H., & de Waal, F. 2005: Audience effects on food calls in captive brown capuchin monkeys, *Cebus apella*. *Animal Behaviour*, *70*(6), 1273–1281.

Puccini, D., & Liszkowski, U. 2012: 15-month-old infants fast map words but not representational gestures of multimodal labels. *Frontiers in Psychology*, *3*, 101, doi: 10.3389/fpsyg.2012.00101

Rendell, L., Fogarty, L., Hoppitt, W. J., Morgan, T. J., Webster, M. M., & Laland, K. N. 2011: Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences*, *15*(2), 68–76.

Roberts, A. I., Roberts, S. G. B., & Vick, S.-J. 2013: The repertoire and intentionality of gestural communication in wild chimpanzees. *Animal Cognition*, *17*(2), 317-336.

Salmi, R., Hammerschmidt, K., & Doran-Sheehy, D. M. 2013: Western Gorilla Vocal Repertoire and Contextual Use of Vocalizations. *Ethology*, *119*(10), 831–847.

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. 2013: Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. *PloS One*, *8*(10), e76674.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. 2011: Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*(7), 878–880.

Seyfarth, R. M., & Cheney, D. L. 2010: Primate vocal communication. In M. L. Platt and A. A. Shazanfar (eds.) *Primate Neuroethology*, Oxford University Press, pp. 84–97.

Sigel, I. E. (1970: The distancing hypothesis: a causal hypothesis for the acquisition of representational thought. In M. R. Jones (Ed.), *Miami Symposium on the prediction of behaviour, 1968: Effect of early experiences* (pp. 99–118). Coral Gables, FL: University of Miami Press.

Slocombe, K. E., Kaller, T., Turman, L., Townsend, S. W., Papworth, S., Squibbs, P., & Zuberbühler, K. 2010: Production of food-associated calls in wild male chimpanzees is dependent on the composition of the audience. *Behavioral Ecology and Sociobiology*, *64*(12), 1959–1966.

Smith, L. B. 2013: It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist*, *68*(8), 618-629.

Smith, L., & Yu, C. 2008: Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Sterelny, K. 2012a: Language, gesture, skill: the co-evolutionary foundations of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2141–2151.

Sterelny, K. 2012b: *The evolved apprentice: How evolution made humans unique (Jean Nicod Lectures)*. Cambridge: MIT Press.

Stout, D. 2011: Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1050–1059.

Stout, D., & Chaminade, T. 2012: Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 75–87.

Striano, T., Rochat, P., & Legerstee, M. 2003: The role of modelling and request type on symbolic comprehension of objects and gestures in young children. *Journal of Child Language*, *30*(1), 27–45.

Taub, S. F. (2000: Iconicity in American Sign Language: concrete and metaphorical applications. *Spatial Cognition and Computation*, *2*(1), 31–50.

Tolar, T. D., Lederberg, A. R., Gokhale, S., & Tomasello, M. 2008: The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education*, *13*(2), 225–240.

Tomasello, M. 2008: *Origins of human communication*. Cambridge MA: MIT Press.

Tomasello, M., Striano, T., & Rochat, P. 1999: Do young children use objects as symbols? *British Journal of Developmental Psychology*, *17*(4), 563–584.

Troseth, G. L., Bloom Pickard, M. E., & DeLoache, J. S. 2007: Young children's use of scale models: testing an alternative to representational insight. *Developmental Science*, *10*(6), 763–769.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. 2013: Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Vermeerbergen, M. 2006: Past and current trends in sign language research. *Language & Communication*, *26*(2), 168–192.

Vlamings, P. H., Uher, J., & Call, J. 2006: How the great apes (*Pan troglodytes, Pongo pygmaeus, Pan paniscus*, and *Gorilla gorilla*) perform on the reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 60-70.

Yu, C., & Smith, L. B. 2012: Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262.