# Language, gesture, skill: the co-evolutionary foundations of language

Kim Sterelny

| | |
|---|---|
| **References** | **This article cites 42 articles, 12 of which can be accessed free**<br>http://rstb.royalsocietypublishing.org/content/367/1599/2141.full.html#ref-list-1<br><br>**Article cited in:**<br>http://rstb.royalsocietypublishing.org/content/367/1599/2141.full.html#related-urls |
| **Subject collections** | Articles on similar topics can be found in the following collections<br><br>behaviour (367 articles)<br>cognition (237 articles)<br>evolution (508 articles) |
| **Email alerting service** | Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here** |

To subscribe to *Phil. Trans. R. Soc. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

*Research*

# Language, gesture, skill: the co-evolutionary foundations of language

## Kim Sterelny[1,2,*]

[1]*Department of Philosophy and Tempo and Mode, Australian National University, Australia, and*
[2]*Philosophy Program, Victoria University of Wellington, Wellington, New Zealand*

This paper defends a gestural origins hypothesis about the evolution of enhanced communication and language in the hominin lineage. The paper shows that we can develop an incremental model of language evolution on that hypothesis, but not if we suppose that language originated in an expansion of great ape vocalization. On the basis of the gestural origins hypothesis, the paper then advances solutions to four classic problems about the evolution of language: (i) why did language evolve only in the hominin lineage? (ii) why is language use an evolutionarily stable form of informational cooperation, despite the fact that hominins have diverging evolutionary interests? (iii) how did stimulus independent symbols emerge? (iv) what were the origins of complex, syntactically organized symbols? The paper concludes by confronting two challenges: those of testability and of explaining the gesture-to-speech transition; crucial issues for any gestural origins hypothesis

**Keywords:** evolution of language; gesture; gestural origins; hominin cooperation; cooperation and communication

## 1. LANGUAGE AND GESTURE

The evolution of language is a remarkably active field (boasting, for example, a dedicated Oxford University Press Monograph Series), but one in which there is extraordinarily little consensus. There is no consensus on the explanatory target: are we trying to explain the origins and stability of a form of cooperative social behaviour, or are we trying to explain the evolutionary construction of a distinctive cognitive mechanism? There are widely divergent views on the crucial adaptation—the crucial evolutionary problem that needed to be solved—before language could emerge. Perhaps the most common view is that language differs from more rudimentary prototypes by having recursive, combinatorial syntax, allowing a finite stock of discrete elements to be assembled into an indefinitely large number of possible messages [1,2]. But there are important alternative views. Dessalles [3] thinks the crucial problem is to explain how, in a competitive world, an honest communication system can use cheap signals. Deacon [4,5] thinks the crucial adaptive breakthrough is the shift from signs that covary with their referent (so sign meanings can be learned by association) to genuine symbols; Bickerton [6] has recently defended a related idea. Date estimates of the origin of language differ widely: some regard 'full' language as a trait unique to our species (or even late members of our species [7]); others place it much deeper in time, dating back at least to the

last common ancestor of modern humans and Neanderthals [8–10].

There is somewhat more agreement on the benchmarks a successful theory of language needs to meet. For example, Odling-Smee & Laland [11] suggest that any serious hypothesis:

— should explain the honesty of early language;
— should depend only on well-understood evolutionary mechanisms;
— should explain the distinctive scope and expressive power of human language;
— should account for the uniqueness of human language; and
— should posit selective forces leading to language that are consistent with the known variability and dynamism of human environments.

Their success criteria mirror others found in the literature (because they are partially borrowed from earlier specifications). These success conditions fall short of full empirical testability, but if met, they push a hypothesis beyond mere story telling. Even so, aside from some partial consensus on what a theory of language evolution needs to explain, there is much debate and little agreement.

In this paper, I take up just one thread of this tangled discussion: the connection between language and gesture. Donald, Corballis, Arbib and Tomasello all defend the idea that language began as a system of gesture or sign; thus language is only secondarily a system of vocal communication (see [8,12–15]). I think they are right. In developing this view, I first outline the distinctive ecology of our hominin ancestors over the past 2.5 Myr, as they evolved as

*kim.sterelny@anu.edu.au

cooperative foragers targeting rich but challenging resources, a way of life that selected for collaboration and for enhanced cognitive capacities. I then turn specifically to gesture, arguing that a gesture-first model is plausible because (i) if language began with the expansion of gesture, an evolutionary transition to enhanced capacities to communicate and coordinate would recruit existing capacities, an existing cognitive platform, (ii) a gesture-first model gives a better model of the emergence of complex, structured signs, and (iii) many gestural symbols were iconic, and hence gesture-first models can more plausibly explain how symbols were learned and used by agents who lacked the full panoply of *sapiens*-grade cognitive equipment. I then move to admittedly more speculative considerations, which link a gesture-first model of language evolution to the origins of stimulus-independent symbols and to syntax.

## 2. THE ADAPTIVE PLATFORM: COOPERATIVE FORAGING

Over the past 2.5 Myr, the hominin archaeological record shows changes in morphology and life history, in brain size, diet and material culture. Hominins lived longer, showing that extrinsic mortality fell [16,17]. They had better quality diets: through some changing combination of scavenging, hunting, extracting rich plant carbohydrates and cooking [18]. Improved resources fuelled an increase in brain size, and a reduction in tooth, jaw and gut mass. Though there is no smooth upward trend, material technology became more sophisticated [19]. The earliest signs of tool-assisted access to meat are found from about 3.4 Mya; stone tools themselves are found from about 2.3 Mya; more sophisticated, Acheulian handaxes (used by larger-brained, erectus-grade hominins) are found from about 1.7 Mya [20,21]. There is clear evidence of the control of fire at about 800 kya [22,23]. Material culture begins to diversify by about the Middle Stone Age, perhaps about 300 kya [24]: by then, hominins had been large game hunters for hundreds of thousands of years [25]. Hunting large game, driving carnivores from their kills, exploiting the valuable but hard to use underground storage organs of plants: all these signal both cooperation and ecological expertise.

I have argued elsewhere that from the late australopithecines through to the very large-brained hominins ancestral to our species (and the Neanderthals), hominins evolved as social foragers. As such, they increasingly combined a reliance on extractive foraging (targeting very valuable but heavily defended resources) with the capacity to cooperate and coordinate [26–29]; for a review of some of the connections between brain evolution and the evolution of skilled foraging, see [30]. This cooperation included informational cooperation across generations. Ecological expertise of the kind needed to kill dangerous game with short-range weapons, or to find and detoxify tubers, is not acquired from scratch each generation. But informational cooperation was also needed for coordination, and eventually for planning. So selection for cooperation included selection for enhanced communication. Hominins were the only primate lineage

that evolved language because hominins were the only great apes that evolved as cooperative extractive foragers, simultaneously under selection for enhanced capacities to coordinate, and enhanced capacities to physically manipulate their environment.

This change to cooperative foraging was probably initially triggered by the increased seasonality and variability of australopithecine environments [31], selecting for both improved cooperation against predation in more exposed environments, and for improved technical capacity to exploit dry-season resources (for example, the underground storage organs of plants [32], but also scavenged bones of large carcasses). As noted above, signs of these changes are found in the record between 2 and 2.5 Mya, with the appearance of the Oldowan stone tool industry, and with the morphological signs of a changed and improved diet [18,33]. But once the change was triggered, and early hominins evolved both enhanced capacities to cooperate and enhanced capacities to manipulate their physical environment, it was driven by co-evolutionary feedback. Positive feedback loops explain the rapidity and extent of the hominin divergence from the great ape stock. One aspect of that feedback was the positive interaction between enhanced communication and technical skill.

## 3. THE ADAPTIVE PLATFORM: GESTURE

Great ape communication systems are very distant indeed from human language. Great ape signals are probably best thought of as Krebs–Dawkins [34,35] signals: as having the function of inducing specific behavioural responses from their audience rather than having representational content, encoding some feature of the common environment or of the signaller's state. If we do take these signals to be representations, they are pushmi-pullyu representations [36]: the vervet's famous alarm calls are hybrids or blends of descriptive and imperative content: neither merely 'leopard about' nor 'to the trees' but some combination of both. So in the lineage that lead from a great ape-like Last Common Ancestor to language equipped hominins, there must have been a very substantial expansion of communication before the emergence of anything like human language. In the early stages of this expansion, whether it was through expanded gesture or through an expanded role for vocal communication, signals would not have had semantic or pragmatic features resembling those of even the simplest utterances. Even so, there are good reasons to suppose that language began with gesture, for gesture recruits pre-existing capacities.

As Arbib and Tomasello, in particular, argue, great ape vocalization is probably not homologous to human language, because great ape vocalization is reflex-like [8,15]. Ape vocalization seems not to be sensitive to the agent's context and purpose: instead, it seems to be under the control of internal and external eliciting stimuli and hence is a largely automatic syndrome of emotional display. Moreover, there is little sign of adaptive plasticity in the overall repertoire. Some human vocal activities are similarly reflex-like responses to circumstance and arousal. But our

moans, yells, hisses and giggles are not part of language. As human speech and song show, it is possible for top-down inhibition and then control to evolve. But the evolutionary changes involved are far from trivial. The morphological and (especially) neurological adaptations for speech are complex and expensive. Talking involves extraordinarily complex control over tongue, mouth shape, inhalation and exhalation. It also depends on the physical reorganization of mouth, tongue and throat [37,38]. These adaptations are not free. Morphologically and neurologically, speech has the marks of an incrementally constructed complex adaptation, and it evolved despite the costs of this re-organization. For that to be possible, selection for enhanced communication must have preceded the capacity to speak. We evolved speech as a result of living in a world in which communication was already important.

Conversely, chimpanzees do use gesture in learned and context-dependent ways, though their range is very limited [39]. Most gestures are requests of various kinds, and chimpanzees do not seem to point informationally [15]. Even so, chimpanzees (and presumably our most recent common ancestor) do have top-down, context-sensitive control over hand movements; control shaped by learning. They need and use this control for gesture, but they need it even more obviously in their ecological interactions with their environment. Great apes are extractive foragers, and they engage in quite complex extractive activities. To do so, they must have visually guided control over their fine motor manipulation. They can see and identify what they are doing. If our prelinguistic ancestors had roughly similar capacities to those of the great apes, they were *poised* for the expansion of gesture; their existing competences sufficed for both the production and comprehension of gesture. Moreover, to the extent that social learning (especially imitation) is important, the hand movements of others are salient. Others will notice and respond to hand movements: what others are doing with their hands matters. So no new kit was needed.

Moreover, the gesture-first model avoids a serious problem that comes with the view that language is homologous to primate vocalization. The call systems of primates are holistic: elements of calls seem to have no discrete, independent significance. As a consequence, there has been considerable attention in the language evolution literature to the conditions under which structured signals emerge from holistic ones (see, [40,41]). Thus Mithen [42] and Wray [43] have suggested that hominin communication systems remained holistic until late in the transition to language. In their view, communication initially expanded to a large system of discrete vocalizations, each of which maps onto a whole situation. Structure, something like syntax, emerges through a process of segmentation, as signal users notice an initially accidental similarity between a set of holistic utterances, and the situations those utterances signify. It might turn out, for example, that the element 'ma' appears in a number of vocalizations, each of which maps onto a situation in which a woman receives resources. So holistic protolanguage speakers come to infer that 'ma' means something like 'female recipient' and a word emerges. Tallerman [44]

and Bickerton [45] both point out that this scenario is desperately implausible, both because of the sophistication of the cognitive mechanisms presupposed, and because (if the similarities involving 'ma' vocalizations really are accidental) there will be false-negatives and false-positives.

This scepticism of Bickerton and Tallerman has broader significance: there are no convincing models of how a system of holistic signals can incrementally turn into a structured system. On the gesture-first view of language evolution, we avoid that problem: gestural communication is primitively structured. Gestures, mimes (and probably iconic representation in general) begin life as structured representations: if an agent is miming ambushing a horse; picking and eating fruit from a ripening tree, the mime will have a sequential structure (often an activity, the target of an activity; and its results) with elements that could be extracted and re-used. In summary: on the assumption that we can use great ape capacities as a baseline, our australopithecine ancestors had a pre-existing potential to deliberately communicate in context-specific ways using gesture. This evolutionary potential was triggered and enhanced as part of the cognitive, ecological and social transformation of the hominin lineage, as they evolved as cooperative foragers. On the view developed here, late australopithecines, habilenes and erectines communicated largely through gesture, building incrementally on great ape capacities. Because they communicated through gesture, communication and skill depended on overlapping cognitive resources: those involved in the memorization and control of complex action sequences. I develop this idea in §4.

## 4. BEHAVIOURAL PROGRAMMES, GESTURE AND SKILL

In the past 15 years, Byrne has argued that the great apes extract resources using complex, often quite precise, multi-stage procedures. Nettle stripping, nut cracking and termite fishing are all examples of such procedures. Sometimes extractive foraging depends on the simultaneous and complementary use of each hand; sometimes it depends on using simple tools. Byrne has suggested that these extractive recipes should be thought of as behavioural programmes: they are organized into subfunctions, rather than being a concatenated sequence of behavioural atoms. Thus he has argued that the Social Intelligence Hypothesis needs to be supplemented by a Technical Intelligence Hypothesis [46]. Byrne is right to emphasize the skilled basis of great ape life, but it is not obvious that we should think of great ape technical competences as behavioural programmes, for the great apes may not represent elements of their own capacities separately. It is not clear that, say, a segment of a chimpanzee or gorilla skill can be redeployed without further ado as a component of another procedure. Nor can elements be taken offline in practice or demonstration.

Whatever the merits of Byrne's view of great apes, hominins have evolved behavioural programmes in this richer sense. Over the past 3.5 Myr, the hominin

lineage has seen a massive enhancement of techno-motor capacities. As a consequence, many human skills are much more complex than any great ape skill; many activities consist of multiple operations organized together. That is true not just of contemporary technology but of foragers' technology. Stout has attempted to make this idea precise by comparing the number of operations required to make Acheulian handaxes with those needed to make Oldowan flakes and scrapers. He argues that Acheulian artefacts are produced not just through long chains of elementary operations: these are hierarchically organized, with (for example) turn and strike sequences repeated, as the knapper turns the roughly shaped core in his hand, trimming and sharpening the edge. This whole multi-part operation is nested in the overall flow from initial preparation of the raw material to a symmetrically shaped handaxe [20].

Moreover, in many cases, humans have some top-down awareness of the structure of these skills. I conjecture that we have such awareness because we have been selected to teach as well as learn, so the skilled need to recognize and diagnose errors in the operations of the less skilled, and because some of these skills are so complex, and have such little error tolerance, that to learn them we need to take crucial elements off line, and autocue their practice [47,48]. Think, for example, of a batsman practising his footwork in front of a mirror, or a young forager practising blowpipe skills by pursing her lips and exhaling explosively but silently. We can demonstrate and practise components of complex operations, as when a bowler demonstrates her grip on the ball, or her follow-through. Furthermore, we can extract and reuse elements of a skill. It is less easy to hammer a nail in straight than it sounds, but once you have acquired this skill, you can redeploy that subprogramme in many contexts.

If language (or protolanguage) evolved as a system of gesture, the evolution of elaborated manual skill and the evolution of gestural communication would support one another. They would depend on the same fundamental cognitive, perceptual and motor capacities (see [49] for an elaboration of this point). Both select for the capacity to learn, memorize and fluently execute increasingly complex sequences. In both cases, we would expect selection for some capacity to represent one's own capacities. For as gesture and skill both elaborated, both involve sequences with structure, and with elements reusable in other contexts. This is an evolutionary two-for-one deal. The costs of the evolutionary innovation are supported by two benefits: the evolution of the capacity to represent and use behavioural programmes upgrades both skill and gestural communication. Gestures are sometimes described as social tools. This view of the relationship between skilled action and gesture adds meat to that metaphor.

## 5. ICONICITY

In accounts of the distinctive features of human language, arbitrariness figures prominently. The idea is that there is no natural relationship between term and referent: 'dog' could just as well refer to cats. With a tiny number of exceptions, this is true of spoken words. That is no accident: only through vocal imitation does language afford the option of a natural correspondence between sound and object, and few referents make a unique sound that humans can easily mimic. But it is not true of sign: not even of the highly developed forms of sign in use by contemporary humans, humans who have all the cognitive machinery that has evolved to facilitate our use of language. Corballis [13, p. 32] remarks that, for example, perhaps 50 per cent of Italian sign terms are iconic (and of course, many that are not now iconic derive from signs that were originally iconic; the same is true of many writing systems). Many other signals systems—for example, many road sign systems—depend on conventionalized icons.

The implication is clear. Even 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. In developing models of the evolution of language, it is important to avoid tacit circularity: to avoid models in which the early stages of the expansion of communication depend on capacities that are built through the evolution of a complex social life. For the expansion of the forms of communication that became language did not begin with humans who had minds adapted to language. It did not begin in cultures adapted to language either. Children are now born into an environment that is saturated with language. Arguably, with the emergence of Motherese, children's early experience is saturated not just with language, but with a special, infant-friendly version of language. But even if we set aside the help Motherese offers children's language learning, they experience language being used *consistently*. In contemporary environments, children are repeatedly exposed to core items of vocabulary being used in a consistent way: 'dog' is not used as a term for cats one day, pigs the next. As the role of communication in the social life of early hominins expanded, their young would not have been so lucky. Stabilized, regularly exploited conventions of sign/object pairings were still in the future, so they would have needed all the help they could get. For agents without specific adaptations for symbol use, and who were not living in a signalling environment with regular and consistent symbol-referent regularities, iconic signals would offer important advantages. Gesture offered those advantages much more freely than sound would.

Thus Donald imagines the expansion of hominin communication beginning with something analogous to mime, or to charades, in contemporary life: whole-body gesture, perhaps using props as well [50]. So, we might imagine an attempt to convey the location of a specific animal might include some mix of directional gesture (perhaps coupled to some simple convention like repetition or intensity to indicate distance as well) linked to a mime of distinctive body motion. Of course, imitation can include the use of sound. Contemporary foragers are often expert at vocal imitation of the local fauna, and that is an important hunting

tool [51]. Deer hunters in New Zealand often hunt by calling stags in by imitating a stag giving a territorial display in the hope that a resident male will arrive to repel the intruder. The advantage of vocal imitation in hunting and in mimetic communication might have been one selective force driving the evolution of elaborate top-down control of our vocal apparatus, control which made the switch from gesture to speech possible.

Even when the essential, minimal cognitive capacities were available, establishing communicative practices of this kind must have been chancy. That is true, even though I am not supposing that gesture would begin with anything as complex as a Donald-style mime. Rather, I imagine that such mimes began on a base established by simpler elements like indicative pointing, perhaps backed by a very simple mime, if that target of pointing was difficult to spot in bush or woodland: perhaps something like pointing, then flapping with one's arms to mimic flight to show that the visual target is a bird. Even from a platform of augmented pointing, a Donald-style proto-conventional mime would be difficult to establish. An agent (or more probably a small group) with something to communicate would have to be highly motivated to attempt to convey such a complex message; an audience would need to be motivated to puzzle it out. If anything like this happened at all, there must have been many failures and false starts. As with other innovations, a practice would not establish unless and until it was profitable enough to induce a lifeway change that is self-reinforcing. But once established, it would be self-reinforcing. Once a particular mime has been read and acted on successfully, second and subsequent uses will be easier; very probably, different mimes will be easier too. Even if they do not reuse elements of existing mimes, they will be salient as attempts to communicate. Once established and reinforced, we would expect to see conventionalization: some iconicity retained while time and energy are saved by abbreviating and simplifying displays, as the system responds both to fluent users' demands for ease of use, and to pressures of ease of entry [52]. Conventionalization apparently happens rapidly with newly invented sign systems used by contemporary humans [53]. They, of course, have the full benefits of minds that evolved under selection for communication, so it is safe to suppose that this process would have been much slower and more hesitant with our ancestors. But it would happen.

So here is the picture. Earlier hominins were evolving into cooperative foragers, probably while retaining an inherited fission–fusion social organization. They had long been bipedal, with range sizes expanded from great ape norms. Meat and marrow were an important, perhaps increasing, part of their diet. These would in part come from low-end scrounging from abandoned carcasses, but increasingly meat and marrow would come from expropriative scavenging and some hunting. Our ancestors were hunting large animals half a million years ago; probably much earlier [9,54]. Such social foraging selects for improved coordination [25,26]. Mime-like communication established in this socio-ecological niche, on the back of increased tolerance, cooperation, learning (including social learning) and theory of mind skills, compared with great ape baselines. Once these practices became a routine part of life, communicative procedures would have simplified and become more conventional. But the pressure exerted by juveniles as they joined the network would select for retaining iconic elements. If these were erectines, they would not have *sapiens*-grade interpretation skills (which depend in part on theory of mind). So their communication tools had to be reliably learnable by agents who were much more intelligent than great apes, but with cognitive systems less powerful than the even larger brained hominins to come.

Suppose all this is right. We have not got language yet; perhaps not even the protolanguage that Bickerton [6,55] and Jackendoff [56] have identified as a late staging post on the trajectory to language. But perhaps by the evolution of *Homo erectus*, hominins lived in cooperative bands that had some quite sophisticated technology; that had and maintained a good deal of information about their habitat and resources, and that relied on a high quality diet. This way of life would depend on reliable (though perhaps not very precise) social learning; on coordination; perhaps on some advanced planning [57]. I conjecture that these agents possessed quite rich, flexible communication skills built around gesture, mime, and perhaps some developing vocal imitation. Communication skills of this order probably do not count as language; perhaps not even word-string protolanguage. But if this were an approximately accurate depiction of *erectus* social lives, they had evolved a unique adaptive platform; an essential preliminary to the evolution of language. Thus the argument so far has linked selection for artisan skills to gesture and expanded communication in general, rather than to language in particular. I now turn to two more speculative arguments, linking expanded skill to two heartland features of language: stimulus independence and syntax.

## 6. STIMULUS INDEPENDENCE AND BEHAVIOURAL PROGRAMMES

One crucial feature of language is that it enables us to escape the here and the now. Words, unlike calls, are stimulus independent [4,6,56]. Terms are not responses to stimuli in the immediate environment; they do not covary with their referent. I sometimes say 'cat' when there is a cat present, but that is not typical: we speak about the elsewhere and the elsewhen. More puzzling still: it can be about the merely possible; it can be about the impossible; it can be about fictions. Donald's scenario of communication elaborating as mime *presupposes* stimulus independence; it does not explain it. On his picture, agents are able to produce iconic displays that resemble their target well enough (given shared circumstances, history and perceptual salience) for the audience to pick up on the target. But they do so independently of the stimulus of the target. How do they come by that capacity? By exaptation from motor skills, in two ways.

Initially, behavioural programmes (or their precursor) probably did not need to be guided by a mental template of the end product of the action sequence: they

could be anchored in the raw materials being transformed: opening a Molongo nut probably did not require a template of the opened nut with the kernel revealed; nor did striking an Oldowan flake from a core. But as action sequences become longer and more complex, and especially as action sequences involved genuinely transformative changes (as in the production of compound weapons involving points, bindings and glue), the sequence as a whole must be guided and initiated by a mental template. Its execution depends on a representation of the intended product, rather than being anchored in the raw materials being processed. It is plausible that human artisan skills were template guided by the late Acheulian. Obviously, making a stone tool depends at least partially on feedback from the world, as the artisan checks intended output of a specific act against the materials being worked. But there were many different raw material starting points, and many different ways the production process might go (depending on the exact details of the fracture pattern). So it is unlikely that the skill could be stored as a series of action-changed substrate-action-further changed substrate chains. But if making a late Acheulian handaxe is initiated by, then guided by, an inner template, then that action sequence *is* stimulus independent: it is driven by internal rather than external cues.

In addition, selection for the capacity for demonstration and offline practice also brings the programme as a whole (and, more probably, some elements of it) under internal control. In some cases, we can represent some aspects of a skill, and produce either the motor behaviour itself, or some critical component, in the absence of any intention to produce its normal product, and sometimes in the absence of its normal material substrate. We do so in autocued practice (musicians practising their breath control, for example, without their instrument); sometimes in teaching those skills to others, for example, in demonstrating a striking angle in making a tool (and occasionally, of course, in the innovative reuse of a behavioural component in a new context). We do not know when humans acquired this capacity to decouple the skilled execution of a motor programme from its normal substrate and product. But just as the skilled craftsmanship of the late Acheulian makes it plausible to attribute mental templates, it also makes it plausible to attribute the capacity to take skill offline. Stout and others argue that late Acheulian technology (that is, the technology of 8 kya) and certainly its successor technology of the Middle Stone Age depended for its uptake on active teaching and practise [20,47,58]. Thus stimulus independence is related to metacognition [59]. Once we have that capacity to demonstrate and to practise, it is available for the stimulus independent production of iconic representations. It is available for the mimes and charades Donald posits. Here as before, I think the process is likely to be co-evolutionary: gesture was elaborated and became more stimulus independent in partnership with the evolution of the capacity to take motor performance offline. Importantly, if this hypothesis is right, the co-evolution of gesture with skilled action explains not just a general expansion of communicative options, it explains why those options include a core feature of language, the capacity to communicate beyond the here and now.

## 7. SYNTAX

Human language is open-ended in ways that animal communication is not. In part, this depends on the productivity of the lexicon: we can coin new words. But it is also true that we can combine words into a hyper-astronomical number of sentences, by specifying the basic constituents of sentences in increasingly elaborate ways: thus the sentence of a subject can be a basic descriptor ('The man') or a modified version of that basic descriptor, and there is no sharp upper limit to the number of modifiers that can be attached. As linguists put it, the noun phrase (NP) that specifies the subject role in the sentence can be expanded without limit. Thus received wisdom is that an autonomous utterance—a sentence—is not just a beads-on-a-string concatenation of words. Sentences have hierarchically structured organization, and this is central to their expressive power. In a famous paper in 2002, Hauser, Chomsky, and Fitch took up this idea, suggesting that recursive syntax—the procedure of building structured signals without upper limit to their complexity—was the distinctive, uniquely human computational innovation that makes human language different from all other forms of communication, though it does so in partnership with the expansion of our conceptual repertoire [1].

Syntax may indeed distinguish human languages from other communication systems. But the contemporary human capacity (i) to expand a repertoire of atomic elements, and (ii) to combine those atomic elements into hierarchically organized sequences in an open-ended way is not limited to communication. Motor capacities are organized as behavioural programmes. And humans acquire new behavioural programmes in part by mastering new atomic skills (sharpening a stick, striking a flint to generate a spark); in part by recombining atomic skills in new ways. Within archaeology, Stout has made the most explicit case for understanding skill in terms of behavioural programmes, demonstrating that late Acheulian technology involved a massive expansion in the depth and complexity of a behavioural programme. He further suggested that these Acheulian programmes are so complex that they could be mastered only with the help of advanced social learning. But the idea is not new: Margaret Boden used knitting to introduce the concept of a recursive algorithm. She points out that knitting patterns are hierarchically organized sets of instructions, with subcomponents that include iterated repetition of specific procedures until a criterion is met, and a new stage of the overall process begins [60].

As knitting depends on domesticated animals, it is unlikely to be an ancient skill. But if anything like the grandmother hypothesis is right, containers woven from flax, pandanus or the like are probably ancient, dating back to erectines [32], for grandmothers were not harvesting feed-as-you-go resources. Underground storage organs need processing, and they are intended for young children, so they are a form of central place foraging. Carrying a large number of items without a

container would have been profoundly inefficient. A basket made from flax or twine will need to be made using a Boden-style behavioural programme: these are transformative technologies requiring multiple stages of processing (many containers use more than one type of material; for example, flax woven around a structure of braced twigs). Such skills probably involve an overall structure controlled by a mental template, with a sequence of steps, each completed by iterating an atomic procedure.

So here is the suggestion. Suppose that from (say) the habilenes to the very large brain hominins of about 5 kya, selection drove the expansion of gesture-based communication and motor skills. Both increased in complexity and importance, as gesture and communicative mime were just special cases of skilled motor capacity. Gestures and mimes were hierarchically organized sequences of elements. Each element in the sequence—say, the element indicating the key action—could be simple, but it could also expand in complexity while leaving the other elements in the mime as before. Such mimetic communication is open-ended both in allowing the expansion of atomic elements, and in allowing those new elements to be exported from one gestural narrative to another. Artisan skills and gesture-based communication co-evolved: effective selection for the capacity to chunk atoms into a more complex unit, to control action sequences using a mental template, and to take chunks offline, enhances both motor skills and communicative capacity. After all, on this view communicative capacity *just is* a special case of a motor skill. Language would be quite literally a communicative technology. I cannot put dates to this process, though if the gesture–skill connection is as crucial as I suppose, it would have been well underway by the late Acheulian, perhaps earlier. Even if the transition to language has deep roots, it remains possible that it was not complete until quite recently. Shultz *et al.* [7] suggest that the final elements of language—multiply embedded clause structures used to report complex scenarios and deeply layered mental ascriptions—were put in place only quite recently. This suggestion is based on their dates for the most recent expansion of hominin brain size (in the past 100 kyr), and their reading of the evidence of the spread of complex modern languages. This very late date for full language might be right, but the suggestion depends on a controversial reanalysis of hominin brain size evolution, and an equally controversial link between brain size and linguistic sophistication. Given the many uncertainties, I remain agnostic about the precise timing of language's emergence.

Let me reinforce this idea about the origins of syntax with a complementary point adapted from Tomasello. He has argued that one of the factors that explains the social difference between humans and the chimpanzee species is the human capacity for collective intentions. Humans do not just act together, they do so knowing that they are acting together with a shared understanding of the common situation. Moreover, they do so in part because they know that others are with them: collaborative activity is intrinsically rewarding, not just materially profitable. Collective intention is partly motivational, partly cognitive. One cognitive aspect is the capacity to form what Tomasello calls a 'bird's eye view' of a collective activity: a third-person representation of its structure in terms of roles, rather than a first person representation in terms of specific agents [61–63].

Chimpanzees do not do well on role-reversal tasks, and Tomasello suggests that this follows from their egocentric representation of task demands. Chimpanzees engage in little collective activity. But our ancestors did: a core feature of the human revolution was our evolution as collective foragers. Collective activity as such does not demand a bird's eye view representation of the activity. Such representations are not needed if there is no role differentiation (as in social carnivores like African wild dogs), or if each agent always takes the same specialized role. But if there is collective action with role specialization, and if agents do not always act in the same role (if sometimes I act to drive the game; if sometimes I am the lookout; if sometimes I wait in ambush), then each team member does need a bird's eye representation of the collective activity. But bird's eye representation is a crucial representational capacity needed for syntax. It is the distinction between a role and the occupant of that role. In representing 'The man hit the horse with a stick' as having an agent–action–patient–instrument structure, we make a role–occupant distinction: we distinguish between the specific term that picks out (say) an instrument from the role in the utterance played by that term. Once we make a role-occupant distinction, a given occupant can be redeployed to other roles, and a role can be occupied in many other ways. So the role-occupant distinction is pivotal to syntax: subject, object and so on are role concepts, which can take many lexical occupants. Again: the organization of action co-evolves with communication, unsurprisingly, since communication is a special case of collective action.

The idea that there might be a 'grammar of behaviour' is not new. Mikhail, for example, in exporting Chomsky's views of language to moral cognition, argues that we represent actions as having hierarchical structure, in just the same way that we represent sentence [64]. What is added here is an account of why hominins might have both needed to represent their own actions in structured ways, and why those representational capacities might have been extracted for use in other domains.

It is time to summarize the argument of this paper. It begins by placing the evolution of language within the context of a broader perspective on the evolution of human life: hominins evolved as cooperative, social foragers, under selection for coordination, information-sharing and social learning. It then argues that an initial expansion of communication skills probably proceeded by the expansion of gesture rather than vocalization. First, gesture was already used in voluntary, context-sensitive, learning-mediated signalling. Second, gesture-based signs are often iconic, hence are more easily interpreted and remembered. Third, gestural communication was primitively structured, and we know that the communication system that emerged from this evolutionary transition was one using complex, structured signs. In the more speculative final two sections of the paper, the argument

connected the elaboration of skilled artisanship with the evolution of context independent and syntactically structured signals, showing that the skilled action–gesture co-evolutionary feedback loop helps us explain not just a general increase in communicative capacity, but features central to language as a communication system. I now turn to two challenges: testability and the gesture-to-speech transition.

## 8. TESTS AND CHALLENGES

This paper has presented a scenario that links the evolution of language to the evolution of enhanced artisan skills, via a connection with gesture. Scenario-building has a poor reputation in evolutionary biology, derided as mere story-telling: see especially [65]. This mocking response understates the value of scenarios. First, a properly constructed scenario for the evolution of language at least identifies a possible pathway for its evolutionary construction, showing that we do not need to posit exotic mechanisms or near-miraculous coincidences. In the case of language, showing that language can evolve without miracles is of some importance, since it has been claimed that language is irreducibly complex, not a system that can be built by small increments. Second, scenarios narrow the search space. Even if a scenario is not itself testable, a scenario identifies broad areas of informational relevance, the materials from which tests can ultimately be built. The scenario defended in this paper supposes that artisan skills are tightly coupled to communicative ones. So it directs a search for linkages both in the developmental and cognitive psychology of skill, and in the archaeological record. Thus, the picture presented predicts that the expansion of technical skill is correlated with the expansion of the capacities to plan, coordinate and inform. That said, these features of social life do not leave unambiguous material traces, especially in the deep past. Third, scenarios are much less easy to construct than the just-so-story sceptics suppose. The critical thought is that we can easily tell many different but equally credible stories so (without rigorous testing) there is no reason to take any of them seriously. Story telling is a more difficult art than this. Writers of historical fiction, and intelligence officers constructing cover stories for their operatives, know that it is difficult to construct a scenario that is internally coherent, intrinsically plausible (that is, not relying on low-probability coincidences) and which articulates smoothly with the known facts. The more detailed the scenario, and the more points of articulation there are with the known facts, the more challenging that construction project becomes. Even granted these constraints, perhaps scenarios are only how-possibly explanations. If so, the array of candidate explanations is much more limited than the just-so-story response suggests.

In sum, then, it is true that the scenario developed here is not unambiguously testable. But it is developed in some detail, and it articulates with both the known facts and the known evolutionary mechanisms. In particular, in §1, I noted five conditions any serious hypothesis should satisfy, and this scenario passes those tests. *Honesty*: the honesty of early language,

and the fact that language use evolved only in the hominin lineage, are both explained by the background model of human social evolution: the evolution of hominins as social foragers. That way of life selects for cooperation, and honest communication is a specific case of cooperation; lying and withholding information, of defection.

In my view, the social, cognitive and environmental factors that support Pleistocene ecological cooperation support honest (enough) talk. The famous 'folk theorem' is a set of game theoretic results that show the stability of reciprocation-based cooperation if certain conditions are satisfied: (i) if helping is high benefit/ low cost; (ii) if agents interact repeatedly; (iii) if free-riders can be detected reliably and cheaply; (iv) if free-riders can be punished cheaply (relative to the benefits of cooperation) [66]. Pleistocene foragers satisfied these conditions. They lived in small stable groups, with high profits to cooperation, both from stag-hunt gains and from managing risk [28,67]. Because they were small and stable, there were few secrets. Since agents exercised a good deal of choice over whom they associated with, reputation was important [68]. So though punishing defection was not free, even when collective, those costs were worth paying, as they signal one's own status as a good cooperator. The honesty of early communication is thus stabilized by just the same factors that stabilized (say) food sharing. There is no special difficulty in detecting or deterring informational failures to cooperate: just as everyone in a village knows who can be counted on for material aid, and who cannot; everyone in a village knows who lies and exaggerates, and who does not.

### (a) *Uniqueness*

Language did not evolve directly from great ape skill sets. It evolved only after, and as a result of, the construction of an adaptive platform: enhanced communication capacities; enhanced theory of mind and working memory (since interpreting protolanguage depends heavily on common knowledge); on the establishment of a cooperative social life. There is a uniqueness problem: explaining why only the hominins among the great apes evolved as collective, technology and technique-dependent foragers. (And why there has been no convergent occupation of that niche in other lineages.) But once that unique aspect of hominin history has been explained, there is no further problem of explaining our unique use of language. Only highly cooperative agents that depend on skill, expertise and coordination, and who have some understanding of one another's points of view, are poised to evolve language.

### (b) *Distinctive scope and expressive power*

The last two sections of the paper speak to this condition, arguing that the gesture–artisan skill co-evolutionary model can explain both the stimulus independence of language, and the fact that utterances are structurally complex, with expandable components. Likewise, the scenario rests on well-understood evolutionary mechanisms: it portrays the evolution of language as gradual, and proceeding through the modification of existing

capacities rather than creating new ones out of nowhere. The scenario depends on positive feedback, but that is part of evolutionary biology's standard toolkit. So there is nothing exotic about the evolutionary mechanisms on which the scenario rests. Nor does the scenario rest on controversial or implausible claims about hominin ecology, though it does depend on the idea that hominins have long confronted environmental challenges collectively, as teams, and often with a division of labour. It depends then on the idea that coordination, not just cooperation, has long been a human challenge. The ethnographic record certainly shows that historically known foraging societies had these characteristics [69]. The picture presented in this paper pushes these characteristics of human life into its deep history. In sum then: while this paper does not present a rigorously testable theory of language evolution, it is more than just a story.

There is an obvious challenge to any theory of the evolution of language which proposes that it began as a system of gesture. Why is the default modality now speech, and how did the transition occur? The objection is less daunting than it seems. (i) In §3, I pointed out that great ape vocalization seems to be a response to arousal, an expression of emotion, rather than being under voluntary control. Hrdy [70] begins by pointing out how different the emotional lives of humans and chimpanzees are: we are both more tolerant of one another, and we are far superior at inhibiting our immediate emotional response. Emotions and their expression are under much more voluntary control. So the early evolution of human cooperation set up an environment that selected for increased tolerance and improved inhibition; for bringing the expressions of emotion, including vocalization, under increasing top-down control. Establishing a cooperative social world made vocalization less automatic, less reflexlike. (ii) Dunbar [71] has long argued persuasively that as hominin social worlds became larger and more complex, conflict and social stress would become increasingly difficult to manage through methods inherited from the great apes. Increasingly, vocal grooming would have to replace or supplement physical grooming. The point about conflict is well taken, but as (for example), Mithen [72] points out, the hypothesis about vocal grooming is more plausibly reinterpreted as explaining the evolution of music and song. Music and song are enormously powerful in shaping affect; talk, including gossip, is much more variable in its emotional impact. Importantly, selection for song-like social grooming and bonding would not just bring vocalization under top-down control, it would build precise control of vocal performance, performance in response to specific social situations. (iii) Once our vocal life was not just under top-down control, but under top-down control that allowed precise execution of a complex vocal sequence, it would naturally be incorporated into mime and gesture. Once that ability is in place, we would expect communicative acts to be a hybrid of manual, bodily and vocal elements. That, indeed, is what they still are, in many situations. Still, the final piece of the puzzle is to explain why vocalization has largely taken over; why it is the default modality.

This *is* a puzzle. One possible solution is as follows: the hands have much to do, so the opportunity cost of using hands as the primary communication tool is high, especially once conversation became an incessant feature of human social life. So once the vocal channel became available, and given that the shift could be slow and incremental, economics favoured a shift to the vocal channel. In summary, then, I see the shift from gesture to speech as a four-stage process: an initial stage in which selection for inhibition of emotional response makes vocalization less reflexive; a Dunbarian stage in which minimal top-down control of vocalization expands and becomes much more precise, as vocalization becomes increasingly recruited as a tool of social cohesion and social bonding; a third stage (which probably overlaps in time with the Dunbarian period) in which the gesture-based system of information sharing and coordination is enriched through recruiting vocal elements; a final stage in which the hybrid system becomes vocal, perhaps driven by the opportunity costs of having one's hands tied up as communicative tools. This is far from a full theory of the transition, but it is enough to show that it is no insoluble mystery, either.

Let me recap briefly. There are many puzzling features of language and its evolution. These puzzles in part flow from continuing controversies about the nature of language as it now is; in part from the great difference between language and other communication systems; in part from the lack of direct evidence about, and a plausible model of, its incremental construction. I have argued that some of these puzzles are less daunting if we adopt a gesture-first view of the origins of language, for this view delivers a plausible model of the emergence of syntax and of stimulus independence, two uncontroversially distinctive and important features of language.

## REFERENCES

1 Hauser, M., Chomsky, N. & Fitch, T. 2002 The faculty of language: what is it, who has it, and how does it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)

2 Pinker, S. & Bloom, P. 1990 Natural language and natural selection. *Behav. Brain Sci.* **13**, 707–784. (doi:10.1017/S0140525X00081061)

3 Dessalles, J.-L. 2007 *Why we talk: the evolutionary origin of language*. Oxford, UK: Oxford University Press.

4 Deacon, T. 1997 *The symbolic species: the co-evolution of language and the brain*. New York, NY: W. W. Norton.

5 Deacon, T. W. 2003 Universal grammar and semiotic constraints. In *Language evolution* (eds M. Christiansen & S. Kirby), pp. 111–139. Oxford, UK: Oxford University Press.

6 Bickerton, D. 2009 *Adam's tongue: how humans made language, how language made humans*. New York, NY: Hill and Wang.

7 Shultz, S., Nelson, E. & Dunbar, R. I. M. 2012 Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record. *Phil. Trans. R. Soc. B* **367**, 2130–2140. (doi:10.1098/rstb.2012.0115)

8 Arbib, M. 2005 From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* **28**, 105–167. (doi:10.1017/S0140525X05000038)

9   Foley, R. & Gamble, C. 2009 The ecology of social transitions in human evolution. *Phil. Trans. R. Soc. B* **364**, 3267–3279. (doi:10.1098/rstb.2009.0136)

10  Klein, R. & Edgar, B. 2002 *The dawn of human culture.* New York, NY: Wiley.

11  Odling-Smee, J. & Laland, K. 2009 Cultural niche construction: evolution's cradle of language. In *The prehistory of language* (eds R. Botha & C. Knight), pp. 99–121. Oxford, UK: Oxford University Press.

12  Corballis, M. 2003 *From hand to mouth: the origins of language.* Princeton, NJ: Princeton University Press.

13  Corballis, M. 2009 The evolution of language. *Ann. NY Acad. Sci.* **1156**, 19–43. (doi:10.1111/j.1749-6632.2009.04423.x)

14  Donald, M. 2005 Imitation and mimesis. In *Imitation, human development, and culture* (eds S. Hurley & N. Chater), pp. 282–300. Cambridge, MA: MIT Press.

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

16  Kaplan, H., Hill, K., Lancaster, J. & Hurtado, M. 2000 A theory of human life history evolution: diet, intelligence and longevity. *Evol. Anthropol.* **9**, 156–185. (doi:10.1002/1520-6505(2000)9:4)

17  Mace, R. 2000 The evolutionary ecology of human life history. *Anim. Behav.* **59**, 1–10. (doi:10.1006/anbe.1999.1287)

18  Wrangham, R. W. 2001 Out of the pan, into the fire: how our ancestors' evolution depended on what they ate. In *Tree of life* (ed. F. de Waal), pp. 121–143. Cambridge, MA: Harvard University Press.

19  Foley, R. & Lahr, M. M. 2003 On stony ground: lithic technology, human evolution and the emergence of culture. *Evol. Anthropol.* **12**, 109–122. (doi:10.1002/evan.10108)

20  Stout, D. 2011 Stone toolmaking and the evolution of human culture and cognition. *Phil. Trans. R. Soc. B* **366**, 1050–1059. (doi:10.1098/rstb.2010.0369)

21  Toth, N. & Schick, K. 2009 The Oldowan: the tool making of early hominins and chimpanzees compared. *Annu. Rev. Anthropol.* **38**, 289–305. (doi:10.1146/annurev-anthro-091908-164521)

22  Alperson-Afil, N., Richter, D. & Goren-Inbar, N. 2007 Phantom hearths and the use of fire at Gesher Benot Ya'Aqov, Israel (2007). *PaleoAnthropology.* **3**, 1–15.

23  Roebroeks, W. & Villa, P. 2011 On the earliest evidence for habitual use of fire in Europe. *Proc. Natl Acad. Sci. USA* **108**, 5209–5214. (doi:10.1073/pnas.1018116108)

24  Hiscock, P. & O'Conner, S. 2006 An Australian perspective on modern behaviour and artefact assemblages. *Before Farming* 2, article 4.

25  Stiner, M. C. 2002 Carnivory, coevolution, and the geographic spread of the genus *Homo*. *J. Archaeol. Res.* **10**, 1–63. (doi:10.1023/A:1014588307174)

26  Sterelny, K. 2007 Social intelligence, human intelligence and niche construction. *Proc. R. Soc. B* **362**, 719–730. (doi:10.1098/rstb.2006.2006)

27  Sterelny, K. 2011 From hominins to humans: how *sapiens* became behaviourally modern. *Phil. Trans. R. Soc. B* **366**, 809–822. (doi:10.1098/rstb.2010.0301)

28  Sterelny, K. 2012 *The evolved apprentice.* Cambridge, MA: MIT Press.

29  Whiten, A. & Erdal, D. 2012 The human socio-cognitive niche and its evolutionary origins. *Phil. Trans. R. Soc. B* **367**, 2119–2129. (doi:10.1098/rstb.2012.0114)

30  Barton, R. 2012 Embodied cognitive evolution and the cerebellum. *Phil. Trans. R. Soc. B* **367**, 2097–2107. (doi:10.1098/rstb.2012.0112)

31  Potts, R. 1998 Variability selection in hominid evolution. *Evol. Anthropol.* **7**, 81–96. (doi:10.1002/(SICI)1520-6505(1998)7:3)

32  O'Connell, J. F., Hawkes, K. & Blurton Jones, N. 1999 Grandmothering and the evolution of *Homo erectus*. *J. Hum. Evol.* **36**, 461–485. (doi:10.1006/jhev.1998.0285)

33  Wrangham, R. 2009 *Catching fire: how cooking made us human.* London, UK: Profile Books.

34  Krebs, J. & Dawkins, R. 1984 Animal signals, mind-reading and manipulation. In *Behavioural ecology: an evolutionary approach*, 2nd edn (eds J. R. Krebs & N. B. Davies), pp. 380–402. Oxford, UK: Blackwell Scientific.

35  Owren, M., Rendall, D. & Ryan, M. 2010 Redefining animal signaling: influence versus information in communication. *Biol. Phil.* **25**, 755–780. (doi:10.1007/s10539-010-9224-4)

36  Millikan, R. 1989 Biosemantics. *J. Phil.* **86**, 281–297. (doi:10.2307/2027123)

37  Fitch, W. T. 2009 Fossil clues to the evolution of speech. In *The cradle of language* (eds R. Botha & C. Knight), pp. 112–134. Oxford, UK: Oxford University Press.

38  Fitch, W. T. 2010 *The evolution of language.* Cambridge, UK: Cambridge University Press.

39  Pollick, A. & de Waal, F. 2007 Ape gestures and language evolution. *Proc. Natl Acad. Sci. USA* **104**, 8184–8189. (doi:10.1073/pnas.0702624104)

40  Brighton, H., Kirby, S. & Smith, K. 2005 Cultural selection for learnability: three principles underlying the view that language adapts to be learnable. In *Language origins: perspectives on evolution* (ed. M. Tallerman). Oxford, UK: Oxford University Press.

41  Smith, K., Kirby, S. & Brighton, H. 2003 Iterated learning: a framework for the emergence of language. *Artif. Life.* **9**, 371–386. (doi:10.1162/106454603322694825)

42  Mithen, S. 2005 *The singing Neanderthals: the origins of music, language, mind and body.* London, UK: Weidenfeld & Nicholson.

43  Wray, A. 1998 Protolanguage as a holistic system for social interaction. *Lang. Commun.* **18**, 47–67. (doi:10.1016/S0271-5309(97)00033-5)

44  Bickerton, D. 2007 Language evolution: a brief guide for linguists. *Lingua* **117**, 510–526. (doi:10.1016/j.lingua.2005.02.006)

45  Tallerman, M. 2006 Abracadabra! Early hominin for 'I think my humming's out of tune with the rest of the world!' Review feature: Mithen, S. 2005. The singing Neanderthals. *Cambr. Archaeol. J.* **16**, 106–107.

46  Byrne, R. 2003 Imitation as behaviour parsing. *Phil. Trans. R. Soc. Lond. B* **358**, 529–536. (doi:10.1098/rstb.2002.1219)

47  Stout, D. 2002 Skill and cognition in stone tool production: an ethnographic case study from Irian Jaya. *Curr. Anthropol.* **43**, 693–722. (doi:10.1086/342638)

48  Wadley, L. 2010 Compound-adhesive manufacture as a behavioural proxy for complex cognition in the middle stone age. *Curr. Anthropol.* **51**(Suppl. 1), S111–S119. (doi:10.1086/649836)

49  Stout, D. & Chaminade, T. 2012 Stone tools, language and the brain in human evolution. *Phil. Trans. R. Soc. B* **367**, 75–87. (doi:10.1098/rstb.2011.0099)

50  Donald, M. 2001 *A mind so rare: the evolution of human consciousness.* New York, NY: W. W. Norton.

51  Lewis, J. 2009 As well as words: Congo Pygmy hunting, mimicry and play. In *The cradle of language* (eds R. Botha & C. Knight), pp. 236–256. Oxford, UK: Oxford University Press.

52  Wray, A. & Grace, G. 2007 The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* **117**, 543–578. (doi:10.1016/j.lingua.2005.05.005)

53  Senghas, A., Kita, S. & Ozyurek, A. 2004 Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science* **305**, 1779–1782. (doi:10.1126/science.1100199)

54 Jones, M. 2007 *Feast: why humans share food*. Oxford, UK: Oxford University Press.

55 Bickerton, D. 2002 From protolanguage to language. In *The speciation of modern Homo sapiens* (ed. T. J. Crow), pp. 193–120. Oxford, UK: Oxford University Press.

56 Jackendoff, R. 1999 Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.* **3**, 272–279. (doi:10.1016/S1364-6613(99)01333-9)

57 Nowell, A. & White, M. 2010 Growing up in the middle Pleistocene. In *Stone tools and the evolution of human cognition* (eds A. Nowell & I. Davidson), pp. 67–82. Boulder, CO: University of Colorado Press.

58 Shea, J. 2006 Child's play: reflections on the invisiblity of children in the Paleolithic record. *Evol. Anthropol.* **15**, 212–216. (doi:10.1002/evan.20112)

59 Frith, C. D. 2012 The role of metacognition in human social interactions. *Phil. Trans. R. Soc. B* **367**, 2213–2223. (doi:10.1098/rstb.2012.0123)

60 Boden, M. 1989 *Artificial intelligence in psychology: interdisciplinary essays*. Cambridge, MA: MIT Press.

61 Tomasello, M. 2009 *Why we cooperate*. Cambridge, MA: MIT Press.

62 Tomasello, M. & Carpenter, M. 2007 Shared intentionality. *Dev. Sci.* **10**, 121–125. (doi:10.1111/j.1467-7687.2007.00573.x)

63 Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. 2005 Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* **28**, 675–691.

64 Mikhail, J. 2007 Universal moral grammar: theory, evidence, and the future. *Trends Cogn. Sci.* **11**, 143–152. (doi:10.1016/j.tics.2006.12.007)

65 Gould, S. J. & Lewontin, R. 1978 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)

66 Binmore, K. 2006 Why do people cooperate? *Polit. Phil. Econ.* **5**, 81–96. (doi:10.1177/1470594X06060620)

67 Sterelny, K. In press. Life in interesting times: co-operation and collective action in the Holocene. In *Evolution, cooperation, and complexity* (eds B. Calcott, B. Fraser, R. Joyce & K. Sterelny). Cambridge, MA: MIT Press.

68 Smith, E. A. *et al.* 2010 Wealth transmission and inequality among hunter–gatherers. *Curr. Anthropol.* **51**, 19–34. (doi:10.1086/648530)

69 Boehm, C. 1999 *Hierarchy in the forest*. Cambridge, MA: Harvard University Press.

70 Hrdy, S. B. 2009 *Mothers and others: the evolutionary origins of mutual understanding*. Cambridge, MA: Harvard University Press.

71 Dunbar, R. 2009 Why only humans have language. In *The prehistory of language* (eds R. Botha & C. Knight), pp. 12–35. Oxford, UK: Oxford University Press.

72 Mithen, S. 2009 Holistic communication and the coevolution of language and music: resurrecting an old idea. In *The prehistory of language* (eds R. Botha & C. Knight), pp. 58–76. Oxford, UK: Oxford University Press.