Biol Philos (2017) 32:819–838
https://doi.org/10.1007/s10539-017-9597-8

From code to speaker meaning

Kim Sterelny[1]

Received: 4 May 2017 / Accepted: 6 October 2017 / Published online: 19 October 2017

Abstract This paper has two aims. One is to defend an incrementalist view of the evolution of language, not from those who think that syntax could not evolve incrementally, but from those who defend a fundamental distinction between Gricean communication or ostensive inferential communication (Scott-Phillips, Sperber, Tomasello, originally based on Grice) and code-based communication. The paper argues against this dichotomy, and sketches ways in which a code-based system could evolve into Gricean communication. The second is to assess the merits of the Sender–Receiver Framework, originally formulated by David Lewis, and much elaborated and set into an evolutionary context by Brian Skyrms and colleagues, as a framework for thinking about the evolution of language. Despite the great strengths of that framework, and despite the great value of a framework that is both general and formally tractable, I argue that there are critical features of language that it fails to capture .

Keywords Codes · Sender–Receiver models · Gricean communication · Evolution of pragmatic competence · Evolution of speaker meaning

Ostensive communication and incremental evolution

This paper has two aims. One is methodological: to discuss the value and limits of one mainstream approach to modelling the evolution of communication; in particular, to consider its applicability to hominin communication in the transition from hominins with communicative capacities typical of the great apes to those with capacities closer to language. The other is substantive, to defend an incremental

✉ Kim Sterelny
Kim.Sterelny@anu.edu.au

[1] School of Philosophy, Australian National University, Canberra, Australia

view of the evolution of language. Despite the apparent organisational complexity of language, and its obvious utility, incremental views of the evolution of language are controversial, primarily because Chomsky and his intellectual allies have argued that there can be no coherent account of the incremental emergence of syntax (mostly recently in Berwick and Chomsky 2016). But there is also a semantic challenge to incremental models of language evolution, most recently articulated by Thomas Scott-Phillips (Scott-Phillips 2015), building in the first instance on the work of Michael Tomasello, Deidre Wilson and Dan Sperber, and more remotely on that of Paul Grice (Grice 1957; Sperber and Wilson 1986; Tomasello 2008, 2014). This challenge places pragmatics—the fact that talking requires us to interpret one another, not just to know the meaning of words—at the heart of language. In particular, Scott-Phillips defends a fundamental and qualitative distinction between communication that depends on codes (for example, the famous vervet warning call system: more on codes in "Codes and the Sender–Receiver framework" section), and communication that depends on overtly assisted interpretation; that is, on communication that depends on an interpreter recognising that a sender is attempting to communicate, and using that recognition, together with common ground and general contextual information, to work out what the speaker intends to communicate. Amongst the pragmatically competent in normal circumstances, this inference (like much human inference) is typically so efficient that it is both rapid and tacit. I catch a friend's eye at a party and tap my glass; the bottle is passed my way. My friend recognises eye capture and glass tap as an intent to communicate; making the glass salient; its emptiness obvious; my well-known preferences then fuel an inference to the content of my message.

In Scott-Phillip's terminology, this is *ostensive inferential communication*. Ostensive, because it depends on the sender making his/her intention to communicate public, overt, obvious. Inferential, because it depends on the interpreter working out what the sender wants to communicate, once he or she has recognised that the sender wants to communicate something. Amongst those competent to communicate this way, ostensive communication is superbly flexible. Just about anything the parties involved can think, they can communicate. In particular, communicating about, say, the need for more wine does not depend on the prior establishment of any arrangement for wine delivery. But that very flexibility seems to depend on great cognitive sophistication. A critical feature of this line of thought is that normal language use exemplifies this pattern. For what the speaker intends to communicate in speaking is almost never exhausted by, and sometimes does not include, the literal meaning of the words used. I say "Not the salt" at the dinner table to get my conversational partner to pass the bowl of chopped chillies instead. I never mentioned chillies, but the look down the table, the exclusion of salt, and contextual information allows that partner to zero in on the fact that I intend to communicate the desire that the chillies be passed. This is typical of the way humans use language. Scott-Phillips and his colleagues think that this gap between the conventional meaning of what is said, and what the speaker intends to convey, is of great theoretical significance. Language has code-like elements, through conventional word meaning and through systematic syntax, but Scott-Phillips and his allies take these facts about the gap between intended message and literal meaning to show that the use of language cannot be reduced to mastering a code, no matter how sophisticated. Ostensive inferential communication is not merely context-sensitive, in the way that the interpretation of "now" or "here" are context-sensitive. No simple rule captures the ways an audience reconstructs the speaker's communicative intention (Scott-Phillips 2017). As he sees it, the critical feature of codes—and animal communication systems are typically or invariably codes—is that signals have fixed meanings (though sometimes course-grained ones).

As this group of theorists see it, ostensive-inferential communication depends on sophisticated theory of mind skills, though their precise nature depends on the specific account of ostensive inferential communication. Analyses differ, though as Richard Moore nicely shows, they have a common pattern (Moore 2016). The speaker has a communicative intention coupled to an informational intention. Thus Moore depicts overt intentional communication (as Scott-Phillips and the Griceans see it) as having the following shape:

A sender *S* means something by a signal *u* if and only if *S* sends *u* to *R* intending:

1. *R* to produce a particular response *w*, and
2. *R* to recognise that *S* intends (1).

The intended response *w* determines what *u* means (Moore 2015, 2016). On this account, *u* is meaningful because *S's* intention is overt—*S* wants the target audience to know what he/she is doing (via clause 2). Tapping my glass is a meaningful signal because I want you to understand that this tapping is intended to tell you something.

In the standard versions of this view, ones that include the full Gricean apparatus, *w* is itself a complex, meta-intentional cognitive response; the audience is intended to represent a complex state of the speaker's mind: that I intend you to believe that I want you to pass me wine. (Or something similar). As Scott-Philips writes "it is the heart of Grice's account … I should intend that my audience believes it, and they should believe it at least in part because they recognise that this was my very intention" (Scott-Phillips 2015, p. 23). The resulting picture can look formidably complex when all this is stated explicitly, as one of Grice's own formulations shows:

'U meant something by x' is true if and only if U uttered x intended thereby

1. that A should produce response r,
2. that A should at least party on the basis of x, think that U intended (1),
3. that A should think that U intended 2,
4. that A's production of r should be based (at least in part) on A's thought that U intended (1),
5. that A should think U intended (4) (Grice 1969, p. 156).

Perhaps not all this is necessary, and in recent version of the idea, equivalents of the final two clauses have been dropped. Even so, in the eyes of many, treating *w* itself as such a rich cognitive response seems an implausibly over-intellectualised and complex view of ordinary conversational interchange (Millikan 1998); a view to which I shall return. But even if that is right, there is an important insight in this picture. It distinguishes two roles in typical human communication. One is the management of audience attention by the speaker. The other is the flow of information or instruction from speaker to audience. In developing an incremental picture of the emergence of ostensive inferential communication, I shall retain this distinction between attention management and information flow, but suggest that in earlier hominins these roles were occupied by simpler cognitive mechanisms than those identified in typical analyses of contemporary ostensive inferential communication.

Scott-Phillips does not think ostensive-inferential communication evolved incrementally from code-based communication, though once agents communicate by interpreting one another, the scope, efficiency and precision of ostensive inferential communication can and did expand incrementally. Instead, on this picture, hominin theory of mind and inferential capacities expanded incrementally. Once they reached a threshold, and only when they reached a threshold, ostensive-inferential communication became possible, and with it, its great flexibility. Ostensive-inferential communication can and does tap into purpose-built communicative resources, words, sentences, conventionalised gestures. It does so, in normal conversation. But it need not. Even when it does, it is not limited to the conventional repertoire those resources make available. Rather, that repertoire is an inferential scaffold; a way of vastly expanding the common ground that makes interpretative inference remarkably, though of course not perfectly, reliable.[1] Even when there are no tigers or pictures of tigers around, I can use the word "tiger" to get the audience to think about why I might be thinking about tigers with communicative hopes. Without language, that would be very difficult.

Codes and the Sender–Receiver framework

The default form of human communication is (on this analysis) ostensive inferential communication. In contrast, most, perhaps all, animals communicate using codes. In contrast to language, we have rich theoretical models of codes and their evolution. The most general and helpful of these is the Sender–Receiver framework that Brian Skyrms and his colleagues (building originally on Lewis 1969) have used as their workhorse (Skyrms 2010). The Sender–Receiver framework models communication in terms of a set of congruent, coadapted Sender–Receiver rules. This form of communication pays its way when the state of the world can vary in respects that are relevant to the interests of both sender and receiver. The sender observes the specific state of the world; sends a signal to the receiver who acts (all going well) in a way that is both tuned to the state of the world, and which is beneficial to both sender and receiver. Specific Sender–Receiver systems are therefore specified by their Sender–Receiver menu: the set of world state → sender → signal → receiver → act → payoff

[1] It may well be that this pragmatics-focused view of language understates the role of syntax, and the ways modifiers and adjectives affect their heads in unobvious ways: a mirror-image of the neglect of pragmatics and semantics by syntax-focused views of language.

options. The menu might be very small: a single generalised alarm call in response to danger of any kind; a single form of escape behaviour (or a small menu of such calls and responses, as with the vervets). It might be limited but continuous: the amplitude of the call might signal the immediacy or seriousness of the danger, and be reflected in the urgency of the escape behaviour. It might be structured: the honeybee system signals direction, distance and value of the resource in systematic ways. Even so, these are bounded systems: each signal is linked to a specific response, and the expressive power of the system as a whole is fixed at a time.

The Sender–Receiver framework is conceptually clear and formally tractable, and has thus been a very productive way of thinking about the evolution of communication. This modelling work suggests that the evolution of signalling is robust. The formal work suggests that partial or full signalling equilibria emerges between agents (a) even if agents have extremely limited cognitive resources; (b) with very partial overlap of interests; (c) even if the receiver has a compromise option (analogous to hunt-the-hare in a Stag Hunt game) that does not depend on a reliable discriminating signal from the sender; (d) even if there is noise and ambiguity in the signalling; (e) in the face of environmental biases that reduce but do not eliminate the value of the sender's signal; (f) in multi-sender, multi-receiver networks. In particular, in the evolution of communication literature, there has been an enormous focus on the evolution of honest signals in situations of conflict of interest, and the role of signal cost as a guarantee of signal honesty (Maynard Smith and Harper 2003; Searcy and Nowicki 2005). This work on signal costs and honesty has not always been explicitly framed in the Sender–Receiver framework, but these analyses, are, I think, special cases of that more general framework. They show that in an important class of conflict of interest cases, for the system to be stable, the choice of signal is constrained: only a small class of equally costly alternatives are possible (see for example Laidre and Johnstone 2013; Biernaskie et al. 2014).

Simple models are often at their most informative by making salient and sharp mis-matches between intuitive expectation and real-world phenomena. The link between costs and honesty noted above is such a case, for in the simplest version of Sender–Receiver models, talk is cheap. So one way this framework reveals its power is by showing that this apparently innocent assumption about signal costs conceals something important. Here in another example: the Sender–Receiver framework makes salient a feature of the natural history of animal communication that would otherwise be easy to miss. Most animal signalling involves agents signalling about their own state rather than signalling about the external environment. Warning calls are exceptions; these signal a state of that environment. However, the bulk of animal communication consists in mating displays, contact calls, juvenile begging calls, territorial calls. These all regulate agent/agent interactions, signalling agent intentions and qualities. This form of signalling uncontroversially fits into the formal structure of the Sender–Receiver framework. Senders are flagging a feature of the environment unobservable by, but relevant to, the receiver; namely one of their own internal states. Even so, it is puzzling that in learning from one another about their common environment, animals tend to rely on cues[2] rather than signals, despite the modelling results that suggest that signals should readily emerge. That seems to be true even of those animals, like social carnivores, that rely on collective action. Lions and wolves do not seem to communicate about their external environment [it is possible, I suppose, that they do so cryptically: an instance of Dawkins' and Krebs' "conspiratorial whispering' (Dawkins and Krebs 1978)]. The framework reveals a puzzle. Perhaps the costs and benefits of communication have been misconceived and there is a selection-based constraint on signals about the world that hominins somehow escaped[3]; perhaps the information gradient in groups of social animals is typically low.

With some reservations then, we have a decent theoretical and empirical grip on the evolution and stability of animal codes; one which makes sense of seemingly puzzling cases, as in prey signalling to predators. Even so, human languages are manifestly immensely richer in their expressive power, and vastly more flexible, than any animal code. Indeed, it is very striking that in no other lineage has an opened-ended system of learned signals evolved. Two questions these facts raise are: (1) Did human language, or human proto-language,[4] evolve incrementally from an older, code-like system? (2) If so, are those intermediate stages, and the transitions between them, well captured in a Sender–Receiver framework? How well do our theoretical tools for explaining the evolution of animal signalling systems extend to hominins? Scott-Phillips defends a negative answer to the first question, so the second hardly arises. I shall defend a positive answer to the first, and incline towards a negative on the second.

Scott-Phillips suggests that the features that make the Sender–Receiver framework such a good model of animal communication militate against its being a good model of language: the fact that codes do not rely on intelligent agency, and that the signals that form a code are tied to the immediate context of perception and action. It is true of course that the application of this framework to communication is compatible with the agents being intelligent and cognitively sophisticated. David Lewis originally developed the framework to apply to human interaction, and our automatized expression of basic emotion—fear, horror and the like—can probably be analysed as a code. But one important insight of this modelling tradition is that intelligence is not necessary for code-based communication: codes do not rely on intelligent agency. Thus Scott-Phillips argues that codes are inflexible both synchronically and diachronically. However cognitively sophisticated the sender and receiver may be, in communicating with a code, the sender chooses from a pre-defined set of options, each of which maps onto a specific state of the world, and

[2] A cue is an act (or state) of an agent that can carry information for another agent, but which is not designed or intended to do so. The escape behaviour of one bird is an indication of danger to another, but the bird is not signalling, it is escaping.

[3] Information-sharing is a form of cooperation, and there is analogous puzzle about reciprocal cooperation that suggests that the costs and benefits have not been properly identified. Just as with signalling, models of cooperation also indicate that reciprocal cooperation should evolve quite readily. Models suggest that reciprocal cooperation does not depend on ecologically implausible cost–benefit ratios, or implausible rates of future interaction. Yet there are very few clear examples (see Boyd 2016 for a discussion of this puzzle).

[4] I mention proto-language here to side-step the issue of syntax, and whether it could evolve incrementally.

calls for a specific response. In David Lewis's example from the American revolution, the sexton of the Old North Church in Boston had two signals from which to choose, each with a fixed meaning about the direction of the British invasion. If the British command had divided their forces, and come by both land and sea, the sexton would have had no appropriate signal to send. Diachronically, codes grow through the associative connection between proto-signal and state of the world. A proto-signal often begins life as a cue, a sender's response to the state of the world which is adaptive independently of any audience response, and it becomes a signal, as its production becomes reinforced by audience response, in either individual life history, or over evolutionary time.

No-one doubts that Sender–Receiver models are idealised. Obviously, in the life of actual organisms signalling to one another, the relation between world state and signal; between signal and response; between response and payoff is not clean. There will be misfires at each stage; there is noise in the system. But noise is not adaptive flexibility. As a consequence of this analysis, Scott-Phillips makes two strong anti-incrementalist claims. First: he suggests that the taxonomy of communication systems into codes and into ostensive-inferential communication is exhaustive. There are only two kinds of communication systems. Second, he argues that there can be no incremental transition from codes to ostensive-inferential communication. I disagree. In the next section I discuss ways in which hominin communication might have become more flexible and less code-like over time, and then consider how flexibility at a time—the agile use of social resources discussed in "Ostensive communication and incremental evolution" section—might evolve incrementally. But I shall first prepare the ground with a brief discussion of great ape gestural communication; our best model of communication amongst the first hominins. In one important way, great ape gesture is inflexible in the way Scott-Phillips has in mind. Even so, already with the great apes (and hence presumably our last common ancestor with them) there are aspects of their communicative capacities which could, and I think did, expand incrementally.

Great ape gestural repertoires are quite large: for example, it is claimed that gorillas have (at least) 102 distinct gestures. But the repertoire does not seem to be readily expandable, and so has some of the inflexibility of typical animal codes. For example, in Genty and colleagues' analysis of gorilla gesture, the menu seems to be largely drawn from the species-typical behavioural repertoire of gorillas, but bought under intentional, top-down control (Genty et al. 2009). Gorillas seem to have evolved the capacity to use their gestural options more flexibility and less reflexively, but have limited ways of adding new gestures to their communicative toolkit. These codes seem to have expanded only by associative learning, as directly functional action patterns become abbreviated and ritualised as signals. There is some controversy over this, for there are claims for a much more creative use of gesture in pantomimes by great apes, including the pantomime by a female chimp to her child of a nut opening sequence (Russon and Andrews 2011a, b). So there are reports of on the fly signal expansion. These reports suggest that when an initial signal failed to elicit the desired response, the signal was not just repeated or even repeated with amplification; it was extended and elaborated; for example the target of a request is offered a different and better tool with which to comply (for example, a better back-scratching stick). However, these are mostly cases of captive great apes, and they rely on observer interpretation of one-off events, so obviously they need to be treated with much caution.[5]

The best guess then is that great ape diachronic flexibility is very limited. On the other hand, some important aspects of great ape communication are not code-like. For observation of great ape gesture strongly suggests that there is nothing like a simple, one-to-one relationship between gesture type and desired response. In a helpful review, Catherine Hobaiter and Richard Byrne systematise reports on chimp gesture, and that systematisation shows some gesture-to-response specificity[6]—gestures do not just mean: pay attention and do something[7] (Hobaiter and Byrne 2014). But there are several gestural routes to each desired response,[8] quite often given in series, when the communicative partner is initially unresponsive. And the one gestural type can be and is used to solicit different responses. So arguably, in great ape gestural communication, we have the dawn of pragmatics in our lineage. The sender has to choose a gesture from a range of options, at the very minimum choosing one that is appropriate for the sensory modality. Great apes manage that: "Gesturing of great apes is appropriately adjusted to the attentional state of the recipient. Silent, visual gestures are given mainly when recipients are looking; audible, visual gestures less so; and tactile (contact) gestures are given indiscriminately of the audience's attention" (Genty et al. 2009, p. 528). They seem to show at least some awareness of what others already know (Crockford et al. 2012). Given the ambiguity of gesture, the target of the gesture has to use contextual cues—some of which might be quite subtle—to identify the expected response. Observational data suggests that great ape audiences manage this quite well. There are plenty of reports of gestures being ignored, but there does not seem to be reports of frustration and conflict being caused by responding, but in an unintended way. Despite many limitations, if our early hominin ancestors were like these great apes, they were beginning to tune into one another. Moreover, there is some evidence that great apes are sensitive to common ground: a recent experiment gave chimpanzees the choice between requesting a moderately desirable food, or requesting better food by pointing to a plate where such food had been. Would they make such points only to an audience who knew what had been on that plate? There was some evidence of that sensitivity (Bohn et al. 2016).

[5] Moreover, some of the reports are at best very marginal cases of pantomime "Orangutans groomed a partner briefly to solicit grooming; so do chimpanzees and gorillas" (Russon and Andrews 2011a, p. 315).

[6] See in particular their tabulated summary on Hobaiter and Byrne (2014, p. 1598).

[7] Or so they argue. However, Richard Moore points out that they have not excluded the possibility that some of the signals are just attention grabbing, with response specificity depending on context and the audience's best guess at what the target wants (Moore 2014).

[8] Identifying the desired response is not at all trivial. If the target of the gesture responds in a way that causes the gesturing agent to stop gesturing, and if that agent does not show obvious signs of frustration and anger, then the response is deemed to have been the intended outcome of the gesture.

Growing a code

Let me begin by pointing out that codes do not expand only through associative mechanisms and their kin. Consider the following few examples:

These road signs are part of a code: they signal a quite specific feature of the local environment and they call for a quite specific response from each receiver, though in contrast to codes with a Sender–Receiver dynamics, the code is not stabilised by positive payoffs to sender and receiver when the receiver acts on the signal. In particular, these signs are not ostensive-inferential icons. Though (presumably) each token icon was erected in place by some identifiable agent, the intention of that agent in erecting the sign is irrelevant to the sign's communicative function. Indeed, almost certainly, those intentions were not communicative at all:

the signs are put in place by tradesmen carrying out a work order. There is, doubtless, a committee of the Department of Main Roads that determines the location of these signs, but unless one is prepared to make very bold claims about the intentions of institutional agents, Departments and their committees do not have Gricean intentions. Yet these signs clearly did not establish as part of the road sign code system by anything like an associative process, and nor did they have an earlier life as cues.

The reader might at this point be tempted to think that these examples are irrelevant. Readily expandable codes do indeed exist, but their existence depends on language, and so their existence is irrelevant to language's evolution. I shall suggest that this is the wrong way to read these examples. The road code illustrates two dimensions of codes underplayed in Scott-Phillips discussion. First: some communication systems are institutional or collective, and the highly intentionalised analysis of ostensive-inferential signals do not apply to such collective signals. As noted above, institutions do not have Gricean intentions. Moreover, there is no reason to suppose that collective signalling—that is, signalling by institutional agents—emerged only very recently in human history. Very plausibly, much ritual is a collective signal.

Whether the second dancer on the left would rather be elsewhere, or hopes no-one will recognise him, is irrelevant. The communicative function of the joint display is independent of his specific attitudes to it. In giving an account of the emergence of language from much simpler systems, one change is an increase in sophistication and flexibility of individual-to-individual interaction. But a second factor is that communication no longer depends solely on the traits of individual agents. Ritual is a simple example of collective communication, but it's an aspect of language as well, as Putnam pointed out in describing the division of linguistic labour (Putnam 1975).

The second and most critical point illustrated by these examples is that as the cognitive sophistication of agents increases, new options for expanding a code come into play. On everyone's view, there were incremental changes in human cognitive capacities over the last three or four million years. Our ancestors acquired technical skills; navigational skills as home ranges increased; greater capacities to read their physical environment, as new resources became important; a richer understanding of their local natural history; and of course greater social skills as well. We do not have good frameworks for identifying the capacities of agents who are not just association engines but who do not have the full sapiens toolkit. Even so, we can reasonably suppose that over this period, there were increases in our ancestors' capacities for causal reasoning, abductive inference, the ability to combine information from different sources and from memory; planning and conditional reasoning, as well as theory of mind. If so, I suggest that it is extremely implausible to suggest that hominin codes remained diachronically inflexible—expanding only by association—until the full sapiens repertoire of language and cognition came on stream.

Here are two potential levers of expansion. First, gesture and mime. Many of the road signs have an iconic element. That is no accident; it means that visitors with no English can make a guess at sign meaning. What are the minimum cognitive capacities a sender and a receiver would need, for them to use a readily expandable iconic code? This question is quite important, as there is a quite persuasive line of thought suggesting that hominin communication initially expanded as gesture and mime, in part because gesture and mime can be iconic.[9] The resemblance between signal and target can support signal comprehension. What cognitive capacities would an agent need, in order for them to be able to introduce, with reasonable prospects for success, miming a hunt as a recruiting signal for a hunt?

Here is a first pass (it is no more) at specifying those requirements: (1) the agents must have some experience with communication. Each realises that sometimes other agents orient towards them, invading their field of perceptual attention, and act so as to induce a response (2) they must have the motor routines involved in hunting under top-down, executive control; they must be able to act as if hunting without the physical stimulus of the presence of prey. Moreover, when they want to hunt, the motor routines of hunting must be salient to them: they are cognitively ready to hand for attempted communication. (3) They must have the intention to direct their acts at others, to be noticed, and believe that their acts are apt to influence audience behaviour, in ways that they want. (4) They must be prepared to experiment. (5) They clearly must have some theory of mind capacities: they must track their audience's field of view, and have the capacity to influence that perceptual focus. They need the notion of an agent having a goal or a purpose, for they aim to

[9] In Genty and Zuberbuhler (2014) it is claimed that bonobos sometimes solicit sexual partners with a beckoning gesture allied with orienting their body to a specific location, and that this has iconic as well as spatial elements. But at best the iconicity is minimal, and its not clear that uptake depends on the iconic character of the gesture (supposing that it is iconic). There is also some suggestion that auditory iconicity (or sound symbolism) may also have helped bootstrap language (see Imai and Kita 2014).

influence those goals. (6) Perhaps all else they need is causal reasoning: perhaps they have noticed that when A prepares to hunt, B is induced to do the same. That is to say: perhaps they have noticed contagion effects and on the fly, in the moment co-ordination triggered by mutual sensitivity to one another's the goal-directed behaviour. And so they come to believe: if I act as if I am hunting, and make sure that others notice this, perhaps others will come and hunt with me.

On the audience side: uptake depends on the similarity between the sender's acts and actually hunting, and perhaps also the associations between the mime and preparing to go hunting: for example, collecting and checking the equipment. Intentions leak: agents with fully modern minds engage in vacuum activity when they are forming a plan, partially deploying their motor routines in such circumstances. Hikers, climbers, birders fiddle with their equipment, re-check their gear and so on. And if Anatomically Modern Humans act this way, we can suppose our ancestors, with less well developed executive control did the same, fiddling with their spear, retouching the blade on their handaxe—"vacuum hunting", before actually hunting. These associations between preparation and activity can make the mimed activity salient. Communication is scaffolded by cues. The audience thinks "what does he want me to do" (so again, they need to concept of a goal or a purpose) and the mime makes hunting salient. Thus the audience recognises both the agent's goal of claiming their attention; their goal of influencing their (the audience's) behaviour, and their goal of hunting in company. Second order intentional capacities are in play. Presumably, once communicative mimes of this kind become an established practice in the community, it becomes less difficult to add a new mime: the fact that the sender is trying to influence your behaviour somehow will be more salient to the audience. The cognitive capacities in play here are quite sophisticated. They very likely exceed those of great apes. But the practice of depictive mimes requires nothing like the facility with nested intentional states (beliefs about intentions to induce beliefs) that ostensive intentional communication supposedly involves.

Here is another possible route to an expandable code. Kim Shaw-Williams has made a persuasive case that trackway reading—following the trails of humans and animals—was a preadaptation for language (Shaw-Williams 2014). Tracks are natural sign system; they carry information about the agents that made them. But they are an unusual system. For they are displaced in space and time from the track-maker about whom they carry information. In contrast to scent trails, they are directional, and in favourable conditions they are persistent. To use a trail, the tracker needs to read both directionality and age. On favourable substrates (like muddy foreshores) they have some combinatorial structure, as one print is superimposed upon another; so there is information about the order of passage. Tracks can tell an experienced tracker a lot about the age, condition and motivational state of the track-maker (the gait changes if the agent is tired or hurt). To those who know what to look for, the footprints of everyone in one's group are individually distinctive and recognisable. There is therefore a lot of ecologically and socially important information in tracks, and there is evidence from the Laetoli footprints that hominins have noticed and attended to footprints for more than three million years (great apes seem not to).

Trackers do not just read signs; they can make them [it is one way of teaching (Morrison 1981)], and of course they can unmake them, should they wish to conceal their own passage. In the simplest version of a shift from tracks (a cue) to signals, one's own tracks can be co-opted as signals, by deliberately making and even exaggerating them (making sure your prints are clear and distinct) as an add to navigation and group re-formation. An agent can use his or her own deliberately placed tracks to find the way back to base. Likewise, other natural signs of passage—branches and foliage broken as trails go through undergrowth—can be co-opted and exaggerated as signals of passage. Many bushwalkers will have used this technique themselves when exploring unfamiliar areas. Finding your way home is not a trivial problem and failure can be very costly. So in the very simplest version of the cue to signal transition, agents are sending signals to their future selves. But it is a very moderate expansion of this capacity for agents to send signals to those who are following on behind. There seems no reason why a simple coding system of this kind could not be readily expanded: for instance, adding directionality for others following by (for example) laying branches on a trail fork, pointing in the direction of travel. Likewise, a small party could readily indicate its membership, and changes in its membership if some members split off, by making their own prints distinct (rather than stepping into others' prints, as at Laetoli).

The fact that great apes do not exploit, let alone reproduce, visual tracks suggests that trackway reading, even in its elementary forms, is quite cognitively sophisticated. The use of such a code requires senders to know what others are likely to notice and how they are likely to respond. Readers (if they see the trail marking as signal rather than cue) need some theory of mind. They need to be able to think that the others want us to travel on this fork, not that one; or that four of us have gone one way, two the other. But there is no need for the full Gricean apparatus. We can see here two incremental shifts. One is from cue to sign to ostensive sign: the branches forming an arrow at the trail fork have to be read as a signal, not a cue. The footprint that results from the deliberate choice of a soft substrate is a signal, but could be read as a cue. But we also see the hint of a transition from fixed-response signals to decoupled signals: to signals providing information in a cooperative social context that scaffold the navigation and decision making of others, but without specifying a fixed response. Knowing the way to base camp can be useful, even if you are not currently heading back home; likewise, knowing the identity and direction of movement of a small foraging group, in making your own decisions about where to search. Let me emphasise that this analysis is intended to illustrate plausible possibilities, not paleoanthropological certainties. I think it is likely that tracks and trackway reading played some quite important role in our cognitive evolution, and think it even more likely that mimetic communication did. The point, though, is that the examples illustrate how minds intermediate between those of Anatomically Modern Humans and of great apes could develop and use a diachronically flexible code.

As the use of these systems becomes more flexible, the Sender–Receiver framework applies less cleanly. In principle, a fair bit of flexibility can be accommodated. A response to a signal can be disjunctive, probabilistic, or conditional on other inputs. Likewise, an agent might have only a tendency to signal in response to information about the environment; a tendency that might be modulated by other environmental, social, or internal factors. The framework does not presuppose that signalling and responding is reflexlike. But in an important sense, the framework is atomistic. The signs can be part of a system of signs, or even have signs composed out of other signs. But at least in the way this framework is formally modelled, each sign is stabilised, one by one, by the particular costs and benefits that flow from using that particular sign. The sign system is not stabilised by selection for its benefits as a whole. Rather, Sender–Receiver models represent the benefit of signalling, benefits that stabilise the system, as accumulating from each atomic interaction. By the time language has evolved, or anything that approximates language has evolved, this representation is obviously not realistic. Many, perhaps most, interactions do not issue in any identifiable action. Moreover the basic atom of language use is a two-way, or many way, conversational interaction, often quite extended, not a unidirectional flow of information or instruction from speaker to audience. I suspect a shift away from payoffs as captured in the Sender–Receiver model happened quite early in the expansion of hominin communication: informational pointing is not a Sender–Receiver interaction, for there is no specified response. Likewise, if there is anything to the suggestion that trackway reading preadapts for signalling with tracks, it too begins to stretch the Sender–Receiver framework for analysing the emergence and stability of signals. As specific signals become less systematically associated with specific responses (and hence specific payoffs), selection[10] begins to act on signalling and responding behaviour as a whole: the evolutionary grain changes from selection on signal/response couplings to selection on communicative behaviour as a whole.

Pragmatic competence

I noted in the introductory section that the pragmatics-focussed view of the evolution of language puts great emphasis on our capacity to co-opt signals with conventional meanings to communicate more than (and sometimes other than) those meanings, and on our capacity to manufacture meaning in the absence of signs with fixed meetings. The church sexton would probably not have given way to despair and done nothing; perhaps he would have manufactured a new sign, hanging out three lanterns to signal that something was amiss. Scott-Phillips and his allies do not just stress the importance of this capacity, they claim that it depends on very rich cognitive capacities; ones that probably appeared late in hominin history. If so, speaker meaning and the decoupling of signals from their immediate context was, likewise, a recent and abrupt innovation in hominin history. On their picture, for the sexton at the Old North Church to mean that the British are coming by both land and sea, he needed rich metarepresentational capacities. He had to intend that his audience recognise that he intended to communicate by hanging out three lanterns; that he intended them to recognise that he was intentionally not using one of the pre-

[10] The same is true of reinforcement or reward, if we think of these changes as happening within the lifespan of individual agents rather than in their lineage.

arranged signals, but that he was signalling something relevant to those signals; that he intended is audience to infer that he intended them to believe that he believed that the British were coming by both land and sea.

Adult, fully competent humans are indeed able to modify signals on the fly in unanticipated circumstances with fair prospects of success. The sexton's three-lamp signal would not have been hopeless. Moreover, in more routine circumstances they are able to signal to one another, often in speech but not always, with great success and little apparent effort. That said, it is not obvious that routine adult interaction is as cognitively demanding as the Gricean analysis represents supposes. Furthermore, competent adults are not the only agents who speak and listen. Quite young children—3 year olds—are linguistically competent. No doubt they do not have the fluent adult mastery of pragmatics, or the extensive adult lexicon and full mastery of syntax. But their communicative powers vastly outrun those of the great apes, and that poses a challenge to the idea that code-like communication cannot become fully fledged ostensive inferential communication by increments. Ontogeny does not recapitulate phylogeny. Three year olds are not living models of erectine or Heidelbergensian communication. However, arguably, they are proof of concept: they are not using a code, but there is no evidence that their utterances have speaker meaning, as the Griceans conceive of speaker meaning. For even if infant theory of mind experiments show that toddlers have a fully-fledged concept of belief, there is no evidence that they can represented nested beliefs and intentions. On my view, they intend to elicit a response from their audience, and this intention is overt, and often directs interventions on their audiences' attention. But they do not have the looping meta-intentions of the Gricean analysis. Arguably, the same is true of much ordinary adult communication.

Scott-Phillips and his allies understate the extraordinary cognitive complexity implicit in their picture of routine linguistic communication. If they are right, in routine speech a speaker has two high order intentional states (a communicative and an informational intention), and chooses linguistically and contextually appropriate clues that enable the audience to inferentially recognise and respond to those clues, recognising the informational intention. That context is rich: it includes common knowledge, not just spatial setting. All this very fast, as speakers process phonological signals above their maximal rate for non-linguistic acoustic information (Christiansen and Chater 2016). It is shifting, as topic and context change as conversation proceeds; apparently effortless; with a low error rate. Humans have little introspective access to their own cognitive mechanisms, so perhaps not we cannot put much weight on the idea that this picture seems extraordinarily complex. That said, the Griceans often use special notational devices, displays and the like, to make their analyses of communicative and informational intentions more user-friendly. Moreover, while Scott-Phillips presents experimental evidence in one study to show that adults participants do significantly better than chance at problem solving depending on high order mentalising, and in another study, they do surprisingly well even at very deep levels of recursion, this hardly suggests that it is effortless and routine, in the ways that it would have to be, if their picture of conversation were right (Cathleen et al. 2015; Scott-Phillips 2015, pp. 73–74). But the most serious empirical challenge is posed by the fact that 3 and 4 year old children have quite good conversational skills but struggle with the false belief test. That struggle suggests that processing multiply embedded intentional representations of the form “X intends that Y believes that X intends that Y believe that p” is not for them fast and effortless. This developmental evidence suggests that such children are not capable of formulating or recognising Gricean intentions; or, at least, that doing so is so cognitively demanding that it disrupts other tasks.

Scott-Phillips resists this whole line of argument. The cognitive psychology of vision makes it seem like an extraordinarily demanding challenge: constructing, under extreme time pressure, a three dimensional representational representation of the world from an intrinsically ambiguous two dimensional retinal image. Yet we pull of the trick effortlessly and reliably. Likewise, he suggests that high order mentalising might be like perception: fast, automatic, effortless for the agent, even though it seems cognitively complex to the theorists. He says: “More likely, it is, like simple mindreading, something we do habitually and subconsciously, as part of our everyday, low-level perceptions of the world around us” (p. 73). It is indeed quite plausible that children, perhaps even very young children, have some kind of automatic, subdoxastic, relatively encapsulated quasi-perceptual, quasi-modular system that registers the perceptual field of those that they interact with, and enables them to anticipate behaviour in the light of other’s perceptual field (even when their own perceptual field is saliently different). Such a system would be automatic and relatively effortless for the agent in question. It would not, for example, impose heavy demands on attention. Encapsulated mutual tracking could then be integrated smoothly with coordinated or interactive behaviour which is less routine, and more demanding of cognitive resources.

Scott-Phillips supports the analogy between mind reading and vision by appeal to the implicit mentalising capacities apparently shown by very young children, through the differential looking time paradigm (Onishi and Baillargeon 2005). He suggests that these infant theory of mind experiments show that the necessary mentalising capacities are in place very early (and to deny that high order intentionality is cognitively demanding). As he reads these experiments, they show that 3 year olds do have the theory of mind capacities needed for ostensive inferential communication. However, even if one were to accept the toddler looking time experiments [and there are sceptics: (Heyes 2014)], the Gricean analysis of routine linguistic communication does not just depend on high order mentalising, it depends on very fluent and skilled high order mentalising: the capacity in real time, under the temporal constraints imposed by the speech stream, to smoothly integrate high order mentalising, registration of the context of interaction, and the speaker’s awareness of common knowledge and its limits. For only thus (given the analysis) can the speaker select highly relevant linguistic clues to her communicative and informational intentions; only thus can the audience recognise those clues and recover the speaker’s informational intention. The Griceans cannot suggest that language processing interferes with or disrupts mentalising, for on their view, interpretation must smoothly mesh with language processing.[11] A processing load

[11] Ron Planer has pointed out to me that some defenders of the ostensive intentional conception of communication think that children younger than four fail language-based versions of the false belief task

explanation of a 3 year old's failure to pass the language-based version of the false belief test is inconsistent with their basic model.

So even if, as I think likely, we have an encapsulated perceptual tracking-action anticipation system, that is quite different from pragmatic competence as described. Providing and reading clues to communicative and informational intentions goes far beyond registering one another's perceptual field and taking into account the constraints of perceptual field on the control of action. Social context; previous history of interaction; non-perceptual common knowledge; all of these are important to the choice and interpretation of clues. Fodor was right to suggest in Modularity of Mind that while phonology and syntax were candidates to be encapsulated modules, pragmatic competence could not depend on an encapsulated system (Fodor 1983). Ian Apperly and Stephen Butterfill have argued that human mind reading depends on two systems. There is a developmentally canalised, early emerging, and fast acting module-like system, with roughly the characteristics noted above. It enables agents to track field of view, differences in field of view, and to anticipate actions based on those differences. In addition, adults do have a full-blown theory of mind, but that is a system that emerges more slowly (and is probably significantly affected by learning and differences in culture); it is more accessible to introspection, and in challenging cases using it requires attentional focus. In principle it enables us to track other's mental states on the basis of anything we know, not just what is available in the immediate scene (Apperly and Butterfill 2009; Apperly 2011).

Three year old children are reasonably competent pragmatically: they understand jokes and other non-literal uses of language; in more obvious cases they understand what a speaker means, not just what he or she says. In view of the fact that they usually fail linguistic versions of the false belief task, we can reasonably conjecture that their competence is largely based on the automatic system, supplemented at most by a rudimentary theory of mind. If so, they are an existence proof of intermediate possibility. Ron Planer and Richard Moore have developed explicit models of those intermediate possibilities; of quasi-overt inferential communication (Moore 2014, 2015, 2016; Planer 2017a, b). In sketching one version of their idea, the cognitive science distinction between system one and system two cognitive processes will be useful. System one processes are fast, automatic, demand little attention (and hence can be parallel), not available to conscious introspection. System two processes are slower, under executive control, demand attention (hence are serial), available to conscious introspection (Kahneman 2011). Using that distinction, here is an illustrative possibility: the speaker makes his or her communicative act obvious: this depends on the speaker tracking the perceptual fields of the audience through the quasi-modular system. By saying *u*, the speaker intends the audience to come to believe p (or do p). And they in no way hide or conceal that intention; to the contrary, they make it obvious. In Richard Moore's terminology, they know how to address a communicative act to their target

Footnote 11 continued
because they cannot perform multiple mindreading tasks in the one interaction, rather than because they cannot integrate language use and mindreading. But that predicts that children younger than four are incapable of conversational interaction in groups of three or more, for all such interactions, on their own analysis, involve multiple mindreading requirements.

audience, and they do so with the aim of eliciting a response (Moore 2016). So as in the fully Gricean case, the speaker is both managing attention and managing information flow. But none of this need be explicitly intended (as a system two representations). Rather, their system-one guided behaviour engineers that causal connection, by guiding their communicative behaviour, ensuring that it is perceptually salient. An action can be overt without an explicit intention to make it overt. Very modest system two theory of mind capacities grafted on top of an encapsulated system could yield quite reasonable pragmatic capacities to use communicative tools with some flexibility; a flexibility that would increase as system two capacities became more powerful and fluent.

Conclusion

The message of this paper is both simple and intuitive. It does not minimise the vast gap between paradigm codes and contemporary human communication. Language (and protolanguage) use cannot be reduced to mastery of a code. Neither language nor many of its precursors are captured by the formal tools that explore the evolutionary dynamics of codes rather well. The set of possible signals is not fixed prior to the communicative interaction; agents attempt to and sometimes succeed in sculpting new signals on the fly, and the thoughts and intentions conveyed are not sharply constrained by the conventional meanings of those signs in regular use. As we have seen, the exact theory of mind and inferential capacities adults need to use language adroitly is a matter of active debate. But it seems clear that a good deal of cognitive sophisticated is required. It is my view that these cognitive capacities (whatever their exact character) evolved in part because our ancestors developed communicative systems that were much richer and more flexible that the codes described by Scott-Phillips and modelled in the Sender–Receiver framework. Three year olds have some theory of mind capacity and quite significant language skills. But they are not fluent masters of conversational interaction; they are not fully competent in the pragmatics of language. Our ancestors were not 3 year olds, but they developed ways of adding fairly readily to their signal repertoire, and also came to use existing signals in somewhat novel ways to communicate a message related to, but not quite the same as, an established regularity. Their response to signals was not stereotyped, but modulated by speaker and context. Hominin social lives became more complex, thus selecting for greater social intelligence, in many ways. But one was by living in increasingly communicatively complex environments. This added complexity was not just scale; not just a larger menu of situation-signal-response options. The point is banal, but it requires abandoning a dichotomised view of communication, recognising only codes and ostensive inferential acts.

Acknowledgements Thanks to Stephen Mann, Richard Moore, Ron Planer and Matt Spike for their comments on an earlier version of this paper; thanks also to the Australian Research Council for their generous funding of my research on human social and cognitive evolution.

References

Apperly IA (2011) Mindreaders: the cognitive basis of ‘theory of mind’. Psychology Press, Hove and New York

Apperly IA, Butterfill S (2009) Do humans have two systems to track beliefs and belief-like states? Psychol Rev 116(4):953–970

Berwick RC, Chomsky N (2016) Why only us: language and its evolution. MIT Press, Cambridge

Biernaskie JM, Grafen A, Perry JC (2014) The evolution of index signals to avoid the cost of dishonesty. Proc R Soc B 281(1790). https://doi.org/10.1098/rspb.2014.0876

Bohn M, Call J, Tomasello M (2016) The role of past interactions in great apes’ communication about absent entities. J Comp Psychol 130(4):351–357

Boyd R (2016) A different kind of animal: how culture made humans exceptionally adaptable and cooperative. Princeton University Press, Princeton

Cathleen OG, Kliesch C, Smith K, Scott-Phillips TC (2015) The ease and extent of recursive mindreading, across implicit and explicit tasks. Evol Hum Behav 36(4):313–322

Christiansen MH, Chater N (2016) The Now-or-Never bottleneck: a fundamental constraint on language. Behav Brain Sci 39:e62

Crockford C, Wittig RM, Mundry R, Zuberbuhler K (2012) Wild chimpanzees inform ignorant group members of danger. Curr Biol 22(January 24):142–146

Dawkins R, Krebs J (1978) Animal signals: information or manipulation. In: Krebs JR, Davies NB (eds) Behavioural ecology. Blackwell, Oxford, pp 282–309

Fodor JA (1983) The modularity of mind. MIT Press, Cambridge

Genty E, Zuberbuhler K (2014) Spatial reference in a bonobo gesture. Curr Biol 24(July 21):1601–1605

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

Grice HP (1957) Meaning. Philos Rev 66:377–388

Grice HP (1969) Utterer’s meaning and intentions. Philos Rev 78:147–177

Heyes C (2014) False belief in infancy: a fresh look. Dev Sci 17:647–659

Hobaiter C, Byrne R (2014) The meanings of chimpanzee gestures. Curr Biol 24(July 21):1596–1600

Imai M, Kita S (2014) The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philos Trans R Soc Ser B 369(1651):20130298

Kahneman D (2011) Thinking, fast and slow. Macmillan, London

Laidre M, Johnstone R (2013) Animal signals. Curr Biol 23(18):R829–R833

Lewis DK (1969) Convention. Harvard University Press, Cambridge

Maynard Smith J, Harper D (2003) Animal signals. Oxford University Press, Oxford

Millikan R (1998) Language conventions made simple. J Philos 94(4):161–180

Moore R (2014) Ape gestures: interpreting chimpanzee and bonobo minds. Curr Biol 24(14):R645–R647

Moore R (2015) Meaning and ostension in great ape gestural communication. Anim Cogn 19(1):223–231

Moore R (2016) Gricean communication and cognitive development. Philos Q 67(267):303–326

Morrison R (1981) A field guide to the tracks and traces of Australian animals. Adelaide, Rigby

Onishi KH, Baillargeon R (2005) Do 15-month-old infants understand false beliefs? Science 308(5719):255–258

Planer R (2017a) Protolanguage might have evolved before ostensive communication. Biol Theory 12:72–84

Planer R (2017b) Talking about tools: Did early pleistocene hominins have a protolanguage? Biol Theory. https//doi.org/10.1007/s13752-017-0279-1

Putnam H (1975) Mind, language and reality: philosophical papers, vol 2. Cambridge University Press, Cambridge

Russon A, Andrews K (2011a) Pantomime in great apes: Evidence and implications. Commun Integr Biol 4(3):315–317

Russon A, Andrews K (2011b) Orangutan pantomime: elaborating the message. Biol Lett 7(4):627–630. https://doi.org/10.1098/rsbl.2010.0564

Scott-Phillips T (2015) Speaking our minds. Palgrave-Macmillan, London

Scott-Phillips T (2017) Pragmatics and the aims of language evolution. Psychon Bull Rev 24:186–189

Searcy W, Nowicki S (2005) The evolution of animal communication: reliability and deception in signaling systems. Princeton University Press, Princeton

Shaw-Williams K (2014) The social trackways theory of the evolution of human cognition. Biol Theory 9(1):16–26
Skyrms B (2010) Signals: evolution, learning, and information. Oxford University Press, New York
Sperber D, Wilson D (1986) Relevance: communication and cognition. Blackwell, Oxford
Tomasello M (2008) Origins of human communication. MIT Press, Cambridge
Tomasello M (2014) A natural history of human thinking. Harvard University Press, Cambridge