

From monkey alarm calls to human language:
how simulations can fill the gap

Short title: Alarm calls to language

Jason Noble

University of Southampton
School of Electronics and Computer Science

J. P. de Ruiter

Max Planck Institute for Psycholinguistics
Nijmegen

and

University of Bielefeld
Faculty of Linguistics and Literary Sciences

Kate Arnold

University of St. Andrews
School of Psychology

Abstract

Observations of alarm calling behaviour in putty-nosed monkeys are suggestive of a link with human language evolution. However, as is often the case in studies of animal behaviour and cognition, competing theories are under-determined by the available data. We argue that computational modelling, and in particular the use of individual-based simulations, is an effective way to reduce the size of the pool of candidate explanations. Simulation achieves this both through the classification of evolutionary trajectories as either plausible or implausible, and by putting lower bounds on the cognitive complexity required to perform particular behaviours. A case is made for using both of these strategies to understand the extent to which the alarm calls of putty-nosed monkeys are likely to be a good model for human language evolution.

Keywords: individual-based modelling, language, communication, evolution, alarm calls, primates.

1 Prologue

Picture the scene: in a remote valley in the rainforests of eastern Nigeria, a group of putty-nosed monkeys (*Cercopithecus nictitans*) are feeding on fruits and seeds in the forest canopy. The group includes a single dominant male, eight females, and several juveniles. The male stops feeding and looks around, checking on the positions of the other monkeys, and scanning the area for danger. He sees a leopard in the fork of a tree, and calls loudly to warn the rest of the group: “Pyow! Pyow! Hack! Hack! Hack!” The monkeys respond immediately, looking around in alarm, grabbing and carrying juveniles, and scrambling away through the treetops. They travel about a hundred and fifty metres before deciding that this is a safe distance from the leopard. Gradually they return to feeding behaviour.

What is really happening in this story? There is much here of interest to students of the evolution of language. At first glance it seems that one monkey has become aware of a dangerous predator, and has decided to protect the rest of the group by warning them about it. The sequence of calls seems to carry meaning, as it certainly produces an appropriate reaction. It’s tempting to gloss the call sequence as meaning “Look out, there’s a leopard”, and to see the other monkeys in the group as having understood this message. It looks like a paradigmatic case of successful animal communication, but moreover — and perhaps this is partly because the animals involved are our primate cousins, and the channel for communication is vocal — it looks a lot like proto-language (Bickerton, 1990) in action. If this is genuinely proto-linguistic communication, it would follow that language-evolution theorists should be paying a lot of attention to these monkeys and other species with similar behaviours.

The bare bones of the above story are true (Arnold & Zuberbühler, 2006; Arnold et al., 2008) but much depends on the words used to express it. For

example, we used the verb “call” to describe what the male did after seeing the leopard, and we implied that he acted with the intention of warning the other monkeys in the group. We also used quotation marks around the transcription of the sounds he made, a convention normally reserved for human speech. If we were to make a less generous interpretation, we might have used the word “vocalize” instead of “call”, and studiously avoided ascribing any plan or goal to the vocalizing male. This would flag the possibility that the male monkey’s behaviour was somehow automatic, rather than the intentional use of a signalling convention in order to get the warning message across to the others. The story would start to look less relevant to language evolution.

On the other hand of course, we could have followed the lead of some researchers in primatology (Savage-Rumbaugh et al., 1998) and animal behaviour (Griffin, 1992) and cast off the careful scientific language that focuses on behaviour rather than mental states. We could have described the scene much as we would describe a conversation between human beings: “one monkey noticed the danger; he warned the others about it and the group decided to move away.” If we take this approach to its logical extreme, it suggests not so much that monkey communication is relevant to language evolution, but more that the monkeys already possess a communication system that is functionally equivalent to language.

There is a twist in the tale that is even more suggestive. It turns out that the sequence of calls made by the male is important, in a way that could be taken to suggest syntactic structure (Arnold & Zuberbühler, 2006). The male could have produced a sequence made up only of “pyow!” calls, which would often, but not always, indicate a leopard. Or he could have made a string of “hack!” calls, which would be associated (again, not reliably) with the presence of an eagle, another predator of the putty-nosed monkey that requires different

escape behaviour. The fact that the male produced a pyow-hack sequence seems to mean something like “Look out everyone, there’s a leopard and we urgently need to move away from it.” Arnold & Zuberbühler established this by playing back different call sequences and measuring the distance moved by groups of monkeys in response. The ordering of the two call types in the sequence of calls was found to be important in predicting the group’s reaction.

2 Outline of our approach

The situation with putty-nosed monkey alarm calls is typical of a wider problem in studies of animal behaviour and cognition: we have a growing body of observational and experimental data, and a range of competing theoretical accounts attempting to make sense of it all. But the theories are under-determined by the available data — consider the difficulty of deciding whether the pyow-hack sequence is descriptive (“there’s a leopard here”) or imperative (“run!”) based only on the kinds of experiments conducted by Arnold & Zuberbühler.

In keeping with the theme of this special issue, and with earlier arguments (see for example Todd, 1996; Di Paolo et al., 2000), we believe that the best way forward is to construct individual-based simulation models of the behaviour in question. By exploring the results of such models and comparing simulation outputs with real data, we hope to be able to show, for example, that theory A is a more plausible explanation of the data than theory B, that theory C is more complex than it needs to be, that theory D contains a logical flaw, that theory E could be tested with a novel experimental design, etc.

In section 3 we will show how simulation modelling can be used to examine evolutionary trajectories. We concur with Dobzhansky (1964) that “nothing in biology makes sense except in the light of evolution.” Good theories about animal communication systems (including human language) need to make evo-

lutionary sense, and, similarly, if a theory proposes a phylogenetic stage that is evolutionarily implausible, it should be dropped. It is not always easy to tell from the verbal statement of a theory, however, whether it is consistent with the theory of evolution. A notable example is the long-held belief by many intelligent biologists that communication could evolve “for the good of the species”. We will show how simulation models can help with this “quality control” process of ensuring that theories are evolutionarily plausible. This focus on the importance of evolutionary constraints leads us to a view on biological function and a definition of communication in evolutionary terms.

In section 4 we look at a second, complementary role for computational modelling: investigating the complexity of the cognitive mechanisms involved in performing a particular behaviour. It is a long-standing problem in the behavioural sciences that whereas the available data involve descriptions of behaviour, the objects of most interest to investigators tend to be the mechanisms underlying those behaviours. Inferring the nature of the enabling mechanisms from their behavioural products is another example of the under-specification of theory by data. Braitenberg (1984) showed that we probably have a tendency to over-estimate the complexity of the internal machinery required to produce a particular behaviour, but fortunately he also pointed to the usefulness of building simple models as a way of, at the very least, putting a lower bound on the internal complexity implied by an observed behaviour.

The way forward that we are advocating is not entirely novel, and indeed the existence of this special issue is evidence that some language-evolution researchers are already well aware of the possibilities of a productive marriage between empirical research and computational modelling (see also Kirby, 2002; Cangelosi & Parisi, 2002; Lyon et al., 2007). In section 5 we will return to our starting point with a detailed description of how simulation models could shed

light on alarm-calling behaviour in putty-nosed monkeys.

3 Communicative behaviour as the result of a history of selection

Why do we need formal models to help us show whether a theory about language or communication is evolutionarily plausible? Darwin (1859) gave us the theory of evolution 150 years ago; it can be adequately summarized as the effects of the combined processes of variation, selection and heredity. You might think that by now we would be used to expressing our theories about biological systems in evolutionarily consistent terms. The problem is that although Darwin's idea is easily communicated, and some intuitions about animal behaviour can be drawn directly from it, its full implications are not so easily seen. The evolution of communication or indeed of any social behaviour will be a complicated story involving the gradual elaboration of two or more complementary behavioural roles, e.g., signaller and receiver. Unaided human intuition has a poor track record in teasing out the various factors involved in this kind of historical process and accounting for an evolutionary trajectory from some hypothesized original state to a currently observable behaviour. Mathematical and, later, computational models have proven tremendously useful in showing why one trajectory and not another is likely to be taken by an evolving population (Maynard Smith, 1982, 1989; Belew & Mitchell, 1996; Grimm, 1999).

The evolution of communication seems to be an especially problematic case as we are, of course, language-using creatures ourselves. In other words, we are embedded in a highly evolved communication system which we use constantly and which shapes our thinking about what counts as a normal or typical communicative act. We are all familiar with the experience of having a thought or a

desire, finding words with which to express it, and more-or-less successfully getting the message across to other language users. We are guided by what Reddy (1979) and Lakoff & Johnson (1980) called “the conduit metaphor”: a mental state in the sender’s head must be accurately transferred to the mind of the receiver via a physical vehicle such as a spoken or written sentence. Note some of the assumptions embedded in this sketch: that both sender and receiver possess internal representations of the world, that sender and receiver will normally have a common interest in getting the message across, and that communication is at heart about information transfer, i.e., reducing the receiver’s uncertainty about the state of the world. It is this mindset that makes us so readily interpret the behaviour of the putty-nosed monkeys in linguistic terms. However, the conduit metaphor can get us into trouble when it suggests a communication system that is not evolutionarily plausible.

The early ethologists (Selous, 1901, 1933; Huxley, 1923; Tinbergen, 1952) pointed out that the evolutionary history of an animal communication system must begin with behaviours that don’t yet have a signalling function. This idea was sound and has been supported by later theoretical and modelling efforts (notably Quinn, 2001). Given that evolution is a non-directed process, it makes sense that a behaviour that later evolves into a signal cannot be functioning as a signal when it first appears, nor can it be selected for its future value to the animal. The behaviour must originally have some other function (e.g., panting to lower body temperature) or potentially no function at all (e.g., a nervous tic) such that the evolutionary process can elaborate it. Ethologists referred to these behaviours as “derived activities”. Tinbergen (1964) and Lorenz (1967) were particularly interested in a specific case: “intention movements” in which one behaviour logically had to come before another, such as the baring of teeth in preparation for biting. Bared teeth would thus be a predictor of biting and

would be excellent raw material for later elaboration into a signal. Tinbergen and Lorenz noted that many animal threat displays were of this form.

The ethologists were also interested in “ritualization”, the hypothesized process by which derived activities were elaborated, over evolutionary time, into true signals. This is of particular interest to us because this is where the ethologists — in the absence of a formal modelling framework — departed from evolutionarily sound reasoning. The assumption was that the accurate transmission of information would be of benefit to the species. Even in obviously competitive situations such as contests over resources, the ethologists reasoned that a species that successfully communicated information (like the intention, determination, and capacity to fight) would do better than a species without such communication, and that therefore the evolutionary process would favour the communicative species. For them, it followed that any mutation that made a nascent signal clearer, more obvious, or more exaggerated, would be selected for. This reasoning is now regarded as deeply problematic.

The logical error in the ethologists’ reasoning was not detected until some decades after their arguments were introduced. The framework that made it possible to identify this kind of flawed evolutionary argument was the *modern synthesis*. This was the joint effort by Fisher (1930), Wright (1931), Haldane (1932), and others to bring together Darwin’s theory and the genetic discoveries of Mendel in a mathematical framework. Why was this needed? The problem of blending inheritance was a keen one for Darwin; he knew that his theory would not work if inheritance led to offspring that were in every way a mixture of their parents characteristics, as any single advantageous mutation would eventually be lost in a sea of averaging. The particulate inheritance that Mendel observed in his “factors” — later genes — was exactly what was needed if an evolving species was to accumulate beneficial mutations. The new mathematical frame-

work allowed questions such as “under what conditions will this mutation go to fixation in the population?” to be asked and answered. The modern synthesis led to the gene-centred view of evolution (epitomized by Dawkins, 1976) and to the modern discipline of population genetics. Notably, it provided a framework for Williams (1966) to criticize the careless use of group-selectionist explanations that had become common in biological thought. The argument of the ethologists, that ritualization can be expected in a species because it will facilitate better communication, is just such an argument.

Another theoretical breakthrough of great relevance to understanding the evolution of communication was the introduction, by Maynard Smith (1979, 1982), of game-theoretic thinking to biology. Game theory was devised by von Neumann & Morgenstern (1947) and was originally intended as a model of human strategic behaviour, asking what constitutes optimal behaviour when two or more rational opponents face each other in a well-defined game. Game theory requires a scale for measuring the preferences of each player across the possible outcomes of the game; in economic contexts, this measure can be money or the more abstract concept of utility. Maynard Smith’s great insight was to see that fitness, measured in numbers of offspring, was the required success measure in the biological world, and that the mathematics of game theory could therefore be applied to the strategic interactions that occur between animals as they pursue goals like feeding, fighting, fleeing and reproducing. There was no need for an assumption that the interacting parties were both rational: the role of rationality was replaced by evolutionary history. Over many generations, animals practicing sub-optimal strategies would have fewer offspring and thus those strategies would be weeded out.

In economic game theory one of the major theoretical advances had been the Nash equilibrium concept (Nash, 1951). The key way to think about optimal

strategies in games was to identify optimal *pairs* of strategies for which neither party had any incentive to deviate from that strategy. For example, if two drivers approach each other on a country road and both are driving on the left, neither has any incentive to deviate and drive on the right or there will be a collision. Thus the strategy set (drive on left, drive on left) is a Nash equilibrium. Maynard Smith took the Nash equilibrium concept and applied it to the evolutionary setting, devising the related idea of an *evolutionarily stable strategy* or ESS. An ESS is a strategy or set of strategies that cannot be invaded by any mutant strategy — because the mutant player achieves a lower fitness than the ESS player when the two strategies interact — and is thus a stable point for evolution. The ESS concept is an excellent example of the way formal modelling techniques can help weed out bad theories: if it can be shown that a hypothesized communication strategy would not be an ESS, there is no point in proposing it to account for observations of animal behaviour.

When evolutionary game theory was applied to the arguments made by ethologists, i.e., that we can expect selection to favour increasingly communicative behaviour between members of the same species, the flaw in the logic became apparent. Suppose that the genetic interests of two animals are directly opposed, e.g., two males competing for access to the same female. Could animals in this situation nevertheless be expected to evolve a signalling system indicating to each other their level of determination to mate? The answer is no: a strategy that involves honestly signalling one's intentions in this case would not be an ESS. That is because mutant strategies of either lying about one's intentions or remaining poker-faced and giving nothing away will be able to invade a population of truth-tellers. Indeed, the evolutionary-game-theory perspective totally changed the landscape of thought about animal signals: honest signalling went from being something that would “obviously” happen for the good of the

species to being an unusual phenomenon in need of special explanation.

The advance of game-theoretic and population-genetic modelling led Dawkins & Krebs (1978) to propose a radical new “mind reading and manipulation” view of signalling (see also Krebs & Dawkins, 1984). Dawkins and Krebs note that it is often in an animal’s interest to manipulate objects in its world. Often the object in question is inanimate or immobile, as when a rabbit displaces earth while digging a burrow. On the other hand, sometimes the object is another animal: for instance, it is in the interests of predators to ingest prey, males to inseminate females, and territory holders to repel intruders. In manipulating the inanimate environment, an animal generally has no choice but to use its own muscle power. However, when one animal seeks to manipulate another, it can stimulate the other’s sensory system, thereby exploiting the muscle power of the second animal, and causing it to behave in a way that benefits the first. For example, a male frog does not actively move about seeking females, but instead sits in one place and makes sounds that cause females to approach him. His croaking can be seen as a way of exploiting the females’ locomotive muscle power.

Dawkins & Krebs (1978) suggested that animal communication should be defined in this way; that communication or signalling is what happens when one animal, the actor, has been selected to produce a response in a second animal, the reactor, such that the reactor’s behaviour (on average) changes to the advantage of the actor. Dawkins and Krebs acknowledge that such a definition moves a long way from our everyday understanding of the word “communication”, and that there is no implication, in their view, that animals should be transmitting information to each other in order to qualify as communicating. They admit (p. 283) that they are “tempted to abandon the word communication altogether.” Dawkins and Krebs argued that animal communication serves

not to inform but to persuade, and that advertising and propaganda are more apt metaphors than language for what goes on in the animal kingdom.

Mathematical modelling and evolutionary game theory in particular have been productive tools in helping to identify flawed reasoning in theories of the evolution of communication, and in suggesting new perspectives on what communication really is. However, they suffer from a serious drawback: for a mathematical model to be of use, it must remain simple enough to be tractable. Many game-theoretic models, for example, consider only a handful of alternative possible strategies — often just two. This is not because of any lack of imagination on the part of the model’s authors. The relevant equations for extremely complex games and strategic situations can be written down, but they cannot be usefully manipulated to arrive at a general solution. (Many of the games people play for fun, such as chess or poker, are too complex for game-theoretic analysis in their complete forms.) This is why computational models are also required (Di Paolo et al., 2000). Computational models have their own difficulties, including a potentially vast number of parameters and the potential for epistemological confusion with real data (for discussion of these issues see for example Grimm, 1999; Di Paolo et al., 2000; Bryden & Noble, 2006). However, with sufficient computing power computational models can be used to tackle far more complex scenarios than can be expressed mathematically. Indeed, we see no fundamental difference between mathematical and computational modelling: both are examples of a broader set of formal models, and mathematical models represent the subset that is amenable to analytic techniques. Computational models, in contrast, admit of no (currently known) short cuts to a solution and must, like a recipe, be carried out in full to achieve a result (Bullock, 1997). The two types of modelling are complementary rather than opposed.

Over the last two decades, work in computational modelling, and specifically

the use of evolutionary simulation models, has extended the game-theoretic perspective and told us much about how communication can evolve. Early work focused on simple, abstract communication systems (MacLennan, 1992; MacLennan & Burghardt, 1994) and then branched out to cover different contexts for animal communication such as sexual display (Werner & Todd, 1997; Bullock, 1998; Noble, 1999b) and aggressive signalling (Wheeler & de Bourcier, 1995; Noble, 2000; Quinn & Noble, 2001). Simulation models have been used to question the standard paradigm of communication as information-transfer, emphasizing coordinated action instead (Di Paolo, 1997; Quinn, 2001; Williams et al., 2008). Researchers interested in human language have looked at the multiple adaptive processes involved (Kirby & Hurford, 1997; Smith et al., 2003) and at the origins of reference (Cangelosi & Harnad, 2000; Vogt, 2006; Vogt & Divina, 2007; Donaldson et al., 2007). Kirby (2002) provides an excellent review of simulation work relevant to language evolution; Lyon et al. (2007) is a recent collection. An important general result has been that the evolutionary path to communication can be counter-intuitive: sometimes things that seem easy to evolve are in fact hard (such as alarm calls, e.g., Noble, 1999a) and sometimes things that initially look impossible turn out to work under the right conditions (e.g., honest sexual signalling under the handicap principle: Bullock, 1998). We conclude from this that building formal models is essential to making sure that theories about how communication evolved really do make evolutionary sense.

Some readers may feel that the argument presented so far is an abstract one about mathematical results in evolutionary theory and that it is not relevant to such traditionally social-science domains as language and culture. This would be a mistake. Although the human animal is certainly exceptional, and phenomena like “ratchet culture” (Tomasello, 1999) appear to be unique to our species, continuity between ourselves and the rest of the biological world is a

non-optional consequence of Darwin's theory (see Darwin, 1871; Tooby & Cosmides, 1992; Pinker, 1994, among many others, for more detailed arguments). The need for a plausible evolutionary history applies as much to haiku and political parties as it does to hacks and pyows.

Furthermore, we believe that in trying to explain the origin of phenomena such as meaning, reference, and intentionality, evolution is ultimately the only game in town. In her book *Language, Thought, and Other Biological Categories*, Millikan (1984) argues that, given a physicalist view of the universe, the only process that can give rise to something like purpose is natural selection. Millikan claims that the function of a biological phenomenon is determined not by looking at its place in a causal network in the here and now, but by examining its evolutionary history. Specifically, the purpose of a trait, or, to use Millikan's terminology, its *proper function*, is to do that which gave a fitness advantage to ancestral holders of the trait. In other words, the proper function of a trait — *what it's for* — is to do whatever it did in the past that led to its being here today. For example, suppose that a herbivore has a tendency to run from any sudden movement, and that this tendency leads, over many generations, to the differential survival of those who possess it because they are more likely to escape attacks by predators. The proper function of this tendency is therefore to assist the animal in evading predators, whether or not it actually achieves this on average, or indeed ever, in the modern environment. The modern trait is to be explained with reference to its evolutionary history.

Consider these positions together: Millikan on evolved function, Dawkins and Krebs on a new view of communication, mathematical and computational approaches to modelling evolutionary trajectories, and the insight from ethology that communication must begin with non-communicative roots. We are led to a typology of communicative and related phenomena (figure 1) where the

emphasis is on the history of selection behind the behaviours involved. We suggest the general term “influence interaction” to refer to an event where one animal acts in such a way as to influence the perceived states of the world, and thus alter the subsequent behavioural response, of a second animal. Note that the first animal’s action is itself a response to *its* perceived states of the world, and that the action’s effect on the second animal is mediated by the environment. In any given influence interaction, we can ask whether the actions of the first and of the second animal are fulfilling their proper functions in Millikan’s sense. The possible answers to these two questions (yes or no in each case) constitute four distinct situations.

[Figure 1 about here.]

First, it may be the case that influencing the behaviour of the second animal is not the proper function of the first animal’s action, and nor is the second animal’s response fulfilling its proper function. We refer to these cases as examples of “accidental influence” (the lowest labelled arrow in figure 1). For example, the vibration and noise caused by a pig rooting for truffles might prompt a mole to flee because it believed that a predator was approaching. The proper function of the pig’s behaviour is to uncover truffles; the proper function of the mole’s behaviour is to help it evade predators. The fact that the pig has influenced the mole in this way is in line with neither of these two functions.

Second, it is possible that the first animal’s action is fulfilling its proper function, but the second animal’s response is not. For instance, when the lure of an angler fish attracts a smaller fish, causing it to approach and be eaten, the lure display is fulfilling its proper function. The smaller fish’s approach response is not: its proper function is to guide it towards its own prey. Krebs & Dawkins (1984) call this “manipulation” and we will adopt the same terminology here. Specifically, the first animal is manipulating a response of the second that has

evolved for some other purpose.

The third possibility is that it is not the proper function of the first animal's action to influence the behaviour of the second, but that the second animal's response is in line with its proper function. In Krebs & Dawkins's (1984) terms this is "mind-reading". However, an example suggests that "exploitation" is a more general term: if the wind changes when a cheetah is stalking a herd of antelope, and they catch her smell and flee, then the antelope have exploited natural information about the cheetah. It is not the proper function of the cheetah's body to produce smells that will scare off antelope, but the proper function of such a response in the antelope is surely to keep them out of danger. Figure 1 shows both manipulation and exploitation as arrows in the centre of the diagram.

Finally, the behaviour of each animal in an influence interaction may be fulfilling its proper function. This is most easily seen in cases where the outcome is mutually beneficial: the dance of a returning bee and the subsequent directed foraging behaviour of its hive-mates are both fulfilling their proper functions. When both the action and the response are performed in accordance with their proper functions, the aspect of the first animal's behaviour that influences the second qualifies as being genuinely representational in the sense discussed by Harvey (1996) — i.e., that it matches the sense in which representation is a four-place predicate with P using Q to represent R to S). We refer to this class of interactions as "proper signalling" (uppermost labelled arrow in figure 1). This definition is not entirely original: Bullock (1997) defines "full-blooded signalling" in a similar fashion, and Oliphant (1997) is getting at much the same idea when he says that true signalling is what happens when an interaction is simultaneously exploitative and manipulative.

We present this typology because it emphasizes the explanatory template for

successful models of the evolutionary history of a behaviour like alarm calling or language use. The ethologists original point about “derived activities” means that all signalling systems must begin as cases of accidental influence: for no particular reason it happens that one animal’s behaviour influences another. The next development will be towards either exploitation or manipulation (or just possibly both simultaneously) depending on which party gains a selective advantage from the interaction. For example, the origins of the bee dance may lie in exploitation, whereby a novel strategy of following other bees when they left the hive was successful. Or it may lie in manipulation, in which mutant dancers managed to affect the take-off directions of other bees. The final step in explaining the emergence of proper signalling is to show how what was previously an exploitative or manipulative relationship comes to be in the interests of both parties. (Note that an immediate jump from accidental influence to proper signalling is inherently unlikely as it requires two fortuitous mutations to occur simultaneously.) The task for simulation modellers interested in a behaviour like monkey alarm calls or human language is to demonstrate an evolutionary trajectory along these lines.

4 Using simulations to make inferences about cognitive complexity

When we see an animal or a robot that exhibits complex behaviour, we tend to suppose that there are complex mechanisms behind it — but we can be wrong. Braitenberg’s (1984) book *Vehicles* begins with this point, and shows that surprisingly complicated behaviour can be the result of very simple internal mechanisms. Braitenberg is ostensibly talking about robots but his book is really a parable about the evolution of nervous systems. Simon (1981) uses a story

about an ant travelling across a rocky beach to make a similar point: apparent complexity in the ant's trajectory need not come from any complex cognitive mechanisms in the ant's head, but from the interaction between potentially very simple mechanisms (e.g., "turn left forty-five degrees when encountering an obstacle, otherwise go straight") and a complex environment.

We know from these kinds of thought experiments and from practical results in fields such as artificial life and evolutionary robotics that simple mechanisms can produce complex behaviour. We also know that evolution is in general a gradual process and, to a first approximation, can construct complex mechanisms only by the slow accumulation of successive adaptations. We therefore have good reason to model the evolution of cognition "from the bottom up", i.e., by starting with theories that propose the minimum possible complexity in the cognitive mechanisms behind a given behaviour. Braitenberg makes an additional point in favour of starting with simple models with his "law of uphill analysis and downhill invention": it is easier to reproduce a behaviour by starting with a simple mechanism and tinkering than it is to do so by analyzing behaviour and constructing inferences about what must have caused it.

What kinds of cognitive mechanisms are we talking about? What are the important differences in the cognitive faculties of real organisms, and how could they be captured in computational models? Dennett's (1996) book *Kinds of Minds* provides a hierarchical classification scheme that we can use to get started.

Darwinian creatures: simple automatons whose behaviour is governed by hard-wired sensorimotor connections. They are not capable of individual learning although the whole species may adapt over evolutionary time. An example: bacteria.

Skinnerian creatures: inherit the hard-wired responses of their Darwinian

ancestors but are also capable of individual learning. Positive and negative reinforcement will shape their behaviour accordingly. Named for the psychologist B. F. Skinner. An example: insects.

Popperian creatures: can learn by trial-and-error as Skinnerian creatures do, but also possess some kind of model or representation of their external environment such that they can test their plans before carrying them out in the real world. Named for the philosopher of science Karl Popper who suggested that this sort of mental modelling “allows our hypotheses to die in our stead.” An example: most mammals and birds.

Gregorian creatures: are Popperian creatures that have transformed their environment through the use of tools, technology, and culture. Dennett regards language as a tool in this sense. Named for the psychologist Richard Gregory. An example: human beings.

Dennett intended this classification as a general description of the different types of cognition in the natural world, but we note that it is specific enough to suggest implementable algorithms for different types of agents in a simulation model.

It is clear from years of empirical research (e.g., Cheney & Seyfarth, 1990) that mammals such as putty-nosed monkeys are capable of reinforcement learning, and moreover that they have at least basic kinds of internal models of the world, which puts them at the Popperian level. (Some non-human primates may even be at the Gregorian level, but this is contentious and exactly the kind of issue we hope to shed light on with computational modelling.) We therefore need to expand the potentially overcrowded Popperian level somewhat, by focusing on the degree of sophistication of the animals’ internal models of the world. Some of the simplest world models might be along the lines of a basic map, enabling navigation between a home base and distant feeding sites. More

complex models start to include representations of other animals, and even of the content of the world models of those animals.

This connects directly with a long-standing debate in primatology on *theory of mind* Premack & Woodruff (1978). Experiments with chimpanzees — involving food-sharing, hidden food, begging, and experimental confederates who may or may not be aware of food locations — have provided a fascinating window onto how the chimpanzee mind works when reasoning about other agents, but there remain very different interpretations of the results. On one side are sceptical researchers like Povinelli & Vonk (2003) who grant that chimpanzee world models include other chimpanzees, but not their cognitive states. Other researchers such as Tomasello & Call (1997) insist that the world models of chimpanzees include *beliefs about the beliefs* of other chimpanzees. The controversy is a perfect example of the under-determination of theory by currently available data (as Povinelli & Vonk, 2004, now recognize).

There is also a connection with studies of language. Grice (1969) put forward the case (later developed by Bennett, 1976) that considering different levels of sophistication of world-models allows us to pick out a special kind of communication. Grice and Bennett rely on the intuition that there is a difference worth marking between a situation in which causal automatons exchange signals (note that there is nothing preventing the simplest Darwinian creatures from having a hard-wired communication system) and a communication system in which participants really mean what they say. Their argument is that “real” communication can be roughly equated with human speech acts, and must involve, at a minimum, *third-order intentionality*.

To have first-order intentionality is to be a basic intentional system, i.e., to have beliefs and desires concerning the world, such as “I *believe* there is a predator nearby”, or “I *want* to mate with this animal”, but not to have any

beliefs or desires that are themselves *about* beliefs or desires. In other words, this is the simplest kind of Popperian creature. Second-order intentionality is to have beliefs (or desires) that can be about beliefs (or desires), such as “I *want* this animal to *believe* that there is a predator nearby.” Finally, third-order intentionality means being able to hold beliefs about beliefs about beliefs (and desires about desires about desires, etc.). Thus, we come to Grice’s formulation for a true speech act: that the speaker *intends* the hearer to *recognize* that the speaker *wants* the hearer to produce a particular response. For instance, if one person asks another to “please pass the salt”, then although the speaker wants the salt, she does not intend to exploit some salt-passing reflex in the listener, but rather that the listener should come to believe that the speaker wants the salt and therefore pass it to her. Grice and Bennett claim that this sophisticated form of communication is what distinguishes true language from simple signalling systems.

De Ruiter & Levinson (2008) (see also Levinson, 2006) argue for the idea that the evolution of communicative skills, i.e., the ability to map signals onto communicative intentions and vice versa, which underlies the Gricean process, necessarily *precedes* the evolution of language. There is indeed some evidence that these skills are “implemented” in humans in the form of an internal simulation model of the intentions of others (De Ruiter et al., 2007). Dennett (1988) also believes that only third-order intentional systems (or better) can “really” communicate. He gives an example (p. 188) of second-order intentionality that fails to qualify: “I *want* you to *believe* I am not in my office; so I sit very quietly and don’t answer your knock. That is not communicating.”

In his earlier work, Dennett (1987) proposed that we take an “intentional stance” towards cognitive agents that we are studying. This means assuming that the agents of interest (e.g., animals) are rational, and trying to predict

their behaviour from that assumption combined with educated guesses as to their beliefs and desires. Dennett contrasted this approach with an insistence on studying only low-level mechanisms, as neuroscientists or ultra-reductionist philosophers would do. There was no empirical claim that the animal *was* rational, only a pragmatic suggestion that we could achieve good predictive success by treating it as if it were. Dennett argues that this approach would allow us to test hypotheses about the order of intentionality involved in a communication system. For example, there may be some debate as to whether monkey alarm calls exhibit first- or second-order intentionality. If the former, then a calling monkey *wants* its hearers to run to safety, for example. If second-order intentionality is involved, then the caller may *want* its hearers to *believe* that there is a leopard approaching. Dennett suggests that careful experimental work could distinguish between these two hypotheses. Note that the second-order hypothesis, for instance, implies that the monkeys have some conception that other agents in their environment can have beliefs. If the monkeys never exhibit this ability — perhaps their occasional attempts at “deception” are always completely unsophisticated, indicating a failure to appreciate that other monkeys can see for themselves that things are not as the would-be deceiver would have them — then we must fall back on the first-order hypothesis to explain their behaviour.

Standing beneath even this, argues Dennett, is the “killjoy” null hypothesis of zero-order intentionality. This is the prospect that the monkeys do not even have first-order beliefs, but behave in accordance with simple tropisms. In Dennett’s later terminology (1996) this means being a Darwinian or at best a Skinnerian creature. In the presence of leopards, a monkey would experience an instinctive or learned “leopard anxiety” and automatically make a certain sound; those who hear the sound experience an equally blind reflex compelling

them to flee.

There is a problem, however, to do with the assumption of rationality in Dennett's intentional stance. Allen & Bekoff (1997) compare Dennett's and Millikan's notions of intentionality or meaning in natural systems, and remind us that Dennett's intentional stance is supposed to be effective to the degree that the organism being studied conforms to an idealized notion of rationality. The animal under investigation is supposed to have certain beliefs and desires, and is predicted to behave in a manner consistent with the logical pursuit of those desires given those beliefs. For example, if a monkey *wants* food currently in the possession of another, and *believes* that the other would abandon the food if it thought there was imminent danger, we could predict that the first animal might try a false alarm call.

On the other hand, Millikan's ideas on intentionality appeal entirely to evolutionary history and make no assumptions about rationality. Millikan's position, according to Allen & Bekoff, allows us to recognize that animals can be highly specialized in their ability to infer intentional states, and that the ability to do so in a particular context does not imply a general ability. Thus it is entirely possible that an animal might behave in Machiavellian third-order ways but only in specific contexts. For example, a monkey might be capable of pretending not to notice that it was being observed by another, in the context of a deceptive food-hiding scheme. Millikan sees no reason why this could not occur despite a complete failure on the part of the same monkey to exhibit third-order intentionality in other situations: she argues that natural selection tends to produce cognitive capacities that match specific ecologically relevant tasks, rather than an all-encompassing reasoning ability. Therefore a definition of "real communication" in terms of higher-order intentionality would be founded on the dubious premise that animals either unambiguously did or unambiguously did

not possess such intentional capabilities.

Millikan's position appears to have been borne out by developments in primatology. The recent consensus in the literature (Seyfarth & Cheney, 2003; Maynard Smith & Harper, 2003) is that although receivers put together information from signallers in such a way as to suggest complex world-models, and possibly higher-order representations, the behaviour of signallers can usually be accounted for by more straightforward (i.e., Darwinian or Skinnerian) mechanisms. The same animal can at different times be both a signaller and a receiver, of course, and so Millikan's point that a cognitive system can be more sophisticated in one aspect and less sophisticated in another is demonstrated.

We need to make a brief cautionary comment here: we have used terms such as "meaning" and "intentionality" but always with a view to remaining good materialists, which is why we have spent time explaining Millikan's position on evolved function. We use the term "intentional" in approximately Brentano's (1874) sense of "aboutness" and do not mean to license in any way the alternative sense meaning roughly "on purpose" or "deliberately", as it is often used in both folk and academic psychology. Authors who use this term seem to be gesturing towards an implicit two-level theory of cognition, associating the good stuff with voluntary control and conscious deliberation. In studying animal behaviour this is completely unhelpful given that non-linguistic animals are not going to be able to tell us which of their decisions were voluntary. We feel that the reasons for steering clear of consciousness-talk were neatly summed up nearly 100 years ago by Watson (1913) and remain true today: "One can assume the presence or absence of consciousness anywhere in the phylogenetic scale without affecting the problems of behavior by one jot or tittle and without influencing in any way the mode of experimental attack upon them."

At this point we have a recognition that simple cognitive mechanisms can

achieve complex behaviour, a rough hierarchy of the kinds of cognitive architectures animals may employ, and a focus on the degree of sophistication of internal world models as being key to differentiating between language-like communication and simpler forms. What is the way forward for computational models of cognition? The most basic role for models in this field is to stand as existence proofs that a simple mechanism can enable a particular behaviour, just as Braitenberg's (1984) "Vehicle 2B" shows that phototaxis is possible without any internal world model at all. Of course, these sorts of demonstrations do not establish that real animals use that particular mechanism to enable the behaviour in question, but they do serve to put a useful lower bound on the complexity required to produce it.

In a sense, though, simple models like Braitenberg's have succeeded too well. Researchers such as Brooks (1991) and Cliff et al. (1993) have demonstrated that robots can be either hand-designed or evolved to perform moderately sophisticated tasks, all without the use of internal representations. (This is quite similar to the argument used in artificial intelligence and philosophy of mind that any given behavioural strategy could in theory be represented by a look-up table.) In Dennett's terminology, this is a demonstration that much can be achieved by Darwinian and Skinnerian creatures. One conclusion that could be drawn from this work is that animals are mere automatons and that the Popperian and Gregorian categories are empty. However, this does not do justice to evolutionary continuity: we know that in the human case we are capable of scheming, planning and even gossiping in ways that require third- or higher-order intentionality, and that we use spoken and written language as tools to gain access to concepts that we could not arrive at alone. It is entirely reasonable to suppose that these abilities did not spring forth in a single mutational leap, and that therefore we should expect to find transitional forms amongst the primates, the

mammals in general, and possibly further afield. (Indeed, we see it as a key long-term question for modelling to ask what small set of genetic changes in the line leading to *Homo sapiens* sparked our Gregorian revolution and has led to such obvious differences between our behaviour and that of other primates.)

Rather than just demonstrating that animal communication systems *could* be implemented using simple Darwinian architectures, we believe that a more constructive and exciting use of computational modelling is to look simultaneously at cognitive complexity and at plausible evolutionary trajectories (as discussed in section 3). We can then ask how strong the evolutionary advantage of a more complex architecture might be over a simpler one. In other words, given the environmental problem it faces, what metabolic costs should an animal be prepared to pay for more cognitive horsepower? In cases like an alarm call system, modelling could show that a simple hard-wired Darwinian scheme would get the job done fairly well (assume a reflex for signalling in response to danger and a reflex for responding to that signal by fleeing). And yet there are features of the observed behaviour that lead some observers to suspect something more Popperian. The question for the model builder becomes: how much of a selective advantage would it be to have the more sophisticated system? Is there a plausible evolutionary route leading from inevitably Darwinian origins to a modern Popperian or Gregorian strategy? Godfrey-Smith (1998) demonstrates that these kinds of increases in cognitive complexity are not inevitable (i.e., much *can* be done with Darwinian / Skinnerian architectures) but he notes that the main evolutionary driver for more sophisticated cognitive equipment is an increase in environmental complexity. Godfrey-Smith also points out that the most challenging feature of the environment will eventually be the behaviour of other animals, and so we need to model the co-evolutionary relationship between (in the alarm-call case) the strategies of signallers, receivers, and their

predators.

Noble et al. (2001) is an example of the approach we advocate. This model deals with social learning in rats: the animals implicitly communicate about which food types are safe to eat by smelling each other's breath. Some cognitive faculties, such as an ability to remember a range of foods, are built into the model. Others are left open to an evolutionary process: whether the rats take the risk of trying totally novel food types, whether they smell the breath of other rats, and whether they're capable of discriminating between sick and healthy rats when doing so. These genetic options spell out increasingly complex behavioural strategies: some rats may simply try out new food for themselves and eschew social learning, others may eat novel food at a low rate and exhibit a "dumb" social learning strategy that is not sensitive to the condition of other rats, and finally some may have a sophisticated social learning strategy and avoid novel food almost entirely. Through the use of very small energetic costs for each cognitive ability Noble et al. ensured that more complex strategies would not evolve through genetic drift when the fitness of a simple and a complex strategy was equal. The logic here was that although it's very difficult to put a number (in joules or in units of predicted offspring) on the cost of some extra cognitive ability, it's clear that more cognitive power costs *something* and so a minimal cost in the simulation is justified. The somewhat surprising result of the study was to show that the dumb social learning strategy (i.e., not the most complex of the available strategies) was evolutionarily stable given some realistic assumptions about the frequency of toxic substances in the environment.

We remain very much aware that the "why?" and the "how?" of any observed animal behaviour are ultimately empirical questions, and we do not believe for a moment that computational modelling can be a substitute for field and laboratory work. The job of a model is to show which of several compet-

ing theories is the most plausible and conceptually economical. In doing this, model builders must rely on empirical data, but one of their outputs can often be productive suggestions for future empirical work. We also believe that is very important that a model does not have hoped-for conclusions built into it. A good model must cover a space of possible evolutionary outcomes such that a phenomenon of interest, such as higher-order intentionality, either might or might not emerge depending on the selective pressures imposed by the environment and by the co-evolution of the relevant strategies. If a model somehow takes it for granted that, for instance, higher-order intentionality is involved in monkey alarm calls, then it can shed no light on the evolutionary origins of the phenomenon.

5 How should we model monkey alarm calling behaviour?

Our plan for modelling the alarm-call system of putty-nosed monkeys follows from our position on simulations as being most useful for modelling evolutionary trajectories and for looking at the mechanisms underlying particular behaviours. There is no argument that these monkeys are Popperian creatures, i.e., that they are at least first-order intentional systems with respect to at least some of their abilities. The questions we want to ask are about the extent to which first- or higher-order intentionality is reflected in the monkey's signalling behaviour in response to predators, and in their responses to the signals of other monkeys.

As described in section 1, male monkeys have two distinct call types, pyows and hacks. Hacks appear to be true alarm calls, probably reserved for the most alarming events. They are given in response to eagles (real and experimenter-simulated), in response to other males hacking in the vicinity, and to noisy

disturbances such as tree falls, baboons fighting, etc. Pyows are elicited by leopard stimuli, the same range of noisy events that elicit hacks although at greater distances (quieter versions), other males hacking in the distance, other males pyowing, and are sometimes given entirely spontaneously. These calls are not reliably referential but there is a bias in favour of hacks being given to eagle stimuli and pyows being given to terrestrial disturbances and predators. The same stimuli can also elicit hacks if nearby and pyows if far away, e.g., trees falling (Arnold & Zuberbühler, 2006; Arnold et al., 2008).

Listeners look up if they hear hacks and look toward the caller if they hear pyows. However, if playbacks of hacks are preceded by a loud noise (thunder or a tree falling) females look up briefly but then direct attention at the source of the sound. Listeners can thus disambiguate meaning by taking into account contextual, usually audible, information. If a male pyows after hearing another male pyowing, then the cause of the pyows are the pyows of the neighbouring male. If there is no audible information, listeners look for other forms of information, e.g., if the male pyows because he has seen a terrestrial predator, his body posture will direct the attention of females within visual contact toward the predator. Pyows given entirely spontaneously are probably male advertisements of presence or location and the sight of the relaxed male and no available audible information indicates this.

All this suggests that calling strategies could be simple Darwinian or Skinnerian capacities (there is no evidence of callers taking their audience into account for example) whereas signal reception strategies are more sophisticated, with the possibility of higher-order intentionality if females are capable of something like “Pyow? There might be a predator...no, I think he only intended for me to see where he was.”

So how do we turn these observations into a productive model? The most

basic strategic problem faced by the monkeys is how much time to spend feeding and how much to spend looking out for predators. The core simulation should thus be a model of anti-predator vigilance (see e.g., Bekoff, 1995) coupled with a signalling model. That is, individuals have some sort of strategy for balancing foraging with looking out for predators. This could be extremely simple, such as a proportion of time spent checking for predators and a probability, when checking, of spotting a nearby predator per unit of time. Individuals would also have a signalling strategy (what call to make, if any, in what circumstance) and a receiving strategy (how to respond to different calls). The interesting part of the model happens when we set up the signalling and reception strategies as being modulated by different levels of intentional sophistication: in a zero-order system, the signalling strategy is based only on directly perceived events. In a first-order system, the signalling strategy is instead based on some kind of represented state of the world. In a second-order system, the signalling strategy is based on a representation that includes the beliefs of others, and so we have possibilities such as “calling because you believe the others have not yet seen the predator”, etc.

The ecological niche of the monkeys must be modelled as plausibly as possible, although this is not the same thing as modelling it in exhaustive detail. Important aspects to get right (informed by field data) include the frequency of predatory encounters, the typical success rate of an encounter from the predator’s point of view, the appropriate response to each type of predator, and the difference that an appropriate response (e.g., hiding in the forest canopy) makes to an individual’s chances of being taken by the predator. Modelling the physical structure of the environment could be done at an arbitrary level of detail, but the important thing is to capture the notion that life in the rainforest means short sight lines and the possibility of predators getting quite close before they

are seen. Similarly the visual system of the monkeys could be modelled at various levels of realism: from abstracting vision to a sphere centred on the monkey's position with a given probability of detecting any predators within range, to an optically realistic model including occlusion by vegetation, foveal and peripheral vision, and head movement. The responsible thing to do here is to implement several different vision systems, starting with the simplest, and to determine whether they have any effect on the conclusions. In initial versions of the model the typical group size and group composition for the monkeys should be an assumption, although in later iterations of the model this could be something that emerges from individual decision-making about whether to stay with or leave a group, and would be based on both foraging success rates and life history strategy with regard to mating.

The evolutionary structure of the model would involve leaving the major variables (i.e., vigilance strategy, signalling and response strategies, intentional system level) open to adaptation, and running a variety of simulations to get an idea of where natural selection tends to go given different assumptions about the environment. The initial population would have reasonable ancestral strategies, such as no signalling and no specific reception policy, such that communication had to build up from accidental influence (e.g., a random genetic tendency to move in response to the noise made by other monkeys moving) through manipulation and exploitation to proper signalling. Initial benchmark runs would implement no special costs for different levels of intentional sophistication (and so we might guess that second- or third-order systems would come about if all that extra cognitive power was free) and then we could introduce modest costs and look at whether higher-order intentional cleverness is worth having when you don't get it for nothing. Building on the idea of evolutionarily stable strategies, invasion studies would also be useful, in which we set up signalling

systems at each level of sophistication and look at whether they're open to invasion by either more or less sophisticated strategies. These techniques will help to establish which picture of the monkeys' world-modelling abilities is the most plausible given their observed behaviour.

Early, simple versions of our model may be open to the criticism that we have a sceptical bias about the intentional capabilities of putty-nosed monkeys as we have built a simple simulated environment for them in which the only significant event is the arrival of a predator, and therefore there is no need for sophistication about the mental states of others as the only interesting message that anyone will ever be trying to get across is "Look out, there's a predator!". The solution to this is of course to enrich the lives of our simulated monkeys: to build a more complex environment in which it makes sense to engage in food calls, contact calls, sexual display, sexual dimorphism in signalling strategies (which the real monkeys certainly have), intraspecific aggression, etc. As Godfrey-Smith (1998) argues, there's no need for cognitive sophistication in a simple world. In an environment where there is more than one message that the speaker might be trying to get across it makes more sense that a listener might want to try to reconstruct the speaker's intention. This would be a pragmatic question for our model though. That is, we would start with a simple environment and gradually add environmental features that could be communicated about, noting the point at which selective pressure for higher-order intentionality emerges. An increasingly rich model might also shed light on the evolutionary origins of syntactic communication, as hinted at in the results of Arnold & Zuberbühler (2006), and the type of intention-modelling cognitive module suggested by De Ruiter et al. (2007).

References

- Allen, C. & Bekoff, M. (1997). *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. MIT Press, Cambridge, MA.
- Arnold, K., Pohlner, Y., & Zuberbühler, K. (2008). A forest monkey's alarm call series to predator models. *Behavioral Ecology and Sociobiology*, 62(4), 549–559.
- Arnold, K. & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, 441(7091), 303.
- Bekoff, M. (1995). Vigilance, flock size, and flock geometry: Information gathering by western evening grosbeaks (Aves, fringillidae). *Ethology*, 99, 150–161.
- Belew, R. & Mitchell, M. (eds.) (1996). *Adaptive Individuals in Evolving Populations: Models and Algorithms*. Addison-Wesley, Reading, MA.
- Bennett, J. (1976). *Linguistic Behaviour*. Cambridge University Press, Cambridge.
- Bickerton, D. (1990). *Language and Species*. University of Chicago Press.
- Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.
- Brentano, F. (1874). *Psychology from an Empirical Standpoint*. Duncker und Humbolt, Leipzig.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139–159.
- Bryden, J. & Noble, J. (2006). Computational modelling, explicit mathematical treatments, and scientific explanation. In L. M. Rocha, L. S. Yaeger, M. A.

- Bedau, D. Floreano, R. L. Goldstone, & A. Vespignani (eds.), *Artificial Life X: Proceedings of the Tenth International Conference on Artificial Life*, pp. 520–526, MIT Press, Cambridge, MA.
- Bullock, S. (1997). *Evolutionary Simulation Models: On their Character, and Application to Problems Concerning the Evolution of Natural Signalling Systems*. Ph.D. thesis, School of Cognitive and Computing Sciences, University of Sussex, Brighton, UK.
- Bullock, S. (1998). A continuous evolutionary simulation model of the attainability of honest signalling equilibria. In C. Adami, R. Belew, H. Kitano, & C. Taylor (eds.), *Artificial Life VI*, pp. 339–348, MIT Press, Cambridge, MA.
- Cangelosi, A. & Harnad, S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4(1), 117–142.
- Cangelosi, A. & Parisi, D. (eds.) (2002). *Computational Approaches to the Evolution of Language and Communication*. Springer, Berlin.
- Cheney, D. L. & Seyfarth, R. M. (1990). *How Monkeys See the World*. University of Chicago Press, Chicago.
- Cliff, D., Harvey, I., & Husbands, P. (1993). Explorations in evolutionary robotics. *Adaptive Behavior*, 2(1), 73–110.
- Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*. John Murray, London.
- Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. John Murray, London.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

- Dawkins, R. & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (eds.), *Behavioural Ecology: An Evolutionary Approach*, pp. 282–309, Blackwell, Oxford.
- De Ruiter, J. & Levinson, S. (2008). A biological infrastructure for communication underlies the cultural evolution of languages. *Behavioral and Brain Sciences*, 30(4), 518, commentary.
- De Ruiter, J. P., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In Y. Rossetto, P. Haggard, & M. Kawato (eds.), *Sensorimotor Foundations of Higher Cognition*, Oxford University Press.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press / Bradford Books, Cambridge, MA.
- Dennett, D. C. (1988). *The intentional stance in theory and practice*. Clarendon Press, Oxford.
- Dennett, D. C. (1996). *Kinds of Minds: Towards an Understanding of Consciousness*. Weidenfeld & Nicolson, London.
- Di Paolo, E. A. (1997). An investigation into the evolution of communication. *Adaptive Behavior*, 6(2), 285–324.
- Di Paolo, E. A., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments. In M. A. Bedau, J. S. McCaskill, N. H. Packard, & S. Rasmussen (eds.), *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life*, pp. 497–506, MIT Press, Cambridge, MA.
- Dobzhansky, T. (1964). Biology, molecular and organismic. *American Zoologist*, 4, 443–452.

- Donaldson, M. C., Lachmann, M., & Bergstrom, C. T. (2007). The evolution of functionally referential meaning in a structured world. *Journal of Theoretical Biology*, 246(2), 225–233.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, London.
- Godfrey-Smith, P. (1998). *Complexity and the Function of Mind in Nature*. Cambridge University Press.
- Grice, H. P. (1969). Utterer's meaning and intention. *Philosophical Review*, 68, 147–177.
- Griffin, D. R. (1992). *Animal Minds*. University of Chicago Press.
- Grimm, V. (1999). Ten years of individual-based modelling in ecology: What have we learned and what could we learn in the future? *Ecological Modelling*, 115, 129–148.
- Haldane, J. B. S. (1932). *The Causes of Evolution*. Longmans, London.
- Harvey, I. (1996). Untimed and misrepresented: Connectionism and the computer metaphor. *AISB Quarterly*, 96, 20–27.
- Huxley, J. S. (1923). Courtship activities of the red-throated diver (*Colymbus stellatus* Pontopp.); together with a discussion on the evolution of courtship in birds. *Journal of the Linnaean Society*, 35, 253–293.
- Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, 8, 185–215.
- Kirby, S. & Hurford, J. (1997). Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands & I. Harvey (eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97)*, pp. 493–502, MIT Press / Bradford Books, Cambridge, MA.

- Krebs, J. R. & Dawkins, R. (1984). Animal signals: Mind reading and manipulation. In J. R. Krebs & N. B. Davies (eds.), *Behavioural Ecology: An Evolutionary Approach*, pp. 380–402, Blackwell, Oxford, second edn.
- Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- Levinson, S. C. (2006). On the human “interaction engine”. In N. J. Enfield & S. C. Levinson (eds.), *Roots of Human Sociality: Culture, Cognition and Interaction*, pp. 39–69, Berg, Oxford.
- Lorenz, K. (1967). *On Aggression*. Methuen, London, translated by M. Latzke.
- Lyon, C., Nehaniv, C. L., & Cangelosi, A. (eds.) (2007). *Emergence of Communication and Language*. Springer, London.
- MacLennan, B. (1992). Synthetic ethology: An approach to the study of communication. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (eds.), *Artificial Life II*, pp. 631–658, Addison-Wesley, Redwood City, CA.
- MacLennan, B. J. & Burghardt, G. M. (1994). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior*, 2(2), 161–188.
- Maynard Smith, J. (1979). Game theory and the evolution of behaviour. *Proceedings of the Royal Society of London: Biological Sciences*, 205, 475–488.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.
- Maynard Smith, J. (1989). *Evolutionary Genetics*. Oxford University Press, Oxford.
- Maynard Smith, J. & Harper, D. (2003). *Animal Signals*. Oxford University Press.

- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press / Bradford Books, Cambridge, MA.
- Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics*, 54(2), 286–295.
- Noble, J. (1999a). Cooperation, conflict and the evolution of communication. *Adaptive Behavior*, 7(3/4), 349–370.
- Noble, J. (1999b). Sexual signalling in an artificial population: When does the handicap principle work? In D. Floreano, J.-D. Nicoud, & F. Mondada (eds.), *Advances in Artificial Life: Fifth European Conference on Artificial Life (ECAL'99)*, vol. 1674 of *Lecture Notes in Artificial Intelligence*, pp. 644–653, Springer, Berlin.
- Noble, J. (2000). Talk is cheap: Evolved strategies for communication and action in asymmetrical animal contests. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, & S. W. Wilson (eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pp. 481–490, MIT Press, Cambridge, MA.
- Noble, J., Todd, P. M., & Tuci, E. (2001). Explaining social learning of food preferences without aversions: An evolutionary simulation model of Norway rats. *Proceedings of the Royal Society of London: Biological Sciences*, 268(1463), 141–149.
- Oliphant, M. (1997). *Formal Approaches to Innate and Learned Communication: Laying the Foundations for Language*. Ph.D. thesis, Department of Cognitive Science, University of California, San Diego.
- Pinker, S. (1994). *The Language Instinct*. Penguin, London.

- Povinelli, D. J. & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? *Trends in Cognitive Science*, 7, 157–160.
- Povinelli, D. J. & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind and Language*, 19(1), 1–28.
- Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, 1(4), 515–526.
- Quinn, M. (2001). Evolving communication without dedicated communication channels. In J. Kelemen & P. Sosík (eds.), *Advances in Artificial Life: Sixth European Conference on Artificial Life (ECAL'01)*, Lecture Notes in Artificial Intelligence, pp. 357–366, Springer, Berlin.
- Quinn, M. & Noble, J. (2001). Modelling animal behaviour in contests: Conventions for resource allocation. In J. Kelemen & P. Sosík (eds.), *Advances in Artificial Life: Sixth European Conference on Artificial Life (ECAL'01)*, Lecture Notes in Artificial Intelligence, pp. 367–376, Springer, Berlin.
- Reddy, M. J. (1979). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (ed.), *Metaphor and Thought*, Cambridge University Press.
- Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, Language, and the Human Mind*. Oxford University Press.
- Selous, E. (1901). *Bird Watching*. Constable, London.
- Selous, E. (1933). *Evolution of Habit in Birds*. Constable, London.
- Seyfarth, R. M. & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, 54, 145–173.

- Simon, H. A. (1981). *The Sciences of the Artificial*. MIT Press, Cambridge, MA, second edn.
- Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems*, 6(4), 537–558.
- Tinbergen, N. (1952). “Derived” activities; their causation, biological significance, origin, and emancipation during evolution. *Quarterly Review of Biology*, 27(1), 1–32.
- Tinbergen, N. (1964). The evolution of signalling devices. In W. Etkin (ed.), *Social Behavior and Organization Among Vertebrates*, pp. 206–230, University of Chicago Press, Chicago.
- Todd, P. M. (1996). The causes and effects of evolutionary simulation in the behavioral sciences. In Belew & Mitchell (1996), pp. 211–224.
- Tomasello, M. (1999). *The Cultural Origins Of Human Cognition*. Harvard University Press, Cambridge, MA.
- Tomasello, M. & Call, J. (1997). *Primate Cognition*. Oxford University Press, New York.
- Tooby, J. & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pp. 19–136, Oxford University Press, New York.
- Vogt, P. (2006). Language evolution and robotics: issues on symbol grounding and language acquisition. In A. Loula, R. Gudwin, & J. Queiroz (eds.), *Artificial Cognition Systems*, pp. 145–173, Idea Group, Hershey, PA.

- Vogt, P. & Divina, F. (2007). Social symbol grounding and language evolution. *Interaction Studies*, 8(1), 31–52.
- von Neumann, J. & Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press, second edn.
- Watson, J. B. (1913). Psychology as the behaviorist sees it. *Psychological Review*, 20, 158–177.
- Werner, G. M. & Todd, P. M. (1997). Too many love songs: Sexual selection and the evolution of communication. In P. Husbands & I. Harvey (eds.), *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97)*, pp. 434–443, MIT Press / Bradford Books, Cambridge, MA.
- Wheeler, M. & de Bourcier, P. (1995). How not to murder your neighbor: Using synthetic behavioral ecology to study aggressive signaling. *Adaptive Behavior*, 3(3), 273–309.
- Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton University Press.
- Williams, P. L., Beer, R. D., & Gasser, M. (2008). Evolving referential communication in embodied dynamical agents. In S. Bullock, J. Noble, R. Watson, & M. A. Bedau (eds.), *ALife XI: Proceedings of the Eleventh International Conference on Artificial Life*, pp. 702–709, MIT Press.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97–158.

List of Figures

- 1 Possible influence interactions between two animals 44

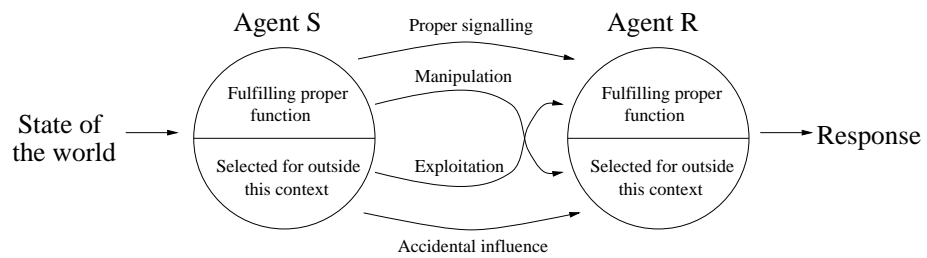


Figure 1: Possible influence interactions between two animals