ON THE CORRECT APPLICATION OF ANIMAL SIGNALLING THEORY TO HUMAN COMMUNICATION

THOMAS C. SCOTT-PHILLIPS

Language Evolution and Computation Research Unit, University of Edinburgh
thom@ling.ed.ac.uk

The defining problem of animal signalling theory is how reliable communication systems remain stable. The problem comes into sharp focus when signals take an arbitrary form, as human words do. Many researchers, including many in evolutionary linguistics, assume that the Handicap Principle is the only recognised solution to this paradox, and hence conclude that the process that underpins reliability in humans must be exceptional. However, this assumption is false: there are many examples of cheap yet reliable signals in nature, and corresponding evolutionary processes that might explain such examples have been identified. This paper briefly reviews the various processes that ay stabilise communication and hence suggests a three-way classification: signals may be kept honest either by (i) being an index, where meaning is tied to form; (ii) handicaps, in which costs are paid by the honest; or (iii) deterrents, in which costs are paid by the dishonest. Of these, the latter seems by far the most likely: humans are able to assess individual reputation, and hence hold the threat of social exclusion against those who signal unreliably.

1. The Problem of Reliability

The ethological question of what keeps signals reliable in the face of the evolutionary pressure to do otherwise is generally regarded as *the* defining problem in animal communication theory (Maynard Smith & Harper, 2003). It is typically cast in the following terms. If one can gain through the use of an unreliable signal then we should expect natural selection to favour such behaviour. Consequently, signals will cease to be of value, since receivers have no guarantee of their reliability. This will, in turn, produce listeners who do not attend to signals, and the system will thus collapse in an evolutionary retelling of Aesop's fable of the boy who cried wolf. What processes keep communication systems stable, and which might apply to human communication? This problem has, somewhat surprisingly, received only limited attention from language evolution researchers, and too often only the most well-known solution – the Handicap Principle (Grafen, 1990; Zahavi, 1975) – or its variants (e.g. Zahavi &

Zahavi, 1997) have been considered. However, contrary to a popular belief both within and outwith evolutionary linguistics, several alternatives to the Handicap Principle are recognised by animal signalling theorists (Maynard Smith & Harper, 2003); there are a number of other well-recognised processes by which signals may be arbitrary yet cheap. This paper's purpose is therefore to briefly consider these alternatives and hence show that we can explain the stability of human communication systems within a traditional behavioural ecology framework and without recourse to post-hoc evolutionary stories.

A brief terminological aside is merited at the outset. In its everyday use, *honesty* makes reference to the relationship between a proposition and its truth value. Although this is roughly the meaning used in animal signalling theory, an obvious but very important caveat is required; namely, that the term *honesty* is necessarily metaphorical. That is, no assumption is made that an animal has 'meanings' that are either true or false. The term is instead used simply as a convenient shorthand to describe animal communicative behaviour. We assign an 'intended' 'meaning' to the behaviour and this allows us to subject it to evolutionary analysis, but this does not at all suppose that the animal necessarily has 'intentions' or 'meanings' in any psychologically real sense. Such shorthand is mostly harmless in the case of animal behaviour (Dennett, 1995; Grafen, 1999), but risks confusion when applied to humans. For that reason, I suggest that the term *reliability* be preferred, and I use it hereafter.

2. The Handicap Principle

The logic of the Handicap Principle is that costs are paid by the signaller as a guarantee of their honesty (Zahavi, 1975). The paradigmatic example is the peacock's tail. Bigger tails leave the peacock less dexterous and less agile, and hence appear to be evolutionarily costly. However, peahens choose to mate with the peacocks with the biggest tails. Why? Because only those peacocks who are of very high quality can afford the cost – the 'handicap' – of big tails.

A distinction should be drawn between *efficacy costs* and *strategic costs* (Maynard Smith & Harper, 1995). Efficacy costs are costs that are necessary for the physical production of the signal. These may be minimal but they are never entirely cost-free; if nothing else there is the opportunity cost of the time spent in production. Strategic costs, on the other hand, are those additional costs that the Handicap Principle imposes on an organism as a guarantee of reliability.

3. Alternatives to the Handicap Principle

Although undeniably important, the Handicap Principle cannot explain all instances of animal signalling: there are many signalling systems that impose no strategic costs on signallers. Many male passerines, for example sparrows, typically display dominance badges on their plumage; the larger the badge, the greater the bird's Resource Holding Potential (an index of all factors that influence fighting ability (Parker, 1974)). However, there appears to be no cost associated with the badge, and no obvious barrier to falsification (Rohwer, 1975; Whitfield, 1987). What alternatives to the Handicap Principle might explain this and other examples? Broadly speaking, four possibilities have been identified by animal signalling theorists.

3.1. *Indices*

An index is a signal in which meaning is fundamentally tied to form, thus preventing even the possibility of unreliability. The classic example is the roar of Red Deer, in which formant dispersion is reliably (negatively) correlated with the deer's size (Reby & McComb, 2003).

3.2. *Coordination games*

In a coordination game each party has a different preference for the outcome of the interaction, but some overriding common interest is shared (Maynard Smith, 1994). An example is the female fruit fly, which mates only once in its lifetime. If a male attempts to court her after this mating she will display her ovipositor towards him, at which point the male immediately ceases courtship (Maynard Smith, 1956). And so although both parties may have conflicting interests (over their desire to mate with one another) both share an overriding common interest: not to waste time if the female has already mated.

3.3. *Repeated interactions*

If individuals meet each other repeatedly over time it may be in both parties' longer-term interests to communicate reliably rather than take whatever short-term payoff may be available through dishonesty (Silk, Kaldor, & Boyd, 2000). This is the essential logic behind reciprocal altruism. Depending upon the specifics of the relationship, the most optimal strategy may be generally honest with occasional deception (Axelrod, 1995; Axelrod & Hamilton, 1981).

3.4. *Punishment of false signals*

If dishonesty is punished then that will obviously reduce or nullify any possible benefit of unreliability (Clutton-Brock & Parker, 1995). Many examples exist; one is the interaction between chicks of the blue-footed booby, in which older chicks will aggressively peck and jostle any younger chicks that signal any attempt to challenge them (Drummond & Osorno, 1992). This does of course raise the second-order problem of why punishing behaviour will evolve if it is itself costly.

4. Three Routes to Stability

Although these processes are often treated as distinct in the animal communication literature, the last three share a common framework: all describe scenarios in which *un*reliable signals incur costs. With regard to coordination games, this will prevent the shared interest from overriding other considerations: the female fruit fly would not display her ovipositor and hence the male would continue to court her, which is a waste of his time and a distraction for her. In repeated interactions unreliability would result in non-cooperation in the future. This would remove the expected future benefits of the relationship; or, put another way, would incur costs relative to the expected payoff over time. Finally, the imposition of costs as a consequence of unreliability is precisely what punishment is.

In general, then, all of these processes describe *deterrents*. We may hence define a three-way classification of the different ways in which signals are kept reliable:

- Indices, in which meaning is causally related to form
- Handicaps, in which costs are incurred by reliable signallers
- Deterrents, in which costs are incurred by unreliable signallers

5. Reputation as Deterrent

Which of the above most likely applied to human communication, and especially language? Indices are clearly not appropriate: linguistic symbols – words – are famously unrelated to form. Some scholars have suggest ways in which the Handicap Principle might apply to human language. For example, handicaps have been used to explore politeness phenomena (van Rooij, 2003), but even if this is correct, it is only concerned with one (rather small) aspect of language. Another suggestion is that ritualised performance acts as a costly signal of commitment to the group, and thus helps to build trust and ultimately ensure reliable communication (Knight, 1998; Power, 2000). But this is a hypothesis about the reliability of the ritualised behaviour, not about words themselves. In general, it is hard to argue that there are any strategic costs associated with utterance production, a point recognised by the inventor of the Handicap Principle: "Language does not contain any component that ensures reliability. It is easy to lie with words" (Zahavi & Zahavi, 1997, p.223).

That leaves us with deterrents. The idea of a deterrent has been formalised in a paper (Lachmann, Számadó, & Bergstrom, 2001) that, given that it explicitly addresses human language as an application of its ideas, has received bafflingly little attention from evolutionary linguists. It has not, for example, received a single citation in any of the collections of work that have arisen from the Evolang conferences that have taken place since the article's publication (Cangelosi, Smith, & Smith, 2006; Tallerman, 2005; Wray, 2002). The basic logic is that although it is cheap and easy to deceive, there are costs to be paid for doing so. In game-theoretic terms, the costs are paid *away* from the equilibrium; they are paid by those who deviate from the evolutionarily stable strategy (ESS). This contrasts with costly signalling, in which the costs are paid *as part of* the ESS. (See also Gintis, Alden Smith, & Bowles, 2001, who show that signalling can be a Nash equilibrium if unreliability is costly.)

Under what circumstances will this logic of deterrents be preferred over the logic of handicaps? Sufficient conditions for cost-free signalling in which reliability is ensured through deterrents are that signals be verified with relative ease (if they are not verifiable then individuals will not know who is and who is not worthy of future attention) and that costs be incurred when unreliable signalling is revealed.

These conditions are fulfilled in the human case: individuals are able to remember the past behaviour of others in sufficient detail to make informed judgements about whether or not to engage in future interactions; and refusal to engage in such interactions produces costs for the excluded individual. At the extreme, social isolation is a very undesirable outcome for a species like humans, in which interactions with others are crucial for our day-to-day survival. This is not, of course, punishment in the conventional sense, but the functional logic is the same: individuals who do not conform will incur prohibitive costs, in this case social exclusion. Moreover, this process would snowball once off the ground, as individuals would be able to exchange information – gossip – about whether others were reliable communication partners (Enquist & Leimar, 1993); and that exchange would itself be kept reliable by the very same mechanisms.

Importantly, the imposition of these costs – the refusal to engage with unreliable individuals – is *not* costly, and hence the second-order problem does not arise. Indeed, such refusal is the most adaptive response if there is good reason to believe that the individual will be unreliable.

It should be explicitly noted that this process allows signals to take an arbitrary form (Lachmann, Számadó, & Bergstrom, 2001). The fact that utterances are cheap yet arbitrary is too often taken to be paradoxical: "resistance to deception has *always* selected against conventional [arbitrary – TSP] signals – with the one puzzling exception of humans" (Knight, 1998, p.72, italics added). This is, as the passerine example and the analysis above both show, simply not true. Instead, once we remove the requirement that costs be causally associated with signal form, as we do if we place the onus of payment on the *dis*honest individual, then the signal is free to take whatever form the signaller wishes. This paves the way for an explosion of symbol use.

6. Concluding Remarks

This necessarily brief survey suggests that there is a single most likely explanation for the stability of human communication: that individuals are deterred from the production of unreliable signals because of the social consequences of doing so. This explanation places a heavy load on the mechanism of reputation, a conclusion that chimes nicely with the emerging consensus from the literature on the evolution of cooperation that reputation is crucial to the stability of human sociality (e.g. Fehr, 2004; Milinski, Semmann, & Krambeck, 2002).

More generally, we should recognise that this process allows us to explain the stability of human communication with the existing tools of animal signalling theory. Evolutionary linguistics has too often resorted to intellectual white flags: the willing abandonment of traditional Darwinian thinking when faced with the heady puissance of natural language. A chronic example of this trend is the suggestion that a capacity for grammar could only have come about via some macro-mutational event (Bickerton, 1990)[1]. The assumption that cheap yet arbitrary signals can only be stabilised by the Handicap Principle is not of the same magnitude, but it is the same type of error. A more learned survey of the animal signalling literature offers a number of alternatives, one of which fits tightly with our intuitive ideas of how social contracts work. Future research should therefore focus on the empirical testing of such ideas rather than the generation of additional post-hoc hypotheses in which language is treated as a special case.

[1] To his credit, Bickerton has since (2003) recognised the implausibility of this suggestion.

Acknowledgement

TSP is funded by a grant from the Arts and Humanities Council of Great Britain.

References

Axelrod, R. (1995). *The evolution of cooperation*. New York: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390-1396.

Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.

Bickerton, D. (2003). Symbol and structure. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 77-93). Oxford: Oxford University Press.

Cangelosi, A., Smith, K., & Smith, A. D. M. (Eds.). (2006). *The evolution of language*. Singapore: World scientific publishing company.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209-216.

Dennett, D. C. (1995). *Darwin's dangerous idea*. London: Penguin.

Drummond, H., & Osorno, J. L. (1992). Training siblings to be submissive losers: dominance between booby nestlings. *Animal behaviour, 44*, 881-893.

Enquist, M., & Leimar, O. (1993). The evolution of cooperation in mobile organisms. *Animal Behaviour, 45*(4), 747-757.

Fehr, E. (2004). Don't lose your reputation. *Nature, 432*, 449-450.

Gintis, H., Alden Smith, E., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of theoretical biology, 213*, 103-119.

Grafen, A. (1990). Biological signals as handicaps. *Journal of theoretical biology, 144*, 517-546.

Grafen, A. (1999). Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proceedings of the Royal Society of London, series B, 266*, 799-803.

Knight, C. (1998). Ritual/speech coevolution: a solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 68-91). Cambridge: Cambridge University Press.

Lachmann, M., Számadó, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences, 98*(23), 13189-13194.

Maynard Smith, J. (1956). Fertility, mating behaviour and sexual selection in *Drosophila subobscura*. *Journal of genetics, 54*, 261-279.

Maynard Smith, J. (1994). Must reliable signals always be costly? *Animal behaviour, 47*, 1115-1120.

Maynard Smith, J., & Harper, D. G. C. (1995). Animal signals: Models and terminology. *Journal of theoretical biology, 177*, 305-311.

Maynard Smith, J., & Harper, D. G. C. (2003). *Animal signals*. Oxford: Oxford University Press.

Milinski, M., Semmann, D., & Krambeck, H.-J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature, 415*, 424-426.

Parker, G. A. (1974). Assessment strategy and the evolution of animal conflicts. *Journal of theoretical biology, 47*, 223-243.

Power, C. (2000). Secret language use at female initiation. In C. Knight, M. Studdert-Kennedy & J. R. Hurford (Eds.), *The evolutionary emergence of language* (pp. 81-98). Cambridge: Cambridge University Press.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of Red Deer stags. *Animal behaviour, 65*, 317-329.

Rohwer, S. (1975). The social significance of avian winter plumage variability. *Evolution, 29*, 593-610.

Silk, J. B., Kaldor, E., & Boyd, R. (2000). Cheap talk when interests conflict. *Animal behaviour, 59*, 423-432.

Tallerman, M. (Ed.). (2005). *Language origins: Perspectives on evolution*. Oxford: Oxford University Press.

van Rooij, R. (2003). *Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behaviour.* Paper presented at the 9th conference on the theoretical aspects of rationality and knowledge.

Whitfield, D. P. (1987). Plumage variability, status signalling and individual recognition in avian flocks. *Trends in ecology and evolution, 2*, 13-18.

Wray, A. (Ed.). (2002). *The transition to language*. Oxford: Oxford University Press.

Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of theoretical biology, 53*, 205-214.

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. Oxford: Oxford University Press.