# Something to talk about: Conflict and coincidence of interest in the evolution of shared meaning

Seth Bullock, School of Computer Studies, University of Leeds

## Introduction

If we are seriously to consider the possibility that human language arose from animal communication through a process of evolutionary change, even if such consideration is merely in order to discount this possibility, we must address the conceptual problems at the heart of our current understanding of animal signalling. In doing so we may throw light upon not only the origins of human language, but also its character.

In this paper I will identify a problem in understanding the semantics of evolved communication. This problem stems from the fact that whilst the meaning of human language is typically considered to be "shared" across a community of language users, the interests of evolving creatures are often not.

Biologists have been aware of the fact that conflicts of interest may prohibit the establishment and maintenance of honest signalling for some time, and have proposed various attempts to deal with this problem (see Johnstone, 1997, for a recent review of this literature). They have also pointed out that the same conflicts of interest prohibit a straightforward reading of meaning in such signalling systems (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984; Maynard Smith & Harper, 1988, 1995). I will argue that even if the problems concerning how honesty may be established and maintained in an evolving system are solved or avoided, the latter problems concerning the semantics of evolved signalling systems remain. I will suggest that a reanalysis of the issue of conflicting interests points to a solution to these problems.

## Stability and Meaning

As group-selectionist accounts lost currency in evolutionary biology (Hamilton, 1964), those interested in the evolution of animal signalling confronted the fact that since the interests of different animals often conflicted, it was not clear why their signals should be honest, and as a result not clear why their signals should be attended to. Despite this, it appeared that signalling was near ubiquitous across the natural world.

Commentators have presented this issue of conflicting interests as essentially dichotomous (e.g., Dawkins & Krebs, 1978; Krebs & Dawkins, 1984; Maynard Smith & Harper, 1988, 1995). Either one is like a bee, living in a community enjoying entirely homogeneous interests due to its peculiar genetics, or one lives in a world of conflict stemming from the struggle to out-reproduce one's competitors. Whilst the adaptiveness of communication is trivially apparent in the former case, how on earth could stable communication survive the tumult of competition implied by the latter?

Faced with the task of building honest signalling systems from deceitful beasts, biologists have taken two approaches. Some have claimed that the notion of stable honest signalling between agents with conflicting interests is a myth. Under this reading, most natural signalling systems are in a constant state of flux — signallers attempting to *manipulate* receivers to their own ends, whilst receivers attempt to *mind-read* the secret thoughts and intentions of signallers (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984). Others have sought to shoe-horn competitive animals into co-operative harmony by identifying restraints on signallers (or receivers) of such gravity that their self-interest can only be served by kowtowing to the yoke of receiver (or signaller) interests. The costs of signalling, for instance, may preclude deceitfulness which would otherwise have been adaptive (Zahavi & Zahavi, 1997).

Both of these approaches have problematic consequences for the semantics of signalling systems. The gravest of these consequences is the possibility that the behaviours picked out by the theories do not count as attending to, or producing, signals at all.

If a vulture spots some carrion which is as yet undetected by its conspecifics (with whom it competes for food), and descends from the sky to devour it, can this act of descending, as seen by other vultures, be taken as a signal from the first vulture to its competitors to the effect that "there is some food to be eaten over here"?

Similarly, consider a species which likes to swim towards magnetic north. If an aquatic predator could generate an electro-magnetic field about itself which attracted its prey by mimicking magnetic north, in what sense is this magnetic field acting as a signal?

Ruth Millikan (1984, 1993) has claimed that neither of the above examples qualify as instances of signalling. The behaviour of one party is merely adapted to exploit the character of the other. No more meaning can be attached to these manipulations or mind-readings than to the pushing or pulling of animals attempting to manually manipulate each other with specially adapted pincers.

In contrast, for a system to be a signalling system, Millikan demands that both agents involved have been adapted by natural selection to play their role in a co-operative signalling episode. Millikan uses the bee dance as a paradigmatic example of such a signalling system. The interests of the dancing bees and their audience are coincident. This coincidence is one way to guarantee that the system is evolutionarily stable, and is the only way to ensure that it counts as a *signalling* system and hence involves *meaningful signals*.

Millikan's account implies that a system may be evolutionarily stable and may even be understood to involve the traffic of information, yet may still not count as a signalling system. The efforts of biologists to stabilize natural signalling systems using some kind of restraint on the agents involved would, under this reading, not result in stable *signalling* systems, since the conflict of interest at their heart would prevent them meeting Millikan's criteria.

For instance, sensory exploitation, in which a "signaller" exploits a tendency in an observer to respond to a particular stimulus (Ryan, 1990), may be evolutionarily stable if the observer's tendency is so valuable in some other context that it is not worthwhile extinguishing, despite the cost of being exploited. For example, the magnetic-north-seeking organism may continue to seek magnetic north despite the risk of predation if this is a successful foraging strategy. Although evolutionarily stable, this system still suffers the problems identified by Millikan. What is being signalled by the exploiter? What is the meaning of the exploitational behaviour?

Similarly, a handicap signalling scenario (Zahavi & Zahavi, 1997) in which, for instance, only some males can afford to exhibit a sexual ornament, will not count straightforwardly as a signalling system, since the signallers and receivers involved have not coevolved to achieve a common aim. Rather, they are fundamentally at odds — signallers being selected to obtain as many mates as possible, while receivers are selected to get the best mates possible. The meaning of the mating display is not shared by signaller and receiver since the job of the sexual ornament is not the same for both parties.

## Human Language

Within models of the evolution of human language (e.g., Kirby & Hurford, 1997; Kirby, 1998, 2000; Batali, 1998; Steels, 1998), there is often little attention to the possibility of deceit (but see Noble, 2000). These accounts either implicitly or explicitly presuppose that the agents involved are taking part in an essentially co-operative enterprise. Language is used by these agents to transmit information. Disregarding some perspectival differences, the information conveyed in an utterance is shared between speaker and listener. Can we justify this conception of human language given that (i) the semantics of animal signalling are threatened by the conflicts of interest they suffer, (ii) this threat is not extinguished when such signalling systems are stabilized by, for instance, signaller costs, and (iii) human signalling evolved from animal signalling. If we cannot square these three assertions, must we abandon continuity between animal signalling and human language?

## Conflict and Coincidence

Figure 1 presents three games. Each is a simple action-response game (Hurd, 1995; Bullock, 1997). In each game there are two possible states of the world, $S \in \{s_1, s_2\}$. Player 1 (P1) is aware of the state of the world and must make one of two possible actions, $A \in \{a_1, a_2\}$. Player 2 (P2) sees the action of P1, but is unaware of the state of the world. She must make one of two possible responses, $R \in \{r_1, r_2\}$. In each game the state of the world makes a difference to at least one of the players in terms of which outcomes $(S, A, R)$ they prefer.

In the first game, P1 and P2 experience a coincidence of interest. Whenever P1 wins, P2 does also. Whenever P1 loses, so does P2. In the second game the players suffer a complete conflict of interest. Whenever P1 wins, P2 loses, and vice versa. The potential for honest signalling (defined, *sensu* Enquist (1985), as $A$ dependent on $S$, and $R$ dependent on $A$, at equilibrium) in these two games is clear — in the first game the players have everything to talk about, in the second, nothing.

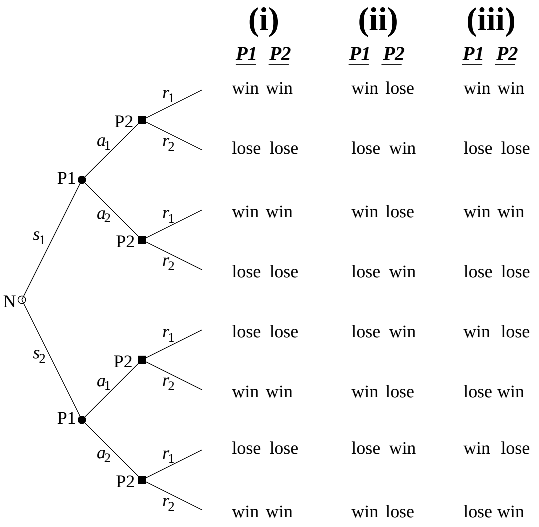

Figure 1: Three games represented in extended form. See text for details.

The first game is capable of supporting a *co-operative signalling system* — one which satisfies Millikan's criteria as outlined above. Since the interests of the two parties coincide, P1 can, and should, use his action as a signal conveying the state of the world to P2. The meaning of this signal is shared by both players — it is both produced and consumed as a representation of $S$.

In the second game P1's action cannot serve any purpose. In fact, P1 should not allow his action to in any way predict the state of the world, since this correlation could be exploited to his disadvantage by P2.

The third game is more interesting precisely because the potential for honest signalling is not immediately clear — do the players have something to talk about? In this game there is a *partial* conflict of interest, or equivalently a partial coincidence of interest. Given that $s_1$ obtains, the interests of P1 and P2 coincide — they both prefer the outcome $(s_1, A, r_1)$ over $(s_1, A, r_2)$. Given that $s_2$ obtains, the interests of P1 and P2 conflict — P1 prefers $(s_2, A, r_1)$, while player 2 prefers $(s_2, A, r_2)$.

In fact, when biologists discuss signalling between agents suffering a conflict of interest, they are always referring to games of this form. Sexual signalling, aggressive displays, warning coloration, etc., all involve the potential for shared interests (when both parties would prefer the same outcome) but this is not guaranteed (the suitor may be poor, the aggressor weak, the prey palatable, etc.). The former aspect ensures the possibility of a signalling system, the latter undermines this possibility. Thus, the interests of evolving agents are not best characterized as either coinciding (game 1) *or* conflicting (game 2). An important intermediate scenario (game 3) exists.

Given this partial coincidence of interest, can we argue that the behaviour of the players at equilibrium is sometimes best characterized by the notion of a co-operative signalling system being parasitised? For example, in the same way that an *eavesdropper* might gain knowledge of $S$ by spying on P1, or a *fraud* might manipulate the behaviour of P2 by mimicking P1, the behaviour of the players, given that $s_2$ obtains, might be parasitic upon that of themselves given that $s_1$ obtains. Can we separate the two halves of the game, and claim that the co-operative half is a signalling system in its own right, despite the threat to it posed by the parasitic half?

Assume P2 plays the strategy $(a_1 \to r_1, a_2 \to r_2)$. Given that $s_2$ obtains, P1 can either lie $(a_1)$ or admit $(a_2)$. Neither action can be awarded the status of signal. In the former case, P1 manipulates P2 to his own benefit (he lies that he is a high quality suitor, a strong aggressor, a poisonous prey item, etc.). In the latter case, P2 mind-reads P1 to her own benefit (P1 admits that he is a poor suitor, or a weak aggressor, or a palatable prey item). External forces (e.g., signalling costs) are required to make it reasonable for P1 to make this admission, just as external forces (e.g., a valuable sensory bias) are required to make it reasonable for P2 to believe a lie. These forces might stabilize honest signalling, but would fail to give the status of signalling episode to either the sequence $s_2 \to a_2 \to r_2$, or $s_2 \to a_1 \to r_1$.

This analysis suggests that the relationship of P1's actions to the interests of both parties confers upon them one of three possible statuses which may be glossed as signal, manipulation (lie), or mind-reading (admission). Only the first of these is representational, but the meaning involved is meaning that is shared between signaller and receiver.

## Conclusion

What impact does this reanalysis have on our understanding of the semantics of animal signalling, and the evolution of language. First, it allows that there may be a valid way to talk about signals between agents lacking entirely coincident interests, which preserves the idea of shared meaning. In the same way that the bee dance is both produced and understood as a representation of the location of nectar, signals between prey and predator, suitor and prospect, attacker and defender, offspring and parent, etc., may sometimes represent prey palatability, mate quality, strength, hunger, etc. despite the fact that sometimes the interests of the agents involved may differ.

Second, this potential for shared meaning to exist in such animal signalling systems allows that our conception of human language as involving utterances with a meaning which is common across a community is commensurable with the notion that there is evolutionary continuity between animal signalling and human language.

# References

Batali, J. (1998). Computational simulations of the emergence of grammar. In Hurford et al. (1998).

Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In Husbands & Harvey (1997), pp. 454–463.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation?. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach*, pp. 282–309. Blackwell, Oxford.

Enquist, M. (1985). Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Animal Behaviour*, *33*, 1152–1161.

Hamilton, W. D. (1964). The genetical evolution of social behaviour, I and II. *Journal of Theoretical Biology*, *7*, 1–16; 17–32.

Hurd, P. L. (1995). Communication in discrete action-response games. *Journal of Theoretical Biology*, *174*, 217–222.

Hurford, J. R., Studdert-Kennedy, M., & Knight, C. (Eds.). (1998). *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge Univeristy Press, Cambridge.

Husbands, P., & Harvey, I. (Eds.). (1997). *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97)*. MIT Press / Bradford Books, Cambridge, MA.

Johnstone, R. A. (1997). The evolution of animal signals. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (Fourth edition)., pp. 155–178. Blackwell, Oxford.

Kirby, S. (1998). Fitness and the selective adaptation of language. In Hurford et al. (1998).

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight et al. (2000).

Kirby, S., & Hurford, J. (1997). Learning, culture and evolution in the origin of linguistic constraints. In Husbands & Harvey (1997), pp. 493–502.

Knight, C., Hurford, J. R., & Studdert-Kennedy, M. (Eds.). (2000). *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge Univeristy Press, Cambridge.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind reading and manipulation. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2nd edition)., pp. 380–402. Blackwell, Oxford.

Maynard Smith, J., & Harper, D. G. C. (1988). The evolution of aggression: Can evolution generate variability?. *Philosophical Transactions of the Royal Society of London, Series* B, *319*, 557–570.

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

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press / Bradford Books, Cambridge, MA.

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. MIT Press / Bradford Books, Cambridge, MA.

Noble, J. (2000). Co-operation, competition and the evolution of pre-linguistic communication. In Knight et al. (2000).

Ryan, M. J. (1990). Sexual selection, sensory systems, and sensory exploitation. *Oxford Survey of Evolutionary Biology*, *7*, 157–195.

Steels, L. (1998). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In Hurford et al. (1998).

Zahavi, A., & Zahavi, A. (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford University Press, Oxford.