Received 10 May 2004
Accepted 16 June 2004
Published online 25 October 2004

The disadvantage of combinatorial communication

Michael Lachmann[1*] **and Carl T. Bergstrom**[2]

[1]*Max Planck Institute for Mathematics in the Sciences, Leipzig, D-04103, Germany*
[2]*Department of Biology, University of Washington, Seattle, WA 98115, USA*

Combinatorial communication allows rapid and efficient transfer of detailed information, yet combinatorial communication is used by few, if any, non-human species. To complement recent studies illustrating the advantages of combinatorial communication, we highlight a critical disadvantage. We use the concept of information value to show that deception poses a greater and qualitatively different threat to combinatorial signalling than to non-combinatorial systems. This additional potential for deception may represent a strategic barrier that has prevented widespread evolution of combinatorial communication. Our approach has the additional benefit of drawing clear distinctions among several types of deception that can occur in communication systems.

Keywords: deception; value of information; game theory; signalling; combinatorial syntax; evolution

> The lie is the specific evil which man has introduced into nature. All our deeds of violence and our misdeeds are only, as it were, a highly-bred development of what this and that creature of nature is able to achieve in its own way. But the lie is our very own invention, different in kind from every deceit that the animals can produce.
>
> (Martin Buber, *Good and evil*, p. 7.)

1. INTRODUCTION

Human language encodes meaning in a combinatorial fashion: complex messages (e.g. sentences) are assembled from simpler units of meaning (e.g. words). This structure confers numerous advantages, which have been highlighted in a set of recent mathematical models (Nowak & Krakauer 1999; Nowak *et al.* 2000). However, combinatorial communication is rare among non-human animals. Why?

A satisfactory account of the evolution of human language needs to be able to account both for why language evolved in humans as it did and why language and its various parts did not evolve in any other animal. Several authors have made considerable progress in describing a path by which language can evolve through gradual change driven by natural selection (Hurford 1989; Nowak & Krakauer 1999; Nowak *et al.* 2000; Rubinstein 2000). The next step is to identify adaptive barriers that are sufficient to prevent other species from achieving this transition.

In a previous article (Lachmann *et al.* 2001), we proposed one potential barrier. We showed that honest combinatorial communication cannot be stabilized by a classical handicap mechanism. Any associated costs necessary to prevent deception must arise from some source other than signal production. Thus, when signaller and receiver have conflicting interests, one barrier is created: the potential for deception impedes the development of combinatorial signals by restricting the conditions under which costly signalling mechanisms can ensure honesty.

In the present paper, we illustrate a more fundamental disadvantage to combinatorial signalling. We show that deception poses a greater and qualitatively different threat to combinatorial systems than to non-combinatorial signalling systems. To do this, we extend the concept of information value (Stephens & Krebs 1986; Stephens 1989) to treat the strategic or game-theoretic aspects of communication (Getty 1997). In particular, we develop a simple machinery to quantify the consequences at an equilibrium state of signals for those who receive them. Using this machinery we are able to highlight the difference between deception as exhibited in animal signals and deception as exercised through human language. Our approach has the additional benefit of drawing clear distinctions among several types of deception that can occur in communication systems.

2. THE VALUE OF INFORMATION IN SIGNALLING GAMES

(a) *The value of information*

L. J. Savage (1954), I. J. Good (1966) and J. P. Gould (1974) address the question of how to determine the *value of information* to an individual facing a decision problem (Stephens 1989). The basic idea is to compare two quantities: the expected payoff that a decision-maker could obtain by using certain information about the world, and the expected payoff that could be obtained without this information. The difference in these two payoffs gives the expected value of obtaining information, and thus represents the maximum amount that one should pay to obtain it.

We can define this measure formally (Gould 1974). Suppose that the world is in some state α from a set of possibilities $\{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n\}$ with known occurrence probabilities. The decision-maker needs to choose the value of a control variable X, and receives a payoff $h(X, \alpha)$ that depends on both the choice of X and the state of the world. Let X_* be the choice of X that maximizes the expected value of h when nothing is known about α beyond the occurrence probabilities, and let X_α be the choice of X that maximizes the expected value of h when α is known exactly. The value to the decision-maker of being fully informed

* Author for correspondence (lachmann@eva.mpg.de).

about α is then

$$E[h(X_\alpha, \alpha)] - E[h(X_*, \alpha)].$$

Here, the expectations are taken over the possible values of α weighted by their occurrence probabilities. This quantity will always be non-negative; having more information never reduces the expected payoff of the decision-maker (Savage 1954; Good 1966; Gould 1974). As a historical aside, posthumously published notes of the mathematician and philosopher Frank Ramsey (1903–1930) reveal that he formulated an analogous measure and derived the non-negativity result several decades earlier (Ramsey 1990).

(b) *The value of information at signalling equilibrium*

In this section, we extend the value of information concept from decision problems to signalling games. We define a trio of value measures and consider the conditions under which each measure can be negative; these correspond to different forms of deception treated in the animal communication literature.

We define a simple signalling game as in previous papers (Lachmann *et al.* 2001; Bergstrom *et al.* 2002). The game, shown in figure 1, has two players, a signaller and signal receiver. To begin the game, 'nature' chooses a state of the world q from some set of possible values Q. Next, the signaller observes the state of the world, and then chooses a signal s from a set of possible signals S. Finally, the receiver observes the signal chosen by the signaller, but not the actual state of the world, and chooses a response r from some set of possible responses R. A *signalling equilibrium* is a pair of signaller and receiver strategies for which neither player can gain an advantage by switching unilaterally to an alternative strategy (Bergstrom & Lachmann 1998).

For any signalling equilibrium, we can define the value to the receiver of the information conveyed in the signal. Let $s = T(q)$ be the signaller's signalling strategy and $r = R(s)$ be the receiver's response strategy at this equilibrium. Let r_* be the receiver's best response in the absence of any knowledge of either the state q of the world or of the signal s that was sent. (Imagine that the signal receiver failed to hear the signal, for example.) Let $w_r(q, s, r)$ be the receiver's fitness. The value of information for the signalling game as a whole, V_g, is the difference of expectations:

$$\begin{aligned} V_g &= \text{expected fitness given signal} \\ &\quad - \text{expected fitness without signal} \\ &= E[w_r(q, T(q), R(T(q)))] - E[w_r(q, \varnothing, r_*)]. \qquad (2.1) \end{aligned}$$

In Appendix A, we prove the following:

Proposition 1: *At any signalling equilibrium in a signalling game, the value of information for the game as a whole,* V_g, *is always non-negative.*

At any equilibrium, signal receivers never do worse overall by heeding the signals. In this sense, equilibrium signals are never entirely deceptive.

(c) *Value of information away from signalling equilibrium*

Out of equilibrium, the value of information for the game as a whole can be positive or negative. For example, signaller and receiver may be engaged in an evolutionary arms race to exploit one another (Krebs & Dawkins 1984). In the course of this arms race, signallers may adjust their signalling strategies to the detriment of receivers (Getty 1999). Until receivers evolve the appropriate counter-adjustments, signallers will be able to 'manipulate' receivers into taking suboptimal actions. In this case receiver responses may reduce receiver fitness on average or even uniformly. The game can have a negative value of information analogous to V_g defined for equilibrium signalling systems above. In this sense, out-of-equilibrium signalling systems can be seen as deceptive (Johnstone & Grafen 1993).

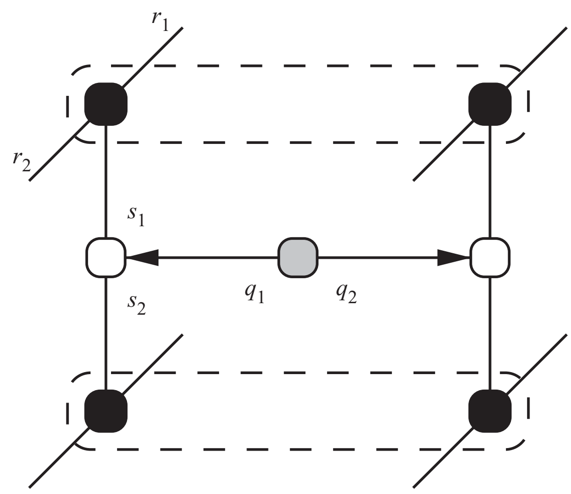

Figure 1. A simple signalling game with two players: signaller (open squares) and receiver (black squares). The game begins at the node represented by the grey square, where nature randomly chooses one of two states q_1 or q_2. The signaller moves next, choosing a signal s_1 or s_2. Finally, the signal receiver chooses a response r_1 or r_2. (Adapted from Fudenberg & Tirole (1991).)

(d) *Value of a particular signal*

We can easily calculate the average value of each particular signal within a specific signalling equilibrium. Define the value of a particular signal V_s as follows. Let Q_s be the set of all world states q in which the signal s is sent at the signalling equilibrium of interest. The value $V_s(s)$ of this signal is the difference between the receiver's expected payoff over Q_s at the signalling equilibrium, and the receiver's expected payoff over Q_s in the absence of any information about s or q.

$$V_s(s) = E_{q\in Q_s}[w_r(q, s, R(s))] - E_{q\in Q_s}[w_r(q, s, r_*)].$$

This definition is effectively a comparison between expected payoffs at two different equilibria: the signalling equilibrium of interest, and the equilibrium in which the signaller provides no information to the receiver. (As an alternative to comparing the value of a signal *across* equilibria, we could look at the importance of the signal *within* a particular signalling equilibrium. To do so, we compare the receiver's expected payoff if this signal is sent, with the receiver's expected payoff if the next-best signal is sent. We refer to the difference in these expected payoffs as the *marginal value* (MV_s) of information for a particular signal:

$$\begin{aligned} MV_s(s) = {} & E_{q\in Q_s}[w_r(q, s, R(s))] \\ & - \max_{s'\neq s}\left[E_{q\in Q_s}[w_r(q, s', R(s'))]\right]. \end{aligned}$$

Using this measure, we can obtain an analogous result to

Proposition 2.2. At any equilibrium in a signalling game, the marginal signal value MV_s of every signal used is always non-negative.)

We prove the following proposition in Appendix A:

Proposition 2.2*: At any equilibrium in a signalling game, the signal value* V_s *of every signal used is always non-negative.*

At a signalling equilibrium, no signal can harm the receiver on average, because otherwise the receiver could simply ignore that signal or treat it as if it were some other signal instead. Proposition 2 is a formal way of expressing the common assertion that signalling will be honest on average because 'it is not evolutionarily stable for the receiver to alter its behaviour [in response to a signal] unless, on average, the signal carries information of value to it.' Maynard Smith & Harper 1995, p. 305 (see also Hasson 1994).

(e) *Value of information conditional on world state*

Thus far we have looked at the value of information for an entire game, and at the value of information for particular signals. In many signalling systems, each signal may be sent in several world states. Further narrowing our focus, we can compute the value of information conditional on the occurrence of a particular world state q with reference to a particular signalling equilibrium. The conditional value of information $V_q(q)$ for world state q is

$$V_q(q) = w_r(q,\ T(q),\ R(s)) - w_r(q,\ T(q),\ r_*).$$

At any *separating equilibrium*, each world state corresponds to a unique signal (Bergstrom *et al.* 2002), and Proposition 2 thus implies that V_q is non-negative for every world state q.

However, for *semi-pooling equilibria*, in which certain world states q_i and q_j elicit a shared signal (Bergstrom & Lachmann 1998):

Proposition 2.3*: The conditional value* $V_q(q)$ *can be negative for some q at a semi-pooling equilibrium.*

We can easily construct an example, and of course any example proves the proposition. Consider a game in which states of the world (signaller qualities, in this case) are uniformly distributed on [0,1]. Signaller fitness is given by $w_s(q,\ s,\ r) = r$ and receiver fitness is given by $w_r(q,\ s,\ r) = 1 - (q - r)^2$. Signallers always benefit from receiving higher responses, whereas receivers maximize fitness by choosing a response r equal to signaller quality q.

Figure 2 shows the equilibrium responses for three different signalling equilibria: a separating equilibrium (dotted line), a pooling equilibrium in which no signalling occurs (dashed line), and one of the many possible semi-pooling equilibria (solid line). At the separating equilibrium, receivers will be completely informed about the state of the world (Bergstrom *et al.* 2002) and thus will be able to choose the optimal response $r = q$. At the no-signalling equilibrium, receivers have no information about world state and thus maximize fitness by choosing the best response to the mean world state: $r = E[q] = 1/2$ (Bergstrom & Lachmann 1997). The semi-pooling equilibrium features two pools. Signallers send one signal when q is in [0, 1/3) and another signal otherwise. Receivers optimize by selecting $r = 1/6$ in response to the former signal and $r = 2/3$ in response to the latter (Lachmann & Bergstrom 1998).

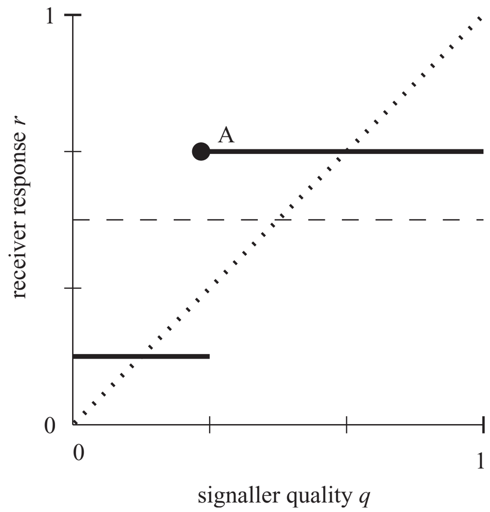

Figure 2. Negative value of information associated with some world states in pooling equilibria. (Note that the axes here are quality and response rather than the more common pairing of signal and response.) Semi-pooling, solid line; separating, dotted line; no signalling, dashed line.

At the semi-pooling equilibrium, the signaller with $q = 1/3$ (point A in the figure) sends the second signal and thereby induces a response $r = 2/3$. The receiver would have received a higher fitness by employing the response $r = 1/2$ used at the no-signalling equilibrium. But because the best response on average to the second signal is $r = 2/3$, and because the receiver cannot distinguish among the signallers who use this signal, $r = 2/3$ is employed instead. Thus the conditional value of information $V_q(1/3)$ for this semi-pooling equilibrium is negative. By a similar logic, the conditional value of information will be negative for all q in the region [1/3, 7/12). Statistician I. J. Good exploited this principle to show how obtaining partial information could be disadvantageous to a decision-maker under certain circumstances (Good 1974; van Rooy 2003).

3. THE VALUE OF INFORMATION IN LINGUISTIC COMMUNICATION

Thus far, we have treated communication systems based upon *monolithic signals*: each signal conveys a complete message and takes its meaning from the whole of the signal rather than from some combination of individually meaningful signal components (sometimes called *holistic signals* (Wray 1998, 2000)). A monolithic signalling system can be described as one in which the receiver has some sort of internal 'look-up table' consisting of separate entries providing separate meanings for every single message that can be sent.

Human language works differently, for the most part. Receivers do not have an all-inclusive look-up table. Instead, they combinatorially construct sentence meanings using syntactic rules that involve the grammatical roles and the individual meanings of the component words. Combinatorial communication offers many advantages (Nowak & Krakauer 1999; Nowak *et al.* 2000); we show that these advantages come at the expense of expanded potential for deception.

(a) *A simple combinatorial signalling model*

We will investigate an idealization of a combinatorial signalling system. Signallers convey information about a world that can be in one of 100 possible states $q \in \{0, 1, \ldots, 99\}$. Receiver fitness depends on how closely the receiver response r matches the world state q: the receiver suffers a cost that is the squared difference between the response and the actual state. In this world, signallers require 100 distinct signals to convey full information via a monolithic signalling system. With a combinatorial system, they can manage with fewer signals.

Suppose signallers and receivers both get the same payoff, so that signallers benefit from accurately representing the state of the world to receivers. Simple semantic systems then afford signalling equilibria with highly informative signals.

Example 1: Signallers indicate the world state by selecting a two-part representation of the world, $s_1 s_2$, where s_1 is some signal from the set {A, B, C, D, E, F, G, H, I, J} and s_2 is some signal from the set {K, L, M, N, O, P, Q, R, S, T}. Receivers then 'decode' or interpret the message by selecting a response $r = \eta(s_1) + \eta(s_2)$, where $\eta(\cdot)$ is as follows:

s_1	A	B	C	D	E	F	G	H	I	J
$\eta(s_1)$	0	10	20	30	40	50	60	70	80	90
s_2	K	L	M	N	O	P	Q	R	S	T
$\eta(s_2)$	0	1	2	3	4	5	6	7	8	9

Clearly, this communication system is combinatorial. Receivers assemble meanings from messages, by applying a set of syntactic rules (simple addition, in this case) to meanings (η values) associated with the signal components that make up the message. When signallers and receivers both receive the same payoff, this system of encoding and decoding rules will represent a stable signalling equilibrium with the following properties:

(i) The game as a whole has a positive value of information V_g.
(ii) Every message $s_1 s_2$ has positive value of information V_s.
(iii) Every world state has a positive value of information V_q.

(b) *Deception in combinatorial signalling*

In practice, signaller and receiver will not always have entirely coincident interests (Lachmann *et al.* 2001). What are the consequences for combinatorial communication? To answer this question, consider what happens if signallers use the above encoding, with one slight change.

Example 2: When the world state q is 0, signallers send the message JT instead of AK. In all other world states, signaller behaviour is as in Example 1.

We can always find a signaller fitness function such that the signaller will behave in this way at equilibrium. A trivial (though rather silly) example follows: signaller fitness is entirely independent of receiver response, and takes on a value of 1 when a signaller follows the encoding in Example 2 and 0 otherwise.

How should receivers respond? If the syntactic rules remain as before (simple addition) and only the meanings of the components can change, signallers will have to adjust the meanings associated with message components. In Appendix B, we compute a locally optimal decoding $\eta'(\cdot)$ for the receivers to use:

s_1	A	B	C	D	E	F	G	H	I	J
$\eta'(s_1)$	0	9.90	19.90	29.90	39.90	49.90	59.90	69.90	79.90	81.58
s_2	K	L	M	N	O	P	Q	R	S	T
$\eta'(s_2)$	1.03	1.92	2.92	3.92	4.92	5.92	6.92	7.92	8.92	1.61

But wait! Why do the meanings change for *all* of the message components A–T, instead of just for the 'misused' components J and T? For example, neither A nor K is part of the message JT, and yet their individual meanings have shifted even though the circumstances of their use seem not to have changed at all.

The answer tells us something very interesting about combinatorial systems of communication: *Meanings in a combinatorial system are tightly intertwined*. If the meaning associated with one signal changes, then those associated with other messages also have to change. In Example 2, the receiver's best response to the signal JT is 49.5—but to move in this direction, the values associated with J and T have to be shifted. This shift, in turn, also influences the values associated with every other signal, as well. For T is used not only in combination with J, but also in the signals AT, BT, CT, etc. Consequently if the value associated with T shifts, the values associated with A, B, C, and so on all must all shift in compensation, if the messages AT, BT, CT, etc. are to retain anything close to their original meanings. Similarly the shifted value of J forces compensatory changes in the values of K, L, M, etc. When the meaning of a single message changes, the effects of that change will propagate back throughout the lexicon. This interconnectivity is not unique to the additive coding used in this example, but instead derives from the rule-based assembly of meanings. Meanings are interconnected because message components appear in multiple messages that derive their meanings from common rules.

In Example 2 as before, the value of information V_g for this game as a whole is positive; the receiver on average benefits from utilizing the information provided in the signal. However, in this combinatorial system, the value of information for a particular signal can be negative, and JT is such a signal. Thus we arrive at the following proposition.

Proposition 2.4*: In a combinatorial signalling system, but not in a monolithic signalling system, the value of information conditional on a particular signal can be negative at equilibrium.*

One additional result follows directly: in combinatorial systems as in the non-combinatorial systems considered earlier, the value of information conditional on a particular world state can be negative.

(c) *Deception and constraint*

Why can signals have negative information value in a combinatorial system but not in a monolithic one? The answer lies in the implicit constraints on the rules for assigning meanings to messages in the combinatorial system. Suppose that in Example 2 the receiver could have interpreted messages by the following rule:

If the signal is JT, use the best-response $r = 49.5$. Otherwise, use the response $r = \eta(s_1) + \eta(s_2)$.

Then every single message, including JT, would have a non-negative value of information V_s. The same would hold if receivers interpreted each message $s_1 s_2$ as monolithic signals, via a look-up table with 100 entries. In

Example 2, the message JT ends up with a negative value of information not simply because of the signaller's strategy, but instead because of the signalling strategy combined with constraints on the response. In particular, we have assumed that receivers are unable to construct a look-up table with a separate entry for each of the 100 messages, but instead they must use some rule to generate meanings. In our example, the rule was addition, but this result will be far more general. In general, simple rule-based systems will not have sufficient degrees of freedom to cover the entire space of possible signal-to-meaning mappings. As a result, signallers will often be able to find signalling strategies such that rule-using receivers will be unable to avoid a negative value of information associated with one or more signals.

Why should receivers be constrained to use simple rules instead of complex look-up tables? One possibility is that storage space is limited. For example, a system in which 10-word strings are assembled from a modest vocabulary of 1000 words would require a look-up table with $1000^{10} = 10^{30}$ entries. And the memory limitation is probably not the greatest obstacle. To use language, humans somehow have to actively learn the meanings of messages. Nowak and colleagues (Nowak 2000; Nowak *et al.* 2000) point out that in order to persist in look-up tables from generation to generation, messages have to be used with sufficient frequency that they can be learned. In the aforementioned 1000-word language, individuals would need to encounter and learn all 10^{30} different messages in their lifetime to capture the full look-up table. To learn a combinatorial version of the same language, by contrast, individuals would only need to learn the 1000 basic words. Thus learning constraints strongly limit the size of a monolithic signalling system. The problem is not that one can *never* assign monolithic meanings to phrases. It is simply that one does not encounter most phrases often enough to assign them monolithic meanings and as a result those phrases can be used in ways that confer negative value of information.

4. CONCLUSIONS

We have developed formal measures of the *value of information* for signalling games. We have used these measures to characterize the different forms of deception, as summarized in table 1.

Table 1. Value of information in signalling games. ((+) must be non-negative. (−) can be negative.)

	game V_g	signal V_s	world state V_q
separating equilibrium	+	+	+
semi-pooling equilibrium	+	+	−
symbolic communication	+	−	−
out of equilibrium	−	−	−

At separating equilibria, signals are not deceptive. The signal receiver obtains complete and accurate information about the state of the world and can respond accordingly. Information value will not be negative by any of the measures addressed here.

Semi-pooling signals (Lachmann & Bergstrom 1998) allow a form of deception. At a semi-pooling equilibrium, the value of information on certain world states can be negative, though both individual signals and the game as a whole will always have non-negative information value. Many previous models of deceptive signalling deal with deception of this sort (Johnstone & Grafen 1993; Hasson 1994; Adams & Mesterton-Gibbons 1995; Számadó 2000).

Out-of-equilibrium signals allow the greatest potential for deception. Away from equilibrium, the value of information can be negative for particular world states, for particular signals, and even for the signalling game as a whole. The reason is straightforward: receivers' expectations may not align with signaller behaviour, and as a consequence signallers may manipulate receivers in arbitrary fashion.

Finally, and most critically, the transition from monolithic to combinatorial signalling allows an additional form of deception. In a combinatorial signalling system, the value of information for the entire game will always be non-negative, but the value of information for individual messages can be negative. This new form of deception arises because receivers are constrained in the meanings that they can assign to messages. Signallers have an opportunity for deception that is not available to them in monolithic communication systems: they can exploit these constraints and the resulting interconnections among meanings, occasionally using message components in atypical ways so as to mislead receivers.

This additional potential for deception may represent a major barrier to the evolution of combinatorial communication. Ultimately, the advantage of combinatorial signalling is also its main weakness. Combinatorial communication can efficiently facilitate large numbers of messages because novel messages can be interpreted simply from a familiarity with the message components. Unfortunately, this also means that receivers will assign meanings to messages without first-hand experience of the circumstances of their use—and thus certain messages can be consistently used to the detriment of signal receivers.

The authors thank Matina Donaldson, Tom Getty, William Harms, Szabolcs Számadó, Robert van Rooy, and two anonymous referees for numerous helpful comments. This project was supported in part by a UW RRF award to C.B.

APPENDIX A

(a) *Proof of Proposition 1*

At any equilibrium in a signalling game, the value of information V_g is always non-negative.

The proof is by contradiction. Suppose the contrary, that at some signalling equilibrium (R, T) the value of information for the game V_g is negative. Then the receiver could increase expected fitness by switching from strategy $R(\cdot)$ to strategy $R'(\cdot) = r_*$, that is, by ignoring the signal altogether and simply playing the best response in the absence of information, r_*, under all circumstances. But here we have a contradiction, since by the definition of signalling equilibrium, at (R, T) the receiver cannot increase fitness by unilaterally switching to an alternative strategy. Therefore information value must be non-negative at any signalling equilibrium.

(b) *Proof of Proposition 2*

At any equilibrium in a signalling game, the measure of signal value V_s is always non-negative.

Again the proof is by contradiction. Suppose the contrary; then at some signalling equilibrium (R, T) there must exist a signal s_i for which $V_s(s_i) < 0$. Then the response r_* is a better response to signal s_i than is $R(s_i)$, and the receiver could increase expected payoff by switching unilaterally from the strategy $R(\cdot)$ to the strategy $R'(\cdot)$ equal to $R(s)$ for $s \neq s_i$ and equal to r_* for $s = s_i$. Thus (R, T) is not a signalling equilibrium, contrary to assumption, and we can conclude that $V_s(s_i) \geqslant 0$ for all signals s_i.

APPENDIX B: FINDING THE OPTIMAL RESPONSE FOR EXAMPLE 2

The signaller can be in a number of world states x, each with probability $p(x)$. In each world state, the signaller sends two signal components $s(x)$ and $t(x)$. We index the possible signal components $s(x)$ as $1, 2, \ldots, \mathrm{N}$ and the possible components $t(x)$ as $(N+1), (N+2), \ldots, \mathrm{M}$. The receiver interprets a message S,T in the following way: each of two signal components is assigned a numerical value r_{S} and q_{T} (in $\mathbb{R}$) and the world state is interpreted as the sum $r_{\mathrm{S}} + q_{\mathrm{T}}$. The receiver suffers a cost that is the square distance of the actual world state from the inferred world state. Thus the average cost of a certain interpretation is

$$\sum_x p(x)(r_{s(x)} + q_{t(x)} - x)^2.$$

An optimal receiver response will be a choice of the vectors $\boldsymbol{r}$ and $\boldsymbol{q}$ of numerical values for the S and T signal components that minimizes the expected cost. To perform the optimization we differentiate with respect to every element of the vector $\boldsymbol{r}$ and set to zero. This gives us a set of N equations of the form

$$\frac{\partial}{\partial r_S}\sum_x p(x)\left(r_{s(x)} + q_{t(x)} - x\right)^2 = 0.$$

Simplifying,

$$\begin{aligned}
&\frac{\partial}{\partial r_S}\sum_x p(x)(r_{s(x)} + q_{t(x)} - x)^2 \\
&\quad = \sum_{x|s(x)=\mathrm{S}} 2\,p(x)(r_{\mathrm{S}} + q_{t(x)} - x) \\
&\quad = 2\left(r_{\mathrm{S}}\sum_{x|s(x)=\mathrm{S}} p(x) + \sum_{x|s(x)=S} p(x)q_{t(x)} - \sum_{x|s(x)=\mathrm{S}} xp(x)\right) \\
&\quad = 2r_{\mathrm{S}}P(x|s(x)=\mathrm{S}) + 2\sum_{\mathrm{T}} q_{\mathrm{T}}P(x|s(x)=\mathrm{S}, t(x)=\mathrm{T}) \\
&\qquad - 2E(x|s(x)=\mathrm{S})P(x|s(x)=\mathrm{S}).
\end{aligned}$$

where $P(x\,|\,\text{condition})$ is the conditional probability on x, and $E(x\,|\,\text{condition})$ is the expectation of x given the condition. So for every S it needs to hold that

$$\begin{aligned}
&2r_S P(x|s(x)=\mathrm{S}) + 2\sum_{\mathrm{T}} q_{\mathrm{T}}P(x|s(x)=\mathrm{S},\ t(x)=\mathrm{T}) \\
&\quad - 2E(x|s(x)=\mathrm{S})P(x|s(x)=\mathrm{S}) = 0.
\end{aligned}$$

Dividing by 2 $P(x|s(x)=\mathrm{S})$, which is strictly positive assuming that all signals are sent:

$$r_S + \sum_{\mathrm{T}} q_{\mathrm{T}}\frac{q_{\mathrm{T}}P(x|s(x)=\mathrm{S}, t(x)=\mathrm{T})}{P(x|s(x)=\mathrm{S})} - E(x|s(x)) = 0.$$

Notice that the expression in the sum is simply the conditional $P(t(x)=\mathrm{T}|s(x)=\mathrm{S})$. So we get

$$r_S + \sum_{\mathrm{T}} q_{\mathrm{T}}P(t(x)=\mathrm{T}|s(x)=\mathrm{S}) = E(x|s(x)=\mathrm{S}).$$

Similarly differentiating with respect to q_{T} gives

$$q_{\mathrm{T}} + \sum_{\mathrm{S}} r_{\mathrm{S}}P(s(x)=\mathrm{S}|t(x)=\mathrm{T}) = E(x|t(x)=\mathrm{T}).$$

Writing $P_{\mathrm{S|T}}$ for the matrix whose element S, T is $P(t(x)=\mathrm{T}|s(x)=\mathrm{S})$ and equivalently $P_{\mathrm{T|S}}$, and writing E_{S} for the vector whose elements are $E(x|s(x)=\mathrm{S})$ and equivalently E_{T} , we can rewrite all these equation as follows:

$$\begin{pmatrix} I & P_{\mathrm{T|S}} \\ P_{\mathrm{S|T}} & I \end{pmatrix}\begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} E_{\mathrm{S}} \\ E_{\mathrm{T}} \end{pmatrix}.$$

It is easy to see that the matrix is singular—just multiply it by the vector $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. We could also have inferred this since it is obvious that these equations do not define $\boldsymbol{r}$ and $\boldsymbol{q}$ uniquely; adding a constant to $\boldsymbol{r}$ and subtracting it from $\boldsymbol{q}$ gives equivalent interpretation of each message. To resolve this, we will arbitrarily add the condition that $r_1 = 0$. We thus replace the first equation in our set by the equation $r_1 = 0$. If we write $P^0_{\mathrm{T|S}}$ for the matrix $P_{\mathrm{T\,|\,S}}$ with its first row set to 0, and E^0_{S} for the vector $\boldsymbol{E}_{\mathrm{S}}$ with its first entry set to 0, we now get

$$\begin{pmatrix} I & P^0_{\mathrm{T|S}} \\ P_{\mathrm{S|T}} & I \end{pmatrix}\begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{q} \end{pmatrix} = \begin{pmatrix} E^0_{\mathrm{S}} \\ E_{\mathrm{T}} \end{pmatrix}.$$

Inverting the matrix yields $\boldsymbol{r}$ and $\boldsymbol{q}$.

REFERENCES

Adams, E. S. & Mesterton-Gibbons, M. 1995 The cost of threat displays and the stability of deceptive communication. *J. Theor. Biol.* **175**, 405–421.

Bergstrom, C. T. & Lachmann, M. 1997 Signalling among relatives. I. Is costly signalling *too* costly? *Phil. Trans. R. Soc. Lond.* B **352**, 609–617. (doi:10.1098.rstb.1997.0041)

Bergstrom, C. T. & Lachmann, M. 1998 Signalling among relatives. III. Talk is cheap. *Proc. Natl Acad. Sci. USA* **95**, 5100–5105.

Bergstrom, C. T., Számadó, S. & Lachmann, M. 2002 Separating equilibria in continuous signalling games. *Phil. Trans. R. Soc. Lond.* B **357**, 1595–1606. (doi:10.1098/rstb.2002.1068)

Buber, M. *Good and Evil*. Upper Saddle River, NJ: Prentice Hall.

Fudenberg, D. & Tirole, J. 1991 *Game theory*. Cambridge, MA: MIT Press.

Getty, T. 1997 Deception: the correct path to enlightenment? *Trends Ecol. Evol.* **12**, 159–160.

Getty, T. 1999 Chase-away sexual selection as noisy reliable signaling. *Evolution* **53**, 299–302.

Good, I. J. 1966 On the principle of total evidence. *Br. J. Phil. Sci.* **17**, 319–321.

Good, I. J. 1974 A little learning can be dangerous. *Br. J. Phil. Sci.* **25**, 340–342.

Gould, J. P. 1974 Risk, stochastic preference, and the value of information. *J. Econ. Theor.* **8**, 64–84.

Hasson, O. 1994 Cheating signals. *J. Theor. Biol.* **167**, 223–238.

Hurford, J. R. 1989 Biological evolution of the saussurean sign as a component of the language acquistion device. *Lingua* **77**, 187–222.

Johnstone, R. A. & Grafen, A. 1993 Dishonesty and the handicap principle. *Anim. Behav.* **46**, 759–764.

Krebs, J. R. & Dawkins, R. 1984 *Animal signals: mind reading and manipulation*. Oxford: Blackwell Scientific.

Lachmann, M. & Bergstrom, C. T. 1998 Signalling among relatives. II. Beyond the Tower of Babel. *Theor. Popul. Biol.* **54**, 146–160.

Lachmann, M., Számadó, S. & Bergstrom, C. T. 2001 Cost and conflict in animal signals and human language. *Proc. Natl Acad. Sci. USA* **98**, 13 189–13 194.

Maynard Smith, J. & Harper, D. G. C. 1995 Animal signals: models and terminology. *J. Theor. Biol.* **177**, 305–311.

Nowak, M. A. 2000 The basic reproductive ratio of a word. *J. Theor. Biol.* **204**, 179–189.

Nowak, M. A. & Krakauer, D. C. 1999 The evolution of language. *Proc. Natl Acad. Sci. USA* **96**, 8028–8033.

Nowak, M. A., Plotkin, J. B. & Jansen, V. A. A. 2000 The evolution of syntactic communication. *Nature* **404**, 495–498.

Ramsey, F. P. 1990 Weight or value of knowledge. *Br. J. Phil. Sci.* **41**, 1–4.

Rubinstein, A. 2000 *Economics and language*. Cambridge University Press.

Savage, L. J. 1954 *The foundations of statistics*. New York: Wiley.

Stephens, D. W. 1989 Variance and the value of information. *Am. Nat.* **134**, 128–140.

Stephens, D. W. & Krebs, J. R. 1986 *Foraging theory*. Princeton University Press.

Számadó, S. 2000 Cheating as a mixed strategy in a simple model of aggressive communication. *Anim. Behav.* **59**, 221–230.

van Rooy, R. 2003 Quality and quantity of information exchange. *J. Logic Lang. Inform.* **12**, 423–451.

Wray, A. 1998 Protolanguage as a holistic system for social interaction. *Lang. Commun.* **18**, 47–67.

Wray, A. 2000 Holistic utterances in protolanguage: the link from primates to humans. In *The evolutionary emergence of language: social function and the origins of linguistic form* (ed. C. Knight, M. Studdert-Kennedy & J. R. Hurford), pp. 285–302. Cambridge University Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.