

# Signalling among relatives. I. Is costly signalling *too* costly?

CARL T. BERGSTROM AND MICHAEL LACHMANN

*Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA*

## CONTENTS

	PAGE
1. Introduction	609
2. The discrete Sir Philip Sidney game	610
3. Johnstone & Grafen's continuous game	613
4. The continuous Sir Philip Sidney game	613
(a) Application to Johnstone & Grafen's model	615
5. Conclusions	615
Appendix 1. Signalling versus no-signalling in the discrete Sir Philip Sidney game	616
References	617

## SUMMARY

Zahavi's handicap principle, originally proposed as an explanation for sexual selection of elaborate male traits, suggests that a sufficient cost to dishonest signals can outweigh the rewards of deception and allow individuals to communicate honestly. Maynard Smith (1991) and Johnstone & Grafen (1992) introduce the Sir Philip Sidney game in order to extend the handicap principle to interactions among related individuals, and to demonstrate that stable costly signalling systems can exist among relatives.

In this paper we demonstrate that despite the benefits associated with honest information transfer, the costs incurred in a stable costly signalling system may leave all participants worse off than they would be in a system with no signalling at all. In both the discrete and continuous forms of the Sir Philip Sidney game, there exist conditions under which costly signalling among relatives, while stable, is so costly that it is disadvantageous compared with no signalling at all. We determine the factors which dictate signal cost and signal benefit in a generalized version of this game, and explain how signal cost can exceed signal value. Such results raise concerns about the evolutionary pathways which could have led to the existence of signalling equilibria in nature. The paper stresses the importance of comparing signalling equilibria with other possible strategies, before drawing conclusions regarding the optimality of signalling.

## 1. INTRODUCTION

How can honest information transfer occur between individuals with different interests? Signalling theory has offered an answer to this question, explaining how the incentive to send deceptive signals might be overcome. Zahavi (1975, 1977) proposed that when there is a conflict of interest between sender and receiver, signals must be costly to be reliable. Previously questioned on theoretical grounds, Zahavi's claims were vindicated by Enquist (1985), Harper (1986) and Grafen (1990*a*). These authors described conditions on signal cost under which signalling could be stable.

While the notion of costly signalling was initially proposed in terms of female-choice sexual selection (Zahavi 1975; Grafen 1990*b*), costly signalling theory was extended to treat interactions among relatives by Maynard Smith (1991, 1994), Godfray (1991, 1995) and Johnstone & Grafen (1992). Maynard Smith (1991) tells the story of the sixteenth-century British

statesman Sir Philip Sidney who, wounded in battle against the Spanish at Zutphen, yielded a precious flask of water to a comrade with the words 'Thy necessity is yet greater than mine'. From this dramatic act of self-sacrifice Maynard Smith derives the Sir Philip Sidney game, in which he assumes relatedness between the players.

Two relatives, Philip and Robert, are in need of one resource, currently controlled by Philip. Both might survive without the resource, but whoever consumes it will surely be saved. In contrast to the original historic example, each individual acts now only to maximize his own inclusive fitness—Philip will consume the resource himself unless he decides that Robert's need is so grave that by handing the resource over he would still profit by reaping a kin selection benefit from his relative's big gain. Robert might try to provide information for this decision by signalling how needy he is, but if Philip simply believes this signal, then cheating and signalling 'my need is very great' would be possible. Notice that

for the present model the two individuals need not be related, but rather must simply share a common interest—as, for example, do the members of a mated pair, each of whom have a stake in the survival of the other.

In the version described by Maynard Smith, the player controlling the resource, hereafter labelled the *donor*, has an initial fitness known to both players. The player with the option of asking for the resource, hereafter labelled the *signaller*, is in one of two states, *needy* or *healthy*, differing in the survival chance without the resource. The signaller's state is known only to himself. To make his state known, he has the option of sending a signal of a certain fixed cost to the donor; the donor can then choose whether or not to respond by transferring the resource. In Johnstone & Grafen's (1992) extension, both donor and signaller fitnesses without the resource are drawn from a uniform distribution over the interval  $[0, 1]$  and are unknown to the other individual; moreover, the signaller is able to pick a signal of any strength between zero and one, where stronger signals are interpreted as signals of greater need.

This model will be analysed under a static game-theoretic approach, with no dynamics. We will refer to equilibrium in the sense of 'Nash equilibrium' for the game between the two players, i.e. as a pair of strategies (donor strategy, signaller strategy) such that neither player can gain by adopting a different strategy. A *signalling* equilibrium will be such an equilibrium, with the additional requirement that the signaller sends a signal which differs based on his condition, and the donor's response differs based on the signal he receives. In this paper we assume, as did Grafen (1990a), that at the signalling equilibrium, different signallers always send different signals. This is not a necessary requirement of signalling equilibria; in Lachmann & Bergstrom (submitted) it is shown that there exist signalling equilibria in which different signallers can send the same signal.

Although a strategy pair meeting these conditions is by definition an equilibrium, it will not necessarily be the *optimal* equilibrium for each player. There might exist an equilibrium under which there is no information transfer between the signaller and the donor, and yet both players are better off than under the signalling equilibrium defined above. In this paper we determine the conditions under which signalling will in this sense be too costly. In the companion paper (Lachmann & Bergstrom, submitted) we take these two equilibria, the signalling equilibrium and the no-signalling equilibrium, as special cases of a wider set of possible equilibria.

While a full genetic model would allow a more thorough analysis of stability conditions, we will follow the lead of the previous signalling literature in using more tractable inclusive fitness methods. This decision imposes several limitations on the results from our model, of which the reader should be aware. The use of inclusive fitness methods restricts the scope of the stability conditions to local stability near the boundary (in this case, near fixation on the signalling

equilibrium) and to additive interaction of fitness effects (see Cavalli-Sforza & Feldman 1978; Aoki 1984). Moreover note that this model treats the coefficient of relatedness,  $k$ , as a constant, in contrast to a genetic model which includes relatedness within its structure without recourse to an external parameter.

In §2 of this paper, we examine the discrete Sir Philip Sidney game as formulated by Maynard Smith (1991) and determine the conditions under which signalling is 'too costly' in that it is more costly than no signalling at all, for either player. In §3, we turn to the continuous Sir Philip Sidney game of Johnstone & Grafen (1992), and demonstrate that in this game (with the uniform distribution of signaller and donor fitnesses used by these authors) signalling is usually 'too costly' in the same sense. Finally, in §4 we examine a general model of signalling and resource transfer among relatives, a Sir Philip Sidney game with non-uniform donor and signaller fitness distributions. Using this model, we explain how the minimal stable signal cost is determined, and provide an analytical method of determining, for these types of games, whether signalling is 'too costly'.

## 2. THE DISCRETE SIR PHILIP SIDNEY GAME

We first consider the Sir Philip Sidney game as presented by Maynard Smith (1991). In this model, there are two players, a signaller and a donor, related by  $k$ . The donor is faced with the decision of whether or not to transfer a resource that he possesses at the start of the game and which guarantees fitness 1 to the individual who consumes it. If the donor chooses to transfer the resource and not consume it himself, he has fitness  $(1 - d)$ . The signaller is in one of two possible states: with probability  $m$  he is needy with fitness  $(1 - a)$  in the absence of the resource; with probability  $(1 - m)$  he is healthy, with fitness  $(1 - b)$  in the absence of the resource. The signaller can send a signal of need and pay an additive cost,  $c$ , which is incurred whether or not the donor responds. Similar results hold when the cost,  $c$ , is multiplicative in nature (Maynard Smith, personal communication). (When donor response lessens signal cost, opportunity arises for signallers to 'blackmail' donors into transferring the resource (see, for example, Eshel & Feldman 1991). We do not consider such cases here.) The model of our present paper is specified in extended form in figure 1. In this section, we demonstrate that in the discrete Sir Philip Sidney game, there exist conditions under which one or both players prefer a no-signalling equilibrium to a signalling equilibrium; a more detailed proof is provided in Appendix 1.

In analysing this game, we will consider donor and signaller strategies as follows. The signaller's strategy specifies what he should do when needy, and what he should do when healthy. Here, there are four possibilities: signal never, signal only when needy, signal only when healthy and signal always. The strategy 'signal only when healthy' will be discarded. (The signal

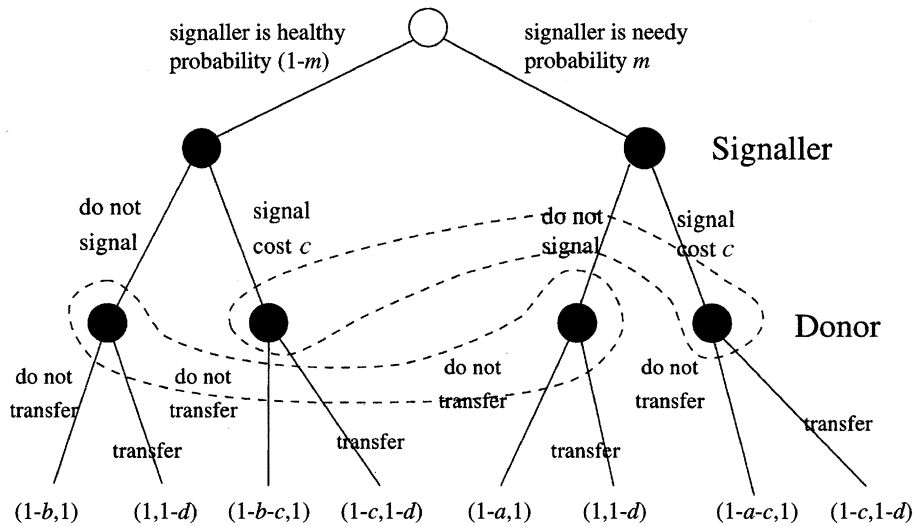


Figure 1. The discrete Sir Philip Sidney game represented in extended form. The regions enclosed by dotted lines are information sets, i.e. the donor is unable to distinguish among points in the same region. The pairs at the base of the tree represent the fitnesses of the signaller and donor, respectively.

Table 1. The discrete Sir Philip Sidney game represented as an asymmetric three-by-three game. The letters **A–I** represent possible strategy pairs

donor	signaller		
	never signal	signal if needy	always signal
never transfer	<b>A</b>	<b>B</b>	<b>C</b>
transfer if signal sent	<b>D</b>	<b>E</b>	<b>F</b>
always transfer	<b>G</b>	<b>H</b>	<b>I</b>

serves only to distinguish a healthy signaller from a needy one, and thus ‘signal only when healthy’ and ‘signal only when needy’ are equivalent given that we are only concerned with the cost associated with these signals.) We examine three out of four donor strategies which specify what to do in the presence and absence of a signal: transfer never, transfer only if a signal is received and transfer always. Again, a fourth possibility exists—transfer only when no signal is received—but this strategy will never be optimal and will also be discarded.

Considered on this level, the discrete Sir Philip Sidney game can be written as an asymmetric three-by-three game with strategies defined as above and (retaining Maynard Smith’s formulation) payoffs given by the expected inclusive fitnesses resulting from interaction between each strategy pair. This game is depicted in table 1. The equilibria of the discrete Sir Philip Sidney game are given by the Nash equilibria of this three-by-three game. The strategy pair at **E** is the only possible signalling equilibrium. When **E** is an equilibrium, the only other possible equilibria are strategy pairs **A** and **G**, of which only one will actually be an equilibrium for given conditions. Let us first consider the conditions under which **E** will be a signalling equilibrium. For this the con-

ditions must be such that the signaller will not raise his average fitness by switching to **D** or **F**, and the donor will not raise his average fitness by switching to **B** or **H**. This gives us the following inequalities, derived in Appendix 1:

$$a \geq c + kd \geq b, \tag{1}$$

$$a \geq (d/k) \geq b. \tag{2}$$

Either **G** or **A** will be an equilibrium as well. As determined in Appendix 1, expression (34), **G** will be an equilibrium if the following expression holds; **A** will be an equilibrium otherwise:

$$d < k(ma + (1 - m)b). \tag{3}$$

In order to demonstrate that even under the most favourable of circumstances, the signalling equilibrium may not be the best equilibrium for either party, we will concern ourselves here with the signal with the cost most favourable to the optimality of signalling. This *least costly believable signal*, which we label  $\hat{c}$ , will be the least costly signal which remains too costly to send when healthy, given that the donor will transfer the resource only in response to a signal. Any smaller signal will allow healthy signallers also to signal and thus prevent honest information transfer; any larger signal will be wasteful in the sense of being more costly than is necessary to prevent deception. From inequality (1),  $\hat{c}$  will be

$$\hat{c} = b - kd. \tag{4}$$

In § 4, we will consider an analogous minimal believable signal for the continuous game. We demonstrate that this signal cost is independent of the probability distribution of the signaller types, just as  $\hat{c}$  above is independent of  $m$ , the probability of a signaller being healthy.

We can compute expected inclusive fitnesses at the signalling equilibrium and the two no-signalling equilibria. The expected inclusive fitnesses for the signaller and the donor at each of these three equilibria are listed in table 2.

Table 2. *Expected fitnesses for signaller and donor at three equilibria: signalling (E), no-signalling where donor always transfers (G) and no-signalling where donor never transfers (A). Necessary conditions for the existence of each type of equilibrium are given in equations (1)–(4)*

equilibrium type	signaller fitness	donor fitness
<b>E</b> —signalling	$1 - b + k + m(b - c - dk)$	$1 - bk + k + m(bk - ck - d)$
<b>G</b> —always transfer	$1 + k(1 - d)$	$k + 1 - d$
<b>A</b> —never transfer	$1 - b + k - am + bm$	$1 - bk + k - akm + bkm$

Table 3. *Conditions under which donor and signaller have higher fitness at the no-signalling equilibrium (G or A) than at the signalling equilibrium (E)*

	donor always transfers	donor never transfers
signaller	$mc > (1 - m)(dk - b)$	$c > a - dk$
donor	$mc > (1 - m)(d/k - b)$	$c > a - (d/k)$

Using these values, we can compare the fitnesses of both players under signalling and no-signalling, respectively. We need to make two separate sets of comparisons—one for the case in which **G**, with the always-transfer strategy, is an equilibrium, and one for the case in which **A**, with the never-transfer strategy, is an equilibrium. These fitness comparisons, derived in Appendix 1, expressions (35)–(38), are summarized in table 3.

Looking first at the case in which the donor never transfers the resource under guesswork, and substituting  $\hat{c}$  for  $c$  in the expressions in table 3, there is a conflict of interests between the signaller and donor when the following condition is met:

$$b < a < d((1/k) - k) + b. \tag{5}$$

Here, the signaller has higher fitness with signalling, but the donor has higher fitness with no-signalling.

In the case in which the donor always transfers the resource, setting  $c = \hat{c}$ , the signaller would be better off under no-signalling whenever  $b/d > k$ . The donor is better off under no-signalling whenever  $b > (1 - m)d/k + m(kd)$ . We can now combine these preference conditions with the conditions for the existence of the always-transfer equilibrium **G** to express the parameter space for which both players prefer the no-signalling equilibrium as a condition on the value of  $k$ :

$$\frac{b - \sqrt{b^2 - 4d^2m(1 - m)}}{2dm} < k < \frac{b + \sqrt{b^2 - 4d^2m(1 - m)}}{2dm}. \tag{6}$$

In Appendix 1, we demonstrate that there exist parameter ranges which satisfy these conditions.

In this section, we have shown that for the discrete Sir Philip Sidney game, there are certain conditions under which one or both players will be better off at the no-signalling equilibrium than at the signalling equilibrium characterized by Maynard Smith (1991).

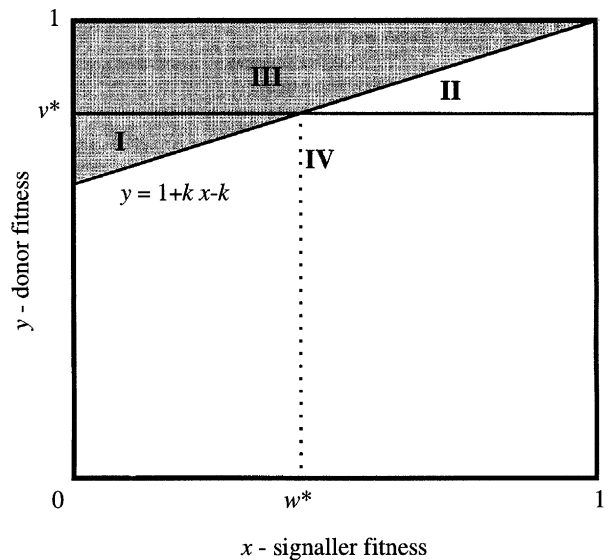


Figure 2. Outcome space for the generalized Sir Philip Sidney game. The signaller and donor fitnesses without the resource are displayed on the axes as labelled. The mean signaller fitness without the resource is labelled  $w^*$ . The minimum quality donor who will respond to  $w^*$  is labelled  $v^*$ . Under guesswork, donors of quality  $v^*$  or higher will transfer the resource. By contrast, donors will transfer the resource under signalling only if they lie above the diagonal line  $y = 1 + kx - k$ . Therefore transfer occurs in regions I and III under signalling, whereas transfer occurs in regions II and III under guesswork.

In particular, when  $d > k(ma + (1 - m)b)$ , the signaller will always prefer the signalling equilibrium, but the donor may prefer the no-signalling equilibrium. When the reverse is true, both players may be better off at the no-signalling equilibrium.

It is worth calling attention to the manner in which the cost of the (minimum believable) signal is set in this model. This signal is simply equal to the gain a healthy individual would receive from lying about his state, given that the donor transfers if and only if the signaller indicates that he is needy. Notice, moreover, that the probabilities that the signaller is needy or healthy play no role whatsoever in determining the signal cost. In § 4, we will find that the signal cost in the continuous Sir Philip Sidney game is determined via a generalization of this principle.

If the signal cost is determined simply by the gain to a healthy signaller of lying about his condition, what determines the *value* of the signal to each player? Under the ‘always-transfer’ no-signal-

ling equilibrium, this value is equal to the amount saved by preventing transfers to a healthy signaller; under the ‘never-transfer’ equilibrium, this value is equal to the amount saved by transferring to a needy signaller. Notice that each of these quantities is weighted by its probability of occurrence and, consequently, the value of the signal, unlike the cost of the signal, is a function of  $m$ . Moreover, notice that under the ‘always-transfer’ conditions, in which both players may prefer the no-signalling equilibrium, the value of the signal decreases as  $m$  increases, whereas the cost of the signal is not altered.

As well as proving the existence of ‘unfavourable’ signalling equilibria in the discrete Sir Philip Sidney game, the analysis of this section anticipates several major results to be seen in the following sections. First, stable systems of costly signalling may be disadvantageous to both donor and signaller, if the two are related. Second, the minimum signal cost under such equilibria is set only by the gain to signallers of lying about their conditions, and is independent of the relative probabilities of being in the different conditions. Third, the signal value depends on the changes in behaviour which the signal induces and consequently does depend on the relative probability that the signaller is in each different state (see Stephens (1989) for a general discussion of the value of information). A more general proof of these results will be provided in § 4.

### 3. JOHNSTONE & GRAFEN’S CONTINUOUS GAME

As in the discrete Sir Philip Sidney game discussed above, signalling may be too costly in the continuous analogue presented by Johnstone & Grafen (1992): both players may be worse off under signalling than under no-signalling, even though each is a stable equilibrium. In fact, costly signalling is always too costly in this sense for  $k < \sqrt{\frac{2}{3}} \approx 0.816$ , i.e. in most biologically plausible cases. Here, we sketch a brief proof of this; a full proof is given by different means in § 4.

In Johnstone & Grafen’s continuous game, there are again two players, a signaller and a donor, related by  $k$ , with respective fitnesses  $x$  and  $y$  chosen independently from a uniform distribution on the interval  $[0, 1]$ . As in the previous section, the donor controls a single indivisible resource. If the donor never transfers the resource, *expected* inclusive fitnesses will be  $1 + \frac{1}{2}k$  for the donor, and  $k + \frac{1}{2}$  for the signaller. This can be used as a baseline for comparisons of alternative patterns of resource transfer. Johnstone & Grafen demonstrate that, under the costly signalling equilibrium, inclusive fitnesses of donor and signaller are  $1 + \frac{1}{2}k + \frac{1}{6}k^4$  and  $\frac{1}{2} + \frac{7}{6}k$ , respectively.

To these fitnesses, we compare the inclusive fitnesses under a ‘guesswork’ equilibrium. This equilibrium is analogous to the no-signalling equilibrium considered in the previous section, in which no signal is sent and the donor simply guesses whether or

not to transfer the resource based on his own condition,  $y$ , alone. The optimal guesswork behaviour will have the donor transferring whenever his expected (inclusive) fitness gain from transferring the resource exceeds his expected fitness gain from retaining the resource, i.e. whenever  $y + k > 1 + kx$ . Under this behaviour, inclusive fitnesses are  $1 + \frac{1}{2}k + \frac{1}{8}k^2$  and  $\frac{1}{2} + \frac{5}{4}k - \frac{1}{8}k^3$  for the donor and signaller, respectively.

We can compare expected fitnesses for the donor and the signaller under signalling and guesswork; algebra reveals that expected fitnesses are higher under guesswork for both the donor and the signaller whenever the relatedness  $k < \sqrt{\frac{2}{3}} \approx 0.816$ . As in the discrete game treated in the previous section, costly signalling among relatives, while a Nash equilibrium, features such a high signal cost that both players are worse off here than at the no-signalling equilibrium.

### 4. THE CONTINUOUS SIR PHILIP SIDNEY GAME

We have shown that signalling can be ‘too costly’ in some versions of the Sir Philip Sidney game. In this section we will try to understand *why* guesswork sometimes turns out to be better than signalling. For this we will construct tools to solve the continuous version of the game described by Johnstone & Grafen, generalized to include non-uniform distributions of signaller and donor qualities.

As in § 3, we have two players, the donor and the signaller, with relatedness  $k$ . The fitnesses with no resource,  $x$  and  $y$ , are now drawn from the probability distribution  $p(x)$  for the signaller and  $q(y)$  for the donor, with ranges  $[0, 1]$  as always. As in Johnstone & Grafen’s model, the cost of signalling condition  $x$  is some function of  $x$ , here denoted  $c(x)$ . Once again, the individual who consumes the resource has fitness 1 (or fitness  $1 - c(x)$ , if he has signalled). In analysing this game, it is useful to examine a two-dimensional fitness space, in which the the signaller’s fitness without the resource is represented by one axis, and the donor’s fitness without the resource by the other. Each encounter between donor and signaller is then represented by a point in this space (see figure 2).

Suppose a donor with fitness  $y$  encounters a signaller with fitness  $x$ . If given perfect information (knowledge of his and the signaller’s fitnesses), he should transfer the resource when  $y + k > 1 + kx$ , i.e. when  $y > 1 + kx - k$ . This region is shaded in figure 2.

When the donor does not know the fitness of the signaller, and is forced to guess whether to transfer, he will do best by transferring the resource only if the average gain from doing so is positive:

$$\int_0^1 (y + k)p(x) dx > \int_0^1 (1 + kx)p(x) dx, \quad (7)$$

or, using the fact that  $\int_0^1 p(x) dy = 1$ ,

$$y + k > 1 + k \int_0^1 xp(x) dx. \quad (8)$$

If the average signaller fitness, in the absence of the resource, is defined by  $w^* \equiv \int_0^1 xp(x) dy$ , then a

donor with resourceless fitness,  $y$ , will donate under guesswork whenever  $y > 1 + kw^* - k$ . This area is also marked in figure 2. This result is generalized by Lachmann & Bergstrom (1996), who demonstrate that in this game, whenever a set of different signallers send the same signal, the donor will respond to this signal as if it was sent by the average signaller in the set.

At the signalling equilibrium the cost  $c(x)$  will be high enough to prevent a signaller with resourceless fitness  $x'$  from misrepresenting himself as having resourceless fitness  $x$ . Thus the average fitness when an individual is misrepresenting himself should always be lower than the average fitness under honest signalling. As mentioned before, when the signaller signals that his resourceless fitness is  $x$ , the donor will transfer the resource if  $y > 1 + kx - k$ . Thus we get

$$\begin{aligned} & \int_0^{1+kx-k} [x' + k]q(y) dy \\ & + \int_{1+kx-k}^1 [1 + ky]q(y) dy - c(x) \\ & \leq \int_0^{1+kx'-k} [x' + k]q(y) dy \\ & + \int_{1+kx'-k}^1 [1 + ky]q(y) dy - c(x'), \end{aligned} \tag{9}$$

or

$$\begin{aligned} c(x') - c(x) & \leq \int_{1+kx-k}^{1+kx'-k} [x' + k]q(y) dy \\ & - \int_{1+kx-k}^{1+kx'-k} [1 + ky]q(y) dy, \end{aligned} \tag{10}$$

i.e.

$$c(x') - c(x) \leq \int_{1+kx-k}^{1+kx'-k} (x' + k - 1 - ky)q(y) dy. \tag{11}$$

Assuming without loss of generality that  $x' > x$ , dividing by  $(x' - x)$ , and taking the limit  $x' \rightarrow x$ , we get

$$\begin{aligned} & \lim_{x' \rightarrow x} \frac{c(x') - c(x)}{x' - x} \\ & \leq \lim_{x' \rightarrow x} \frac{\int_{1+kx-k}^{1+kx'-k} (x' + k - 1 - ky)q(y) dy}{x' - x}, \end{aligned} \tag{12}$$

or

$$\begin{aligned} & \frac{dc(x)}{dx} \\ & \leq \lim_{x' \rightarrow x} \frac{(x' + k^2 - 1 - k^2x)q(1 + kx - k)k(x' - x)}{x' - x}. \end{aligned} \tag{13}$$

From this we can derive a condition on the derivative of  $c(x)$ :

$$q(1 + kx - k)(x - 1)(1 - k^2) \geq \frac{1}{k} \frac{dc(x)}{dx}. \tag{14}$$

As no signaller can gain from representing himself as having fitness 1, a minimal cost function will have

$c(1) = 0$ , and thus we can choose the boundary condition  $c(1) = 0$ . We can then integrate equation (14), to get

$$\int_{1+kx-k}^1 \left(\frac{1}{k} - k\right) (1 - y)q(y) dy \leq c(x). \tag{15}$$

Here we have a condition for the minimum signal cost which keeps everyone honest,  $c(x)$ . Notice, moreover, that this condition on  $c(x)$  depends on the distribution of donors  $q(y)$ , but not on the distribution of signallers. This confirms, in greater generality, the results from §2, namely, that signal cost is independent of signaller distribution. This will be the case whenever the signals chosen are such that no two signallers send the same signal when in different conditions.

At this point, we can compute the average signal cost and average signal value in the signalling system. Assuming that signallers use the minimum believable signals, average signal cost  $\bar{c}$  is simply the integral of the signal cost function  $c(x)$ , integrated over the probability density of signaller types,

$$\bar{c} = \int_0^1 p(x)c(x) dx. \tag{16}$$

When signallers are using the minimum believable signal, this will be

$$\bar{c} = \int_0^1 p(x) \int_{1+kx-k}^1 q(y)(1 - y) \left(\frac{1}{k} - k\right) dy dx. \tag{17}$$

While  $\bar{c}$  gives us the average signal cost (to the signaller—the cost to the donor is  $k\bar{c}$ ), we can also compute the average signal value to the signaller and to the donor, denoted  $\bar{v}_s$  and  $\bar{v}_d$ , respectively. Looking at figure 2, notice that in both the signalling equilibrium and the guesswork equilibrium, transfer always occurs in region III and never occurs in region IV. By contrast, the outcomes in regions I and II differ under the two equilibria, with transfer occurring only under signalling in region I and only under guesswork in region II.

Since the only differences between signalling and guesswork lie in regions I and II, we only need to consider these two regions when comparing the value of signalling. The value of signalling relative to guesswork is equal to the gain from transferring in I minus the gain from transferring in II. The point at which the regions touch is  $(w^*, 1 - kw^* - k)$ , and thus the gain to the signaller from transfer in region I can be written as

$$\int_0^{w^*} p(x) \int_{1+kx-k}^{1+kw^*-k} q(y)((1 + ky) - (x + k)) dy dx. \tag{18}$$

The value to the donor is computed similarly and can be written as

$$\int_0^{w^*} p(x) \int_{1+kx-k}^{1+kw^*-k} q(y)((k + y) - (kx + 1)) dy dx. \tag{19}$$

The value to the signaller from transfer in region II can be written as

$$\int_{w^*}^1 p(x) \int_{1+kx-k}^{1+kx-k} q(y)((x+k) - (1+ky)) dy dx, \tag{20}$$

and the value to the donor can be written as

$$\int_{w^*}^1 p(x) \int_{1+kx-k}^{1+kx-k} q(y)((kx+1) - (k+y)) dy dx. \tag{21}$$

Since these integrals are taken over the same interval, they can be combined:

$$\bar{v}_s = \int_0^1 p(x) \int_{1+kx-k}^{1+kx-k} q(y)(1+ky-x-k) dy dx, \tag{22}$$

$$\bar{v}_d = \int_0^1 p(x) \int_{1+kx-k}^{1+kx-k} q(y)(k+y-kx-1) dy dx. \tag{23}$$

We now have expressions for expected signal cost and signal value (relative to guesswork) for both signaller and donor. The signaller will prefer signalling to guesswork whenever  $\bar{v}_s - \bar{c} > 0$ , and donor will prefer signalling to guesswork when  $\bar{v}_d - k\bar{c} > 0$ . In the following subsection, we use these results to demonstrate that costly signalling gives a lower average fitness than no-signalling for both players in Johnstone & Grafen's (1992) model (for realistic values of  $k$ ).

**(a) Application to Johnstone & Grafen's model**

We can apply the general formulations of signal cost and signal value directly to the continuous Sir Philip Sidney game as proposed by Johnstone & Grafen (1992), i.e. with signaller and donor conditions drawn from a uniform distribution on  $[0, 1]$  so that  $p(x) = 1$  for  $x \in [0, 1]$  and  $q(y) = 1$  for  $y \in [0, 1]$ .

Average signal cost is then

$$\bar{c} = \int_0^1 \int_{1+kx-k}^1 (1-y) \left(\frac{1}{k} - k\right) dy dx. \tag{24}$$

Evaluating the double integral, we arrive at the average signal cost (to the signaller) computed by Johnstone & Grafen,

$$\bar{c} = \frac{1}{6}k(1 - k^2). \tag{25}$$

In a similar fashion, we can use our general formulation above to compute the signal value for this game. Average signal value to signaller and donor are as follows:

$$\bar{v}_s = \int_0^1 \int_{1+kx-k}^{1+kx-k} (1+ky-x-k) dy dx, \tag{26}$$

$$\bar{v}_d = \int_0^1 \int_{1+kx-k}^{1+kx-k} (k+y-kx-1) dy dx. \tag{27}$$

Evaluating these integrals, we get

$$\bar{v}_s = \frac{1}{24}2k - k^3, \tag{28}$$

$$\bar{v}_d = \frac{1}{24}k^2. \tag{29}$$

Now, we can compare signal value and signal cost. No signalling will be better than costly signalling

when  $\bar{c} > \bar{v}_s$  and  $k\bar{c} > \bar{v}_d$ , respectively. For the Johnstone & Grafen game, these conditions simplify to  $k < \sqrt{\frac{2}{3}}$  and  $k < \frac{1}{2}\sqrt{3}$ . These conditions are identical to those computed by other methods in § 3.

**5. CONCLUSIONS**

Costly signalling among relatives is a two-edged sword. Signalling can be useful in that it allows information to be shared, yet ensuring honesty can be difficult when parties have conflicting interests. Sufficient cost to dishonest signals can solve this problem, preventing lies. This cost, however, may actually outweigh whatever advantages were to be gained by honest communication in the first place. Therefore, when considering costly signalling equilibria for interactions among relatives, we must keep in mind the counter-intuitive fact that at a stable signalling equilibrium, both players may be worse off than they would be with no signalling at all. This is shown clearly for both Maynard Smith's (1991) discrete Sir Philip Sidney game and Johnstone & Grafen's (1992) continuous Sir Philip Sidney game.

At least two phenomena demand explanation. First, how does signal cost somehow outgrow signal value? Second, in such situations, why does the signalling equilibrium remain stable despite being inferior, from the perspective of both players, to the no-signalling equilibrium? We will tackle these two questions in turn.

To answer the first question, we refer to the general expression for signal cost  $c(x)$  computed in § 4,

$$c(x) \geq \int_{1+kx-k}^1 (1-y) \left(\frac{1}{k} - k\right) q(y) dy.$$

This expression reveals that signal cost is independent of the distribution of signaller condition. Cost is dictated not by expected signal value but simply by the requirement that no individual, regardless of how rare that particular condition may be among signallers, can gain from misrepresenting his condition. By contrast, signal value *does* depend on the distribution of signaller types in that the expected gain from signalling is a function of the frequencies of different types of individuals in the population. Consequently, types which occur very rarely and have little effect on signal value can nevertheless play a large role in setting signal cost. In this way, signal cost can remain sizeable even when signal value becomes arbitrarily small.

The second question becomes easier to answer when costly signalling interactions are viewed as a coordination game between two players. The players aim either to proceed under mutual assumptions of signalling, or to proceed under mutual agreement that there will be no signalling. If the signaller signals to a donor who is not listening for a signal, the signal cost is wasted and both players suffer. If the signaller does not signal, even when needy, and the donor is expecting a signal, then the donor may fail to transfer a resource that he would have willingly transferred, had he known that the signaller would

not signal regardless of condition. Again both players suffer. When the donor is expecting a signal, the very act of not signalling is also a signal (of health). It is not possible for the signaller to indicate that he is using the ‘no-signalling’ strategy instead of the signalling strategy. Thus, the strategy pairs (signal, expect signal) and (don’t signal, don’t expect signal) are the two equilibria of the coordination game; it never pays a player to move away from either equilibrium.

Under some circumstances (e.g. in the discrete Sir Philip Sidney game when expression (5) is satisfied) this coordination game takes the form of a *battle-of-the-sexes game*, in the sense of Luce & Raiffa (1957). While both players prefer each of the coordinating equilibria to anti-coordination, there is a conflict of interest over which coordinating equilibrium is to be expressed.

As we see in §§ 3 and 4, the signalling/no-signalling system can produce not only a game in the *battle-of-the-sexes* form, but also a *coordination game*, for which the signalling equilibrium is inferior, from the perspective of both players, to the guesswork equilibrium.

An important observation in all cases analysed in this paper is that in a signalling equilibrium, the cost of signals is independent of the signaller *distribution*, depending instead only on the *types* of signallers present. This leads to the disturbing result that even a very rare signaller could cause the signal cost of all of the more common signallers to be high, and thus dramatically lower the average fitness of the signallers. One would expect that in this case the common signallers would signal with lower cost, enabling this rare signaller to cheat. The rare signaller would now ‘cheat always’, by simply signalling as if he were one of the common signallers. Analysis reveals that this can also be a signalling equilibrium, in which different signallers *can* send the same signal. A full investigation of such equilibria in the Sir Philip Sidney game is presented in Lachmann & Bergstrom (1997).

We thank Marcus W. Feldman, Jochen Kumm and Jonathan Pritchard for a number of helpful suggestions. We also thank John Maynard Smith for pointing out that this model applies not only to signalling among relatives, but also to signalling among any two individuals with an interest in one another’s survival. The authors were supported by NIH grant GM 28016 to M. Feldman. C.T.B. is a Howard Hughes Medical Institute Predoctoral Fellow.

#### Appendix 1. SIGNALLING VERSUS NO-SIGNALLING IN THE DISCRETE SIR PHILIP SIDNEY GAME

In this appendix we prove that there exist conditions under which one or both players in the discrete Sir Philip Sidney game (as defined in figure 1) have higher fitness under no-signalling than under signalling. As mentioned in § 2, for strategy pair **E** in table 1 to be an equilibrium, the average fitness of

the signaller at **E** has to be at least as great as at **D** and **F**. This gives

$$\begin{aligned} m((1-c) + k(1-d)) + (1-m)((1-b) + k) \\ \geq m((1-a) + k(1-d)) \\ + (1-m)((1-b) + k(1-d)), \end{aligned} \quad (30)$$

$$\begin{aligned} m((1-c) + k(1-d)) + (1-m)((1-b) + k) \\ \geq m((1-c) + k) \\ + (1-m)((1-c) + k). \end{aligned} \quad (31)$$

$$\begin{aligned} m(k(1-c) + (1-d)) + (1-m)(k(1-b) + 1) \\ \geq m(k(1-a-c) + 1) \\ + (1-m)(k(1-b) + 1), \end{aligned} \quad (32)$$

$$\begin{aligned} m(k(1-c) + (1-d)) + (1-m)(k(1-b) + 1) \\ \geq m(k(1-c) + (1-d)) \\ + (1-m)(k + (1-d)). \end{aligned} \quad (33)$$

Also, the average fitness of the donor at **E** has to be at least as great as at **B** and **H**, see equations (32) and (33).

In addition to the signalling equilibrium, there will be a no-signalling equilibrium either at **A** or at **G**. The average inclusive fitness of the donor determines which of these strategy pairs is the equilibrium. **A** will be the equilibrium when the donor never transfers under no-signalling, i.e. when

$$d > k(ma + (1-m)b). \quad (34)$$

**G** will be the equilibrium otherwise.

We can now compare the fitnesses of both players under signalling and no-signalling, respectively. We need to make two separate sets of comparisons—one for the ‘always-transfer’ case where, as mentioned above,  $d < k(ma + (1-m)b)$ , one for the ‘never-transfer’ case where  $d > k(ma + (1-m)b)$ .

In the first case, where transfer always occurs in the no-signalling system, a comparison of expected fitnesses from table 2 reveals that the signaller and donor, respectively, will have higher fitness in the no-signalling system whenever

$$mc > (1-m)(dk - b), \quad (35)$$

$$mc > (1-m)(d/k - b). \quad (36)$$

In the second case, where transfer never occurs in the no-signalling system, the expressions in table 2 reveal that the signaller and donor, respectively, will have higher fitness in the no-signalling system whenever

$$c > a - dk, \quad (37)$$

$$c > a - d/k. \quad (38)$$

These four inequalities are summarized in table 3 in § 2.

Now we determine whether there exist conditions for which both players have higher fitness under the no-signalling equilibrium. We start with the case in which the donor never transfers the resource under guesswork. Are there conditions under which inequalities (1), (2), (34), (37), and (38) hold? As  $k < 1$ , substituting the minimal signal cost,  $c = \hat{c} = b - kd$  for  $c$ , one can see that condition (37) can never hold, when  $a \geq b$ . In this case, the signaller will never be better off at the no-signalling equilibrium. For the



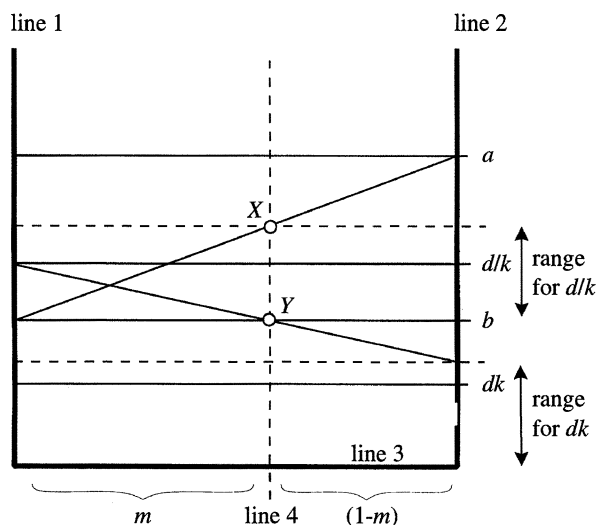


Figure 3. Geometric representation of the conditions required so that when the donor never transfers under no signalling, both donor and signaller are better off at the no-signalling equilibrium than at the signalling equilibrium. Lines 1 and 2 are orthogonal to line 3; on line 2 we mark points at distances  $b$ ,  $a$  and  $d/k$  from line 3, so that condition (44) is satisfied. The length of line 3 is 1, and on it we mark a point at a distance  $m$  from line 1. Point X marks the intersection of a line from  $b$  on line 1 to  $a$  on line 2 with line 4, and therefore has height  $b(1-m) + am$ . We require  $d/k$  to be less than this height of point X, satisfying condition (45). Point Y marks distance  $b$  on line 4. Intersecting a line from  $d/k$  on line 1 to Y with line 2 gives the range for  $dk$  so that condition (46) is satisfied.

donor to prefer no signalling, we are left with the following conditions:

$$a \geq d/k \geq b, \tag{39}$$

$$d > k(ma + (1-m)b), \tag{40}$$

$$b - dk > a - d/k. \tag{41}$$

After some algebraic manipulation, the following conditions on  $k$  have to be satisfied for the donor to be better off at the no-signalling equilibrium

$$\frac{d}{ma + (1-m)b} > k > \frac{d}{a}, \tag{42}$$

$$\frac{-(a-b) + \sqrt{(a-b)^2 + 4d^2}}{2d} > k. \tag{43}$$

In these cases there will be a conflict of interest between the signaller and the donor over which equilibrium is preferred, the signalling or the no-signalling equilibrium.

Now we turn to the case in which the donor always transfers the resource. In this case conditions (1), (2), (34), (35), and (36) have to be satisfied for both the donor and the signaller to be better off at the no-

signalling equilibrium. Again we substitute  $c = \hat{c} = b - kd$  for  $c$ . Because  $k < 1$ , inequality (36) subsumes inequality (35). We are then left with the following conditions:

$$a \geq d/k \geq b, \tag{44}$$

$$d/k < ma + (1-m)b, \tag{45}$$

$$b > mkd + (1-m)d/k. \tag{46}$$

Through a geometric representation of the inequalities, figure 3 shows that there are conditions in which all of these inequalities are satisfied.

## REFERENCES

Aoki, K. 1984 A quantitative genetic model of two-policy games between relatives. *J. Theor. Biol.* **109**, 111–126.  
 Cavalli-Sforza, L. L. & Feldman, M. W. 1978 Darwinian selection and altruism. *Theor. Popul. Biol.* **14**, 268–280.  
 Enquist, M. 1985 Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Anim. Behav.* **33**, 1152–1161.  
 Eshel, I. & Feldman, M. W. 1991 The handicap principle in parent–offspring conflict: comparison of optimality and population genetic analyses. *Am. Nat.* **137**, 167–185.  
 Godfray, H. C. J. 1991 Signalling of need by offspring to their parents. *Nature* **352**, 328–330.  
 Godfray, H. C. J. 1995 Signalling of need between parent and young: parent–offspring conflict and sibling rivalry. *Am. Nat.* **146**, 1–24.  
 Grafen, A. 1990a Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546.  
 Grafen, A. 1990b Sexual selection unhandicapped by the Fisher process. *J. Theor. Biol.* **144**, 473–516.  
 Harper, A. B. 1986 The evolution of begging: sibling competition and parent–offspring conflict. *Am. Nat.* **128**, 99–114.  
 Johnstone, R. A. & Grafen, A. 1992 The continuous Sir Philip Sidney game: a simple model of biological signalling. *J. Theor. Biol.* **156**, 215–234.  
 Lachmann, M. & Bergstrom, C. T. 1997 Signalling among relatives. II. Beyond the Tower of Babel. (In preparation.)  
 Luce, R. D. & Raiffa, H. 1957 *Games and decisions*. New York: Wiley.  
 Maynard Smith, J. 1991 Honest signalling: the Philip Sidney game. *Anim. Behav.* **42**, 1034–1035.  
 Maynard Smith, J. 1994 Must reliable signals always be costly? *Anim. Behav.* **47**, 1115–1120.  
 Stephens, D. W. 1989 Variance and the value of information. *Am. Nat.* **134**, 128–140.  
 Zahavi, A. 1975 Mate selection—a selection for handicap. *J. Theor. Biol.* **53**, 205–214.  
 Zahavi, A. 1977 The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.* **67**, 603–605.

Received 10 June 1996; accepted 1 November 1996