Separating equilibria in continuous signalling games

Carl T. Bergstrom[1]*, Szabolcs Számadó[2] and Michael Lachmann[3]†

[1]*Department of Zoology, University of Washington, Seattle, WA 98115, USA*
[2]*Department of Plant Taxonomy and Ecology, Eotvos Lorand University, Budapest H-1518, Hungary*
[3]*Santa Fe Institute, Santa Fe, NM 87501, USA*

Much of the literature on costly signalling theory concentrates on separating equilibria of continuous signalling games. At such equilibria, every signaller sends a distinct signal, and signal receivers are able to exactly infer the signaller's condition from the signal sent. In this paper, we introduce a vector-field solution method that simplifies the process of solving for separating equilibria. Using this approach, we show that continuous signalling games can have low-cost separating equilibria despite conflicting interests between signaller and receiver. We find that contrary to prior arguments, honesty does not require wasteful signals. Finally, we examine signalling games in which different signallers have different minimal-cost signals, and provide a mathematical justification for the argument that even non-signalling traits will be exaggerated beyond their phenotypic optimum when they are used by other individuals to judge condition or quality.

Keywords: costly signalling; honest signalling; handicap principle; separating equilibrium; indexes; vector field

1. INTRODUCTION

Game theoretic models are commonly used to study the strategic aspects of communication, including how and why signals evolve. In a typical model, a *signaller*, who has private information about the state of the world, sends a signal to a *receiver*, who selects a response to the signal. Communication occurs when the signaller sends informative signals that influence the receiver's choice of response. This communication will be stable when the signaller and receiver pursue strategies (rules for choosing signal and response, respectively) that together comprise a signalling equilibrium: a pair of signaller and receiver strategies such that neither party gains from a unilateral change in strategy (Bergstrom & Lachmann 1998).

In an extremely influential paper, Grafen (1990) formulated a model in which males signal quality to females, and showed that under certain conditions, costly signals facilitate honesty. Grafen's analysis, based on a continuous signalling game with a continuum of possible signaller states and signals, identified a stable *separating* equilibrium in which every difference in signaller state is reflected by a difference in signal. At this equilibrium, the signal receiver can exactly infer the signaller's quality from the signal sent. *Pooling equilibria*, in which signallers in different states share common signals, may also exist (Bergstrom & Lachmann 1998; Lachmann & Bergstrom 1998).

In this paper we focus on separating equilibria, and attempt to address the following questions as generally as possible. What are the basic common properties characterizing separating equilibria? When will separating equilibria exist? How many separating equilibria will exist in a given game? Why might there be no separating equilibria for a given game? How costly do separating equilibrium signals have to be? What predictions for empirical studies can be made from theoretical models of separating equilibria?

In § 2 we present initial definitions and describe the conditions for the existence of separating signalling equilibria. In § 3, we present a simple method for finding the separating equilibria of signalling games as integral curves of a vector field. We also show that this approach provides a useful intuitive picture of various signalling games. In § 4 we show that many signalling equilibria are possible if one varies the cost function, and that some of these have very low signal costs at equilibrium. For these cost functions, honest signalling need not be wasteful even when signaller and receiver interests differ substantially. In § 5 we use the results of the earlier sections to present a mathematical proof of an idea proposed by Lotem *et al.* (1999): when non-signalling characters are used as cues by observers, the individuals exhibiting these characters will be selected to express levels other than the phenotypic optimum.

2. CONDITIONS FOR SIGNALLING EQUILIBRIA

(a) *Definitions*

Appendix A contains a table of all variables and their definitions. We begin by defining the signalling game Γ that we will study in this paper. The game Γ features two players, a signaller and a receiver of the signals. The signaller has some quality $q \in [q_{min}, q_{max}]$ known only to herself. She sends a signal s to the responder, eliciting a

*Author for correspondence (cbergst@u.washington.edu).
† Present address: Max Planck Institute for Mathematics in the Sciences, Leipzig D-04103, Germany.

Phil. Trans. R. Soc. Lond. B (2002) **357**, 1595–1606
DOI 10.1098/rstb.2002.1068

1595

© 2002 The Royal Society

response $r \in [r_{\min}, r_{\max}]$. The signaller receives a pay-off of $H(q,r) - C(q,s)$, where H can be thought of as the value of the response and C as the display cost of the signal s. The receiver gets a pay-off $G(q, r)$ from providing response r to a signaller of quality q. Each signaller will have a signal that elicits the minimal response $r_{\min}$.

Let $s = S(q)$ and $r = R(s)$ be signaller and receiver strategies which specify what signal to send given one's quality and what response to provide given the signal received, respectively. The pair (S, R) is said to be a *signalling equilibrium* if neither signaller nor receiver can benefit or break even from a unilateral change in strategy. It will be a *separating equilibrium* if furthermore the signaller strategy S is one-to-one.

Denote a strategy of the receiver given that the state q of the signaller is known to her by $R_q(q)$. The receiver's optimum response is then given by the curve $R_q^*(q) = \arg_r \max[G(q,r)]$ (the r for which the maximum is achieved). Hence, if the signal receiver knows the true quality q of the signaller, she will optimally choose a response[1] $R_q^*(q)$.

(b) *Properties of separating equilibria*

What are the properties of these separating equilibria? We begin with two lemmas.

Lemma 2.1. *At any separating equilibrium (S,R) in the game Γ, a signaller of quality q will receive the response $R_q^*(q)$.*

In other words, the receiver will always provide exactly the response that she would if there was no signalling interaction but instead she started with perfect information. The proof is simple: by the definition of separating equilibrium, the equilibrium signalling strategy S: $q \to s$ is one-to-one and thus has an inverse $S^{-1}: s \to q$. At equilibrium the receiver must select a response r that maximizes her own net pay-off $G(q,r)$. By definition this response is given by $r = R_q^*(q)$. Having received the signal the receiver can attain this response level by picking $R(s) = R_q^*(S^{-1}(s))$. As all separating equilibria will be characterized by this response level, we call the curve $r = R_q^*(q)$ the *equilibrium path*. Notice that this lemma will not hold for pooling equilibria—in those the receiver will not have perfect information about the signaller's state.

Once we examine a strategy pair (S, R), we can also define a cost function $C_r(q, r)$ that says what cost a signaller has to pay to send a signal that would elicit the response r. This cost function is defined as $C_r(q,r) = C(q,R^{-1}(r))$. Notice that while the cost function $C(q, s)$ does not change as we move from strategy to strategy, the cost function $C_r(q,r)$ does. Nöldeke & Samuelson (1999) point out an enlightening corollary.

Corollary 2.1. *At any separating equilibrium (S,R), the signal cost function C_r serves to bring the optimization problems of signaller and responder into accord.*

More specifically, at any separating equilibrium (S,R), the following equality holds for all q:

$$\arg_r\max[H(q,r) - C_r(q,r)] = \arg_r\max[G(q,r)].$$

The logic underlying this result is as follows. At any separating equilibrium the signaller can essentially 'pick' her desired response and send the signal necessary to get it. None the less, lemma 2.1 tells us that at any separating equilibrium, the signal receiver always provides exactly the response that maximizes her fitness, $R_q^*(q)$ (assuming that the signal cost is independent of the response). Therefore, the optimal choice for the signaller must be to 'ask for' precisely the response that the responder would want to provide.

Next we can derive a set of conditions for the slope of signal cost on separating equilibria.

Lemma 2.2. *At any separating equilibrium in the game Γ, $(\partial H(q,r))/(\partial r) = (\partial C_r(q,r))/(\partial r)$ along the equilibrium path.*

The proof of this lemma, and an interesting corollary, are provided in Appendix B.

Lemma 2.2 and corollary 2.1 tell us something interesting about a signalling equilibrium: even though C_r changes as we move from strategy to strategy, every time these strategies are a signalling equilibrium, C_r will have very distinct properties: it will align the maximum of $H - C_r$ with that of G, and its slope will be equal to the slope of H.

Example 2.1. *Throughout this paper, we will examine an example loosely based on the original model of Grafen (1990). A male bird of quality q, where q is drawn from some distribution with a range $[0,1)$, signals his quality to a female, using a signal of intensity s. The display cost, C, of producing this signal depends on both the signaller's quality and the signal's intensity: $C(q,s) = s(1 - q)$. Upon observing his signal, the female selects a response r. Her pay-off is then given by the function $G(q,r) = \exp[-(q - r)^2]$; the closer in value is her response r to his quality q, the higher is her pay-off (this function was arbitrarily chosen as a function that has its maxima along the diagonal). Thus the receiver's optimal response is given by the function $R_q^*(q) = q$. The signaller receives a benefit that increases as the intensity r of the receiver's response increases: $H(q,r) = r$. We assume that the cost necessary to elicit the least-desirable response $r_{\min} = 0$ is zero for all signallers q.*

In figure 1a, we show the signaller's benefit function $H(q,r)$, along with the signaller's desired response (dashed line) and the receiver's desired response (solid line). By corollary 2.1, at a separating equilibrium, the signal cost must serve to bring these two different paths into accord. In figure 1b, we show the function $C(q, s)$ giving signal cost as a function of signaller quality q and signal intensity s. These two surfaces—the benefit surface H and the cost surface C—can be seen as defining the signalling game. Our task is then to determine optimum signaller and receiver strategies $S(q)$ and $R(s)$, and from these to compute the cost $C(q,r)$ of eliciting a given response and the net fitness consequences $H(q,r) - C_r(q,r)$ to the signaller.

For this particular example, we can find the separating equilibrium as follows. By the definition of separating equilibrium, for every signaller q', signaller's choice of signal $S(q')$ maximizes the signaller's fitness. A necessary condition is that for all q' in $[0,1)$

$$\frac{\mathrm{d}H(q',R(s'))}{\mathrm{d}s} - \frac{\mathrm{d}C(q',s')}{\mathrm{d}s}\bigg|_{s' = S(q')} = 0.$$
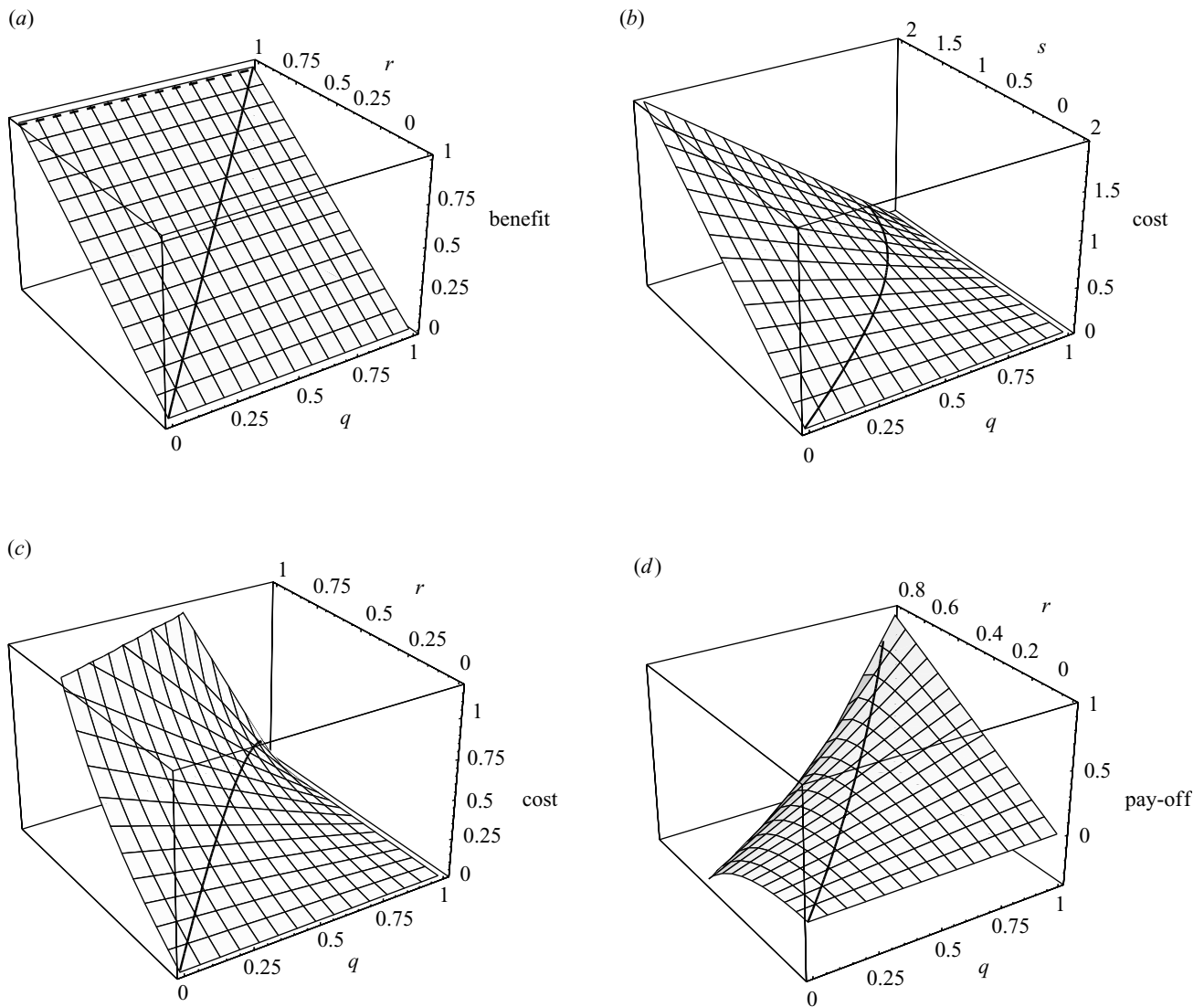
(a)

(b)

(c)

(d)

Figure 1. Benefit, cost and net fitness functions for example 2.1. The separating equilibrium is represented by the dark curves across the surfaces. (a) Benefit function $H(q,r)$: dashed line, signaller's desired response; solid line, receiver's desired response; (b) cost $C(q,s)$ as a function of quality and signal sent; (c) cost $C_r(q,r)$ as a function of quality and response elicited; (d) signaller's net fitness $H(q,r) - C_r(q,r)$.

By lemma 2.1, at this equilibrium the signaller will receive the response that optimizes the receiver's fitness, i.e. $R(S(q')) = q'$. Therefore we have

$$\left[\frac{dH(q',q')}{dr}\frac{dR(s')}{ds} - \frac{dC(q',s')}{ds}\right]\Bigg|_{s'=S(q')} = 0.$$

By the inverse rule for derivatives,

$$\frac{dR(s')}{ds}\Bigg|_{s'=S(q')} = 1\Bigg/\frac{dS(q')}{dq}\Bigg|_{q=q'}.$$

As $(dH/dr) = 1$ everywhere, we have simply

$$\frac{dS(q')}{dq}\Bigg|_{q=q'} = 1\Bigg/\frac{dC(q',s')}{ds}\Bigg|_{s'=T(q')}.$$

It follows that the signaller's strategy S is the solution to the differential equation $dS/dq = 1/(1 - q)$, i.e. $S(q) = \log[1/(1 - q)] + c$. Substituting our initial condition $C(q_0,s_0) = 0$, where $s_0 = 0$ is the signal necessary to elicit

response r_{min}, we find that $c = 0$. Receivers will then respond to a signal s with a response $R(s) = S^{-1}(s) = 1 - e^{-s}$. This gives us a candidate solution; we still have to check that it truly is an equilibrium, i.e. that no signaller can benefit from sending an alternate signal. In § 3 we will see an easier method for arriving at the equilibrium.

At this equilibrium, the signallers send signals of intensity $S(q) = -\log[1 - q]$ and receivers respond to a signal s with a response $R(s) = 1 - e^{-s}$. The equilibrium cost to a signaller q of eliciting a response r is then $C_r(q,r) = (q - 1)\log[1 - r]$. This cost function is shown in figure 1c. Figure 1d shows the signaller's net fitness $H(q,r) - C_r(q,r)$ as a function of quality and response induced. Figure 1d illustrates the consequence of lemma 2.1 and corollary 2.1: a signaller of quality q—who seeks to maximize $H(q, r) - C_r(q, r)$ with respect to r—ends up choosing the point along receiver's optimum path $q = r$. The signaller's and the responder's optimization problems were brought into accord. Moreover, simple calculus verifies that lemma 2.2 is met, as everywhere along the

equilibrium path the signaller's fitness is maximized with respect to r, i.e. $\partial H(q,r)/\partial r = \partial C_r(q,r)/\partial r$.

3. A VECTOR-FIELD APPROACH TO SIGNALLING GAMES

The task of finding a separating equilibrium for a given signalling game can be difficult, since for each candidate signalling strategy $S(q)$ we have to compute a corresponding response strategy $R(s)$ to use in determining whether the strategy $S(q)$ is actually stable. In this section, we present a method that simplifies this process. We represent equilibria as integral curves of a vector field, and in this way we are able to find all the possible separating equilibria directly from the basic functions by which the game Γ is defined: the signaller's benefit function $H(q,r)$, signal cost function $C(q,s)$, and the equilibrium path $R_q^*(q)$, which is easily derived from the receiver's benefit function $G(q,r)$ using lemma 2.1. Not only does this approach give us an easy way to find the equilibrium strategies S and R, but it also provides a way of visualizing the form of various signalling games, which serves to highlight the differences and similarities among them.

The following proposition provides the groundwork for the vector-field representation. A proof is provided in Appendix C.

Proposition 3.1. *Define the vector field V as[2]*

$$V(s',q') = \left(\left. \frac{\mathrm{d}}{\mathrm{d}s}C(q',s) \right|_{s=s'}, \left. \frac{\mathrm{d}}{\mathrm{d}q}H(q',R_q^*(q)) \right|_{q=q'} \right).$$

If a separating equilibrium exists for the game Γ with $S(q_0) = s_0$, the integral curve of V through (q_0,s_0) will be an equilibrium signalling strategy $S(q)$, provided that everywhere along this integral curve, the following inequality is satisfied:

$$\left. \frac{\mathrm{d}}{\mathrm{d}q}\frac{\mathrm{d}}{\mathrm{d}p}H(q,R_q^*(p)) \right|_{p=q',q=q'}$$

$$> \frac{\left(\left. \frac{\partial^2 C}{\partial s \partial q}(q,s) \right|_{q=q',s=s'} \right)\left(\left. \frac{\mathrm{d}}{\mathrm{d}q}H(q',R_q^*(p)) \right|_{q=q'} \right)}{\left. \frac{\mathrm{d}}{\mathrm{d}s}C(q',s) \right|_{s=s'}}. \quad (3.1)$$

The equilibrium receiver strategy is given by $R(s) = R_q^*(S^{-1}(s))$, where $S^{-1}(s)$ is the inverse of S.

Integral curves including points of the (q,s) plane at which condition 3.1 is not satisfied will not be stable separating equilibria.

The game described in example 2.1 has a corresponding vector field of $V = (1 - q,1)$, shown in figure 2. The cross derivative condition 3.1 holds everywhere. Assuming that a minimal signal of $s_{\min} = 0$ is sufficient to induce a minimum response $r_{\min}$, the equilibrium signalling strategy $S(q)$ is simply the integral curve of V through the point $(0,0)$. This curve is represented by the dark curve in figure 2. Notice that the cost of a signal s to a signaller of quality q is simply the path integral of the x-axis component of the vector field $(\partial C/\partial s)$ along the path given by the vertical line from $y = 0$ to $y = s$, as illustrated by the dashed line in figure 2. This exemplifies that the signal cost depends
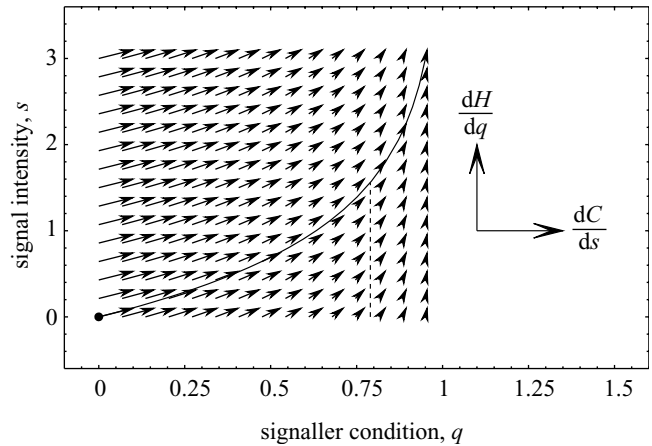
Figure 2. The vector field $V(s,q)$ for the game treated in example 2.1. Each vector has magnitude $(\mathrm{d}/\mathrm{d}q) H(q',R_q^*(q))|_{q=q'}$ direction and $(\mathrm{d}/\mathrm{d}s) C(q',s)|_{s=s'}$ in the horizontal direction. The solid curve is the equilibrium signalling strategy $S(q)$.

on the properties of the vector field off the equilibrium path. One can also see that we could have chosen many other integral curves that lie off the point $(0,0)$, for example starting at a higher point $(0,0.1)$. These would also define signalling equilibria, except that now no signaller signals the lowest cost signals, such as the signal $s = 0$. The stability of these equilibria depends on how the receiver responds to signals that the signaller never sends. If the receiver 'generalizes' and responds to a signal that is lower than 0.1 as she would to a signaller of quality 0, the equilibrium would not be stable. However, this does not have to be the case, and is part of the question of equilibrium selection. For this paper we will not consider the stability of a signalling strategy versus signals that no signaller employs.

Figure 4a is a contour plot of the signaller's total fitness, $H(q, R(s)) - C(q,s)$, when the receivers use the equilibrium response strategy R appropriate to the signalling strategy S for which $S(0) = 0$. (This signalling strategy corresponds to the integral curve through $(0,0)$, and is illustrated by the heavy yellow line.) Notice that the signalling strategy always intersects the contour lines when they are vertical. This means that the signalling strategy $S(q)$ extremizes (maximizes, in this case) the signaller's fitness function for every signalling quality q.

The vector-field approach can also help us to understand why a separating equilibrium might fail to exist. Siller (1998) presents an example of a game similar to the model of Grafen (1990), for which no separating equilibrium exists. While the mathematics behind the model of Siller (1998) may appear somewhat daunting, the vector-field approach clarifies the problem that Siller (1998) describes.

Example 3.1. *As in example 2.1, a male bird of quality $q \in [0, 1)$ signals his quality using a signal s and receives a response r. Benefit functions are as before: the receiver gets a pay-off $G(q,r) = \exp[-(q - r)^2]$ and the signaller gets a pay-off $H(q,r) = r$, with the signal necessary to receive response $r_{\min} = 0$ costing zero for all signallers q. In this example, define the signal cost as follows:*
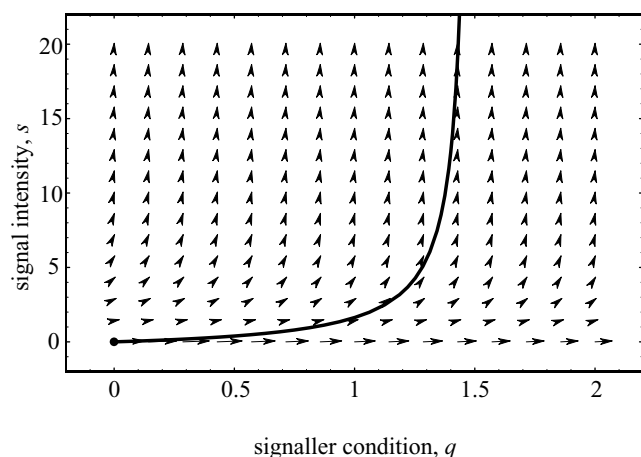
Figure 3. The vector field and integral curve through (0,0) for example 3.1.

$$C(q,s) = 2 - \frac{e^{-q} + 1}{s + 1}.$$

This is an unusual signal cost function: the cost of any particular signal increases with signaller quality, but the marginal cost of signalling decreases both with signaller quality and with signal intensity. Consider signallers with qualities q on the interval [0,2], and assume that the signaller of quality q_{min} = 0 sends a signal of intensity s_{min} = 0.

Figure 3 shows the vector field for this model. As is clear from the figure, the signal intensity 'blows up' near $q = 1.5$, as the vector field along the integral curve describing the signal path approaches an infinite slope. Indeed, as Siller (1998) points out, all signallers beyond some limiting point will send a signal of infinite intensity. The maximal integral curve through the point (0,0) does not afford an inverse on the domain [0,2], and so no separating equilibrium exists. Siller points out that this situation may be reasonable when signal intensity measured by a ratio such as that of pigmented surface area to non-pigment surface area.

4. MANY COST FUNCTIONS ALLOW STABLE SIGNALLING

In § 1, we assumed that the cost function was specified as part of the game Γ. However, many different cost functions can allow honest signalling for particular benefit functions $H(q,r)$ and $G(q,r)$. In this section, we will explore the properties of the separating equilibria that correspond to different cost functions C. We will find that signal costs at these equilibria may be higher or lower than those found above.

Example 4.1. *As before, a male bird of quality $q \in [0,1]$ signals his quality using a signal s and receives a response r. Benefit functions are in example 2.1: the receiver gets a pay-off $G(q,r) = \exp[-(q - r)^2]$ and the signaller gets a pay-off $H(q,r) = r$. The signal needed to elicit response $r_{min} = 0$ costs zero for all signallers. In this example, however, we consider a range of different signal cost functions, represented by the function $C(q, s) = s^\alpha(1 - q)^\alpha$, where α is a strictly positive constant.*

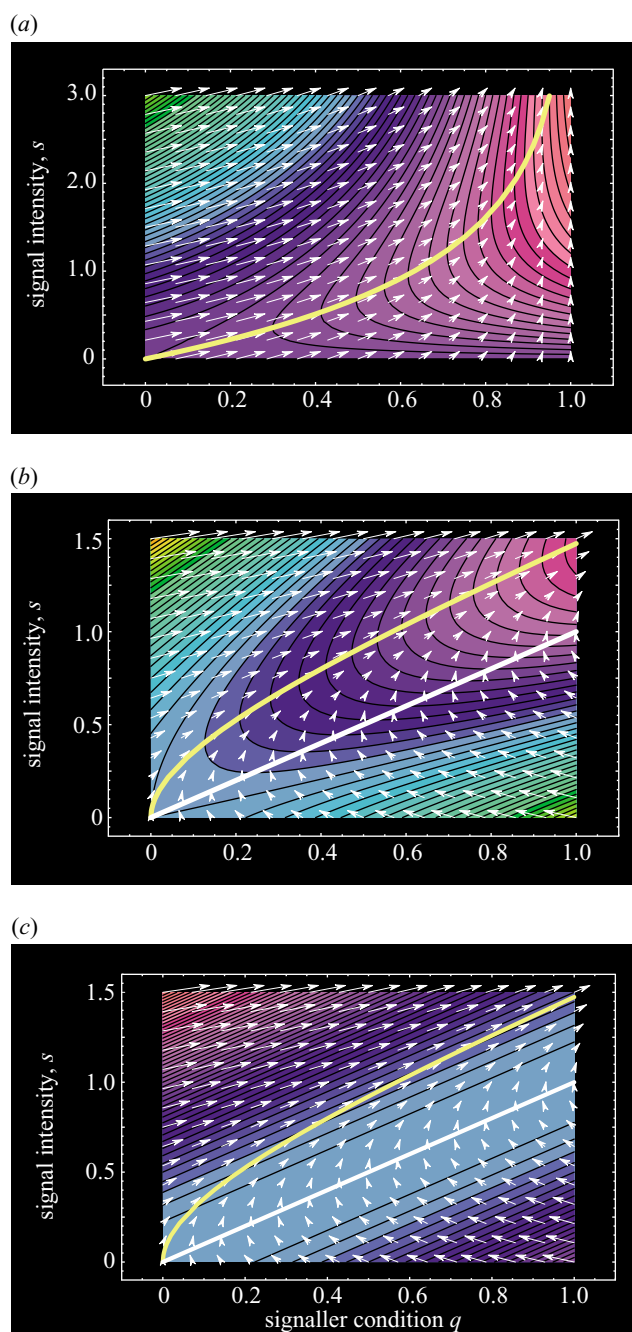Figure 5 shows the signal cost for three different cost

(a)

(b)

(c)

Figure 4. Vector fields, signalling equilibrium and contour plots of fitness or cost. (a–c) Signaller quality q versus signal intensity s. Arrows represent the vector field V. A signalling equilibrium strategy S(q) that passes through the point (0,0) is plotted in yellow, over a contour plot of the fitness function $H(q, R(s)) - C(q,s)$. (a) Example 2.1, V, S and total fitness; (b) example 5.1, V, S and total fitness; (c) example 5.1, V, S and cost C; vector field and signalling strategy overlaid over the cost $C(q,s)$.

functions, corresponding to $\alpha = 1/4$, $\alpha = 1$, and $\alpha = 4$. Figure 5a shows cost as a function of signaller quality and signal intensity, figure 5b shows cost as a function of signaller quality and response received. Each cost function affords a separating equilibrium; equilibrium signal cost at this equilibrium is shown as a function of signaller quality in figure 5c. This equilibrium signalling strategy can be found as in example 2.1, and for $\alpha \neq 1$ is given by
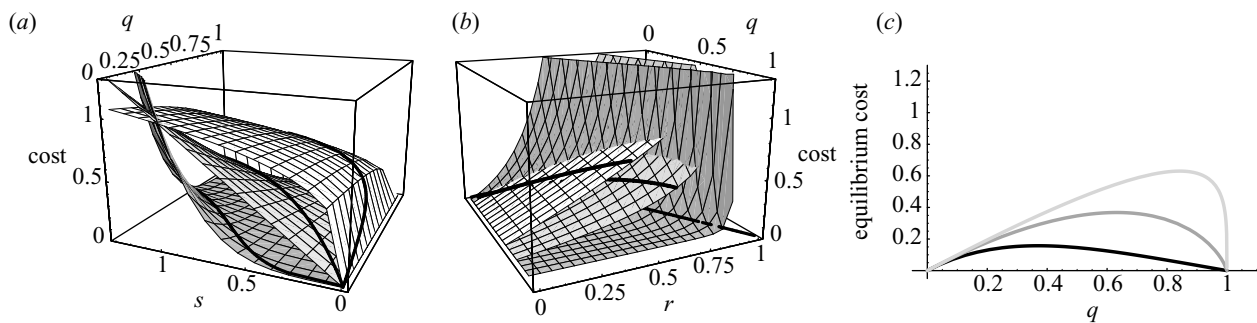
Figure 5. Signal costs for cost functions $C(q,s) = s^\alpha(1-q)^\alpha$. Light grey, $\alpha = \frac{1}{4}$; medium grey, $\alpha = 1$; dark grey, $\alpha = 4$. Dark curves show the equilibrium path for each cost function. (*a*) Cost of signalling, as a function of signaller's quality and signal intensity. (*b*) Cost of signalling, as a function of signaller's quality and response received given equilibrium behaviour by the receiver. The upper cost surfaces have been partially cut away to reveal those beneath them. (*c*) Signal cost at equilibrium, as a function of signaller quality.

$$S(q) = \left(\frac{(1-q)^{1-\alpha} - 1}{\alpha - 1} \right)^{1/\alpha}.$$

The signalling costs at equilibrium for $\alpha \neq 1$ are given by

$$C(q, S(q)) = \frac{(1-q)^\alpha - (1-q)}{1-\alpha}$$

(when $\alpha = 1$, the equilibrium becomes that found in example 2.1). Thus, the signal costs at equilibrium differ dramatically for the three cost functions. Unless we know the exact shape of the cost function—which typically will depend on the biology of the organism and the mechanics of signalling—signalling theory will not offer precise predictions about the necessary signal costs at equilibrium.

Indeed, we can even find a signal cost function C of this form so that equilibrium signal costs are arbitrarily cheap. Calculus reveals that the equilibrium cost paid by a signaller of quality q will be less than δ provided that α is sufficiently large.[3] For example, $\alpha = 8$ is sufficient to ensure that equilibrium signal costs are <0.1 for all signallers. In this way, we can construct cost functions that allow arbitrarily cheap separating equilibria. Lachmann *et al.* (2001) used a similar approach (but a different functional form) to derive analogous low-cost equilibria for the paradigmatic signalling games of Grafen (1990) and Godfray (1991). These models—which were designed to show that costly signalling is indeed *feasible*—should not be interpreted as evidence that costly signalling is *necessary*, even when signallers' and receivers' interests conflict. Indeed, contrary to the principal claim of the handicap principle (Zahavi 1975, 1977), these results show that reliable signalling does not require waste.

We can get an intuitive understanding of *why* different cost functions allow different equilibrium costs, by examining the form of the equilibrium conditions. Consider figure 5*b*: lemma 2.2 specifies the slope of the cost surface $C_r(q,r)$ in the '*r* direction', everywhere along the equilibrium path, but says nothing about the slope of the cost surface in the *q* direction. If the cost surface were flat in the '*q* direction', then there would be no flexibility in the shape of the cost function; absolute signal cost along the equilibrium path would be determined as well. But the surface need not be flat in the *q* direction. Thus, as one moves along the equilibrium path, marginal increases in signal cost from movement in the *r* direction can be offset by marginal decreases in signal cost from movement in the *q* direction. In fact, figure 5*c* shows that equilibrium signal cost decreases for high-quality signallers. This is because in this case the increase necessary for the equilibrium to be stable in the *s* direction is more than offset by the change of the cost in the *q* direction. Thus, we see that cheap signalling is facilitated by the *interaction* between signal and quality in determining signal cost. If there is no interaction—for example, if signal cost is independent of signaller quality *q*—cheap equilibria will not exist.

But should we really expect signal cost to interact with signaller quality and the signal sent? In particular, is it realistic to expect real signal cost functions to take forms that will allow low-cost signalling? Perhaps unsurprisingly, this depends on the biology of the system in question. In some circumstances, signal costs will probably not depend on signaller quality much at all. For begging baby birds, much of the signal cost may be associated with increased predation risk; given a particular level of begging intensity, this risk is unlikely to vary much with signaller condition. In other circumstances, signal costs may depend on signaller quality, but not in the right way to facilitate low-cost separating equilibria. None the less, certain scenarios may provide the proper forms of signalling costs. In particular, when signal costs arise through social punishment of dishonest signallers, rather than through the production costs themselves, honest signals can be free while even relatively small deceptions can provoke retaliation and substantial costs. Lachmann *et al.* (2001) explore signal costs of this form in further detail, with a particular focus on human language.

Empirical studies that measure the absolute costs of equilibrium signals will usually not suffice to confirm or reject costly signalling hypotheses. Measurements of net expenditure on signals tell us little, because many different costs (or even zero cost) are possible at equilibrium. Furthermore, comparisons of equilibrium signal costs tell us little because these costs can increase or decrease with signaller quality (Getty 1998). If we cannot easily predict absolute cost, what predictions can we make from signalling theory? To answer this, we seek common features of these equilibria. Lemma 2.2 tells us that for a given signaller quality q, the slope of cost with respect to a response that is induced around equilibrium $(\partial C_r(q,r))/\partial r$ is the

same regardless of which cost function we are looking at. Indeed, here $(\partial C_r(q,r))/\partial r = (1 - q)^\alpha/(1 - r)^\alpha = 1$ along the equilibrium path $R_q^*(q) = r$ for all positive values of α. To phrase this another way, at equilibrium the marginal costs of signalling are the same regardless of the shape of the signalling function. As the theory's strong and unambiguous predictions involve the consequences of deviations from equilibrium, empirical studies of honest signalling may require signal magnitudes to be experimentally altered. We are not the first to address the marginal–absolute cost distinction (Zahavi 1981; Nur & Hasson 1984; Michod & Hasson 1990; Hasson 1997; Getty 1998; Lachmann *et al.* 2001). Nevertheless, we mention it here because its importance is not easily understated.

5. WHEN SIGNALS HAVE INDEPENDENT MEANING

In almost all models of costly signalling, signal cost is assumed to increase monotonically with the intensity of the signal (the only exceptions are the discrete models in Hurd 1995; Számadó 1999). Holding the response of the signal receiver constant, any signaller—regardless of quality—will optimize fitnesses by expressing the minimal signal intensity.

In practice, however, some signals may be structured so that different signallers have different minimal-cost signal intensities. Consider, for example, the ornamental tails exhibited by many bird species. While these extended tails undoubtedly play a signalling role, they may also influence flight performance. If so, different quality signallers would have different optimal tail lengths even if the tail had no signalling function (Hurd 1995).

We can use the framework that we have developed thus far to explore the properties of separating equilibria for signals of this type. Let us move directly to an example.

Example 5.1. *As in previous examples, a male bird of quality $q \in [0,1)$ signals his quality using a signal s and receives a response r. Benefit functions are as always: the receiver gets a pay-off $G(q,r) = e^{-(q-r)^2}$ and the signaller gets a pay-off $H(q,r) = r$. Here, however, let us assume that a signaller of quality q can most cheaply send a signal of intensity $s = q$; any deviation imposes a fitness cost. In particular, we will define $C(q,s) = (q - s)^2$. One might think that this sort of trait could facilitate honest cost-free signalling (of q), between individuals with conflicting interests, but this turns out not to be the case.*

The optimal signaller and receiver strategies for this game are as follows, where W is the Lambert W function,

$$S(q) = \frac{1}{2}(1 + 2q + W(-e^{-1-2q}))$$

$$R(s) = s + \frac{e^{-2s} - 1}{2}. \tag{5.1}$$

Figure 4*b* shows the net signaller pay-off (coloured contours), equilibrium signaller strategy (yellow curve) and minimum-cost signalling strategy (white curve). The cross-derivative condition in this case is valid everywhere above the diagonal ($s = q$). The signalling equilibrium in this case starts on the optimal signal line $s = q$ at $q = 0$, but immediately heads up in signal intensity (with an infinite derivative) . This is because the slope of the cost function at the optimum is zero—one needs to move very quickly

off this optimum to achieve any rise in cost. Figure 4*c* shows the same signalling strategy curves, this time overlaid on a contour plot of signal cost. From this figure we can see that the separating equilibrium does not follow the 'zero-cost' valley of the cost surface (indicated by the white curve), but rather heads off it, with each signaller paying a bit extra to signal.

Another way to think of this is that the marginal cost of signalling along the zero-cost valley is zero everywhere. If the marginal gain is non-zero, then by lemma 2.2 the equilibrium cannot possibly lie along this ridge but rather it must lie up off the valley floor, where the marginal cost is also non-zero.

How far above the valley floor does the signal have to be? In Appendix D it is shown that for H, which is only a function of r and not of q, the following equation holds:

$$\Delta C = \Delta H + \iint \frac{\partial^2 C}{\partial s \partial q}, \tag{5.2}$$

where ΔC is the cost difference for signaller q of the signal sent minus the optimal signal with ΔH being the fitness difference between the signaller q and the worst signaller. This means that $\partial^2 C/\partial s \partial q$ is the only factor that causes the cost of signalling to be not more than the benefit of being a higher quality signaller (notice that in our example this integral is negative). This result is shown by using Gauss's Theorem on the vector field V.

So far, we have applied this mathematical model to a scenario in which different signallers have different minimal-cost signals. The same analysis can also be applied to non-signalling traits. As Lotem *et al.* (1999) point out, we should expect organisms to exaggerate non-signalling traits once observers begin to use these traits (or associated cues: Maynard Smith & Harper (1995)) as sources of information. By interpreting the 'signal intensity' in our model as the quantitative value of the non-signalling trait, the model provides a mathematical justification for the argument of Lotem *et al.*: non-signalling traits will deviate from phenotypically optimal levels when observers use these traits in assessing an individual's quality. Lachmann & Bergstrom (1999) derive an analogous result by alternative means.

6. CONCLUSIONS

By constructing a pair of ingenious game-theoretical models, Grafen (1990) and Godfray (1991) demonstrated that the basic idea underlying the handicap principle of Zahavi (1975) is at least feasible. Individuals with divergent interests can communicate honestly using costly signals in an evolutionarily stable manner. Godfray (1991) and Grafen (1990) each presented specific examples of continuous signalling games, and found that for these examples, costly signals allow a stable separating equilibrium. Their results have often been interpreted as implicit support for the stronger claim made by Zahavi (1975): signal cost is *necessary* for honest communication when interests diverge.

In this paper, we have used a new vector-field solution method to show that this interpretation is not valid. When interests conflict, costly signals allow—but are not necessarily essential for—honest signalling. Depending on the

form of the function mapping from signaller quality and signal intensity to signal cost, an honest signalling system can exhibit any of a broad range of equilibrium signal costs. Separating equilibria can even have extremely low equilibrium signal costs.

These findings have important implications for empirical testing of costly signalling theory. Studies that measure the absolute costs of equilibrium signals will usually not suffice to confirm or reject costly signalling hypotheses, because many different costs are possible at equilibrium. *Even the complete absence of signal cost at equilibrium (where most empirical measurements are taken) is insufficient to reject the hypothesis that signal cost plays a strategic role in facilitating signal honesty.* Furthermore, comparisons of equilibrium signal costs tell us little because these costs can change with signaller quality (Getty 1998). If we cannot easily predict absolute cost, what is left for signalling theory? What predictions *can* we make?

To answer these questions, we seek common features of signalling equilibria, and find that signalling theory's strong and unambiguous predictions involve not equilibrium costs but rather the costs and consequences of deviations from equilibrium. For this reason, we suggest that empirical studies of honest signalling may need to focus on the *marginal* costs of increased signal intensity, by experimentally perturbing signal levels or otherwise inducing individuals to alter the normal choices of signal intensity.

ENDNOTES

[1] Note that without loss of generality, 'quality' q can be defined so that the function $R_q^*(q)$ is monotone. We will take this approach here, and further assume that $R_q^*(q)$ is strictly increasing, i.e., that the receiver cares about quality differences across the entire range $q \in [q_{min}, q_{max}]$. For an alternative approach, see Proulx (2001).

[2] Strictly speaking, the vector field V should be defined as

$$v(s',q') = \begin{cases} \left(\left. \dfrac{d}{ds}C(q',s) \right|_{s=s'}, \left. \dfrac{d}{dq}H(q',R_q^*(q)) \right|_{q=q'} \right) & \text{if } \left. \dfrac{d}{ds}C(q',s) \right|_{s=s'} \geqslant 0 \\[2ex] \left(-\left. \dfrac{d}{ds}C(q',s) \right|_{s=s'}, -\left. \dfrac{d}{dq}H(q',R_q^*(q)) \right|_{q=q'} \right) & \text{if } \left. \dfrac{d}{ds}C(q',s) \right|_{s=s'} < 0 \end{cases}.$$

However, if we are willing to treat the integral curves as undirected, the simpler formulation proves adequate.

[3] Specifically, equilibrium cost to a signaller q will be less than δ provided that

$$\alpha > \frac{1-q+\delta}{\delta} - \frac{W((1/\delta)(1-q)^{(1-q+\delta)/\delta}\log[1-q])}{\log[1-q]},$$

where W is the Lambert W-function.

APPENDIX A: DEFINITION OF VARIABLES AND FUNCTIONS

variable	definition
$q \in [q_{min}, q_{max}]$	signaller quality
s	signal sent
$r \in [r_{min}, r_{max}]$	receiver's response
$H(q,r)$	fitness value to signaller of quality q when receiving response r
$G(q,r)$	fitness value to receiver for providing response r to a signaller of quality q
$C(q,s)$	fitness cost for signaller of quality q of sending signal s
$C_r(q,x)$	fitness cost for signaller of quality q when sending a signal that would elicit the response x
$R(s)$	receiver strategy specifying what response to provide signaller who sends signal s
$R_q(x)$	receiver strategy specifying what response to provide to a signaller of quality x
$R_q^*(x)$	receiver's optimal response to a signaller of quality x
$S(q)$	signaller strategy specifying signal to send when the quality is q
$V(s,q)$	two-dimensional vector field representing partial derivatives of C and H, and used for finding signalling equilibria

APPENDIX B: PROOF OF LEMMA 2.2

(a) *Proof*

Basically, we want to show that at signalling equilibrium, the receiver strategy automatically adjusts for the marginal cost of signalling along the equilibrium path, so that only the signal cost—and not the actual signal intensity—affects the equilibrium response level. We can do this as follows. Let (S,R) be a separating equilibrium, where $R(s)$ is the receiver's equilibrium response to a sig-

nal s and $S(q)$ is the signaller's equilibrium signal when in condition q. As (S,R) is an equilibrium, the signaller's strategy S will maximize his own net pay-off $H(q,r)-C(q,s)$ for every q. Because all component functions are smooth and since at equilibrium $C(q,s) = C_r(q,R(s))$, this requires that for every q,

$$\frac{\partial[H(q,R(s)) - C_r(q,R(s))]}{\partial s}\bigg|_{s=S(q)} = 0. \tag{B 1}$$

By lemma 2.1, if a signalling equilibrium (S,R) exists,

the equilibrium path will be given by $R(S(q)) = R_q^*(q)$. Therefore, for every signalling condition q',

$$\left.\frac{\partial H(q',r)}{\partial r}\right|_{r=R_q^*(q')} \left.\frac{\mathrm{d}R}{\mathrm{d}s}\right|_{s=S(q')} \left.\frac{\mathrm{d}S}{\mathrm{d}q}\right|_{q=q'}$$
$$-\left.\frac{\partial C_r(q,r)}{\partial r}\right|_{r=R_q^*(q')} \left.\frac{\mathrm{d}R}{\mathrm{d}s}\right|_{s=S(q')} \left.\frac{\mathrm{d}S}{\mathrm{d}q}\right|_{q=q'} = 0. \tag{B 2}$$

As we assumed that the receiver's response $R_q^*(q)$ is strictly increasing, it follows that for every q', at separating equilibrium $(\mathrm{d}R(s)/\mathrm{d}s)(\mathrm{d}s/\mathrm{d}q) \neq 0$. It then follows immediately from equation (B 2) that for every q,

$$\left.\frac{\partial H(q,r)}{\partial r}\right|_{r=R_q^*(q)} = \left.\frac{\partial C_r(q,r)}{\partial r}\right|_{r=R_q^*(q)}, \tag{B 3}$$

where $r = R_q^*(q)$ is the equilibrium path.□

In several signalling models (e.g. Godfray 1991; Johnstone & Grafen 1992), the cost of signalling $C(q,s)$, is assumed to be independent of the signaller's quality q. In models of this type, the cost $C_r(q,r)$ of eliciting a response r is also independent of q; the cost functions can thus be written as $C(s)$ and $C_r(r)$. The following corollary then applies.

Corollary B 1. *When signal costs are independent of the signaller quality, the equilibrium signal cost $C_r(r)$ of eliciting a response r does not depend on the relationship between signal level and signal cost $C(s)$, provided that a separating equilibrium exists for the cost function C.*

Suppose that from a game Γ, we generate a new game Γ' by altering the display cost function $C(s)$ by a monotone transformation θ to generate a new display cost function $C'(s) = \theta(C(s))$. If the game Γ had a separating equilibrium, the game Γ' will have a separating equilibrium (S', R') for which the signal costs $C_r(r)$ of eliciting responses r will be exactly as for the game Γ.

Several important models of costly begging signals (Godfray 1991; Johnstone & Grafen 1992; Godfray 1995) assume that signal cost is a monotone function of signal intensity, effectively equating signal intensity with signal cost. This corollary provides a justification for this assumption under certain circumstances, namely, for the separating equilibria of games in which signal costs are independent of signaller quality. In these models, there will always be a one-to-one mapping between intensity and cost at separating equilibrium, and either scale can be used in analysing the system. However, Lachmann and Bergstrom (1998) show that signal and cost cannot be equated when considering pooling equilibria. Getty (1998) provides an example in which signal costs depend on signaller quality and in which higher-intensity signals actually cost *less* to their producers than lower-intensity signals cost to their producers; here intensity and cost clearly cannot be equated either.

APPENDIX C: PROOF OF PROPOSITION 3.1

Now to proposition 3.1. Define the vector field V as

$$V(s',q') = \left(\left.\frac{\mathrm{d}}{\mathrm{d}s}C(q',s)\right|_{s=s'}, \left.\frac{\mathrm{d}}{\mathrm{d}q}H(q',R_q^*(q))\right|_{q=q'}\right).$$

If a separating equilibrium exists for the game Γ with $S(q_0) = s_0$, the equilibrium signalling strategy $S(q)$ will be the maximal integral curve of V through (q_0, s_0). The equilibrium receiver strategy is given by $R(s) = R_q^*(S^{-1}(s))$ where $S^{-1}(s)$ is the inverse of S.

To prove this, we first notice that since we are dealing with a separating equilibrium, the integral curve of V can be expressed as a function of q. We will now show that if (i) signallers of qualities at and around q_0 signal $T(q)$, (ii) the signalling function $T(q)$ has all necessary derivatives, and (iii) this strategy is stable to small changes of the signaller strategy, then the direction of $(q, S(q))$ at q_0 is equal to the direction of $(q, T(q))$ there.

As usual, we focus on the game Γ and assume that the signal strategy $T(q)$ together with the response strategy $R(s)$ constitute a separating signalling equilibrium in a neighbourhood of q_0. The signal costs $C(q,s)$ are assumed to be non-negative for all q and s. We do not need to look at an alternative response strategy R', because by lemma 2.1 the responder will use a response on the equilibrium path $R_q^*(q)$ such that $R(T(q)) = R_q^*(q)$.

For T to be an equilibrium, we require that the following expression hold for all q around q_0 and for all $s \neq T(q)$:

$$H(q,R(T(q))) - C(q,T(q)) \geqslant H(q,R(s)) - C(q,s). \tag{C 1}$$

Therefore, if T is to be an equilibrium, it is necessary that for all q,

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}[H(q,R(s)) - C(q,s)]\right|_{s=T(q)} = 0, \tag{C 2}$$

i.e. for all q,

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}H(q,R(s))\right|_{s=T(q)} = \left.\frac{\mathrm{d}}{\mathrm{d}s}C(q,s)\right|_{s=T(q)}. \tag{C 3}$$

Expanding the left-hand side, for all q

$$\left.\frac{\partial}{\partial r}H(q,r)\right|_{r=R(T(q))} \left.\frac{\mathrm{d}}{\mathrm{d}s}R(s)\right|_{s=T(q)} = \left.\frac{\mathrm{d}}{\mathrm{d}s}C(q,s)\right|_{s=S^*(q)}. \tag{C 4}$$

As we said before, at equilibrium, we know by lemma 2.1 that the responder will use a response on the equilibrium path $R_q^*(q)$ and therefore $R(S(q)) = R_q^*(q)$:

$$\left(\left.\frac{\partial}{\partial r}H(q,r)\right|_{r=R_q^*(q)}\right)\left(\left.\frac{\mathrm{d}}{\mathrm{d}s}R(s)\right|_{s=T(q)}\right) = \left.\frac{\mathrm{d}}{\mathrm{d}s}C(q,s)\right|_{s=T(q)}, \tag{C 5}$$

i.e. for all q,

$$\left.\frac{\partial}{\partial r}H(q,r)\right|_{r=R_q^*(q)} = \left.\frac{\mathrm{d}}{\mathrm{d}s}C(q,s)\right|_{s=T(q)} \bigg/ \left.\frac{\partial}{\partial s}R(s)\right|_{s=T(q)}. \tag{C 6}$$

This equation is equivalent to expression (2) in Grafen (1990). At equilibrium, the responder's response is $R(s) = R_q^*(T^{-1}(s))$. Differentiating this expression with respect to s, we see that

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}R(s)\right|_{s=T(q)} = \left.\frac{\mathrm{d}}{\mathrm{d}s}R_q^*(T^{-1}(s))\right|_{s=T(q)}$$

$$= \left(\left.\frac{\mathrm{d}}{\mathrm{d}q}R_q^*(T^{-1}(s))\right|_{s=T(q)}\right)\left(\left.\frac{\mathrm{d}}{\mathrm{d}s}T^{-1}(s)\right|_{s=T(q)}\right). \tag{C 7}$$

The integrals above are each evaluated along the equilibrium signal path $s = T(q)$, or equivalently $q = T^{-1}(s)$. Applying this substitution

$$\frac{\mathrm{d}}{\mathrm{d}s}R(s)\bigg|_{s=T(q)} = \left(\frac{\mathrm{d}}{\mathrm{d}q}R_q^*(q)\right)\left(\frac{\mathrm{d}}{\mathrm{d}s}T^{-1}(s)\bigg|_{s=T(q)}\right). \qquad \text{(C 8)}$$

By the inverse rule for derivatives

$$\frac{\mathrm{d}}{\mathrm{d}s}R(s)\bigg|_{s=T(q)} = \frac{\mathrm{d}}{\mathrm{d}q}R_q^*(q)\bigg/\frac{\mathrm{d}}{\mathrm{d}q}T(q). \qquad \text{(C 9)}$$

Substituting this into equation (C 6), we get an expression for the relationship between marginal costs and gains of signalling at equilibrium

$$\left(\frac{\mathrm{d}}{\mathrm{d}q}R_q^*(q)\right)\left(\frac{\partial}{\partial r}H(q,r)\bigg|_{r=R_q^*(q)}\right) = \left(\frac{\mathrm{d}}{\mathrm{d}q}T(q)\right)\left(\frac{\partial}{\partial r}C(q,s)\bigg|_{s=T(q)}\right). \qquad \text{(C 10)}$$

This says that at equilibrium, the product of the marginal change of response with respect to a change in signaller's condition, and the marginal gain from response with respect to a change in the response is equal to the product of the marginal change in signal with respect to a change in the signaller's condition and the marginal change in cost with respect to a change in the signal.

Alternatively, equation (C 10) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}\bar{q}}H(\bar{q},R_q^*(\bar{q})) - \frac{\partial}{\partial q}H(\bar{q},R_q^*(\bar{q})) = \left(\frac{\mathrm{d}}{\mathrm{d}q}T(q)\right)\left(\frac{\mathrm{d}}{\mathrm{d}s}C(q,s)\bigg|_{s=T(q)}\right). \qquad \text{(C 11)}$$

If we are interested in the decision problem facing a particular signaller q' (who cannot, after all, change q'), then we can evaluate $H(q,r)$ and $C(q,s)$ at $q = q'$ only, treating these as functions of a single variable and writing them as $H_{(q')}(r)$ and $C_{(q')}(s)$, respectively.

For this particular signaller q', equation (C 12) becomes simply

$$\frac{\mathrm{d}}{\mathrm{d}q}H_{(q')}(R_q^*(q))\bigg|_{q=q'} = \frac{\mathrm{d}}{\mathrm{d}q}C_{(q')}(T(q))\bigg|_{q=q'}. \qquad \text{(C 12)}$$

Qualitatively, this means that for a given signaller q', the marginal fitness gain from the response induced by pretending to be a slightly better (or slightly worse) signaller $q' + \delta$ is equal to the marginal fitness loss from the cost of pretending to be this signaller. We have re-derived our old result that at equilibrium, the marginal cost of exaggerating one's condition equals the marginal benefit of doing so, only now we have included the signal s and the signalling strategy $T(q)$ explicitly. We can now solve for this signal function directly. Carrying over the derivative in equation (C 12) with respect to s, and rearranging, we find that

$$\frac{\mathrm{d}}{\mathrm{d}q}T(q)\bigg|_{q=q'} = \frac{\mathrm{d}}{\mathrm{d}q}H_{(q')}(R_q^*(q))\bigg|_{q=q'}\bigg/\frac{\mathrm{d}}{\mathrm{d}s}C_{(q')}(s)\bigg|_{q=T(q')}. \qquad \text{(C 13)}$$

So, after much work, we arrived at the expression we wanted. This expression says that at q_0 the direction of $(q,T(q))$ is equal to the direction of the curve $S(q)$ defined by the vector field V.
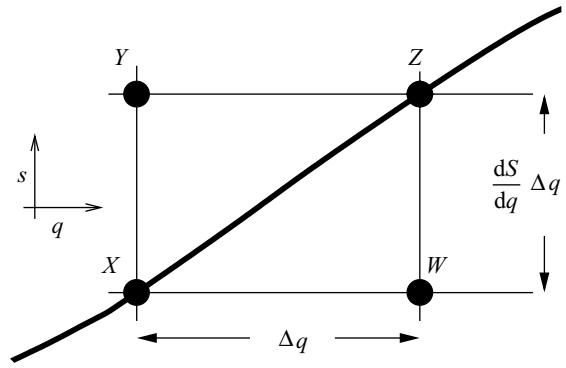
Figure 6. Schematic representation of a point X on the equilibrium signalling path $S(q)$, and three points in close proximity.

(a) Cross derivative condition

We want to ensure that the signalling equilibrium is indeed a local maximum of the total fitness function.

Figure 6 shows a local neighbourhood of point X, which lies along a signalling path $S(q)$. Point Z which is a small distance Δq away also lies on the signalling path. Points Y and W lie off the signalling path. For S to be a stable signalling strategy, requirement (C 1) has to hold, and in particular

$$H(X) - C(X) \geqslant H(Y) - C(Y), \qquad \text{(C 14)}$$

$$H(Z) - C(Z) \geqslant H(W) - C(W). \qquad \text{(C 15)}$$

From which we get

$$[H(Z) - H(W)] - [H(Y) - H(X)] \geqslant \\ [C(Z) - C(W)] - [C(Y) - C(X)]. \qquad \text{(C 16)}$$

Expanding to the second order in Δq, we get

$$C(Y) - C(X) = \frac{\partial C}{\partial s}\frac{\mathrm{d}S}{\mathrm{d}q}\Delta q + \frac{1}{2}\frac{\partial^2 C}{\partial s^2}\left(\frac{\mathrm{d}S}{\mathrm{d}q}\right)^2(\Delta q)^2 \qquad \text{(C 17)}$$

$$C(Z) - C(W) = \frac{\partial C}{\partial s}\frac{\mathrm{d}S}{\mathrm{d}q}\Delta q + \frac{1}{2}\frac{\partial^2 C}{\partial s^2}\left(\frac{\mathrm{d}S}{\mathrm{d}q}\right)^2(\Delta q)^2 \\ + \frac{\partial^2 C}{\partial q \partial s}\frac{\mathrm{d}S}{\mathrm{d}q}(\Delta q)^2. \qquad \text{(C 18)}$$

We can do the same for H, noticing that a signaller of quality q will get the response $R_q^*(q + \Delta q)$ when signalling at point Y. From all this we get the condition

$$\frac{\mathrm{d}}{\mathrm{d}q}\frac{\mathrm{d}}{\mathrm{d}p}H(q,R_q^*(p)) \geqslant \frac{\partial^2 C}{\partial q \partial s}\frac{\mathrm{d}S}{\mathrm{d}q} \qquad \text{(C 19)}$$

and if we expand $\mathrm{d}S/\mathrm{d}q$ we get finally

$$\frac{\mathrm{d}}{\mathrm{d}q}\frac{\mathrm{d}}{\mathrm{d}p}H(q,R_q^*(p)) \geqslant \left(\frac{\partial^2 C}{\partial q \partial s}\right)\left(\frac{\mathrm{d}}{\mathrm{d}p}H(q,R_q^*(p))\right)\bigg/\frac{\partial C}{\partial s}, \qquad \text{(C 20)}$$

rearranging a bit will give us

$$\frac{\partial C}{\partial s}\left(\frac{\frac{\mathrm{d}}{\mathrm{d}q}\frac{\mathrm{d}}{\mathrm{d}p}H(q,R_q^*(p))\frac{\partial C}{\partial s} - \left(\frac{\partial^2 C}{\partial q \partial s}\right)\left(\frac{\mathrm{d}}{\mathrm{d}p}H(q,R_q^*(p))\right)}{\left(\frac{\partial C}{\partial s}\right)^2}\right) \geqslant 0 \qquad \text{(C 21)}$$
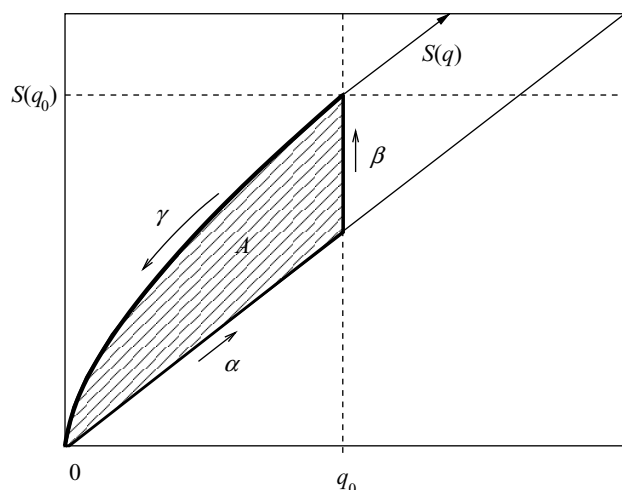
or

Figure 7. Schematic representation of the line α, β, γ, and the area A, that the integral is taken over. Shown is an integral curve of the vector field V, with the optimum of the cost function on the diagonal, as in example 5.1 and figure 4c.

$$\frac{\partial C}{\partial s}\frac{\mathrm{d}}{\mathrm{d}q}\frac{\mathrm{d}S}{\mathrm{d}q} \geqslant 0. \tag{C 22}$$

Notice that if this expression is not met at some point (q,s), then no stable separating equilibrium can assign signal s to quality q.

One can interpret this expression as saying that in areas in which C increases with s, the derivative of $S(q)$ has to increase, i.e. the vector V should rotate to the left.

APPENDIX D: CALCULATING THE COST DIFFERENCE FROM THE OPTIMUM

Figure 7 shows a schematic representation of the situation described in example 5.1. The signalling strategy $S(q)$ lies along an integral curve of the vector field V, and the diagonal describes the optimum of the signal cost function C. We want to calculate the cost difference for signaller q_0 between the cost on the signalling strategy $C(q_0, S(q_0))$ and the optimal cost (which in this case is on the diagonal).

In figure 7 we see the area A, which is enclosed by the curves α, β and γ. α is on the optimum of C, from 0 to q, β is vertical at q_0, and γ goes back along the signalling strategy S. Gauss's theorem tells us that

$$\int_{\partial A} V \cdot n \, \mathrm{d}s = \iint_A \nabla \cdot V \, \mathrm{d}a, \tag{D 1}$$

where n is the normal to the curve along which the integral is taken.

So we have

$$\int_{\partial A}\left(\frac{\mathrm{d}}{\mathrm{d}s}C(q',s)\Big|_{s=s'}, \frac{\mathrm{d}}{\mathrm{d}p}H(q',R_q^*(p))\Big|_{p=q'}\right)\cdot n \, \mathrm{d}l =$$
$$\iint_A\left(\frac{\partial}{\partial q'},\frac{\partial}{\partial s}\right)\left(\frac{\mathrm{d}}{\mathrm{d}s}C(q',s)\Big|_{s=s'}, \frac{\mathrm{d}}{\mathrm{d}p}H(q',R_q^*(p))\Big|_{p=q'}\right)\mathrm{d}v. \tag{D 2}$$

First let us look at the first integral. We have to integrate over α, β and γ. Over γ the curve follows the vector field exactly, therefore $V \cdot n = 0$, so that part of the integral is

0. Over part β, the normal is $(1,0)$, and therefore we only get an integral of $\mathrm{d}C/\mathrm{d}s$, which gives exactly ΔC, the cost difference for signaller q_0. In the integral over α, $\mathrm{d}C/\mathrm{d}s = 0$, because we are talking about an optimum in C, so we only integrate over $\mathrm{d}H/\mathrm{d}p$. As the integral curve points up along the diagonal, and the normal points down, we get $(-\Delta H)$, for

$$\Delta H = \int_0^{q_0}\frac{\mathrm{d}}{\mathrm{d}p}H(q',R_q^*(p))\,|_{p=q'} \, \mathrm{d}q' \tag{D 3}$$

and when H is independent of q we get $\Delta H = H(q_0,R_q^*(q_0)) - H(0,R_q^*(0))$. Now let us turn to the area integral, it is

$$\iint_A \frac{\partial^2 C}{\partial q \partial s} + \frac{\partial^2 H}{\partial q \partial p} \, \mathrm{d}v.$$

But the second term is 0 because $\mathrm{d}H/\mathrm{d}p$ is constant across s. Finally we get

$$\Delta C - \Delta H = \iint_A \frac{\partial^2 C}{\partial q \partial s} \, \mathrm{d}s \mathrm{d}q$$

or

$$\Delta C = \Delta H + \iint_A \frac{\partial^2 C}{\partial q \partial s} \, \mathrm{d}s \mathrm{d}q,$$

which tells us how the cost relates to how desirable it is to be low-cost versus high-cost signallers.

REFERENCES

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

Getty, T. 1998 Reliable signalling need not be a handicap. *Anim. Behav.* **56**, 253–255.

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 parents and young: Parent–offspring conflict and sibling rivalry. *Am. Nat.* **146**, 1–24.

Grafen, A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546.

Hasson, O. 1997 Towards a general theory of biological signaling. *J. Theor. Biol.* **185**, 139–156.

Hurd, P. L. 1995 Communication in discrete action-response games. *J. Theor. Biol.* **174**, 217–222.

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. 1998 Signalling among relatives. II. Beyond the Tower of Babel. *Theor. Pop. Biol.* **54**, 146–160.

Lachmann, M. & Bergstrom, C. T. 1999 When honest signals must be costly. Santa Fe Institute Working Paper Series.

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.

Lotem, A., Wagner, R. H. & Balshine-Earn, S. 1999 The overlooked signaling component of nonsignaling behavior. *Behav. Ecol.* **10**, 209–212.

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

Michod, R. E. & Hasson, O. 1990 On the evolution of reliable indicators of fitness. *Am. Nat.* **135**, 788–808.

Noldeke, G. & Samuelson, L. 1999 How costly is the honest signalling of need. *J. Theor. Biol.* **97**, 527–539.

Nur, N. & Hasson, O. 1984 Phenotypic plasticity and the handicap principle. *J. Theor. Biol.* **110**, 275–297.

Proulx, S. R. 2001 Can behavioural constraints alter the stability of signalling equilibria. *Proc. R. Soc. Lond.* B **268**, 2307–2313. (DOI 10.1098/rspb.2001.1804.)

Siller, S. 1998 A note on errors in grafen's strategic handicap models. *J. Theor. Biol.* **195**, 413–417.

Számadó, S. 1999 The validity of the handicap principle in discrete action-response games. *J. Theor. Biol.* **198**, 593–602.

Zahavi, A. 1975 Mate selection—a selection for a 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.

Zahavi, A. 1981 Natural selection, sexual selection, and the selection of signals. In *Evolution today, Proceedings of the Second International Congress of Systematic and Evolutionary Biology* (ed. G. G. E. Scudder & J. L. Reveal), pp. 133–138.