Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Theoretical Biology 224 (2003) 139–152

Journal of Theoretical Biology

www.elsevier.com/locate/jtbi

Error-proneness as a handicap signal

Kris De Jaegher*

Vrije Universiteit Brussel, MICE, Pleinlaan 2, 1050 Brussel, Belgium

Received 2 October 2000; received in revised form 13 March 2003; accepted 14 March 2003

Abstract

This paper describes two discrete signalling models in which the error-proneness of signals can serve as a handicap signal. In the *first* model, the direct handicap of sending a high-quality signal is not large enough to assure that a low-quality signaller will not send it. However, if the receiver sometimes mistakes a high-quality signal for a low-quality one, then there is an indirect handicap to sending a high-quality signal. The total handicap of sending such a signal may then still be such that a low-quality signaller would not want to send it. In the *second* model, there is no direct handicap of sending signals, so that nothing would seem to stop a signaller from always sending a high-quality signal. However, the receiver sometimes fails to detect signals, and this causes an indirect handicap of sending a high-quality signal that still stops the low-quality signaller of sending such a signal. The conditions for honesty are that the probability of an error of detection is higher for a high-quality than for a low-quality signal, and that the signaller who does not detect a signal adopts a response that is bad to the signaller. In both our models, we thus obtain the result that signal accuracy should not lie above a certain level in order for honest signalling to be possible. Moreover, we show that the maximal accuracy that can be achieved is higher the lower the degree of conflict between signaller and receiver. As well, we show that it is the conditions for honest signalling that may be constraining signal accuracy, rather than the signaller trying to make honest signals as effective as possible given receiver psychology, or the signaller adapting the accuracy of honest signals depending on his interests.

Keywords: Signalling theory; Handicap principle; Error-prone signalling

1. Introduction

By Zahavi's handicap principle Zahavi (1975, 1977), biological signals must be honest, costly and costly in a way related to true quality. Some basic game-theoretical models (Grafen, 1990; Maynard Smith, 1991; Hurd, 1995) corroborate Zahavi's verbal argument, but assume that biological signals are without error. This is an unlikely assumption for two reasons (Johnstone, 1998, p. 45). First, the environment in which the signal is sent will cause the signal to be attenuated and degraded by the time it reaches the receiver. Second, a receiver with limited resources will never be able to fully eliminate errors caused by his failure to detect signals, or by his failure to discriminate one signal from the other.

Given now that signals are unlikely to be without error, two issues have been treated in signalling literature. First, it has been investigated whether for a fixed level of noise, an evolutionary stable equilibrium with honest signalling still exists. Second, it has been investigated how, for given honest signals, the design of these signals will depend on the receiver's psychology, or on the signaller's interests. The *first* issue has been treated by Johnstone and Grafen (1992) and Johnstone (1998), who look at the existence of signalling equilibria in the presence of noise for respectively Grafen's (1990) model of sexual selection, and Maynard Smith (1991) Sir Philip Sydney Game. Bradbury and Vehrencamp (2000) have looked at the problem of existence from a different angle by asking themselves whether, for a given level of noise, it is still in the signaller's interest to participate in signalling, and for the receiver to pay attention to signals.

The *second* issue has been investigated by Guilford and Dawkins (1991), and by Hasson (1989, 1990, 1991).

*Tel.: +32-2-629-2214; fax: +32-2-629-2282.
E-mail address: Kris.De.Jaegher@vub.ac.be (K. De Jaegher).

0022-5193/03/$ - see front matter
doi:10.1016/S0022-5193(03)00135-8

Guilford and Dawkins (1991) stress that to an important extent, the design of a signal will depend on the psychology of the receiver. If accuracy is in the interest of the signaller, signals will evolve to be such that the receiver can detect them as well as possible, and can discriminate between them as well as possible. But as the receiver's resources are necessarily limited, signals will never be perfectly accurately interpreted. Hasson (1989, 1990, 1991) stresses that accuracy also depends on the signaller's interests. If it is in the interest of the signaller that the receiver gets more accurate information, the signaller will send an *amplifier* to assure that the signals are more likely to be detected, or are more likely to be discriminated from one another. But if it is not in the interest of the signaller that the receiver gets more accurate information, the signaller will send an *attenuator*, to assure that signals are detected less often, or discriminated one from the other less often.

Though we will touch upon both these issues treated in the literature on error-prone signalling, our focus is on a *third* issue, which we introduce. The starting point is that noise, with its errors of detection and/or discrimination, may take such a form that it causes a handicap to the signaller. Given then that error-proneness may cause a handicap, Zahavi's handicap principle could be at work, in that this handicap may cause signals to be honest. We corroborate this reasoning by means of two models. The first model (Section 2) is an extension of Hurd's (1995) and Számadó's, (1999) *minimal costly signalling model* to the case of error-prone signals. We show that in case the receiver makes *errors of discrimination*, evolutionary stable communication equilibria exist for a wider range of parameters in conflict-of-interest games than is the case when the receiver does not make such errors. In particular, when the handicap caused by the direct costs of signals is not high enough to assure honest signalling, an additional handicap caused by the receiver's errors of discrimination may make honest signalling evolutionary stable. A by-product of this model is that we show that, allowing for noise, the conditions for existence of a signalling equilibrium are equivalent in our discrete signalling model and in Grafen's (1990) continuous signalling model. The second model (Section 3) is a model of *conventional* (i.e. cheap) *signals*. We show that *errors of detection* can assure that an evolutionary stable communication equilibrium exists for conflict-of-interest games if failure to detect signals induces the receiver to adopt a response that is bad to the signaller. In particular, if the conflict consists herein that the signaller always prefers the receiver to take a high-quality response, then errors of detection must be made more often by the receiver for a high-quality signal than for a low-quality signal. Reinterpreting the probability that the receiver fails to detect a signal as a handicap to the signaller, one sees that this is a straightforward restatement of Zahavi's handicap principle (Zahavi, 1975, 1977). The paper ends with some conclusions in Section 4.

2. Costly signalling

2.1. Signalling without errors

We start by reviewing a standard model of costly signalling without noise. In order to make information transmission relevant, we must at least have two different types of signallers in our model, at least two different signals, and at least two different responses by receivers to these signals. We follow Hurd (1995) and Számadó (1999) in adopting a minimal signalling model where these conditions are exactly met. With probability π_h, the signaller is of high quality (q_h), and with probability π_l, he is of low quality (q_l) (where of course $\pi_l + \pi_h = 1$). When the signaller is of quality q_h (resp. q_l), the receiver's optimal response is p_h (resp. p_l). Formally, the receiver's fitness equals $w(q_i, p_j)$ with $i, j = h, l$. Following Hurd (1995), we adopt the additional notation $W_i = w(q_i, p_h) - w(q_i, p_l)$ for $i = h, l$. It follows that $W_h > 0$, $W_l < 0$.

The signaller can either take actions a_l or a_h, where we label the actions in this manner because we will investigate whether an evolutionary stable equilibrium exists where the low- (high-) quality signaller adopts response $a_l(a_h)$,[1] which we refer to as a *communication equilibrium*. The signaller's fitness is additively separable in two components, the first component v expressing the signaller's value of the receiver's response, and the second component c expressing the signaller's cost of the action taken. Just as Számadó, (1999), but unlike Hurd (1995), we assume that both the c- and the v-components are functions of the signaller's quality. Formally, signaller fitness equals $v(q_i, p_j) - c(q_i, a_k)$ for $i, j, k = h, l$. Following Hurd (1995) and Számadó (1999), we adopt the additional notation $V_i = v(q_i, p_h) - v(q_i, p_l)$, $C_i = c(q_i, a_h) - c(q_i, a_l)$ for $i = h, l$. It is now easy to check that an evolutionary stable communication equilibrium where type $q_h(q_l)$ takes action a_h (a_l) must meet the conditions $V_h > C_h$, $V_l < C_l$. As long as $C_i > 0$ (i.e. action a_h is always more costly than a_l), it is clear that communication can only be stable either under common interests ($V_h > 0$, $V_l < 0$), or if the conflict of interest takes the form that the signaller always prefers response p_h ($V_h > 0$, $V_l > 0$). This excludes the possibility of communication for the additional types of conflict of interest treated by

[1] It is evident that when such an equilibrium exists, an alternative equilibrium may also exist where the low- (resp. high-) quality signaller adopts response a_h (resp. a_l), as noted by Hurd (1995) and Számadó (1999). But the analysis is then simply analogous.

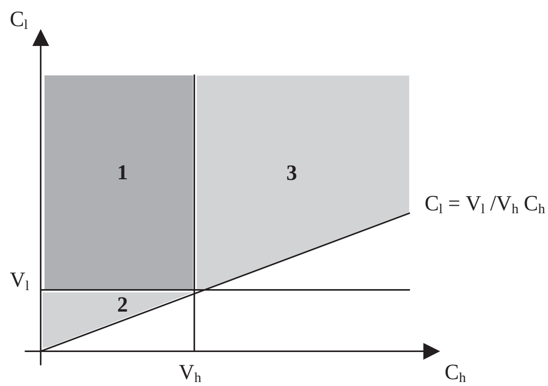

Fig. 1. Conditions on signaller fitness for validity of handicap principle. The figure is based on Hurd (1995). C_i and V_i for $i = h, l$ denote respectively the marginal cost and benefit to a high- (h) and low-quality (l) signaller of pretending to be of high quality rather than low quality. Area 1 denotes the combinations of costs and benefits allowing for a communication equilibrium in a discrete model with noiseless signalling (see Hurd, 1995and Számadó, 1999). The surface above the line $C_l = V_l/V_h C_h$ (areas 1, 2 and 3) is the area where a communication equilibrium exists in a continuous model of signalling (see Grafen, 1990). This surface coincides with the area allowing for a communication equilibrium in a discrete model with error-prone signalling (see Section 2.2).

Bradbury and Vehrencamp (2000), namely $V_h < 0$, $V_l > 0$ (for each type, the signaller prefers a different response to the one preferred by the receiver), and $V_h < 0$, $V_l < 0$ (the signaller always prefers response p_l).[2] The rectangular area 1 in Fig. 1 indicates the levels of change in costs for which a communication equilibrium exists for the case where the signaller always prefers receiver response p_h ($V_h > 0$, $V_l > 0$).This discrete model leads to results about the validity of the handicap principle that are somewhat different from those of a continuous version of this model à la Grafen (1990). Letting q, p and a be continuous, denoting by $\partial p(a)/\partial a$ the slope of the function $p(a)$ giving the receiver's response p as a function of the signaller's action a, and letting V_i and C_i denote respectively $\partial V(q_i, p)/\partial p$ and $\partial C(q_i, a)/\partial a$, for each quality type q_i we obtain a first-order condition $S_i = V_i \partial p/\partial a - C_i = 0$. For this to be an optimum, the second-order condition must be met, i.e. $\partial S_i/\partial a < 0$. Totally differentiating the first-order condition with respect to a and q, one obtains that $\partial S_i/\partial a \, \mathrm{d}a + [\partial V_i/\partial q \, \partial p/\partial a - \partial C_i/\partial q]\mathrm{d}q = 0$. Substituting for $\partial p/\partial a$ using the first-order condition, and given that the second-order condition must be met, it follows that $[\partial V_i/\partial q \, C_i/V_i - \partial C_i/\partial q] > 0$. But this expression has the opposite sign of $\partial(C_i/V_i)/\partial q$, so that one finds that a sufficient condition for an ESS communication equilibrium to exist is that $\partial(C_i/V_i)/\partial q < 0$. Following Grafen, we can conclude from this that (a) signals must be honest; (b) signals must be costly ($C_i > 0$); (c) signals must be costlier for low-quality signallers. The verbal formulation in (c) fits the mathematical condition $\partial(C_i/V_i)/\partial q < 0$ if one considers C_i/V_i as relative cost.

Comparing this to the results from the discrete model summarised in Fig. 1, we conclude along with Számadó (1999) that while in the discrete model $C_l/V_l > C_h/V_h$ is a necessary condition for an evolutionary stable communication equilibrium to exist, it is not a sufficient condition, since such an equilibrium does not exist for the triangular areas 2 and 3 in Fig. 1. We now show, however, that if the model is extended to the case of error-prone signals, for certain levels of noise the condition $C_l/V_l > C_h/V_h$ *is* in fact sufficient for the existence of an evolutionary stable equilibrium even for discrete signals. In other words, error-proneness of signals expands the range of parameters for which an evolutionary stable communication equilibrium exists in the discrete model to include the triangular areas 2 and 3 in Fig. 1. This result adds to the importance of studying error-prone signals. Indeed, at first sight one could claim that we need not formally model noise. Simply, to account for error-prone signalling, we could re-interpret V_i and/or C_i for $i = h, l$ as expected values, and say that all the above results carry through. But this masks the fact that it may actually be the error-proneness of signals that assures the existence of an evolutionary stable communication equilibrium, as we now go on to show.[3]

2.2. Error-prone signalling

We now introduce noise into the model. We continue to keep our model as simple as possible and assume that, just as the signaller sends only two signals, the receiver *perceives* only two signals. This means that, just as is assumed in Johnstone and Grafen (1992), the range of costs that the receiver perceives the signaller to incur is the same with or without errors of perception.[4]

[2] It should be noted finally that Számadó (1999) allows for a coefficient of relatedness r between signaller and receiver. His results are obtained by simply writing $W_i + rV_i$ everywhere where we write W_i, and $V_i + rW_i$ everywhere where we write V_i.

[3] It is well-known in a game theory that the presence of noise expands the range of possible equilibria in signalling games (see Myerson, 1991, Chapter 6).

[4] A model of noise where the range of perceived signals differs with and without noise is found in Johnstone (1998). Specifically, Johnstone assumes that the receiver perceives the true cost of an action taken by the signaller plus an error term that is normally distributed with mean zero. In such a model of noise, the level of noise is necessarily correlated with the cost of a signal, as there is then less noise the larger the difference in cost between two signals. Johnstone thereby obtains that noise should not be too small, because if noise is small, the cost to a low-quality signaller of sending a high-quality signal is small, and the equilibrium may be easily destabilised. For the purposes of our model, where we wish to investigate whether the indirect cost caused by noise signalling can cause the handicap principle to work if direct costs are not sufficient to make it work, it is best to assume that noise can be varied independently from the direct costs of each signal.

Specifically, for $i = h, l$ we let $\mu(i/j)$ denote the probability that, given that the signaller has taken action j, the receiver perceives action i, where of course $\mu(i/j) + \mu(j/j) = 1$. In other words, $\mu(i/j)$ is the probability that an error of discrimination is made when the signaller sends signal j.[5] We further follow Johnstone and Grafen (1992) in making no further assumptions about the level of $\mu(i/j)$ for $i = h, l$, contrary to Lachmann and Bergstrom (1998) and Bradbury and Vehrencamp (2000), who assume that $\mu(h/l) = \mu(l/h)$.

In order to obtain a communication equilibrium, we now need two pairs of conditions, namely that the signaller should be honest, and that the receivers should trust the signals. These conditions are treated in Sections 2.2.1 and 2.2.2. Extensions of the model, in view of the literature on error-prone signalling, are treated in Section 2.2.3, along with an interpretation of the results.

2.2.1. Signaller honesty

In order for the signaller to be honest, the following pair of conditions should be met:

$$\begin{aligned}&\mu(i/i)v(q_i,p_i) + \mu(j,i)v(q_i,p_j) - c(q_i,a_i)\\&\quad \geqslant \mu(i/j)v(q_i,p_i) + \mu(j/j)v(q_i,p_j) - c(q_i,a_j)\end{aligned} \tag{1}$$

for $i, j = h, l$ and $i \neq j$

This implies that

$$[\mu(h/h) - \mu(h/l)] > C_h/V_h, \tag{2a}$$

$$[\mu(h/h) - \mu(h/l)] < C_h/V_h. \tag{2b}$$

Given that $\mu(h/h) - \mu(h/l) \leqslant 1$, by (2a), just as in the case with noiseless signalling, it should still be met that $C_h < V_h$. However, if this is met, then we also see from (2a) and (2b) that as long as $C_l/V_l > C_h/V_h$ an evolutionary stable communication equilibrium is also possible if $V_l > C_l$ and $V_h > C_h$ (see the triangular area 2 in Fig. 1). That is, an evolutionary stable communication equilibrium is possible even if handicaps caused by signals are such that under perfect accuracy, the signaller prefers to always take the high-quality action. It is on this situation that we concentrate in what follows.

While the scope for evolutionary stable communication equilibria is thereby extended, this is conditional on error-proneness of signals taking a particular form. First of all, since $C_h/V_h > 0$ in area 2 of Fig. 1, it must be the case that

$$\mu(h/h) > \mu(h/l). \tag{3}$$

That is, as indicated in the left and right parts of Fig. 2, in the $\mu(h/h)$–$\mu(h/l)$-space, noise must be such that we are situated above the 45° line (we will later see why Fig. 2 has two parts); (3) is in fact trivial, since it simply says that signals should contain information, and is further simply a consequence of the fact that we restrict our analysis to the case where the $h-$ ($l-$) response is taken upon the $h-$ ($l-$) action (one could equally well look for a communication equilibrium where the $h-$ ($l-$) response is taken upon the $l-$ ($h-$) action). Second, rewriting (2a) and (2b), we see that it must be the case that

$$\mu(h/h) > C_h/V_h + \mu(h/l), \tag{4a}$$

$$\mu(h/h) < C_l/V_l + \mu(h/l). \tag{4b}$$

It follows that for the case in area 2 of Fig. 1 ($C_h/V_h < C_l/V_l < 1$), noise levels must be such that we are situated in the grey area in the left and right part of Fig. 2. It is clear from this that constraint (3) is slack. As well, in order to allow for honest signalling, noise should not be reduced to zero.

This result that error-proneness may assure honest signalling can be seen as a straightforward extension of Zahavi's handicap principle. If C_l is lower than V_l, then the direct handicap to the low-quality type of taking a high-quality rather than a low-quality action will be too small to stop him from preferring this action to the low-quality action. However, noise can make the high-quality action less attractive to the low-quality signaller, in that this action will sometimes be perceived by the receiver as a low-quality action. In other words, where the direct handicap of sending a high-quality signal is not high enough to assure honest signalling, the indirect handicap of noise can make the total handicap of sending a high-quality signal large enough to assure honest signalling.

This extension of the handicap principle brings the conditions for the existence of an evolutionary communication equilibrium in a discrete model ($C_h/V_h < C_l/V_l < 1$) closer to the conditions for existence in a continuous model ($C_h/V_h < C_l/V_l$). Yet, for area 3 in Fig. 1, a communication equilibrium would now still not be possible in a discrete model. However, honest signalling is also possible for area 3 in a discrete model if costs become expected values. Here, we may imagine that a signaller taking the high-quality action, while mostly incurring high costs, will sometimes get away with low costs. Similarly, a signaller taking the low-quality action, while mostly incurring low costs, will sometimes be unlucky and incur high costs anyway. In Appendix A, it is shown that this extends the area in Fig. 1 where evolutionary stable communication

[5] Bradbury and Vehrencamp (2000) give a different interpretation to this model of noise. In their interpretation, noise arises not because of errors of perception made by the receiver, but because of uncertainty in the signaller's production function of actions. Concretely, a signaller who incurs the costs to produce action a_h will indeed produce this action most of the time, but will still sometimes be unlucky and only produce action a_l. Similarly, a signaller who incurs the costs to produce action a_l will most of the time only be able to produce action a_l, but will sometimes be lucky, and pull off an action a_h anyway. However, the standard interpretation given to noise in the literature is errors of perception made by the receiver.

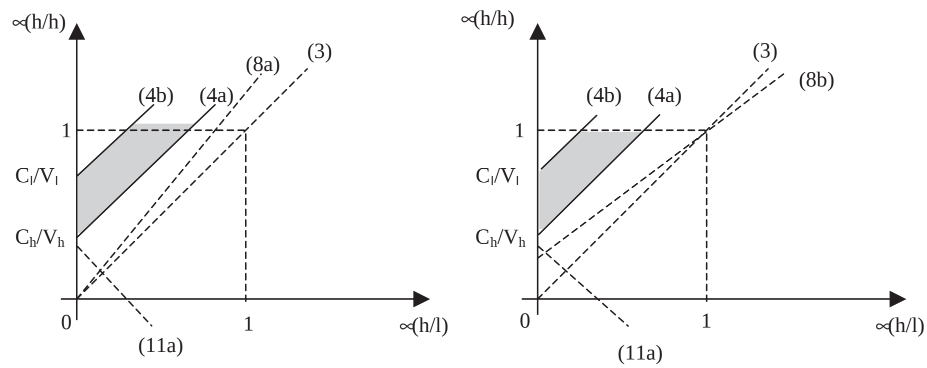

Fig. 2. Conditions on accuracy for validity of handicap principle. The left (right) part of the figure denotes the case where, when not getting any information, the receiver prefers the low- (high-) quality response. $\mu(h/h)$ denotes the probability that the receiver perceives a high-quality signal when the signaller sends a signal that he is of high-quality. $\mu(h/l)$ denotes the probability that the receiver mistakes a signal sent by the signaller telling that he is of low quality for a high-quality signal. The trapezoid area indicates the levels of accuracy for which a communication equilibrium exists. Above the line denoted (4a), the high-quality signaller prefers to send a high-quality signal to a low-quality one, and below the line denoted (4b), the low-quality signaller prefers to send a low-quality signal to a high-quality one. These lines are upward sloping because, starting from a given combination $\mu(h/l)$ and $\mu(h/h)$ for which the signaller is indifferent between sending a high- and a low-quality signal, the signaller can only continue to be indifferent between sending these two signals if either $\mu(h/l)$ and $\mu(h/h)$ are both increased or decreased. The dashed lines represent conditions on accuracy treated elsewhere in the literature. Above the line denoted (3) ($\mu(h/h) = \mu(h/l)$) signals contain information (see Johnstone and Grafen, 1992; technically, the monotone likelihood ratio property is met). Above the line denoted (11a), the signaller prefers to participate in signalling before his type has been revealed (see Bradbury and Vehrencamp, 2000), under the assumption that the receiver adopts the low-quality response when the signaller does not participate (where non-participation is considered as an out-of-equilibrium signal). (11a) is downward sloping because ex-ante to the signaller, high-quality responses adopted for the high- or low-quality types are substitutes. Above the line denoted (8a) in the left figure, the receiver prefers heeding signals to always adopting the low-quality response. Above the line denoted (8b) in the right figure, the receiver prefers heeding signals to always adopting the high-quality response (see Bradbury and Vehrencamp, 2000). These lines are downward sloping because to the receiver high accuracy of the h- or l-signal is substitutes to one another. It is clear from the figure that the conditions for signaller honesty (4a) and (4b) may be the only relevant ones.

equilibria are possible to include area 3. Allowing for appropriate noise, Grafen's conditions (1990) for communication in his continuous model, and those in our discrete model, therefore coincide.

2.2.2. Receiver trustfulness

Under error-proneness, any signal perceived by the signaller may have come from two types of signallers. Therefore, contrary to what was the case for the model without noise, in order for a communication equilibrium to exist, we must make sure that signals are not error-prone to such an extent that the receiver does not let his response depend on the signals perceived. In particular, for any communication to take place, noise must be such that the receiver prefers to adopt response p_h (p_l) when receiving a high-quality (low-quality) signal. To calculate the receiver's expected fitness upon perceiving a signal j, we use Bayes' rule[6] to derive the updated probability that signaller is of quality i, given that the receiver perceives signal j, which equals $\pi_i\mu(j/i)[\pi_i\mu(j/i)+\pi_j\mu(j/j)]^{-1}$. Eliminating the denominator of this updated probability, one obtains the following pair of conditions for the receiver's response of $p_h(p_l)$ to a perceived $a_h(a_l)$ to be an evolutionary stable strategy (ESS):

$$\begin{aligned}&\pi_i\mu(i/i)w(q_i,p_i)+\pi_j\mu(i/j)w(q_j,p_i)\\&\quad>\pi_i\mu(i/i)w(q_i,p_j)+\pi_j\mu(i/j)w(q_j,p_j)\end{aligned}\tag{5}$$

for $i,j = h,l$ and $i \neq j$.

Johnstone and Grafen (1992) derive a necessary condition on noise for Eq. (5) to be met in the following way. Multiplying the left-hand side of (5) for $i = h, j = l$ with the left-hand side of (5) for $i = l, j = h$, doing the same for the right-hand side of the pair of conditions, and reworking one obtains the new condition that

$$\begin{aligned}&[\mu(h/h)\mu(l/l)\ -\mu(l/h)\mu(h/l)]\ [w(q_l,p_l)w(q_h,p_h)\\&\quad-w(q_h,p_1)w(q_1,p_h)]>0.\end{aligned}\tag{6}$$

It can be checked that, given our assumption that $W_h > 0$ and $W_l < 0$, the term in receiver utilities in (6) is necessarily positive. It follows then that the term in probabilities must be positive, or reworking that

$$\mu(l/l)\mu(h/l)^{-1} > \mu(l/h)\mu(h/h)^{-1}.\tag{7}$$

In words, the higher (lower) the advertising level of a signaller, the more likely he should be perceived as advertising strongly (weakly). In economic signalling theory, this property is generally assumed to be met in signalling models, and is known as the *monotone likelihood ratio property* (Milgrom, 1981). In our simple model with two signals sent and two signals perceived,

[6] Bayes' rule is equally well applied for the receiver's decision in Lachmann and Bergstrom (1998, p. 151). In their continuous signalling model, Grafen and Johnstone (1993) on the contrary assume that the receiver minimises the expected sum of squared discrepancies between actual and estimated quality.

reworking (7) we see that it is simply identical to the trivial condition (3), namely $\mu(h/h) > \mu(h/l)$. It follows that constraints (4a) and (4b) remain the relevant ones, as they lie everywhere above the 45° line in the left and right parts of Fig. 2.

In Johnstone and Grafen's (1992) model, the restriction (7) on noise is sufficient to assure that the receiver chooses a different response for each perceived signal. This is because they assume that the receiver can choose from a continuum of possible responses, which assures that the receiver will adopt two different responses to two only very slightly different perceived signals. In our model with discrete responses, however, two perceived signals that are only slightly different could be met with one and the same response. It follows that we still need to check (5), namely whether it is an evolutionary stable strategy for the receiver to adopt response $p_h(p_l)$ when perceiving that the signaller took action $a_h(a_l)$.

A condition equivalent to (5) is investigated in Bradbury and Vehrencamp (2000), though the authors state this condition in a different way. The reason for this is that they do not investigate whether the receiver will adopt the right response to a perceived signal, but instead ask themselves whether, given the presence of noise, a receiver who slavishly adopts the right responses once he has decided to pay attention to signals finds it worth at all to pay attention to signals. With zero attention costs, however, the conditions that the receiver should choose the right response to each perceived signal, and the condition that the receiver should be willing to pay attention to signals, are perfectly equivalent. This can be seen by adding $\pi_i\mu(j/i)w(q_i,p_j) + \pi_j\mu(j/j)w(q_j,p_j)$ to both the left-hand side and the right-hand side in (5). One then obtains that the receiver should prefer to follow each signal, rather than always adopting response p_j, or equivalently, the receiver should prefer to pay attention to signals. Since j can be both l or h, one of the two conditions described in (5) must be redundant, depending on the receiver's optimal response when not paying any attention to signals, which Bradbury and Vehrencamp term the *default response*. Specifically, it is easily calculated that p_l is the default response when $\pi_h W_h < \pi_l |W_l|$. In this case (5) reduces to

$$\mu(h/h) > \pi_l |W_l| [\pi_h W_h]^{-1} \mu(h/l). \tag{8a}$$

This case, with its corresponding constraint (8a), is represented in the left part of Fig. 2. If on the contrary p_h is the default response, i.e. if $\pi_h W_h > \pi_l |W_l|$, (1) reduces to

$$\mu(h/h) > 1 - \pi_1 |W_l| [\pi_h W_h]^{-1} + \pi_l |W_l| [\pi_h W_h]^{-1} \mu(h/l). \tag{8b}$$

This second case, with its corresponding constraint (8b), is represented in the right part of Fig. 2. As is clear from Fig. 2, condition (7) (equivalent to condition (3)) is met everywhere where one of the conditions (8a) or (8b) are met. It should be noted that conditions (8a) and (8b) are easily generalised to the more realistic case where paying attention to signals causes the receiver a cost. This will then simply imply that signalling is only possible if signals are accurate enough to compensate for the attention costs. In Fig. 2, this results in a shift upwards of the lines where (8a) and (8b) are exactly met. But the receiver's choice may not just be, as in Bradbury and Vehrencamp (2000) between either paying attention to certain signals or not paying attention at all, but between paying attention at the current level, or paying less attention. Denoting expected costs in fitness from paying attention at the current level by K, and expected fitness from paying less attention by V, one obtains the condition

$$\pi_l[\mu(l/l)w(q_l,p_l) + \mu(h/l)w(q_l,p_h)] + \pi_h[\mu(l/h)w(q_h,p_l) + \mu(h/h)w(q_h,p_h)] - K > V$$

$$\Leftrightarrow$$

$$\mu(h/h) > [V + K - \pi_l w(q_l,p_l) - \pi_h w(q_h,p_l)][\pi_h W_h]^{-1} + \pi_l |W_l| [\pi_h W_h]^{-1} \mu(h/l) \tag{9}$$

which is again a line parallel to the line where condition (8a) is exactly met in the left part of Fig. 2, and the line where condition (8b) is exactly met in the right part of Fig. 2. Moreover, while Bradbury and Vehrencamp (2000) stress that for noisy communication to be evolutionary stable, signals must not be so noisy that the receiver prefers not to pay attention (or by extension, to pay less attention), if noisy communication is stable it must be that receivers find it prohibitively costly to reduce noise further by paying more attention (Guilford and Dawkins, 1991). Such a condition for evolutionary stability of a certain level of noise can again be expressed by condition (9), where V is now re-interpreted as the total fitness of paying more attention. It follows that the condition that the receiver should not want to pay more attention is again represented by a line parallel to the line where (8a) is exactly met in the left part of Fig. 2, or a line parallel to the line where (8b) is exactly met in the right part of Fig. 2.

It is clear now from Fig. 2 that it is perfectly possible that the constraints that the receiver should take the appropriate responses, and should not want to pay less or more attention, are all slack, and that the constraints for signaller honesty ((4a) and (4b)) are the only relevant constraints on noise for the existence of an evolutionary stable communication equilibrium in case of a conflict of interest. The literature suggests that honesty is still possible when signals are noisy, but this depending on some additional conditions (signals must be informative (Johnstone and Grafen, 1992), and additionally informative to such an extent that it is worth for receivers to pay attention to them (Bradbury and Vehrencamp, 2000)). Our analysis on the contrary suggests that it may be the fact that signals are noisy that assures honest signalling. Antropomorphising, the receiver of a high-quality signal reasons that this signal cannot have been sent by a low-quality signaller, because of both the direct signalling cost, and the indirect cost of noise, that sending a high-quality signal would cause to a low-quality signaller.

In order to further interpret our results, it is of interest to look at what levels of noise are optimal to the receiver. To obtain these optimal levels of noise, we first note that (9) can be reinterpreted as a line where receiver fitness is constant. Higher lines in Fig. 2, parallel to, respectively, lines (8a) and (8b), therefore represent higher receiver fitness. It follows that for the case in the left part of Fig. 2, where l is the default response, the point with $\mu(h/l) = 0$, $\mu(h/h) = C_l/V_l$ is optimal. For the case in the right part of Fig. 2, the point with $\mu(h/l) = 1 - C_l/V_l$, $\mu(h/h) = 1$ is optimal. Simply, the receiver prefers maximal possible accuracy, meaning that he either prefers $\mu(h/l)$ or $\mu(l/h)$ to be reduced to a minimum. Clearly then, the higher C_l/V_l, the higher is the maximal level of accuracy that still allows for an evolutionary stable communication equilibrium. C_l/V_l can now be interpreted as a measure of the extent of the conflict of interest, in that the conflict of interest is smaller the smaller the (relative) cost to the low-quality signaller of taking the high-quality action rather than the low-quality action. We conclude that the *larger* the conflict of interest, the *smaller* is the maximal accuracy that can be achieved without destabilising the communication equilibrium.

2.2.3. Extensions and interpretation

The results above apply for a toy world where the signaller can only take one of two possible actions. In this section, we investigate what happens in the more realistic case where the signaller can choose from additional actions. One such additional action, namely a signaller's possible decision (taken before his type has been revealed) not to participate in signalling at all, is investigated by Bradbury and Vehrencamp (2000). We pay special attention to Bradbury and Vehrencamp's analysis, because they obtain a result that at first sight seems very similar to ours, namely that noise in some cases is a must for the existence of a communication equilibrium—though this result is in fact quite different from ours. Particularities about Bradbury and Vehrencamp's (2000) analysis are that they assume that, once a signaller has decided to participate in signalling, he is slavishly honest, and that a receiver who observes the signaller not to be participating in signalling automatically adopts the default response. Formally, the following conditions should then be met:

$$\begin{aligned}&\pi_l[\mu(l/l)v(q_l,p_l) + \mu(h/l)v(q_l,p_h) - c(q_l,p_l)]\\&\quad + \pi_h[\mu(l/h)v(q_h,p_l) + \mu(h/h)v(q_h,p_h) - c(q_h,p_h)]\\&\qquad > \pi_l v(q_l,p_i) + \pi_h v(q_h,p_i) \ \text{ for } \ i = l,h, \end{aligned} \tag{10}$$

where i is the default response. These conditions reduce to:

$$\begin{aligned}&[\pi_h V_h]\mu(h/h)\\&\quad > [\pi_l c(q_l,p_l) + \pi_h c(q_h,p_h)] - \pi_l V_l \mu(h/l),\end{aligned} \tag{11a}$$

$$\begin{aligned}&[\pi_h V_h]\mu(l/h)\\&\quad < -[\pi_l c(q_l,p_l) + \pi_h c(q_h,p_h)] - \pi_l V_l \mu(l/l).\end{aligned} \tag{11b}$$

From (11b), it is clear that the signaller now would never want to participate if p_h is the default response. This is simply because not participating then guarantees the signaller that his most preferred response is taken. The case in the right part of Fig. 2 would thereby become irrelevant. Concentrating on the case where p_l is the default action then (left part of Fig. 2), we see that for $V_h > 0$, $V_l > 0$, as long as the costs of taking actions are not too high, the signaller will want to participate in noiseless signalling ($\mu(h/h) = 1$, $\mu(h/l) = 0$). Following Bradbury and Vehrencamp's (2000) assumption that the signaller is slavishly honest once having decided to participate in signalling, along with the authors we can also consider the cases $V_h > 0$, $V_l < 0$; $V_h < 0$, $V_l < 0$; $V_h < 0$, $V_l > 0$. For the case $V_h > 0$, $V_l < 0$ (common interests), it is again the case that as long as the costs of taking actions are not too high, the signaller will want to participate in noiseless signalling. If $V_h < 0$, $V_l < 0$ (the signaller prefers the receiver always to adopt response p_l), the signaller never wants to participate in signalling. Finally, if $V_h < 0$, $V_l > 0$ (signaller and receiver prefer different responses for each state), the signaller only wants to participate in signalling if signals contain at least a minimum of noise. Intuitively, given that the low-quality signaller prefers p_h and the high quality signaller p_l, the signaller only wants to participate in signalling if a minimum of errors of discrimination occurs.

For the latter case, Bradbury and Vehrencamp (2000) thereby obtain a result that appears similar to ours, in that there should be a minimum level of noise for an evolutionary stable communication equilibrium to exist. However, when we give up the authors' assumption that signallers are slavishly honest once they have decided to signal, it becomes clear from conditions (4a) and (4b) above that honest signalling is impossible for the cases $V_h < 0$, $V_l > 0$ and $V_h < 0$, $V_l < 0$. The case $V_h > 0$, $V_l > 0$ is therefore the only type of conflict of interest for which honest signalling is possible. For the case where p_l is the default response (left part of Fig. 2), we see that as long as signalling costs are low, the intercept of (11a) will lie everywhere below conditions (4a) and (4b), such that the latter conditions may again be the only relevant ones.

Bradbury and Vehrencamp's (2000) results additionally depend on their assumption that the receiver always adopts the default response when the signaller does not participate in signalling. But non-participation in signalling may itself be interpreted as a signal by the receiver. Starting then from a candidate communication equilibrium as we have described it, non-participation in signalling is an out-of-equilibrium signal. Any receiver response to such a signal is now a best response, including some catastrophic response that assures that the signaller prefers to participate in signalling. By this argument, honest signalling remains possible for the case in the right part of Fig. 2, even though p_h is the default response. Simply, given that not participating is an out-of-equilibrium signal, p_l is a best response to it, and constraint (11a) is relevant even for this case. Again, (4a) and (4b) may remain the only relevant constraints.

While this argument is sound, it should be noted that such stabilising receiver best responses to out-of-equilibrium signals are weak best responses, since to a signal that is never sent, *any* receiver response is a best response. We therefore obtain Nash equilibria, but not the evolutionary stable equilibria we are interested in. One remedy suggested by Johnstone and Grafen (1992, p. 231) is to assume that, while out-of-equilibria signals are never sent by the signaller, because of noise every possible out-of-equilibrium signal is always perceived by the receiver. Thus, while the signaller always participates in signalling, the receiver now with positive probability $\mu(0/j)$ for $j = h, l$ perceives the signaller not to be participating in signalling (where of course $\mu(h/j) + \mu(l/j) + \mu(0/j) = 1$ for $j = h, l$). As long as the probabilities $\mu(0/j)$ for $j = h, l$ happen to take such a form that, in case the signaller is honestly signalling, it is a strict best response for the receiver to adopt response p_l, we obtain an evolutionary stable communication equilibrium for both cases represented in Fig. 2. Noise therefore has the nice property that it allows us to fix a strong best response to signals not sent by the signaller in equilibrium.

Having thus investigated how our results are affected if, in line with Bradbury and Vehrencamp (2000), signallers must also decide whether or not to participate in signalling, we can generalise this analysis to a more realistic world where there are even more actions for the signaller to choose from. In particular, for given levels of attention from the receiver, the signaller may also face the decision between sending more accurate signals at higher encoding costs, or sending less accurate signals at lower encoding costs. Assuming that the signaller must make his decision on the accuracy of signals before his type is determined, it is easy to see that the restrictions assuring that the signaller incurs a certain level of encoding costs, when depicted in Fig. 2, are again lines parallel to (11a). We conclude that it is perfectly possible that (4a) and (4b) remain the only relevant constraints. However, this suggests a coincidental character to the equilibrium we describe; honest signalling under the assumed conflict of interest would now only be possible if signaller encoding costs happen to be such that the signaller chooses a level of error-proneness that can support honest signalling. If the type of equilibrium we suggest is to be more than coincidental, such an equilibrium must also exist when encoding costs are such that the signaller would prefer to send signals with a level of noise too low to support honest signalling.

Let $\tilde{l}$ and $\tilde{h}$ be two such signals, with $\mu(\tilde{h}/\tilde{h})$ and $\mu(\tilde{h}/\tilde{l})$ such that we are situated above (4b) in Fig. 2. If the receiver responds with p_l to signal $\tilde{l}$ and with p_h to signal $\tilde{h}$, then it is clear that a candidate equilibrium with signals l and h meeting constraints (4a) and (4b) cannot be evolutionary stable. However, the responses to signals $\tilde{l}$ and $\tilde{h}$ are not pre-given if the receiver is able to observe the error-proneness of signals, i.e. if the receiver considers $\tilde{l}$ and $\tilde{h}$ as signals different to the signals l and h meeting the conditions summarised in Fig. 2. Starting from our candidate equilibrium, $\tilde{l}$ and $\tilde{h}$ are now out-of-equilibrium signals, and any receiver response is a best response to them, including such a response that makes it optimal to the signaller never to send signals $\tilde{l}$ and $\tilde{h}$.

Again, this argument yields us a Nash equilibrium rather than an evolutionary stable equilibrium, as the receiver's responses to $\tilde{l}$ and $\tilde{h}$ are weak best responses. The remedy, due to Johnstone and Grafen (1992, p. 231), again is to assume that while the signaller never sends signals $\tilde{l}$ and $\tilde{h}$, the receiver sometimes perceives these signals even if the signaller sends only signals l and h. In particular, we assume $\mu(l/i) + \mu(h/i) + \mu(\tilde{l}/i) + \mu(\tilde{h}/i) = 1$ for $i = l, h$. As long as $\mu(\tilde{l}/i)$ and $\mu(\tilde{h}/i)$ for $i = h, l$ are such now that the receiver's strong best response to both $\tilde{l}$ and $\tilde{h}$ is p_l, we have an evolutionary stable communication equilibrium. The same argument applies to any number of possible out-of-equilibrium signals.

But even including all the signals that the signaller could possibly send, does not do complete justice to the richness of possible actions that the signaller can take. As argued by Hasson (1989, 1990, 1991), starting from a given candidate communication equilibrium, the signaller may want to send a so-called *amplifier* along with his signals, in order to increase the probability that the signals are discriminated. Alternatively, he may send an *attenuator*, which decreases the probability that signals are discriminated. In the conflict of interest we have concentrated on, an amplifier benefits the high-quality signaller, and causes a cost for the low-quality signaller; the opposite is true for an attenuator. Assuming that the signaller chooses whether to send amplifiers or attenuators *before* his type has been determined, it is then clear that whether the signaller sends an amplifier or an attenuator depends on the frequency with which the types occur. It is easily checked that such conditions can again be expressed by means of lines parallel to (11a) in Fig. 2. It follows that the signaller would then only choose an accuracy level such that we are situated in the grey area in Fig. 2 if choosing such an accuracy level happens to be in the signaller's interest.

Hasson (1989, 1990, 1991) also considers the case where the signaller can choose to send amplifiers or attenuators *after* his quality has been determined. In this case, it is clear that the high-quality signaller always wants to amplify his signals as much as possible, and our candidate equilibrium would seem to be destabilised. However, we again stress that in our model, the level of error-proneness of a signal is *itself* a signal to the receiver. Thus, even if a signaller would like to amplify or attenuate signals before his quality has been determined, or after it has been determined, he does not do so in our model because signals with a different level of error-proneness are punished. Such punishment by the receiver is not necessarily a weak-best response, because by the same argument as above, noise may assure that out-of-equilibrium amplifiers or attenuators are sometimes perceived, and because noise may take on such a form that it is indeed optimal to punish perceived out-of-equilibrium amplifiers or attenuators.

Summarising, using Johnstone and Grafen's (1992) argument, our results are not changed if we include in our model all actions that the signaller could possibly take. The reasoning is that noise assures that each possible out-of-equilibrium action that the signaller could take is with positive probability perceived by the receiver, even if the signaller in reality takes only one of two possible equilibrium actions. Moreover, the out-of-equilibrium action are perceived with such probabilities that it is a strong best response for the receiver to punish these actions, in turn making it a strong best response for the signaller not to take these actions.

Admittedly, there are several weaknesses in this argument. First, it seems unlikely that the receiver would perceive with positive probability every possible action that the signaller could possibly take, as this would imply that the receiver could never be surprised by a new action taken by the signaller. Second, even if the receiver could never be surprised, then noise would still need to take on a particular form, such that it is optimal to the receiver to punish any out-of-equilibrium signal. There is no particular reason why this would be the case. Third, the same argument may be used to claim that an equilibrium without communication is stable in our model. Indeed, we could imagine that the signaller does not take any action (i.e. does not send any signal), but that because of noise all possible signals are perceived with positive probability, and with such a probability that it is optimal for the receiver to punish any signaller action. Thus, ironically, the same argument could be used to show the evolutionary stability of a complete absence of communication.

If on the contrary we take the more realistic assumption that the receiver can always be surprised, we get the uncomfortable result that no communication equilibrium can be evolutionary stable. New signals can always invade. In Grafen's (1990, p. 538) words, this would seem to 'lend a whimsical and episodic nature to signalling systems'. In the context of our model, imagine e.g. two out-of-equilibrium signals that are too accurate and/or too cheap to support honest signalling under the conflict of interest corresponding to area 2 in Fig. 1. Invasions of such signals may take place in such a way that each type of signaller happens to predominantly send one of to the two signals. It then becomes a best response for the receiver to let his response depend on these signals, and our candidate equilibrium cannot be evolutionary stable. However, our analysis remains relevant if we consider the fact that these invading signals cannot themselves lead to an evolutionary stable equilibrium, and that invasion of these signals may then only present temporary deviations from our equilibrium. This verbal argument deserves a formal analysis using a dynamical setting.

The more realistic assumption that out-of-equilibrium signals always exist allows us to offer an additional interpretation to our results. Guilford and Dawkins' (1991) analysis suggests that signals will evolve to be as accurate as possible, given the receiver's psychology. The receiver's psychology therefore would seem the determining factor in the shape that signals take. However, our analysis suggests that in case of a conflict of interest, signals may not evolve to be as accurate as the receiver's psychology would allow them to be, and this because maximally effective signals may not support honest signalling. Specifically, referring to condition (9) above, which says that the receiver should not prefer to pay more attention to signals, the receiver's psychology may allow for signals such that condition (9) is only met for levels of accuracy such that one is situated to the northwest of the trapezoid area in Fig. 2. While such signals may invade our candidate equilibrium, they cannot themselves lead to an evolutionary stable communication equilibrium, and again may present only temporary deviations from our equilibrium. In this equilibrium, the shape of signals is not determined by receiver's psychology, but by the form that the conflict of interest takes. In particular, as we have seen above, the maximal level of accuracy that can support honest signalling is inversely related to the degree of conflict between the signaller's and the receiver's interests.

We end this section with an application of our analysis to some of the canonical examples of the handicap principle, where examples serve as illustrations, and do not claim to describe reality (cf. Grafen, 1990, p. 527). It may be too large a handicap for a low-quality male to grow a larger tail, or larger antlers not only because of the direct cost of the handicap of such a signal, but also because of the indirect cost that such a signal is still quite often mistaken for a low-quality signal. Antlers or tails may be less accurate signals than they could be, because only inaccurate signals can support honest signalling. Similarly, antilope stotting may be less accurate than it could be, and lead to more errors of discrimination with a lower investment in stotting, because the direct handicap of stotting is not sufficient to discourage a catchable antilope from stotting at a certain level. Finally, nestling begging by a needy chick may not be as accurate as it could be, because the direct handicap of squeaking loudly is not high enough to discourage a less needy chick to squeak at the same level. It may be the indirect handicap consisting of the fact that louder squeaking is still quite often mistaken for less loud squeaking, that discourages the chick from squeaking at a higher level.

3. Conventional signalling

We now treat a variant of the minimal signalling game above, where we make the following modifications. First, we now assume that signals h and l are costless. The essence here is that we want a case where the handicap principle cannot apply through the direct costs of signalling. More generally, we could assume that signals are conventional, defined by Hurd and Enquist (1998, p. 198) as signals in which there is no difference in cost for the use of alternative signals. The assumption of completely costless signals (or *cheap talk*, see Silk et al., 2000) is taken for simplicity. Second, we assume that the error-proneness of signals lies in errors of detection instead of errors of discrimination. Specifically, when the signaller takes action $i = h, l$, then with probability $\mu(0/i)$ the receiver fails to detect the action taken by the signaller; with the complementary probability $\mu(i/i)$, the receiver does perceive the signaller's action. Third, we introduce notation for a third possible response p_0 that the receiver can adopt on top of responses p_l and p_h. Fourth, we include the possibility that the signaller does not send any signal at all, and denote this action as a_0. Since it seems unlikely that the receiver would perceive an action by the signaller if the signaller does not undertake one, we assume that a signal is never perceived when none is sent. We now look at the conditions for signaller honesty (Section 3.1) and for receiver trustfulness (Section 3.2) in this modified model, and end with some extensions and an interpretation of the model (Section 3.3).[7]

[7] A very similar model is treated economics in Pitchik and Schotter (1987) and in De Jaegher and Jegers (2001). However, in these models, there is no noise in the signals. A mixed Nash equilibrium is obtained where, referring to our model, the high-quality signaller always sends a high-quality signal, and the low-quality signaller randomises between sending a low- and a high-quality signaller. The receiver always responds with a low-quality response to a low-quality signal, but randomises between adopting the high- and the low-quality response when receiving a high-quality signal. Such a mixed Nash equilibrium is not evolutionary stable, however. De Jaegher (forthcoming) treats another cheap-talk signalling game where communication is not possible without noise, namely a modified version of the so-called *coordinated attack problem*.

3.1. Signaller honesty

We start by showing that there is a clear reason for including a third response in our model of conventional signalling, since such a third response is a necessary condition for the existence of a communication equilibrium. This becomes clear from the conditions stating that the high-quality (low-quality) signaller should prefer to send a $h-$ ($l-$) signal to an $l-$ ($h-$) signal:

$$\mu(i/i)v(q_i,p_i) + \mu(0/i)v(q_i,p_0) > \mu(j/j)v(q_i,p_j) + \mu(0/j)v(q_i,p_0) \text{ for } i,j = 1,h. \quad (12)$$

Reworking, one obtains that

$$\mu(l/l)[v(q_l,p_l) - v(q_l,p_0)] > \mu(h/h)[v(q_l,p_h) - v(q_1,p_0)], \quad (13a)$$

$$\mu(h/h)[v(q_h,p_h) - v(q_h,p_0)] > \mu(l/l)[v(q_h,p_l) - v(q_h,p_0)]. \quad (13b)$$

It is easily checked that if either $p_0 = p_l$ or $p_0 = p_h$, then communication is only possible if there is no conflict between signaller and receiver about their preferences over responses p_l and p_h. In other words, in order for communication still to be possible under a conflict of interest, when not detecting any signal the receiver must necessarily take a response different from either p_l or p_h.

We now go on to derive for what levels of signaller fitness stable communication is possible under a conflict of interest. We can immediately eliminate the case where p_0 is the most preferred response of either the low- or the high-quality signaller by looking at the condition that the signaller should prefer sending a signal to not sending one:

$$\mu(i/i)v(q_i,p_i) + \mu(0/i)v(q_i,p_0) > v(q_i,p_0) \text{ for } i = h,l. \quad (14)$$

It is clear that these conditions exclude the cases $v(q_l,p_0) \geqslant v(q_l,p_l)$ and $v(q_h,p_0) \geqslant v(q_h,p_h)$. For all other cases (14) is slack. Additionally, a communication equilibrium is not possible when the signaller's and the receiver's interests are completely opposed, i.e. if both $v(q_l,p_h) > v(q_l,p_l)$ and $v(q_h,p_l) > v(q_h,p_h)$ (or in the notation used in Section 2, $V_l > 0$ and $V_h < 0$). From (13a) and (13b) it is clear that the cases $v(q_l,p_h) > v(q_l,p_0) > v(q_l,p_l)$ and $v(q_h,p_l) > v(q_h,p_0) > v(q_h,p_h)$ are excluded, since the right-hand side is then each time positive and the left-hand side negative. Given that we have already excluded that p_0 would be the signaller's most preferred response, the remaining case of completely opposing interests is $v(q_l,p_h) > v(q_l,p_l) > v(q_l,p_0)$ and $v(q_h,p_l) > v(q_h,p_h) > v(q_h,p_0)$. However, by (13a), it should then be met that $\mu(l/l) > \mu(h/h)$ and by (13b) it should be met that $\mu(h/h) > \mu(l/l)$, a contradiction.

It follows that the only type of conflict of interest where stable communication is possible is one where the signaller always has the same preference ordering over p_h and p_l, whatever his type. Since signals are costless, analytically there is no difference between the case where the signaller prefers p_h to p_l, or the opposite case. We further concentrate on the case where the signaller prefers p_h to p_l. Two possible cases allowing for communication under a conflict of interest then remain, namely the case $v(q_l,p_h)>v(q_l,p_l)>v(q_l,p_0)$ and $v(q_h,p_h)>v(q_h,p_0)>v(q_h,p_l)$, and the case $v(q_l,p_h)>v(q_l,p_l)>v(q_l,p_0)$ and $v(q_h,p_h)>v(q_h,p_l)>v(q_h,p_0)$. In the former case, constraint (13b) is slack. We concentrate on the latter case, where each type of signaller has the same preference ordering over the responses. For simplicity, we assume that $v(q_h,p_0)=v(q_l,p_0)=0$. Reworking (13a) and (13b), we obtain:

$$\mu(0/l)<1-v(q_l,p_h)v(q_l,p_l)^{-1} + v(q_l,p_h)v(q_l,p_l)^{-1}\mu(0/h), \tag{15a}$$

$$\mu(0/l)>1-v(q_h,p_h)v(q_h,p_l)^{-1} + v(q_h,p_h)v(q_h,p_l)^{-1}\mu(0/h). \tag{15b}$$

These constraints are depicted in Fig. 3, and reveal a final necessary condition on signaller fitness to allow for a communication equilibrium, namely $v(q_l,p_h)v(q_l,p_l)^{-1}<v(q_h,p_h)v(q_h,p_l)^{-1}$, as is easily checked. In words, the change in benefit from the h-signal being detected relative to the change in benefit from the *l*-signal being detected should be larger for the high-quality signaller than for the low-quality signaller.

Having thus derived the necessary conditions on signaller fitness for a communication equilibrium to exist, we now look at the necessary conditions on noise. As is clear from Fig. 3, a communication equilibrium is only possible $\mu(0/h)>\mu(0/l)$. Intuitively, since sending the *h*-signal as such is better to the signaller than sending the *l*-signal, in compensation, the cost (in the form of the probability of errors of detection) of sending the *h*-signal must be higher than the cost of sending the *l*-signal. This is again in the spirit of Zahavi's handicap principle.

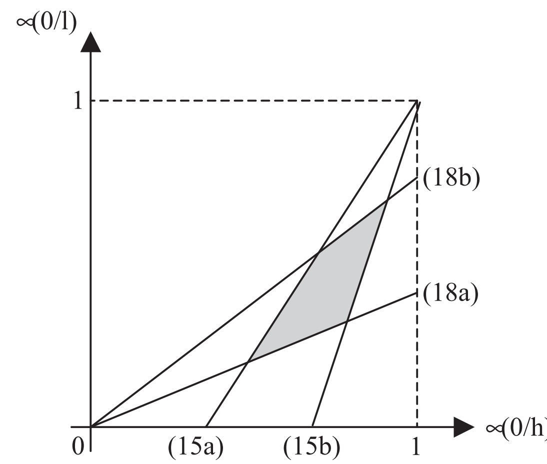

Fig. 3. Conditions on accuracy for existence of communication equilibrium in model of conventional signalling with conflict of interests. The signaller's preferences are ordered such that he prefers a high-quality response to a low-quality one to no response. $\mu(0/l)$ and $\mu(0/h)$ denote the probability that the receiver fails to detect respectively a low-quality and a high-quality signal. The grey area denotes the levels of these probabilities for which a communication equilibrium exists. Below the line (15a), a low-quality signaller prefers sending a low- to a high-quality signal. Above the line (15b), the high-quality signaller prefers sending a high- to a low-quality signal. These lines are upward sloping because in order to keep the signaller indifferent between sending low- and high-quality signals, both types of errors of detection must be increased. Lines (15a) and (15b) cross at $\mu(0/h)=\mu(0/l)=1$, as the signaller is indifferent between sending a high-quality or a low-quality signal if these signals are never detected. Finally, (15a) and (15b) have a slope higher than 1 because, in order for the signaller to be indifferent between sending a high- and a low-quality signal, it must be met that $\mu(0/h)>\mu(0/l)$, since the high-quality response is the signaller's most preferred response. Above (18a), the receiver prefers adopting response p_0 to response p_h when not detecting any signal. Below (18b), the receiver prefers adopting response p_0 to response p_l when not detecting any signal. These lines are upward sloping because, starting from a certain combination of $\mu(0/h)$ and $\mu(0/l)$, the information obtained when not detecting any signal will only be the same if both these probabilities are either increased or decreased. The lines cross at $\mu(0/h)=\mu(0/l)=0$, because a receiver is indifferent about how to respond when no signal is detected if this event never occurs. Finally, the slope of (18a) and (18b) is chosen to be larger than 1, because otherwise an equilibrium is not possible, given the conditions derived for signaller honesty. This implies that a communication equilibrium is only possible if the receiver prefers to adopt response p_0 even if an error of detection is relatively more likely to have come from a high-quality signaller.

3.2. Receiver trustfulness

We next consider the conditions that the receiver should take the appropriate responses. If signals are detected, these conditions are trivial, since there are no errors of discrimination. However, as we have shown above, a communication equilibrium can only exist if there are errors of detection, and if the receiver adopts a response p_0 different from either p_l or p_h when not detecting any signal. But then, noise and receiver payoffs should not take such a form that the receiver always wants to adopt either response p_l or p_h. Thus, contrary to what is the case in the model with costly signals in Section 2.2, the existence of a communication equilibrium imposes specific conditions not only through the constraints that the signaller should take the appropriate actions, but also through the constraints that the receiver should adopt appropriate responses. Specifically, it must be met that

$$\pi_l\mu(0/l)w(q_l,p_0)+\pi_h\mu(0/h)w(q_h,p_0) > \pi_l\mu(0/l)w(q_l,p_i)+\pi_h\mu(0/h)w(q_h,p_i) \quad \text{for } i=l,h, \tag{16}$$

where we have used Bayes' rule, and have cancelled out $[\pi_l\mu(0/l)+\pi_h\mu(0/h)]^{-1}$ for each term. Reworking, we find that

$$\pi_l\mu(0/l)[w(q_l,p_0) - w(q_l,p_h)] > \pi_h\mu(0/h)[w(q_h,p_h) - w(q_h,p_0)], \quad (17a)$$

$$\pi_h\mu(0/h)[w(q_h,p_0) - w(q_h,p_l)] > \pi_l\mu(0/l)[w(q_l,p_l) - w(q_l,p_0)]. \quad (17b)$$

Given that the receiver at least prefers $p_l(p_h)$ to p_0 when the signaller is of low (high) quality, the right-hand sides in both (17a) and (17b) should be positive. It follows then that the left-hand sides should also be positive, i.e. when the signaller is of low (high) quality, the receiver should prefer response p_0 to response $p_h(p_l)$. Assuming this condition to be met, we can rewrite (17a) and (17b) as

$$\mu(0/l) > \pi_h[w(q_h,p_h) - w(q_h,p_0)] \{\pi_l[w(q_l,p_0) - w(q_l,p_h)]\}^{-1}\mu(0/h), \quad (18a)$$

$$\mu(0/l) < \pi_h[w(q_h,p_0) - w(q_h,p_l)] \{\pi_l[w(q_l,p_l) - w(q_l,p_0)]\}^{-1}\mu(0/h). \quad (18b)$$

From (18a), given that it must be met that $\mu(0/h) > \mu(0/l)$, it follows that $\pi_l[w(q_l,p_0) - w(q_l,p_h)] > \pi_h[w(q_h,p_h) - w(q_h,p_0)]$, which means that a receiver without any information should prefer response p_0 to response p_h. Intuitively, given that the receiver should adopt response p_0 when not receiving any signal, and given that it must be met that it should happen more often for the high-quality than for the low-quality type that no signal is detected ($\mu(0/h) > \mu(0/l)$), communication is certainly not possible if p_h is the default response. Moreover, the receiver's fitness values should meet the condition $[w(q_h,p_h) - w(q_h,p_0)][w(q_l,p_0) - w(q_l,p_h)]^{-1} < [w(q_h,p_0) - w(q_h,p_l)][w(q_l,p_l) - w(q_l,p_0)]^{-1}$. It is easy to check that this condition assures that, if the receiver prefers response p_0 to response p_h when not detecting any signal, he will not necessarily prefer response p_l to response p_0.

Conditions (18a) and (18b) are depicted in Fig. 3. The grey area denotes the levels of accuracy which meet the conditions for signaller honesty and receiver trustfulness. It is again of interest to look at the highest level of accuracy that can be achieved. It is clear from Fig. 3 that maximal accuracy is achieved at the point where (15a) and (18a) intersect. Since both the signaller and the receiver prefer response $p_h(p_l)$ for the high- (low-) quality state, this is also the best both signaller and receiver can do in a communication equilibrium.

3.3. Extensions and interpretation

We may investigate, in the spirit of Bradbury and Vehrencamp (2000) whether the receiver has interest in paying attention to signals, and whether the signaller has an interest to participate in signalling. For zero attention costs and encoding costs, it is clear that this is met everywhere where the constraints above are met. It is evident that the receiver strongly prefers to adopt response $p_h(p_l)$ when observing an $h-$ $(l-)$ signal, rather than always adopting a single response p_i for $i = h, l, 0$. Furthermore, (16) tells us that a receiver who does not detect any signal is at least as well off by adopting response p_0 as by adopting response $p_i = h, l, 0$. It follows that under the conditions that we have determined, the receiver also wants to pay attention to signals. The conditions also directly imply that the signaller wants to participate in signalling, since in equilibrium, if the signaller does not participate, the receiver always adopts response p_0, which is the response least preferred by the signaller.

The analysis is easily generalised to the case where participating in signalling involves encoding costs to the signaller and attention costs to the receiver. As long as these costs are not too high, the constraints in Fig. 3 remain the only relevant constraints. But then, just as was the case for the model in Section 2, our results would only be maintained if it happened to be the case that the levels of noise which allow for a communication equilibrium were also optimal to a signaller choosing an optimal level of encoding, and to a receiver choosing an optimal level of attention. Our results would then lose much of their relevance. If on the other hand, in our candidate equilibrium, the signaller finds it optimal to send more accurate signals as long as the receiver remains trustful, our candidate equilibrium cannot be stable. We suggest the same solution to this conundrum as in Section 2. While our candidate equilibrium could be invaded by signals with a level of accuracy that situates us to the southwest of the grey area in Fig. 3, these invading signals themselves cannot lead to a communication equilibrium, and may therefore constitute only temporary deviations from our candidate equilibrium.

If, as in Hasson (1989, 1990, 1991), the signaller can at no cost increase the accuracy of signals by letting these signals be accompanied by what Hasson terms amplifiers, then it would seem that he would always want to do so, as p_0 is his least-preferred response. However, we stress again that in our model, when the accuracy of a signal is changed, the receiver considers this as a different signal. While it is true now that, assuming that accuracy can be increased without cost, our candidate equilibrium can always be invaded by more accurate signals, these more accurate signals will not themselves lead to a new communication equilibrium if they do not meet the conditions summarised in Fig. 3. For instance, let us start from our candidate equilibrium, and assume an invasion of signallers sending a more accurate signal than the original h-signal. It is now a best response for the receiver to adopt response p_h both when the original h-signal and the more accurate h-signal are perceived. However, if the more accurate h-signal does not meet condition (15a), then low-quality signallers will start sending it too. The more accurate h-signal may then no longer contain information in the end, and it may become a best response for the signaller to adopt response p_0 when it is not detected. Once this is the case, it again becomes a best response for the high-quality signaller to send the h-signal, and for the low-quality signaller to send an l-signal. This verbal argument again deserves formal dynamical analysis.

As we have seen in Section 3.2, it is both the signaller and the receiver's interest that the signals are as accurate as possible, within the range of signals allowing for a communication equilibrium. But then, we can again relate the degree of conflict of interest to maximal potential for accuracy of the signals. In particular, we could say that the higher the low-quality signaller's relative benefit of obtaining the high-quality response, i.e. the higher $v(q_l, p_h)v(q_l, p_l)^{-1}$, the higher the degree of conflict between signaller and receiver. An increase in $v(q_l, p_h)v(q_l, p_l)^{-1}$ now results in a counterclockwise tilt of line (15a) in Fig. 3 around the point (1, 1). It follows then that the higher the degree of conflict, the higher both types of errors of detection must at least be to allow for honest signalling.

A particularity about the model in this section is the restrictive conditions on receiver responses. Honest signalling is only possible if the receiver prefers p_h to p_0 to p_l if the signaller is of high quality, and prefers p_l to p_0 to p_h if the signaller is of low quality. In other words, the receiver should prefer a zero response to a response that is optimal to the other type. It is hard to link such a preference ordering to realistic examples. For instance, if we apply the model to sexual display, and interpret p_i as the probability with which the female selects the male, with $p_h > p_l > p_0 = 0$, then the female should prefer not to mate with a high-quality male to mating with him with a low probability, which is not intuitive. Still, a variant of the model in this section with three instead of two states does yield more intuitive examples.[8] With probability π_0, a third state occurs, and it is then optimal for the receiver to respond with p_0. This third state either has nothing to do with the signaller, or has nothing to do with the quality that the signaller is trying to signal. A signalling equilibrium is then still possible under more intuitive receiver preference ordering where p_0 is the least preferred response to each of the two qualities of signallers. Rather than representing an error of detection, the errors $\mu(0/h)$ and $\mu(0/l)$ consists of failure to discriminate an l- or h-signal from a 0-signal. As long as $\mu(0/h)$ and $\mu(0/l)$ are not too high, p_0 continues to be the optimal receiver response to a 0-signal.

Applying this modified interpretation of the model to the case of sexual selection, we could think of three states rather than two being relevant to the receiver. Either no male is detected, or a male is detected of low- or high quality. When no male is present, showing readiness to mate is not a best response. This continues to be so when no male is detected, as long as errors in detecting a signal from a male are not too large. At the same time, the fact that the high-quality male is sending a signal which has a higher probability of not being detected makes it credible to the female that he is a high-quality male, as a low-quality signaller would not want to send such a signal.

[8] I am indebted for this point to one of the referees.

4. Conclusion

Conventional wisdom in the signalling literature tells that signal honesty and signal accuracy are two separate matters. This may be seen in the following quote from Bradbury and Vehrencamp (2000, p. 260): "Signal accuracy is often confounded with signal honesty … it may not pay senders to produce or receivers to require signals that are highly accurate; the costs may exceed the benefits for one or both parties. Sender errors thus may not reflect dishonesty in the game theoretic sense, but are expected even when the two parties have completely identical interests. This view implies a likely upper limit to accuracy we should expect to see in an animal communication system."

However, we have shown that the upper limit to accuracy may lie in the fact that more accurate signals cannot support honest signalling. Moreover, the maximal possible accuracy of signals that can support honest signalling is higher the smaller the degree of conflict.[9] Antropomorphising, the larger the signaller's incentive to lie, the less accurate his signals can be, suggesting a direct link between accuracy and honesty. This result is an extension of Zahavi's handicap principle. If the direct handicap of sending a high-quality signal to the low-quality signaller is not high enough, then the indirect handicap caused by the fact that a high-quality signal is sometimes confounded with a low-quality signal can still prevent the low-quality signaller from sending a high-quality signal. If signals do not involve a direct handicap, then if errors of detection are higher for a high-quality signal than for a low-quality signal, and if errors of detection induce a response that is bad to the signaller, sending a high-quality signal causes the low-quality signaller an indirect handicap that stops him from sending such a signal.

[9] Another instance in which the accuracy of signals is linked to the degree of conflict between signaller and receiver is Crawford and Sobel (1982). However, accuracy does not concern errors of discrimination or errors of detection here, but the fineness with which the signaller partitions a continuum of states. For a similar argument see Enquist et al. (1998).

Acknowledgements

I would like to thank two anonymous referees for their helpful comments, and the Fund for Scientific Research of Flanders (Belgium) for financial support.

Appendix A

Denote by $\rho(i/j)$ the probability that the signaller incurs cost i given that he has taken action j, where we assume that $\rho(i/i) > \rho(i/j)$. The conditions for signaller honesty now become

$$v(q_i, p_i) - \rho(i/i)c(q_i, a_i) - \rho(j/i)c(q_i, a_j) \geqslant v(q_i, p_j) - \rho(i/j)c(q_i, a_i) - \rho(j/j)c(q_i, a_j).$$

Reworking, one obtains that

$$V_h > [\rho(h/h) - \rho(h/l)]C_h,$$

$$V_h < [\rho(h/h) - \rho(h/l)]C_l.$$

It follows that, for an equilibrium to exist, it must be met that $V_l < C_l$, but it may be met that $V_h > C_h$. Additionally, it must be met that $C_l/V_l > C_h/V_h$. This shows that if noise takes the appropriate form, a signalling equilibrium exists for area 3 in Fig. 1.

References

Bradbury, J.W., Vehrencamp, S.L., 2000. Economic models of animal communication. Anim. Behav. 59, 259–268.
Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50, 1431–1451.
De Jaegher, K., Jegers, M., 2001. The physician-patient relationship as a game of strategic information transmission. Health Econ. 10, 651–668.
De Jaegher, K., 2003. A game-theoretic rationale for vagueness. Ling. Philos. in press.
Enquist, M., Ghirlanda, S., Hurd, P.L., 1998. Discrete conventional signalling of a continuous variable. Anim. Behav. 56, 749–754.
Grafen, A., 1990. Biological signals as handicaps. J. Theor. Biol. 144, 517–546.
Grafen, A., Johnstone, R.A., 1993. Why we need ESS signalling theory. Philos. Trans. R. Soc. Lond. B 340, 245–250.
Guilford, T., Dawkins, M.S., 1991. Receiver psychology and the evolution of animal signals. Anim. Behav. 42, 1–14.
Hasson, O., 1989. Amplifiers and the handicap principle in sexual selection: a different emphasis. Proc. R. Soc. Lond. B 235, 383–406.
Hasson, O., 1990. The role of amplifiers in sexual selection: an integration of the amplifying and the fisherian mechanism. Evol. Ecol. 4, 227–289.
Hasson, O., 1991. Sexual displays as amplifiers: practical examples with an emphasis on feather decorations. Behav. Ecol. 2, 189–197.
Hurd, P.L., 1995. Communication in discrete action-response games. J. Theor. Biol. 174, 217–222.
Hurd, P.L., Enquist, M., 1998. Conventional signalling in aggressive interactions: the importance of temporal structure. J. Theor. Biol. 192, 197–212.
Johnstone, R.A., 1998. Efficacy and honesty in communication between relatives. Am. Nat. 152, 45–48.
Johnstone, R.A., Grafen, A., 1992. Error-prone signalling. Philos. Trans. R. Soc. Lond. B 248, 229–233.
Lachmann, M., Bergstrom, C.T., 1998. Signalling among relatives II. Beyond the tower of Babel. Theor. Popul. Biol. 54, 146–160.
Maynard Smith, J., 1991. Honest signalling: the Philip Sidney game. Anim. Behav. 42, 1034–1035.
Milgrom, P., 1981. Good news and bad news: representation theorems and applications. Bell J. Econ. 12, 380–391.
Myerson, R.B., 1991. Game Theory—Analysis of Conflict. Harvard University Press, Cambridge.
Pitchik, C., Schotter, A., 1987. Honesty in a model of strategic information transmission. Am. Econ. Rev. 77, 1032–1036.
Silk, J.B., Kaldor, E., Boyd, R., 2000. Cheap talk when interests conflict. Anim. Behav. 59, 423–432.
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 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.