A Continuous Evolutionary Simulation Model of the Attainability of Honest Signalling Equilibria

Seth Bullock
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, D-14195 Berlin (-Dahlem), GERMANY
Tel: +49 (0)30-82406-350
Fax: +49 (0)30-82406-399
EMail: bullock@mpib-berlin.mpg.de

Abstract

A particular game-theoretic model (Grafen, 1990) of the evolutionary stability of honest signalling, which attempts a formal proof of the validity of Zahavi's (1975, 1977) handicap principle, is generalised and rendered as an evolutionary simulation model. In addition to supporting new theoretical results, this allows the effects of differing initial conditions on the attainability of signalling equilibria to be explored. Furthermore, it allows an examination of the manner in which the character of equilibrium signalling behaviour varies with the model's parameters.

It is demonstrated that (i) non-handicap signalling equilibria exist, (ii) honest signalling equilibria need not involve extravagant signals, and (iii) the basins of attraction for such equilibria are, however, relatively small. General conditions for the existence of honest signalling equilibria (which replace those offered by Zahavi) are provided, and it is demonstrated that previous theoretical results are easily accommodated by these general conditions. It is concluded that the supposed generality of the handicap principle, and the coherence of its terminology, are both suspect.

Keywords: Evolutionary Simulation Models; Coevolution; Communication; Honesty.

Accepted for publication within the Proceedings of the Sixth International Conference on Artificial Life, University of California, June 26–29, 1998, Los Angeles.

A Continuous Evolutionary Simulation Model of the Attainability of Honest Signalling Equilibria

Seth Bullock
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, D-14195 Berlin (-Dahlem)
Tel: 0049-30-82406–350, Fax: 0049-30-82406-399, Email: bullock@mpib-berlin.mpg.de

Abstract

A particular game-theoretic model (Grafen, 1990) of the evolutionary stability of honest signalling, which attempts a formal proof of the validity of Zahavi's (1975, 1977) handicap principle, is generalised and rendered as an evolutionary simulation model. In addition to supporting new theoretical results, this allows the effects of differing initial conditions on the attainability of signalling equilibria to be explored. Furthermore, it allows an examination of the manner in which the character of equilibrium signalling behaviour varies with the model's parameters.

It is demonstrated that (i) non-handicap signalling equilibria exist, (ii) honest signalling equilibria need not involve extravagant signals, and (iii) the basins of attraction for such equilibria are, however, relatively small. General conditions for the existence of honest signalling equilibria (which replace those offered by Zahavi) are provided, and it is demonstrated that previous theoretical results are easily accommodated by these general conditions. It is concluded that the supposed generality of the handicap principle, and the coherence of its terminology, are both suspect.

Models of the evolution of signalling have received renewed interest since the re-assessment of group selection arguments during the mid-sixties encouraged theorists to consider the worth of honest communication to the selfish individual (see Johnstone, 1997, for a recent review of the literature). Initial claims that honest communication could not be stable outside of scenarios in which signallers and receivers enjoy a shared interest in honest information exchange (Dawkins & Krebs, 1978) have been challenged by the development of Zahavi's (1975, 1977) handicap principle.

The evolution of signalling has been of interest within artificial life since its inception (e.g., MacLennan, 1991; Werner & Dyer, 1991). However, with some exceptions (e.g., de Bourcier & Wheeler, 1995; Bullock, 1997), such research has not attempted to address theoretical concerns which are live within theoretical biology. Within this paper, a combination of traditional evolutionary stable strategy (ESS) modelling (Maynard Smith, 1982) and evolutionary simulation modelling (Bullock, 1998) will be applied to a specific theory within current evolutionary biology — the handicap principle (Zahavi, 1975, 1977; Zahavi & Zahavi, 1997).

The handicap principle may be presented in many forms. Indeed the multitude of scenarios which appear to admit of explanation in its terms is one of its strongest attractions. This apparent ubiquity of application has led Zahavi to suggest that his theory might usefully replace the theory of sexual selection suggested by Darwin (1871) as a means of accounting for the specific class of behavioural and morphological adaptations which arise as a result of selective pressure to accumulate mating opportunities.

Here the handicap principle will be cast in terms of courtship display — the context in which it was first described (Zahavi, 1975). Assume that males vary in some respect of interest to choosy females (e.g., in their ability to forage). Females cannot ascertain this male quality directly. However, they are sensitive to an alternative trait. If this alternative trait were to systematically reflect the value of the underlying male quality, females would be selected to exploit it as a cue or advertisement upon which to base their mating choices.

Why should such an 'advertisement' accurately reflect some underlying quality? If females respond favorably to suitors with such an advertisement, what prevents every suitor from investing in this compelling signal, thus rendering it useless? In short, what might maintain the stability of a mate choice system in which males make some courtship display which reveals their quality, and females mediate their mating choices on the basis of the information gained from such a courtship display?

Zahavi's insight was to suggest that the costs incurred in producing courtship displays might enforce honesty amongst suitors if these costs were of a certain character. For example, an honest advertisement of a suitor's ability to forage might be the extent to which the suitor deliberately wastes food items which it has accrued through foraging. Since poor foragers can less afford to waste hard-won prey items than good foragers, a system in which suitors demonstrate their foraging ability through wasting food items cannot be invaded by cheats who exaggerate their foraging ability since the costs involved in such exaggeration are prohibitive of such a strategy.[1]

From this perspective, Zahavi suggests, signals should be regarded as *handicaps* which signallers must bear if they are to demonstrate their true quality. It is through suffering costs that signallers are able to convince their assessors of their status.

The validity of Zahavi's argument has proven hard to establish. However, recent game theoretic models (e.g., Grafen, 1990) have suggested that the central tenets of his argument are sound. Within this paper the phrase "Zahavi's handicap condition" will be used to refer to the stipulation that as signaller quality increases, the cost of making any particular signal decreases.

Johnstone (1997) has usefully characterized the literature concerning the handicap principle as comprising two contrasting classes of account. The first class, described above, attempts to account for the evolutionary stability of the honest advertisement of *quality* as a result of the manner in which the *costs* of signalling vary with quality (e.g., Grafen, 1990; Hurd, 1995). The second class attempts to account for the evolutionary stability of the honest advertisement of *need* as a result of the manner in which signaller *benefits* vary with need (e.g., Godfray, 1991; Maynard Smith, 1991).

The latter class includes models of the kind used by Godfray (1991) to demonstrate the evolutionary stability of a strategy in which nestlings honestly advertise their hunger by varying the strength of their begging calls. Godfray showed that such a strategy is evolutionarily stable if the costs of begging are independent of a chick's hunger, but the value of any particular parental resource to a begging chick increases with the chick's hunger. In such situations no chick will exaggerate its hunger since the value of a parental resource solicited through exaggerated begging will not compensate for the increased cost of begging. Hungry chicks beg more than sated chicks because the resources are worth more to them.

Previous models (Bullock, 1997) have demonstrated that the two classes identified by Johnstone (1997) are special cases of a general class of account in which the interactions between the advertised trait (quality or need) and both costs *and* benefits are such that honest signalling strategies are the best policies.

Here an evolutionary simulation model capable of addressing this superordinate class of scenarios will be implemented. The general conditions under which honest signalling may take place between parties which suffer a conflict of interests will be determined. In addition, the evolutionary attainability of such honest signalling equilibria and the character of the signalling behaviour at such equilibria will be examined.

In the following section, Grafen's (1990) continuous signalling game is presented, its implementation as an evolutionary simulation model is described, and data generated by this simulation are summarized. The satisfaction of Zahavi's handicap condition will be shown to be neither necessary nor sufficient to ensure the evolutionary stability of honest communication. Subsequently, the relationship between the simulation results and those of previous models will be discussed. It will be concluded that these previous results are accommodated as special cases of those presented here. A condition for the presence of honest signalling equilibria will be offered which replaces that proposed by Zahavi. This condition admits the existence of signalling equilibria in which (*contra* Zahavi) low-quality signallers enjoy lower signalling costs than high-quality signallers.

1 An Evolutionary Simulation Model of a Continuous Signalling Game

Grafen (1990) cast his model in terms of mate choice. Male fitness, w_m, was defined as a function of quality, q, level of advertisement, a, and degree of female response, p. This function was constrained such that male fitness decreased with increasing advertisement, and increased with increasing female response. Female fitness, w_f, was defined as increasing with the accuracy with which female response approximated male quality. Briefly, Grafen demonstrated that honest signalling of quality could be an ESS if the negative fitness consequences of male advertisement decrease with increasing quality, i.e., Zahavi's handicap condition is met. However, Grafen's analysis demanded one extra assumption: that the positive fitness consequences of female preference were neutral with respect to male quality, or increased with male quality.

A more general treatment of the model (Bullock, 1997) demonstrated that, once Grafen's assumption concerning the manner in which male quality mediates the positive fitness consequences of female response is relaxed, Zahavi's handicap condition ceases to be either necessary or sufficient for the stability of honest signalling. Here, an evolutionary simulation model will replicate this analytic result, before allowing a more involved examination of the behaviours exhibited by signallers and receivers.

Before an evolutionary simulation model can be attempted, fitness functions which adequately capture the assumptions made during the above analysis must be defined for both signallers and receivers. Particular attention will be paid to the fitness functions' ability to capture the assumptions made by the full range of continuous signalling models under consideration here.

[1] This notion of waste as a signal of quality is reminiscent of the concept of "conspicuous consumption" discussed by Veblen (1899).

In addition, schemes for representing a range of continuous signalling and response strategies must be defined. They must be simple in order that the representation of strategies be amenable to manipulation by a genetic algorithm, yet they must also be able to capture an adequate range of signalling and responding behaviours.

1.1 Fitness Functions

After Grafen (1990), (female) receiver fitness, w_f, may be calculated as

$$w_f = \frac{1}{1 + |p - q|}$$

Receiver fitness increases with the accuracy with which the receiver response, p, approximates signaller quality, q.

Grafen (1990) constructed a specific function determining (male) signaller fitness with which to demonstrate how his general model worked.

$$w_m = p^r q^a$$

This fitness function allows that increases in signaller quality, q, reduce the costs incurred in making an advertisement, a, and that increases in signaller quality increase the positive fitness consequences of female preference, p. The degree to which female preference influences signaller fitness is governed by a parameter, r.[2]

As such, Grafen's function cannot accommodate the possibility that the fitness consequences of receiver responses might vary with signaller quality independently from the manner in which the negative fitness consequences of advertising vary with signaller quality. Furthermore, the function fails to accommodate the possibility that the negative fitness consequences for signallers of advertising might increase with signaller quality.

An alternative function must be constructed before an unconstrained exploration of the various possible signalling scenarios entertained within the literature can be undertaken.

$$w_m(a, p, q) = pq^R - aq^S$$

For this function, w_m, a, p, and q denote, as before, signaller fitness, level of advertising, degree of receiver preference, and level of signaller quality, respectively, whilst R and S are exponents which govern, respectively, the manner in which signaller quality mediates the positive effect of receiver responses and the manner in which signaller quality mediates the negative effect of signaller advertisement. The function is naturally understood as the sum of a positive benefit term and a negative cost term.

[2] Grafen assumes that both q and p lie in the interval [0,1], and that both a and r are greater than or equal to unity.

The first term of the fitness function, pq^R, connotes the benefit of signalling. The receiver response, p, contributes positively to signaller fitness, but the manner in which it contributes may be sensitive to signaller quality. For scenarios in which $R = 0$, the fitness consequences of receiver responses are independent of signaller quality. For scenarios in which $R > 0$, the positive contributions of receiver responses increase with signaller quality. For scenarios in which $R < 0$, the positive contributions of receiver responses decrease with signaller quality.

The second term, aq^S, represents the cost of signalling. The signaller's level of advertisement, a, contributes negatively to signaller fitness, but the manner in which it contributes may be sensitive to signaller quality. For scenarios in which $S = 0$, the fitness consequences of advertising are independent of signaller quality. For scenarios in which $S > 0$, the cost of advertising increases with q. Conversely, for scenarios in which $S < 0$, the cost of advertising decreases with q. This last class of scenarios is asserted by Zahavi (1975, 1977; Zahavi & Zahavi, 1997) to be the only class admitting of honest signalling behaviour.

In order to derive the conditions for the existence of an honest signalling ESS we must derive the conditions under which "better males do better by advertising more" (Grafen, 1990, p.520). Grafen formulated the condition thus:

$$\frac{\partial w_m / \partial a}{\partial w_m / \partial p} \text{ is strictly increasing in q.}$$

For the functions defined above, this yields,

$$(R - S)q^{S-R-1} > 0,$$

which is satisfied exclusively by $R > S$. Thus we can expect honest signalling ESSs to exist for scenarios in which $R > S$, i.e., scenarios in which, naturally enough, the manner in which quality mediates the positive fitness consequences of female preference (R) outweigh the manner in which quality mediates the negative fitness consequences of advertising (S).

Thus, through manipulation of the signaller fitness function's two free parameters, R and S, this continuous model can be made to capture the assumptions of various models within the literature. In addition, a clear prediction concerning the conditions under which honest advertisement is an ESS has been made. These ESS conditions accommodate results presented within, for example, Grafen (1990) and Godfray (1991), whilst allowing the existence of a broader class of honest signalling conditions than predicted under such models (see Figure 4). This broader class of ESS conditions includes scenarios in which Zahavi's handicap condition do not have to be met (i.e., conditions in which $S \not< 0$).

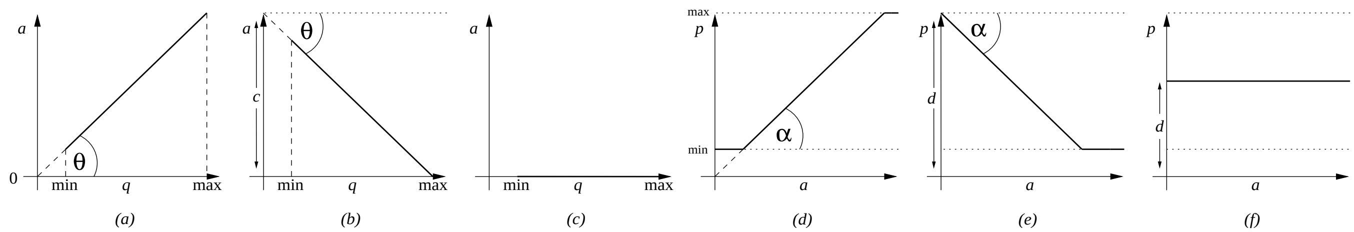

Figure 1: Examples of three continuous signalling strategies mapping signaller quality, q, onto advertisement, a, and three continuous response strategies mapping advertisement, a, onto receiver response, p. Each strategy is defined by a gradient (constrained to lie strictly within the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$) and an intercept (unbounded).

1.2 Signalling and Response Strategies

A population of signallers/receivers was distributed across a 25-by-25 grid. Each cell in the grid contained one signaller and one receiver. Each signaller was allocated an internal state, q, drawn at random from a uniform probability distribution in the range $[q_{min}, q_{max}]$. In addition, each signaller inherited a signalling strategy from its parent. A signalling strategy comprised two real values, $\{\theta, c\}$. An advertisement, a, was calculated as $q \tan \theta + c$. Advertisements of below zero were truncated to zero. Similarly each receiver inherited a response strategy from its parent. A response strategy comprised two real values, $\{\alpha, d\}$. Receiver response, p, was calculated as $a \tan \alpha + d$. Responses lying outside the range $[q_{min}, q_{max}]$ were truncated to their nearest extreme. For all simulations reported here $q_{min} = 0.1$ and $q_{max} = 5.0$.

The honesty of such a signalling strategy cannot be ascertained through consideration of the strategy in isolation. Either of the signalling strategies depicted in Figures 1a and 1b could take part in an honest signalling scenario since they each provide a unique advertisement for each possible value of signaller quality. For example, Figures 1d and 1e depict response strategies which would successfully recover the value of q from advertisements made by signallers adopting the signalling strategies depicted in Figures 1a and 1b respectively. In contrast, the signalling strategy depicted in Figure 1c does not provide unique advertisements for each possible value of signaller quality. The best reply to such signalling strategies is to play the response strategy depicted in Figure 1f, which ensures that each signaller is assessed as of average quality.

Whether communication is deceitful or honest is thus contingent upon the manner in which the signalling and response strategies match up across the population. If a period of adaptation under the selection pressures implemented by the fitness functions outlined above leads to a population of signallers playing the strategy depicted in Figure 1b partnered by a population of receivers playing the strategy depicted in Figure 1e, such populations can be considered, in concert, to be taking part in an honest signalling scenario since receiver prediction error is minimized in such circumstances. In such a population, a mutant signaller playing the alternative signalling strategy depicted in Figure 1a is cheating since the quality of such a mutant would be systematically misjudged by receivers. The classification of such a signaller as a cheat must be made despite the fact that the particular signalling strategy employed by the mutant generates advertisements which are directly proportional to its internal state.

This scheme for the representation of signalling and response strategies compares favourably with alternative schemes proposed within similar models. For example, de Bourcier and Wheeler (1995) construct a model of aggressive signalling with which to explore the handicap principle, and propose that a signalling strategy can be represented as the (positive) gradient, m, of an advertising function of the form $a = mq$. Under such a scheme, although signallers may employ different degrees of exaggeration, no signaller is able to signal more strongly when low quality than when high, and every signaller must make an advertisement of zero when of zero quality. This overly restricts the strategy space and consequently limits the evolutionary dynamics of their model.

1.3 Algorithm and Parameters

The fitnesses of signallers and receivers were calculated as per the fitness functions defined above, each interacting once with four partners chosen randomly (with replacement) from its local neighbourhood. Once each signaller and receiver had been assessed, the whole population was updated synchronously and asexually. One parent from the previous generation was chosen for each offspring cell. The location of a potential parent was chosen through perturbing both the x and y grid co-ordinates of the offspring cell by independent values drawn from a normal probability distribution with standard deviation 1.75 and mean zero. Four potential parents were chosen for each offspring signaller. An offspring signaller inherited its signalling strategy from the fittest of these four. Similarly, an offspring receiver inherited its response strategy from the fittest of four receivers chosen from the previous generation in the same manner.

A mutation operator ensured that offspring sometimes inherited a strategy which differed from that of their parents. For both signallers and receivers each of the two values comprising their inherited strategy were independently exposed to the chance of mutations, which occurred with probability 0.01. Mutations, when they occurred, consisted of perturbations drawn from a normal distribution with mean zero and standard deviation 0.05. Mutated values which lay outside the legal range for the parameter they coded for were truncated to the nearest legal value for that parameter.

Populations were simulated for 1000 generations in this manner, during which time the signalling and response strategies present in the population were recorded. The parameters R and S were varied across simulations but remained constant throughout each. The 441 possible pairs of parameter values, $\{R, S\}$, drawn from the set {-2.0, -1.8, ... 1.8, 2.0}, were exhaustively explored under each of three differing classes of initial condition. Each of the resulting 1323 (3 by 441) conditions were simulated 10 times. The pseudo-random number generator employed by the algorithm was itself seeded randomly for each simulation.

The first class of initial conditions consisted of a population of signallers sharing an 'honest' signalling strategy which mapped q directly onto a, $\{\theta = \frac{\pi}{4}, c = 0\}$, and a population of receivers sharing a 'believing' response strategy, $\{\alpha = \frac{\pi}{4}, d = 0\}$, which faithfully recovers values of q from signaller advertisements produced under the honest signalling strategy. This class of initial conditions will be termed 'Honest' since receivers are able to predict signaller quality accurately from signaller advertisements.

The second class of initial conditions consisted of a population of signallers and receivers, each with a strategy generated by drawing values for θ and α at random from a uniform distribution $[-\frac{\pi}{4}, \frac{\pi}{4}]$, and similarly drawing values for c and d at random from a uniform distribution $[-q_{max}, q_{max}]$. This class of initial conditions will be termed 'Random' since signallers' strategies and receivers' strategies are unrelated and implement a wide range of mappings.

The third class of initial conditions consisted of a population of signallers sharing a signalling strategy which mapped any value of q onto 0, i.e., $\{\theta = 0, c = 0\}$, and a population of receivers sharing a response strategy which mapped any advertisement onto 0, i.e., $\{\alpha = 0, d = 0\}$. This class of initial conditions will be termed 'Cynical' since signallers never make advertisements, whilst receivers never make responses.

1.4 Results

Results were consistent with the predictions arrived at through the analysis presented above. Two measures of performance were utilized in assessing the degree of honesty within a population. Both measures were derived from population summary statistics calculated for a particular generation. First, the average signalling strategy and response strategy were calculated. This was achieved simply by taking the population mean values of θ, c, α, and d.

From the mean signalling strategy, $\{\bar{\theta}, \bar{c}\}$, the mean strategy signal range, $\bar{r}$, was calculated as $(q_{max} - q_{min}) \tan \bar{\theta}$. The mean strategy response error, $\bar{e}$, was calculated as the absolute mean difference between signaller quality, q, and receiver response, p, for signallers using the mean signalling strategy $\{\bar{\theta}, \bar{c}\}$ and receivers using the mean response strategy $\{\bar{\alpha}, \bar{d}\}$, calculated for q ranging from q_{min} to q_{max}.

Since both these metrics are population-level summary statistics, care must be taken to appreciate that many heterogeneous populations could be responsible for any observed value. For example, a value of $\bar{e} = \frac{q_{max}+q_{min}}{4}$ may indicate a homogeneous population of receivers adopting the strategy $\{\alpha = 0, c = \frac{q_{max}+q_{min}}{2}\}$, or a heterogeneous population comprised such that, although each receiver employs a different strategy, *on average* they achieve chance levels of performance. Throughout the following sections, such ambiguity was avoided though recourse to the relevant standard deviations.

Honest Initial Conditions: The equality $R = S$ divided the parameter space into two areas (see Figure 2). The area defined by $R > S$ contained signallers which made advertisements which increased with signaller quality ($\bar{r} > 0$), and receivers which were able to recover signaller quality accurately from such advertisements ($\bar{e} \approx 0$); i.e., honest signalling obtained under these conditions. In contrast, the area defined by $R < S$ contained signallers which made advertisements which did not differ with signaller quality ($\bar{r} \approx 0$), and, as a result, receivers which were unable to accurately recover signaller quality from signaller advertisements ($\bar{e} > 0$); i.e., non-signalling obtained under these conditions.[3]

Furthermore, for scenarios in which $R > S$, mean signal range, $\bar{r}$, increased with $R - S$. For scenarios in which the difference between R and S is small, the range of signals is also small. However, for scenarios in which R far outstrips S, signals given by high quality signallers are orders of magnitude higher than those given by low quality signallers.

Random Initial Conditions: Within the area of parameter space in which honest signalling equilibria are not predicted to exist, $\bar{r} \approx 0$ whilst $\bar{e} \approx \frac{q_{max}+q_{min}}{4}$, i.e., non-signalling strategies, and response strategies which perform at the level of chance are observed.

Within the area of parameter space predicted to admit of honest signalling equilibria, both honest signalling

[3]Mean response error is sometimes higher than that resulting from performance at chance levels. This is due to an artefact of the simulation design (limiting receiver response to fall within the range $[q_{min}, q_{max}]$). A full account is given in Bullock (1998).

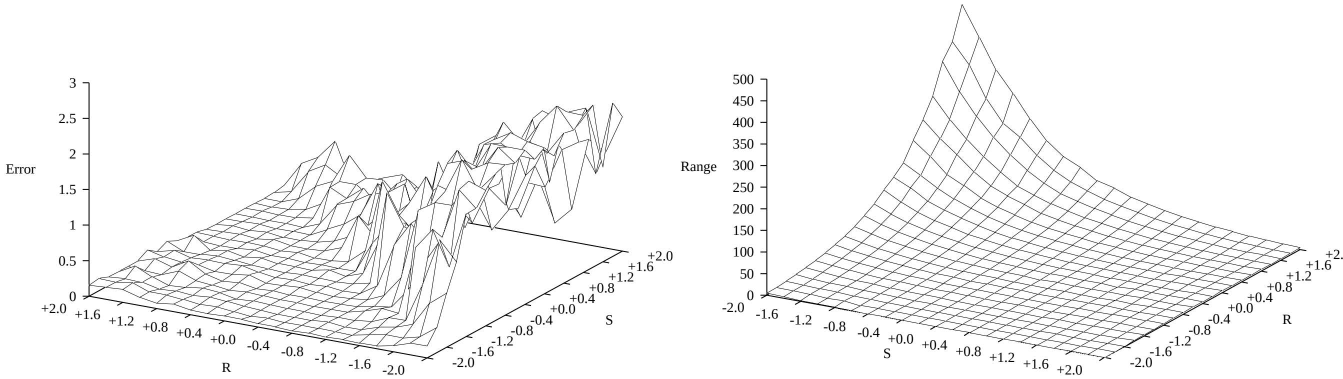

Figure 2: Mean response error (left) and signal range (right) after 1000 generations, averaged across 10 simulation runs from Honest initial conditions. For reasons of clarity the left graph has been rotated 90° anti-clockwise about the vertical axis.

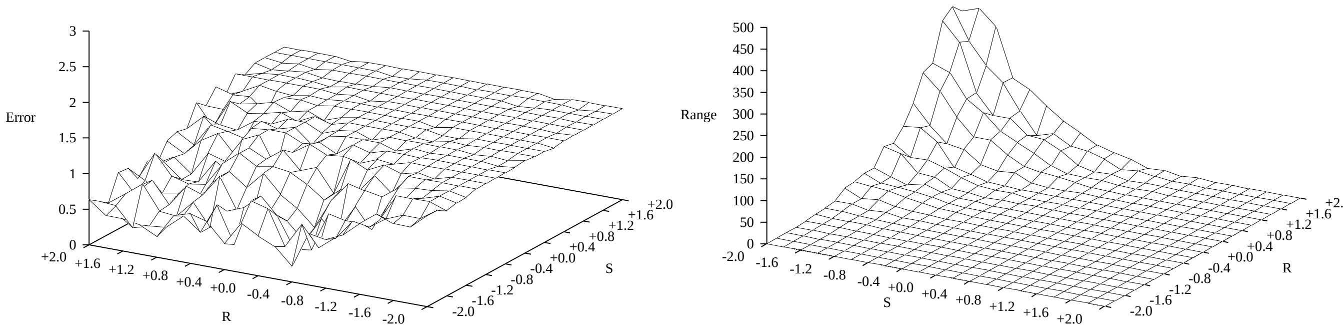

Figure 3: Mean response error (left) and signal range (right) after 1000 generations, averaged across 10 simulation runs from Random initial conditions. For reasons of clarity, the left graph has been rotated 90° anti-clockwise about the vertical axis.

equilibria and non-signalling equilibria were achieved. The frequency with which honest signalling equilibria were achieved from Random initial conditions increases with the magnitude of $R-S$. For simulations in which R is only slightly higher than S, honest signalling equilibria are achieved only rarely. As $R-S$ increases, signalling equilibria are achieved with increasing frequency. This is reflected in the variation, across the parameter space, of both the mean values for $\bar{r}$ and $\bar{e}$ (see Figure 3) and their standard deviations.

Cynical Initial Conditions: The behaviour of the model is similar to that resulting from Random initial conditions. For simulations in which $R < S$, behaviour is indistinguishable from that resulting from Random initial conditions. For simulations in which $R > S$, signalling equilibria are sometimes attained, although the frequency with which this occurs is lower than that observed for simulations from Random initial conditions. As before, the frequency with which signalling equilibria are achieved increases with $R-S$.

1.5 A Note on Equilibria

At several points throughout the preceding sections use is made of the term equilibrium. The honest signalling equilibria described have the general character of point equilibria. However, the stochasticity of the tournament selection process and the allocation of signaller quality, the statistical independence of mutation events, and the co-evolutionary nature of the signaller-receiver relationship all ensure that a population of signallers or receivers will tend to move around the vicinity of its equilibrium state, rather than fix upon it rigidly, as might be expected from an idealized numerical approximation to the dynamic equations of an ESS model. Thus to call the equilibria achieved by the simulation ESSs is not strictly accurate. However, in their defence, the honest signalling equilibria achieved by the simulation are predicted by the ESS model and are characterized by approximately constant trajectories within both the signaller and receiver populations.

By contrast, what might be called the simulation's non-signalling equilibria permit significant amounts of evolutionary drift on the part of both signaller and receiver populations. There are, for example, many signalling strategies which result in signallers never making a signal. If, under certain conditions, selection favours making no signal, evolutionary drift amongst these functionally identical strategies is inevitable.

Similarly, although there is an optimal response strategy in reply to such non-signalling signallers, the relatively small sample of four signallers against which each receiver is assessed ensures that there exists a high degree of variability in fitness scores achieved by strategies in the vicinity of this optimal strategy (which is only strictly optimal if assessed on the basis of an infinite number of trials, each featuring a signaller drawn at random from the entire population). One might conceive of this situation as involving a receiver population which is subject to a rather weak negative feedback from its coevolutionary partner. This feedback keeps the receiver population within a volume of strategy space containing strategies with fitnesses sharing a similar mean and a relatively high variance.

A related but distinct point concerns whether equilibria achieved by the simulation are repeatable, i.e., whether the same population states are always achieved from the same initial conditions under the same parameter values. For the purposes of this model no claim to this effect will be made since the simulation's stochasticity is, at times, quite capable of perturbing trajectories from one basin of attraction to another. Despite this indeterminacy, basins of attraction remain characterizable.

Considerations such as these do no damage to the generality of the results presented here, but should be borne in mind when analyzing the behaviour of any evolutionary simulation model.

2 Summary

These simulation results suggest that both the possibility of a stable honest signalling system, and the extravagance of signals within such a system, depends critically on the difference between the manner in which the advertised trait influences the *benefits* of signalling (R) and the manner in which it influences the *costs* of signalling (S). Where this difference is negative ($R < S$), no honest signalling is possible. Where this difference is positive ($R > S$), honest signalling equilibria exist.

Thus, honest signalling equilibria may exist for scenarios in which Zahavi's handicap condition does not hold (i.e., $S \not< 0$), and conversely honest, signalling equilibria may not exist for scenarios in which Zahavi's handicap condition does hold (i.e., $S < 0$).

Furthermore, the magnitude of the difference, $R - S$, is positively correlated with the extent of the basin of attraction for any signalling equilibrium. Thus, although under conditions in which $R - S$ is positive but small, stable signalling equilibria in which signals are relatively

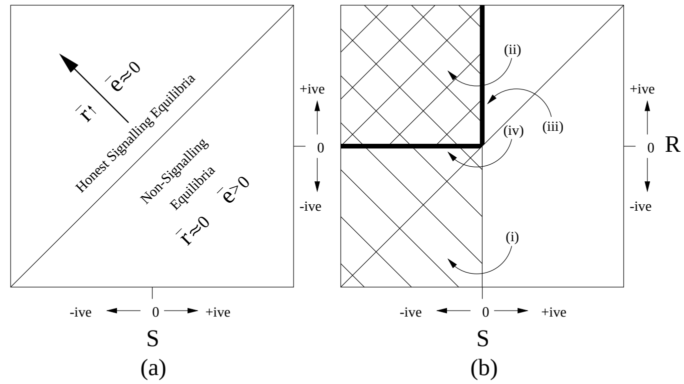

Figure 4: Showing *(a)* the conditions under which honest signalling obtain for the model presented here, and *(b)* the conditions predicted to admit of honest signalling equilibria by (i) Zahavi (1975, 1977): diagonal hatching defined by $S < 0$, (ii) Grafen (1990): cross-hatching defined by $S < 0$ and $R \geq 0$, (iii) Godfray (1991) and Maynard Smith (1991): bold vertical line defined by $R > 0$ and $S = 0$, and (iv) Hurd (1995): bold horizontal line defined by $S < 0$ and $R = 0$.

cheap are viable, such equilibria will seldom be attained through the evolution of non-signalling ancestral populations since the basins of attraction for such equilibria are prohibitively small. This result is derived solely from the simulation behaviour since predictions concerning the attainability of equilibria could not be made on the basis of the analytic findings reported in Bullock (1997).

3 Discussion

It has been demonstrated through the use of an evolutionary simulation model that non-handicap equilibria exist for an extension to Grafen's (1990) continuous model of signal evolution. In this section, the findings reported in this paper will be compared with those reported within previous studies. These previous results are easily accommodated by those presented within this paper, which themselves provide a general formulation of the conditions governing the existence of what have been termed 'handicap' signalling scenarios.

Once this reconciliation of previous results has been described, a reconciliation of the positions which lead to their presentation will be attempted. The handicap principle will be assessed in three regards. The first issue discussed will be the various interpretations of the relationships between costs, benefits, and fitness which appear to motivate models of the handicap principle. Secondly, the validity of the term handicap itself will be considered before, finally, the implications of the results presented here for the supposed generality of the handicap principle will be addressed.

3.1 Reconciliation of Results

Figure 4a depicts the broad conclusion suggested by the continuous model of signalling explored here. Honest signalling is stable for scenarios in which the manner in which the advertised trait mediates the influence of signalling benefits on signaller fitness outweighs the manner in which the advertised trait mediates the influence of signalling costs on signaller fitness, i.e., the net cost of signalling (the cost of honest signalling minus the benefit of an accurate receiver response) decreases monotonically with the advertised trait.

This result is captured graphically in Figure 4a by dividing the space of possible signalling scenarios into two halves, separated by a diagonal line along which the influence of the advertised trait on costs is exactly balanced by its influence on benefits (i.e., $R = S$). Above this line (i.e, for $R > S$), honest signalling equilibria obtain; below it (i.e., for $R < S$), no such honest signalling equilibria exist.

In Figure 4b this graphical device is used to locate previous theoretical results. For example, Zahavi's (1975, 1977; Zahavi & Zahavi, 1997) claim that honest signalling may only exist for scenarios in which the costs of signalling decrease with the trait being advertised may be represented by the area satisfying the inequality $S < 0$. It is plain from the diagram that this inequality is neither necessary nor sufficient for the existence of honest signalling equilibria. Grafen's (1990) contention is hown to be correct. Given that signaller benefits are either unaffected by the advertised trait or increase with the advertised trait ($R \geq 0$), in order that signalling be honest, signalling costs must decrease with the advertised trait ($S < 0$). However, the space of possible signalling scenarios defined by his conditions, does not exhaustively account for all honest signalling equilibria.

Models in which the negative fitness consequences of signal costs are assumed to be independent of the trait being advertised (i.e., models for which $S = 0$) have often concluded that, in order for such signalling to be honest, the positive fitness consequences for signallers of receiver behaviour must increase with the signaller's advertised trait (i.e., $R > 0$). Such models typically take the advertised trait to be signaller need (e.g., Godfray, 1991; Maynard Smith, 1991). They make a claim which can be recast as asserting that honest signalling may be stable for signalling systems which lie along the bold vertical line in Figure 4b.

Similarly, models in which the positive fitness consequences of the benefits accrued by signallers are assumed to be independent of the trait being advertised (i.e., models for which $R = 0$) have often concluded that, in order for such signalling to be honest, the negative fitness consequences (for signallers) of signal cost must increase with the signaller's advertised trait (i.e., $S < 0$). Such models typically take the advertised trait to be signaller quality (e.g., Hurd, 1995), and make a claim which can be rephrased as asserting that honest signalling may be stable for signalling systems which lie along the bold horizontal line in Figure 4b.

3.2 Costs, Benefits, and Fitness

This paper opened with a description of two complementary arguments which each result from Zahavi's handicap signalling notion. The first argument suggested that honest advertisement of quality might be stabilized by differential signaller costs. The second argument suggested that the honest advertisement of need might be stabilized by differential signaller benefits. The results of the model constructed within this paper demonstrate that the honest advertisement of either quality, *or* need, may each be stabilized by differential costs, *and/or* differential benefits. This result is due to the fact that the terms 'cost' and 'benefit' may each be cashed out in the same currency — fitness. Costs are merely negative increments to fitness, whereas benefits are positive increments to fitness.

However, at a less abstract level of description, costs and benefits may come in many different forms. For example, negative fitness consequences may arise as a result of energetic costs, risks of predation, parasitism, or infection, costs of missing a high-quality mating opportunity, of mating with a sub-optimal mate, etc. Although each of these costs has negative fitness consequences, the character of these negative fitness consequences may differ radically across these different forms of cost.

Similarly there are benefits to be gained from obtaining a copulation, a food resource, a territory, an opponent's surrender, etc. Again, although each of these benefits has positive fitness consequences, the character of these positive fitness consequences may not be uniform across these different forms of benefit.

Within evolutionary models, the manner in which costs and benefits influence fitness is formalized identically. Costs, whatever their nature, influence fitness negatively, whereas benefits, whatever their nature, influence fitness positively.

However, theorists constructing models of handicap signalling are faced with a decision concerning the manner in which the influence of signalling costs (and signalling benefits) upon fitness is to *vary* with the trait which signallers are advertising. For example, how does the effect of signal production cost vary with signaller need? What will interest us here are the different decisions which may be made regarding these aspects of handicap modelling.

Consider the example of a begging nestling which is signalling in an attempt to solicit parental resources. We will assume that the trait of interest to parents is a chick's need, and that quality varies inversely with need. For this scenario Godfray (1991) models the cost of signalling as equal across all signallers. Grafen (1990), on the other hand, models cost as decreasing with signaller quality. Godfray (1991) models the benefit of soliciting a particular parental resource as increasing with need, whereas Grafen (1990) models such benefit as either independent of signaller need, or *decreasing* with signaller need.

A second example, also addressed by Grafen (1990), involves an interloper making a signal of aggressive intent to an observing harem holder. Grafen asserts that in such a situation, the costs of signalling decrease with the increasing quality of a signaller. He further claims that the benefits for the signaller of a retreat response by a receiver increase with the quality of a signaller. In contrast, Adams and Mesterton-Gibbons (1995) suggest that in such situations, the benefit of eliciting a retreat response might *decrease* with increasing signaller quality. They reason that "strong animals can win many conflicts without threatening (i.e., by direct fighting), while weak animals cannot. Furthermore, weak animals have more to gain by avoiding direct fights since they are less able to defend against injury." (p. 406).

It is clear from these two examples that the authors of these models have made radically opposed assumptions with respect to the relationship between costs, benefits, and fitness (see Bullock, 1997, for a discussion of possible reasons for these differences). In contrast, the model presented within this paper makes no assumptions concerning the manner in which costs and benefits influence fitness, save that costs are a negative influence, whilst benefits are a positive influence. As a result of this neutrality, a degree of generality has been gained.

3.3 Are Signals Handicaps?

The force of the results presented within this paper is to qualify previous statements of the conditions which must be met before honest handicap signalling may be evolutionarily stable. Rather than merely requiring *gross* signalling costs to vary with signaller quality in some manner, the model presented here requires consideration of the manner in which the *net* cost of signalling varies with signaller quality.

Although Zahavi often appears to consider the net costs involved in signalling when formulating his principle (e.g., "it is reasonable to expect a population in its optimal fitness to benefit from a handicap", and "so long as the offspring ... does not deviate to grow its handicap larger than it can afford, the handicap [may persist] as a marker of honest advertisement", Zahavi, 1977, p. 604), when describing examples of natural signalling he rarely appreciates the benefits which might be accrued from signalling, and the manner in which such benefits might negate the increased costs involved in bluffing.

Zahavi's ambivalence toward the potential benefits of signalling (or bluffing) led Wiley (1983) to characterize Zahavi's (1975) claim as "signals should evolve to become a *net* handicap to signallers" (p. 176, my emphasis), whilst Adams and Mesterton-Gibbons (1995) reach the opposite conclusion, stating that the scenario they consider differs from that prescribed by the handicap principle in that within their model, "the net benefit for a given advertisement may not increase monotonically with the signaller's strength" (p. 406).

Furthermore, the sense of much of Zahavi's verbal argument does not seem to accord with a notion of the handicap principle couched in terms of net costs. For example, as Hurd (1995) points out, if the costs involved in signalling must be acceptable costs (i.e., they must be compensated for by consonant benefits), then in what sense are these costs a 'handicap'? Although the costs incurred by a *bluffer* might be characterized as a handicap, since these costs would not be compensated for by the receiver response, this is not the sense in which Zahavi proposed the term. For Zahavi, *honest* signallers suffer a handicap. This suffering is necessary as a means of demonstrating honesty. However, once one appreciates the role played by benefits in assuaging these costs, the notion that signallers are "suffering" becomes suspect.

3.4 The Generality of the Handicap Principle

The inclusion of a benefit clause in the definition of the handicap principle does not preclude the existence of handicap signalling equilibria. However, it does have implications for the proposed ubiquity of the handicap principle as it has been presented by Zahavi and others.

The condition that signal cost is related to signaller quality in the manner stipulated by Zahavi (i.e., that as signaller quality increases the cost of signalling decreases) appears to be a candidate for very wide application. Indeed, Zahavi has demonstrated the breadth of this application, even going so far as to suggest that the handicap principle accounts for all natural signalling. However, the model constructed here demonstrates that the influence of benefits on signaller behaviour may ensure that despite signal cost being related to signaller quality in the manner prescribed by Zahavi, honesty may never-the-less be unstable. Similarly, some systems, despite failing to meet Zahavi's handicap condition (e.g., systems in which there is no relationship between signal cost and signaller quality) may be stable due to the influence of benefits upon signaller behaviour.

As such, the ease with which these revised conditions for the existence of evolutionarily stable handicap signalling may be confidently predicted to hold across classes of signalling scenario is much reduced. Field biologists charged with the task of establishing whether real signalling systems are handicap signalling systems must characterize both the manner in which signal cost differs with the trait being advertised *and* the manner in which signaller benefits differ with the same trait. This increased burden is compounded by the fact that, as demonstrated above, theorists' predictions concerning the manner in which costs and benefits vary with, for example, quality or need across signalling populations themselves demonstrate a lack of coherence.

From this discussion, it is clear that the model constructed within this paper, in addition to clarifying the conditions under which signalling may be honest and stable, questions the integrity of handicap terminology. It also challenges the handicap principle's supposed ubiquity through highlighting the complications which arise from a consideration of the manner in which costs *and benefits* are mediated by advertised traits.

4 Conclusion

In summary, the satisfaction of Zahavi's handicap condition was demonstrated to be neither necessary nor sufficient for the existence of honest signalling equilibria within a general continuous evolutionary simulation model.

It was demonstrated that in order for a signalling system to be stable, a relationship between signalling costs, signaller quality, and (*contra* Zahavi) signalling *benefits* must hold, not merely a relationship between signalling costs and signaller quality.

Stable signalling systems involving relatively cheap signals were shown to be viable under certain conditions. However, the evolutionary attainability of these equilibria was shown to be compromised by the size of their basins of attraction.

Acknowledgments

This paper was improved immeasurably through the influence of Valerie Chase, Jason Nobel, Ezequiel di Paolo, Henrietta Wilson, and the useful commentary of two anonymous referees.

References

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

Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*, pp. 454–463 Cambridge, MA. MIT Press.

Bullock, S. (1998). Evolutionary simulation models: On their character, and application to problems involving the evolution of natural signalling systems. Unpublished PhD Thesis, University of Sussex, England.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex.* John Murray, London.

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

de Bourcier, P., & Wheeler, M. (1995). Aggressive signalling meets adaptive receiving: Further experiments in Synthetic Behavioural Ecology. In Morán, F., Moreno, A., Morelo, J. J., & Chacón, P. (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life* Berlin and Heidelberg. Springer-Verlag.

Godfray, H. C. J. (1991). Signalling of need by offspring to their parents. *Nature*, *352*, 328–330.

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

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

Johnstone, R. A. (1997). The Evolution of Animal Signals. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (4th edition)., pp. 155–178. Blackwell, Oxford.

MacLennan, B. (1991). Synthetic Ethology: An Approach to the Study of Communication. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 631–658 Redwood City, California. Addison-Wesley.

Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge University Press, Cambridge.

Maynard Smith, J. (1991). Honest signalling: The Philip Sidney game. *Anim. Behav.*, *42*, 1034–1035.

Veblen, T. (1899). The theory of the leisure class. In Lerner, M. (Ed.), *The Portable Veblen*, pp. 53–214. Viking Press, New York. Collection published 1948.

Werner, G. M., & Dyer, M. G. (1991). Evolution of Communication in Artificial Organisms. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 659–687 Redwood City, California. Addison-Wesley.

Wiley, R. H. (1983). The Evolution of Communication: Information and Manipulation. In Halliday, T. R., & Slater, P. J. B. (Eds.), *Communication*, pp. 156–189. Blackwell, Oxford.

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., & Zahavi, A. (1997). *The Handicap Principle: A missing piece of Darwin's puzzle.* OUP.