Costly Signaling and Cooperation

Eric Smith
Samuel Bowles
Herbert Gintis[*]

November 14, 2000

Abstract

We propose an explanation of cooperation among unrelated members of a social group, in which providing group benefits evolves because it constitutes an honest signal of the member's quality as a mate, coalition partner or competitor, and therefore results in advantageous alliances for those signaling in this manner. Our model is framed as an n-player game that involves no repeated or assortative interactions, and assumes a payoff structure that would conform to an n-player public goods game in which non-cooperation would be a dominant strategy if there were no signaling benefits. We show that honest signaling of underlying quality by providing a public good to group members can be evolutionarily stable. We also show that this behavior is capable of proliferating in a population in which it is initially rare. Our model applies to a range of cooperative interactions, including providing individually consumable resources, participating in group raiding or defense, and punishing free-riding or other violations of social norms. Our signaling model is distinctive in applying to group rather than dyadic interactions and in determining endogenously the fraction of the group that signals high quality in equilibrium.

---

[*] Eric Smith, Department of Anthropology, University of Washington, Box 353100 Seattle, WA 9 8195-3100, easmith@u.washington.edu; Samuel Bowles, Department of Economics, University of Massachusetts, Amherst, bowles@econs.umass.edu, http://www-unix.oit.umass.edu/b̃owles; Herbert Gintis, Department of Economics, University of Massachusetts, Amherst, hgintis@mediaone.net, http://www-unix.oit.umass.edu/g̃intis. We would like to thank Michael Alvard, Rebecca Bliege Bird, Robert Boyd, Steven Frank, Peter Richerson, Steven Siller, and Richard Sosis for useful comments, and the MacArthur Foundation for financial support. Please do not quote or cite without the authors' permission.

1 Introduction

Cooperation among unrelated individuals has generally been explained by some form of conditional reciprocity (Trivers 1971, Axelrod and Hamilton 1981). However, there is an increasing interest in examining alternative mechanisms for the evolution of cooperation (Dugatkin 1997, Pusey and Packer 1997), including multilevel selection (Wilson 1977, Bowles 2000, Sober and Wilson 1998, Gintis 2000b) and mutualism (Brown 1983, Connor 1995). This paper presents a game-theoretic model of the evolution of cooperation based on costly signaling theory (Spence 1973, Zahavi 1975, Grafen 1990a). Our signaling model is distinctive in applying to group rather than dyadic interactions and in determining endogenously the fraction of the group that signals high quality in equilibrium.

Several authors (Zahavi 1977, 1995, Roberts 1998, Wright 1999, Smith and Bliege Bird 2000) have proposed that costly signaling could provide an explanation for cooperation and group-beneficial behavior, but this proposal has not been formally modeled. Our model is framed as an n-player game that involves no repeated or assortative interactions, and assumes a payoff structure that would conform to an n-player public goods game if there were no signaling benefits. The unique equilibrium of this game involves universal defection as the dominant strategy, so no player supplies the group benefit. We show that honest signaling of underlying quality by providing a benefit to group members can be evolutionarily stable, and may proliferate when rare as long as high-quality individuals are neither too common nor too rare, as given by equation (1), and the cost of signaling is sufficiently greater for low than for high quality players (see Theorem 1).

In our model, cooperation involves providing a benefit to all members of the group without reciprocation in kind. Given the resulting public goods game payoff structure and the one-shot nature of the interactions, individually costly cooperation could not evolve unless one postulated the group selection of altruistic behavior. Even if interactions among group members were repeated, n-player cooperation requires implausible forms of coordination, particularly if the group contains more than a few individuals (Boyd and Richerson 1988). The model presented here is meant to apply to such cases, where reciprocity is unlikely to emerge and is vulnerable to free-riding.

We propose that cooperating by providing a benefit to group members may be a reliable signal of signaler quality—i.e., phenotypic attributes that are difficult for others to assess directly, yet have important effects on the payoffs from social interactions with the signaler. Those who provide the benefit to others, or who provide more of the benefit (by signaling more intensively), assume costs greater than their personal gain from the benefit provided, but by so doing honestly advertise their quality as allies, mates, or competitors. This information alters the behavior of other group members to act, for purely selfish motives, in ways that provide positive payoffs to signalers—for example, preferring them as allies or mates, or deferring to them in competitive situations (Smith and Bliege Bird 2000).

Our model applies to a range of social interactions. First, the benefit whose provision signals high quality may take the form of individually consumable resources. For example, anthropologists have explained the widespread practice among hunter-gatherers of sharing individually harvested resources among non-kin, as a means of reducing risk (Smith 1988). While risk-reduction effects are plausible and can be demonstrated (Cashdan 1985, Kaplan, Hill and Hurtado 1990), such food-sharing practices produce a Prisoner's Dilemma payoff structure (Smith and Boyd 1990), and thus create incentives to slack off and free-ride on the efforts of others (Blurton Jones 1986, Hawkes 1993). This insight has led to an active debate about how to explain the undisputed fact of extensive food-sharing by hunter-gatherers (Winterhalder 1996). We think it clear that while conditional reciprocity may explain some cases of hunter-gatherer food sharing, it cannot explain them all. When all group members have rights to consume the resource regardless of their past contribution, and the number partaking is dozens or more, the conditions for conditional reciprocity are not met (Hawkes 1992). Such a situation has been described for human societies as diverse as the Ache Indians of the Paraguayan forest (Kaplan and Hill 1985a), Hadza of the East African savanna (Hawkes 1993), and Meriam turtle hunters of Torres Strait in tropical Australia (Bliege Bird and Bird 1997). In each of these cases, at least some types of harvested resources are shared unconditionally with all members of the community, and some hunters consistently provide more than others while sharing equally in the catch. These "altruistic" providers in fact reap higher social status and reproductive success than their less-productive peers, despite the absence of any conditional exchange of "meat for mates" (Kaplan and Hill 1985b, Marlowe 2000, Bliege Bird, Smith and Bird in press).

Recently costly signaling has been proposed as a reason for certain types of food-sharing in human societies, such as providing large and/or difficult-to-harvest game for consumption at ritual feasts (Boone 1998, Gurven, Allen-Arave, Hill and Hurtado 2000, Smith and Bliege Bird 2000, Sosis 2000). The model presented here provides a theoretical foundation for such empirical analyses. Our model may also apply to some cases of food-sharing in other species, including various birds (Brown, Brown and Shaffer 1991, Heinrich and Marzluff 1995, Zahavi 1990), macaques (Dittus 1984, Hauser and Marler 1993), and chimpanzees (Stanford 1999). The last case is interesting because wild chimpanzee sharing only involves hunted resources, the hunters are always males, and the best predictor of hunting frequency is the number of estrous females present in the social group, even though females are not the primary recipients of meat shares (Stanford, Wallis, Mpongo and Goodall 1996).

A second type of cooperative situation to which our model applies are cases where the benefit is a public good that is not individually consumable. For instance, participating in group raiding or defense—an individually costly behavior that is common among chimpanzees as well as human societies (Boehm 1992)—provides benefits that are available to all group members and cannot be hoarded. Another example of such a benefit is punishing those who free ride on the group's cooperative activities. This latter activity is particularly salient, since it is well known that while enforcing cooperation by punishing defectors will solve collective action problems, such enforcement poses a second-order collective action problem (Hardin 1982). Boyd and Richerson (1992) demonstrated that if enforcement takes the form of punishing both non-cooperators and non-punishers, then cooperation (or anything else) can be evolutionarily stable, even in large groups. Our model provides one mechanism for the evolution of such a system. In this version, enforcement—punishment of non-cooperators—itself is the benefit to others that signals high quality, and may be a potent element in stabilizing cooperation in many types of social systems (Boyd and Richerson 1992, Clutton-Brock and Parker 1995, Frank 1995, Michod 1997). The model presented below easily allows such punishment or enforcement to serve as the costly signal, and hence to be maintained when the conditions for evolutionary stability specified in our model are met. Honest signaling of quality need not be group beneficial, of course, and our signaling model applies equally well to socially neutral or harmful forms of costly signaling, such as conspicuous private consumption, brawling and dueling, flouting social norms with impunity, and the like, as discussed further below (Section 4).

The paper is organized as follows. In Section 2 we develop a model showing that there exists an equilibrium in which quality is honestly signaled over a wide range of parameter values in which low quality types must pay more to signal than high quality types. In Section 3 we show that under plausible conditions the costly signaling equilibrium is stable in a replicator dynamic. Section 4 analyzes when costly signaling will take the form of providing benefits to others. In Section 5 we develop a dynamic model determining the movement towards an equilibrium fraction of costly signalers. A final section draws some conclusions and implications from the results.

2 A Model of Costly Signaling

A model of costly signaling must account for why individuals engage in signaling and why those who observe the signal respond in ways favorable to the signaler. The processes by which each of these two behaviors are updated may reflect a fitness based selection of genetically transmitted traits or a process of cultural transmission, or both.

Consider a group consisting of n members. Once in each period each member of the group can perform an action at personal cost $c > 0$ that confers a benefit $g > 0$ on each other member of the group. Since there is a strictly positive cost c to providing the benefit, self-interested individuals will not do so, and those who do provide the benefit will be eliminated by any evolutionary process in which differential replication is monotonic in payoffs. However, providing the benefit may be stable and may evolve if we add two elements to the above structure of social interactions.

First, suppose group members have a personal characteristic, which we will call 'quality,' that can either be high or low. We assume members know their own quality but not that of others in the group. We assume that the expected cost c per period of providing the benefit to the group for the high quality type is less than the corresponding expected cost c' for the low quality type, so $0 < c < c'$. We consider the cases where $c = 0$, $c < 0$, and $c = c'$ below.

Second, suppose at some point in time each member has occasion to enter into a profitable alliance (e.g., mating or political coalition) with one or more of the other $n-1$ group members. This member, whom we will call the Partner, derives a benefit $h > 0$ from choosing a high quality ally, and a benefit (or possibly a loss) $l < h$ from choosing a low quality ally, and has payoff zero if no ally is chosen. Given a Partner, let p be the frequency of high quality members among the group of remaining $n-1$ members. We assume the Partner knows p but not the high quality/low quality type of individuals among the other $n-1$ members. A Signaler receives a payoff $s > 0$ from each of the $n-1$ Partners who chooses to ally with him. Finally, we treat each period as a one period game, as would be the case if periods represent generations, or an agent's quality in one period cannot be inferred from his quality in previous periods.

We construct an n-player game in which the players, whom we will call Signalers, choose independently whether or not to signal by providing the benefit to group members and then each player in turn is designated the Partner, who can choose an ally (a) randomly from the other $n-1$-members; (b) randomly from the subset of other members who provided the benefit; (c) randomly from the subset of other members who did not provide the benefit; or (d) can choose no ally. We will assume that in equilibrium all Signalers on the one hand, and all Partners on the other, follow the same pure strategy, since it is well known that a mixed strategy equilibrium in an asymmetric game of this type is always unstable (Selten 1980), so it may be ignored.

A distinctive aspect of this model is that individuals signal their quality to the group as a whole, prior to interacting with particular Partners within the group. This treatment of signaling reflects the fact that in most empirical cases discussed in the literature, the costly signal is not private to an intended receiver, but is emitted without the signaler knowing exactly with which among a population of possible agents it might influence. For instance, the peacock's tail is seen by all peahens, and the bullfrog's croak is heard by all female frogs in the listening area.

Consider a particular Signaler. We represent the probabilistic character of this Signaler's quality by considering the candidate as a Player 1, considering the Partner as Player 2, and then introducing a third player, Nature, who moves first by choosing a high quality Signaler with probability p and a low quality Signaler with probability $q = 1 - p$. The Signaler, informed of his high or low quality, then chooses either to signal or not, and the Partner chooses whether or not to consider the Signaler in the pool of potential allies.

In the game just described there are therefore four Signaler strategies, which we label $\{ss, sn, ns, nn\}$. Here ss means 'signal if high quality and signal if low quality,' sn means 'signal if high quality and do not signal if low quality,' ns means 'do not signal if high quality and signal if low quality,' and finally nn means 'do not signal if high quality and do not signal if low quality.' We can abbreviate these as {Always Signal, Signal Truthfully, Signal Untruthfully, Never Signal}.[1] Similarly, the Partner has four strategies, which we label $\{aa, ar, ra, rr\}$. Here, using the same convention as with the Signaler, aa means 'Always Accept (whether or not the Signaler signals),' ar means 'Accept if Signaler Signals, Reject if Signaler does not Signal,' ra means 'Reject if Signaler Signals, Accept if Signaler does not Signal,' rr means 'Reject Always.'

It is clear that as long as $ph + ql > 0$, there is a nonsignaling equilibrium (nn,aa) in which no agent signals and Partners choose randomly from all other group members for an ally. Similarly, if $ph + ql < 0$, there is a nonsignaling equilibrium (nn,rr) in which no agent signals and Partners never choose allies. We define an *honest signaling equilibrium* to be a strict Nash equilibrium in which agents signal if and only if they are high quality, and Partners choose randomly among the set of agents who signaled for allies.[2]

To find the conditions under which there is an honest signaling equilibrium (i.e., all signal honestly), we derive the conditions under which honest signaling is a best response for one agent, assuming all other agents signal honestly. This gives rise to the matrix shown in Figure 1. Multiplayer games are notoriously unwieldy, so we

---

[1] Since in some cases a Signaler may always be high quality or low quality, it may seem unnatural to assign to a Signaler a strategy part of which is never used (e.g., a high quality Signaler does not need an option for the case where he or she is low quality). However Harsanyi (1967) has shown that this formality is harmless. It has the advantage of considerably simplifying the analysis.

[2] A Nash equilibrium is *strict* if each player has a unique best response. A pure Nash equilibrium must be strict in order to be stable under a monotonic dynamic, so we consider only strict equilibria from the outset.

have introduced several notational simplifications to reduce the clutter of symbols in Figure 1 and the subsequent analysis dependent upon Figure 1. First, we have assumed p is independent of the particular Partner or, equivalently, we ignore the fact that some Partners may be high quality and others low quality. Unless n is quite small, this simplification is harmless. Second, we have multiplied the payoff to Signalers from alliances by n to reflect the fact that a Signaler pays the cost c only once per period, but has n potential Partners (actually $n-1$ but again we assume this is n to reduce notational overhead), each of whom can independently ally with the Signaler. Third, we have not included the benefit to others that accrues to both players from the actions of the other $n-2$ players, since this payoff merely adds a constant to each row of the game matrix for the Partner, and a constant to each column of the matrix for the Signaler. To see this, suppose k of the $n-2$ remaining players provide the benefit to group members. Then both Signaler and Partner receive payoff kg from the signalers' provision of benefits, no matter what strategies they choose. Moreover, whatever strategy the Signaler chooses, the Partner receives the same benefit from the Signaler. For instance, if the Signaler chooses sn, then the Partner receives a payoff of pg from the Signaler's provision of benefits, whether Partner chooses aa, ar, ra, or rr. Since adding a constant to the payoffs of a player, given the choices of the other players, cannot change the player's best response strategy, we omit these payoffs.

	aa	ar	ra	rr
ss	$s-pc-qc'$ $ph+ql$	$s/p-pc-qc'$ $ph+ql$	$-pc-qc'$ 0	$-pc-qc'$ 0
sn	$s-pc$ $ph+ql$	$s-pc$ h	$s-pc$ l	$-pc$ 0
ns	$s-qc'$ $ph+ql$	$qs/p-qc'$ l	$s-qc'$ h	$-qc'$ 0
nn	s $ph+ql$	0 0	s $ph+ql$	0 0

Figure 1: The Matrix of Payoffs to a Signaler S and a Partner P, assuming all other Signalers play the same strategy as S.

To understand how the entries in Figure 1 are calculated, we will derive them for the honest signaling equilibrium, which is the (sn,ar) box. In this case a high quality Signaler provides the benefit at cost c and receives expected benefit s/pn from each Partner, since each Partner now chooses randomly from among the group of pn

Signalers who provided the benefit. Multiplying the alliance payoff by n, because there are n Partners, and multiplying the net payoff by p, since the Signaler is high quality with probability p, the expected payoff to the Signaler is $p(s/p-c) = s-pc$. Similarly, since a Signaler can have multiple allies, a Partner certainly finds an ally among the high quality members, giving payoff h. A similar argument is used to fill in the other entries in Figure 1.

From the construction of the matrix in Figure 1, we know that a pair of best responses for Signaler and Partner determine a Nash equilibrium of the game. It follows that the conditions $pc' > s > pc$, and $h > l$ are necessary and sufficient for honest signaling (sn,ar) to be a strict Nash equilibrium. The conditions for this signaling equilibrium to exist are easily interpreted. The first, $s > pc$ requires that the benefits of signaling exceed the expected cost of signaling for the high quality type. The second, $s < pc'$, requires exactly the opposite for low quality types. Finally, $h > l$ states that the benefit from allying with a high quality type is greater than with a low quality type.[3]

The reader will note that these equilibrium conditions are *dependent upon the frequency p of high quality types in the group*. This aspect of costly signaling has of course been noted in verbal descriptions of costly signaling, but did not appear in the Grafen's pioneering contribution (Grafen, 1990a,b) or in the more recent papers of Carl Bergstrom and Michael Lachmann (Bergstrom and Lachmann 1997, Lachmann and Bergstrom 1998), although it plays a role in Siller (1998). We have the following theorem.

Theorem 1. *Suppose the cost of signaling is nonnegative for the high quality type ($c \geq 0$), there is a positive payoff to an alliance for the Signaler ($s > 0$), and Partners prefer to ally with high quality types ($h > l$). Then there is a range of frequencies p of high quality types for which there is an honest signaling equilibrium if and only signaling is more costly for low quality than high quality types ($c' > c$), and more costly than the payoff to a single alliance ($c' > s$).*

To prove the theorem, we rewrite the condition $pc' > s > pc$ as

$$\frac{s}{c'} < p < \frac{s}{c}. \tag{1}$$

---

[3] There is also a mixed strategy equilibrium in which

$$\alpha = \alpha^* \equiv \frac{hp+ql}{h}, \qquad \beta = \beta^* \equiv \frac{pc}{s},$$

where α is the probability the agent plays sn and β is the probability the Partner plays ar, provided both α^* and β^* lie between zero and one.. The mixed strategy equilibrium is necessarily unstable in any monotonic dynamic (Selten 1980), so we will not consider it further other than to note that when we consider the dynamics of this model, α^* and β^* will define the boundary of the basin of attraction of the two equilibria.

If (1) holds, then clearly $c' > c$ and since $p < 1$, we must have $s < c'$. Conversely, if these parameter inequalities hold, then (1) holds for some $p < 1$.

Note that this theorem does not require $c > 0$. If $c = 0$, so signaling is not costly for the high quality signaler, there still exists an honest signaling equilibrium for some p, so long as $c' > s$. Indeed, it is easy to see that we can have $c < 0$ and the signaling equilibrium will still exist under the same conditions. This is an important observation because it implies that the cost of signaling for the high quality type cannot be signed, so even signals that are intrinsically beneficial to the Signaler can be part of an honest signaling equilibrium. The nonsignaling equilibrium does not exist in this case, however.

Our model assumes low quality types have higher signaling costs. An alternative assumption that is sometimes more accurate (Johnstone 1995,1997, Getty 1998) is that high quality types do not have low signaling costs, but do have higher benefits from signaling than do low quality types. In terms of our parameters, this means $c = c'$ and $s > s'$, where s' is the value of an alliance to a low quality signaler. If we solve for Nash equilibria using the parameters c, c', s, and s' with $c, c' > 0$, we find that the conditions for an honest signaling equilibrium are simply $s > pc$, and $s' < pc'$. These conditions of course reduce to the above conditions when $s = s'$, and we will not pursue this variant of the model further in this paper.

Among the forms of signaling described by the model is the punishment of those who violate community norms. To see how our model captures this form of signaling, here is a brief account of this mechanism. Suppose that a group of n members can cooperate to provide some group benefit. By cooperating, each member contributes a total benefit of γ_c to others at a fitness cost of δ_c to himself. Thus, the gain from defecting is δ_c and to induce cooperation, members must be punished at least δ_c for defecting. Now suppose that a high quality individual can impose δ_c on defectors at a personal cost of c, whereas a low quality individual must incur cost $c' > c$ to achieve the same effect. Following the model presented above, there will be an equilibrium in which high quality individuals will punish and low quality ones will not, provided certain parameter values obtain.

3 Dynamics

We make this model dynamic by assuming that the fraction of honest Signalers increases when the payoff to honest signaling exceeds that of Never Signal, and the fraction of ar Partners increases when the payoff to ar exceeds that of aa. We model these as a‘replicator dynamic’ (Taylor and Jonker 1978), which can reflect either cultural dynamics, in which members are prone to switch from inferior to superior strategies, or genetic dynamics, in which those who pursue successful strategies have more offspring, who tend to follow their parent's strategies.[4]

When the fraction of sn Signalers is α and the fraction of ar Partners is β, it is easy to calculate that the expected payoffs are

$$\begin{array}{lcll}
\pi_{sn} & = & s - pc & \text{Signal Truthfully} \\
\pi_{nn} & = & (1-\beta)s & \text{Never Signal} \\
\pi_{aa} & = & ph + ql & \text{Always Accept} \\
\pi_{ar} & = & h & \text{Accept Only if Signaler Signals} \\
\overline{\pi}^1 & = & s(1-\beta) + \alpha(s\beta - pc) & \text{Average Signaler} \\
\overline{\pi}^2 & = & (ph+ql)(1-\beta) + \beta h & \text{Average Partner}
\end{array}$$

Note that the equation $\pi_{ar} = h$ assumes that $\alpha > 0$, so that there is at least one signaler, and hence all Partners make successful alliances.

The replicator equations are then

$$\dot{\alpha} = \alpha(\pi_{sn} - \overline{\pi}^1)$$
$$\dot{\beta} = b(\pi_{ar} - \overline{\pi}^2),$$

which reduce to

$$\dot{\alpha} = \alpha(1-\alpha)(\beta s - pc) \tag{2}$$
$$\dot{\beta} = \beta(1-\beta)(h-l)q \qquad (\alpha > 0) \tag{3}$$
$$\dot{\beta} = -\beta(1-\beta)(hp+lq) \qquad (\alpha = 0). \tag{4}$$

These equations express the familiar result that the rate of change of the frequency of a trait in a population varies with the variance of fitness, or equivalently, the variance of the trait times the effect of the trait on fitness. The first says that the rate of increase of honest signaling equals the variance of the frequency of honest signaling, which is $\alpha(1-\alpha)$ times the net gain from Truthful Signaling ($\pi_{sn} - \pi_{nn}$), given by $(\beta s - pc)$. The second says that, assuming $\alpha > 0$, the rate of increase in ar equals the variance of the frequency of ar, which is $\beta(1-\beta)$, times the net gain from ar ($\pi_{ar} - \pi_{aa}$), given by $(h-l)q$.

These replicator equations have four equilibria assuming $c > 0$, corresponding to $\alpha = 0, 1$ and $\beta = 0, 1$. However only one of these equilibria is stable: the honest signaling equilibrium. In particular, the nonsignaling equilibrium is not stable because it is costless for a Partner to play ar as long as there is at least one sr type in the population. Since it is implausible that every possible costly signaling equilibrium is actually realized we will add two additional assumptions, both plausible, that render the nonsignaling equilibrium stable. First, we assume that when a Partner seeks an alliance with a specific member of the group, that alliance

---

[4] For various derivations of the replicator dynamic equations, see Gintis (2000a), Chapter 9.

can fail to materialize with probability $1-\gamma$. We then revise the ar strategy to say that a Partner who fails to form an honest signaling alliance, then seeks an alliance with any member of the group. Second, there is a positive cost ν_c involved in processing the signal and/or differentially seeking alliances only with signalers.

Suppose there are n group members. The probability that an aa type forms an alliance is then

$$1-(1-\gamma)^{n-1}.$$

To simplify notation, we assume that n is sufficiently large that this probability is effectively unity. The probability that an ar type forms an alliance is

$$\delta(\alpha) = 1-(1-\gamma)^{\alpha p(n-1)}.$$

We cannot assume that this is unity, since α is allowed to vary over the unit interval. The payoffs now become

$$\begin{array}{lcll}
\pi_{sn} & = & \beta s/\alpha + (1-\beta)s - pc & \text{Signal Truthfully} \\
\pi_{nn} & = & (1-\beta)s & \text{Never Signal} \\
\pi_{aa} & = & ph + ql & \text{Always Accept} \\
\pi_{ar} & = & \delta(\alpha)h + (1-\delta(\alpha))(ph+ql) & \text{Accept Only if Signaler Signals}
\end{array}$$

The replicator equations are then

$$\dot{\alpha} = \alpha(1-\alpha)(\beta s/\alpha - pc) \tag{5}$$

$$\dot{\beta} = \beta(1-\beta)(\delta(\alpha)(h-l)q - \nu_c) \tag{6}$$

with the proviso that when $\alpha = 0$, $\dot{\alpha} = 0$. These equations have five equilibria assuming $c > 0$. The first four correspond to $\alpha = 0, 1$ and $\beta = 0, 1$, and the fifth sets

$$\alpha_o = \delta^{-1}\left(\frac{\nu_c}{(h-l)q}\right) \tag{7}$$

$$\beta_o = \frac{\alpha_o cp}{s}. \tag{8}$$

The Jacobian matrix of the replicator equations is

$$J = \begin{bmatrix} (2\alpha-1)pc - \beta s & (1-\alpha)s \\ -(h-l)(n-1)pq(1-\beta)\beta(1-\delta(\alpha)\log(1-\gamma)) & (1-2\beta)(h-l)q\delta(\alpha) + \nu_c \end{bmatrix}.$$

Checking the eigenvalues at the equilibria, we find that the equilibria given by $(\alpha, \beta) = (0, 1)$ and $(\alpha, \beta) = (1, 0)$, are unstable, while the honest signaling and nonsignaling equilibria are stable. The mixed strategy equilibrium (α_o, β_o) is also unstable, and hence separates the basin of attraction of the stable signaling and

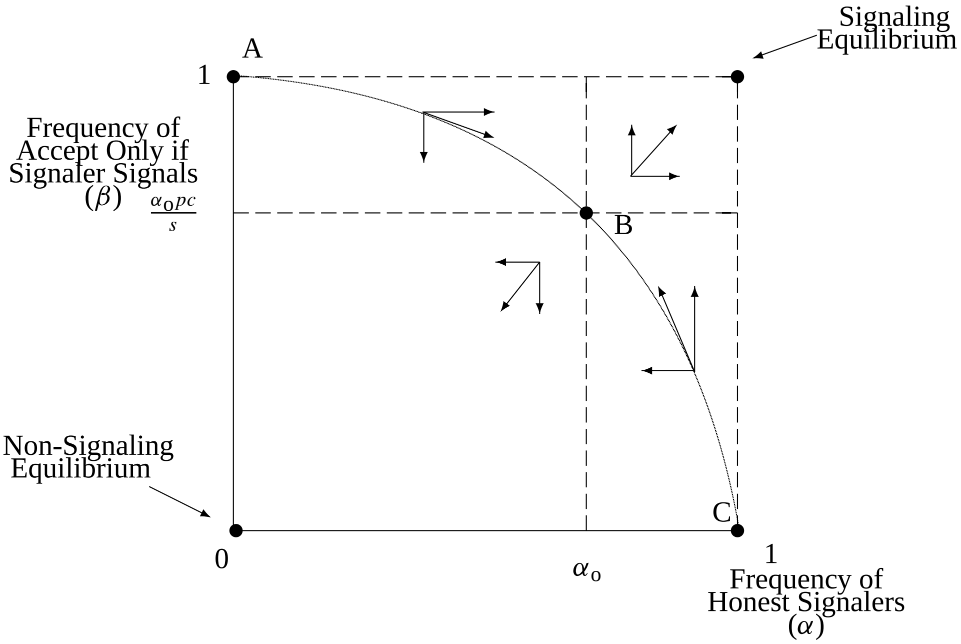

Figure 2: The Phase Diagram. Points below the ridge line ABC are in the basin of attraction of the Nonsignaling equilibrium.

the nonsignaling equilibria. Figure 2 shows the phase diagram for the dynamical system.

Note that the lower the ratio $q(h-l)/v_c$, the larger the basin of attraction of the signalling equilibrium, reflecting the higher signaling costs (v_c), the fact that very high frequencies of high quality types make signaling redundant, since one is then very likely to meet a high quality type by chance, and the fact that the closer l is to h, the more nearly equally valuable as partners are the two types. As expected, higher signaling costs and lower alliance benefits also reduce the size of the basin of attraction of the honest signaling equilibrium.

4 Why Signal by Providing Benefits to Others?

We have shown that costly signaling, in the form of providing benefits to others, can solve the problem of maintaining unilateral cooperation in a group when, in the absence of other mechanisms, self-interested agents would not provide the benefits. But we have not explained why the signal should take the form of benefits to others. Indeed, in the model presented in Section 2 the existence of a costly signaling equilibrium does not depend upon the character or quality of the benefit provided to the group. To see this, note that the per member benefit g could equally well be positive, negative, or zero without altering the conditions for existence of stability of a costly signaling equilibrium. Moreover, when $ph + ql > 0$, whenever there is a stable costly signaling equilibrium, there is another stable equilibrium in which that signal is not used. Therefore, when there are a variety of possible signals, our analysis to this point does not tell us which among them will in fact be used.

Let us call a costly signal *feasible* if it satisfies the conditions for a costly signaling equilibrium developed in Section 2. We then face the question: is there any reason to expect feasible costly signals deployed within a group to be prosocial? While we do not formally model the process of equilibrium selection where many signals are feasible, we can offer three reasons why prosocial signalling may be favored. First, signals that are associated with positive benefits to the group may indicate in the Signaler some personal attributes that enhance the Signaler's value to a potential ally. That is, the relevant traits being signaled may go beyond attributes such as "vigor" and include prosociality as an intrinsic aspect of quality. For instance an agent who punishes wrong-doers within the group may also be more likely to punish enemies of the political alliances of which he is part. Similarly, an individual's benefit from a team effort depends both on the productivity of one's partners and on the share of the returns to the project, so a relevant quality of a partner is willingness and ability to contribute to group productivity and share the proceeds of group effort. For example, an agent who shares meat rather than conspicuously consuming it personally may be more likely to share the burdens of family and politics. In both cases, a high quality agent is more likely to provide the social benefit because the cost of doing so is lower than for a low quality agent. This argument is a variant of, or directly analogous to, the "direct benefits," or "good parent," explanation for female preference of males who signal superior ability to provide parental care or other resources (Johnstone 1995, Iwasa and Pomiankowski 1999).

Formally, suppose there are several signals $\{\sigma_1, \ldots, \sigma_k\}$ that produce benefits $\{g_1, \ldots, g_k\}$ and corresponding alliance payoffs $\{h_1, \ldots, h_k\}$ to Partners. Clearly Partner strategies that form alliances with those emitting the highest h-value signal, say h_1 will have higher payoffs and will proliferate. Thus Signalers who use σ_1 will eventually receive all the alliance benefits, and no other σ_i will be used. The question then becomes: why should h_1 be associated with a positive benefit to the group; i.e., why should $g_1 > 0$? One plausible answer is that $g_1 > 0$ might signal the Signaler's willingness to cooperate in social relationships, and to act in a prosocial manner in bargaining over the distribution of the benefits of alliances, not only with the group as a whole ($g_1 > 0$), but with smaller subgroups as well ($h_1 = \max\{h_1, \ldots, h_k\}$). The quality being signaled in this case might be anything that lowers the cost of behaving in a prosocial manner, such as greater foraging ability or wealth, and hence a lower cost of sharing with others. Alternatively, past or present generosity or cooperation might be an honest signal of intent to behave similarly in the future if a contemplated alliance imposes a cost that would not be worth paying if the signaler did not intend to cooperate in the future. In this case the signal then indicates the capacity and willingness to honor commitments (Schelling 1978, Zahavi and Zahavi 1997).

Second, signaling by providing a benefit to group members may increase "broadcast efficiency" (Bliege Bird 1999, Smith and Bliege Bird 2000), in the sense that it attracts a larger audience to witness the signal. We initially assumed that all members of the group "see" the signal. But suppose there are several signals $\{\sigma_1, \ldots, \sigma_k\}$ each of which is observed by a fraction $\{v_1, \ldots, v_k\}$ of the group. Then the payoffs to the Signaler using σ_i will be reduced by $(1 - v)s$ in a signaling equilibrium, because the Signaler now receives an expected benefit only from the $v(n - 1)$ members who "viewed" the signal. Therefore, everything else being equal, more prosocial Signaler strategies will gain higher payoffs and will increase their share in the population. This "audience effect" might, of course, be more likely to obtain when social groups are relatively aggregated rather than dispersed, relatively stable in membership (allowing reputation effects), and in ecological situations where cooperative behavior enhances fitness (e.g., where there are limited opportunities for individually harvesting large game, or where there is chronic inter-group aggression).

To summarize these two reasons, group beneficial signalling may evolve because the group benefit, g covaries with either h, the quality of the signaler, or with v, broadcast efficiency. Thus while a variety of signals may be feasible if considered singly, only prosocial signals can constitute a signaling equilibrium where there is a choice of signals. This is because where g and h covary, it cannot be a best response for a Partner to respond to any but the most prosocial signal, and where g and v covary it cannot be a best response for the Signaler to emit any but the most prosocial signal.

A third plausible process favoring the selection of prosocial costly signaling, though not modeled in this paper, is that groups with a high level of prosocial costly signaling will have members who, on average, are more fit than groups in which such behavior is absent. Such groups, by withstanding extinction and dispersion, and by having superior strength in hostile interactions with other groups, can spread the prosocial practices beyond their original boundaries (Gintis 2000b). Our model provides a possible basis for a more general understanding of which among the multiplicity of signals are likely to evolve and persist, namely, those characterized by a large basin of attraction for the associated equilibria.

5 The Evolution of Signaling

We turn finally to the question: could signaling—both sending and conditioning behavior on the signal—proliferate in a population if initially rare? We may answer in two ways. First a population at the nonsignaling equilibrium might be displaced by a series of stochastic events into the basin of attraction of the signaling equilibrium, should the underlying parameters be such that the honest signaling equilibrium exists. But the model reveals a second path to a signaling equilibrium. Suppose that periodically, say for reasons of adverse climate variation, h and/or l is reduced and the expected value of alliances falls, so $ph + ql < 0$. It is then only profitable to ally with an agent who reveals high quality through signaling.

If $pc' > s > pc$, there are now two pure strategy equilibria, the honest signaling equilibrium (sn,ar) and the nonsignaling equilibrium (nn,rr), where no alliances are made. But now the honest signaling equilibrium is the only possibly stable equilibrium. This is because Partners are indifferent between rr and ar, so stochastic drift will render the fraction of Partners using ar positive, in which case nn is no longer a best response, so all high quality types will shift to sn, and the system will evolve to the honest signaling equilibrium. When conditions change rendering $ph + ql > 0$ once more, there will be no tendency to abandon this equilibrium.

Thus it is possible for the signaling equilibrium to become generalized in the population by each group independently switching from a nonsignaling to a signaling equilibrium. Moreover, for those cases where the net payoff to signaling—the numerator in equation (9)—is positive, intergroup competition may contribute to the evolutionary success of signaling. Once having proliferated within a single group or a few groups, signaling equilibria may proliferate in a larger population through multilevel selection. To see this, suppose that every member of the population is both a Partner and a Signaler in each period. Then average payoffs for the two stable equilibria are just the sum of the entries in the reduced normal form matrix for the relevant strategy profile, plus the gains to all parties from the provision of Signaler benefits. Thus the honest signaling equilibrium has higher average payoffs if $h + (n-1)pg + p(s-c) > hp + ql$, which reduces to

$$p\left(1 - \frac{(n-1)g + s - c}{h - l}\right) < 1 \tag{9}$$

indicating that the honest signaling equilibrium will have higher payoffs than the Nonsignaling equilibrium when, holding all other parameters of the model fixed, (a) high quality types are sufficiently rare (p is small); (b) the gain g from the Signaler's provision of benefits is sufficiently large; (c) the advantage of allying with high quality types, $h - l$, is sufficiently large; and (d) the cost c of signaling is sufficiently small. It may seem counterintuitive that the Nonsignaling equilibrium could ever support higher payoffs, but this is a direct result of the fact that signaling is a costly activity, and when high quality types are sufficiently common, signaling is unnecessary to gain a large fraction of high quality alliances.

We have shown that signaling by providing benefits to members could proliferate when rare, and would under plausible conditions be sustained in a population in which behaviors evolved according to the payoff-monotonic updating described in equations (2), (3), and (4). But this is not sufficient to ensure the evolutionary success of such signaling. Recall that the existence of the signaling equilibrium requires that the high quality types not be excessively prevalent in the population. But if high quality types have higher fitness than low quality types in the signaling equilibrium their frequency p may increase over time, thus undermining the signaling equilibrium.

The payoff difference between high and low quality types in the signaling equilibrium is $s/p - c > 0$ so in the absence of any other influence on p, high quality would evolve to fixation. But we have modeled only a subset of the influences on p, and we may suppose other influences to be at work. The relevant differential equation is

$$\dot{p} = pq(s/p - c) - zp + wq \tag{10}$$

where the first term expresses the rate at which the differential fitness of high quality types is translated into offspring, $z > 0$ is the proportion of offspring produced by high quality parents that are low quality, and $w > 0$ is the proportion of produced by low quality parents that are high quality. Equation (10) can be simplified to

$$\dot{p} = cp^2 - (c + s + w + z)p + s + w. \tag{11}$$

Since the left hand side of (11) is $s + w > 0$ at $p = 0$ and $-z < 0$ at $p = 1$, there is surely a stable equilibrium p^*, where $0 < p^* < 1$. We have

$$p^* = \frac{c + s + w + z - \sqrt{(c + s + w + z)^2 - 4c(s + w)}}{2c}.$$

This expression is complicated, but it does have some intuitive implications. Setting the right hand side of (10) to zero and totally differentiating to see how p^* varies in response to changes in our parameters, we find that (a) an increase in the cost c to the high quality type of signaling leads to a lower equilibrium value of p^*; (b) an increase in the benefit s of an alliance to the Signaler raises the equilibrium value of p^*; (c) an increase in rate w at which low quality agents produce high quality offspring raises the equilibrium value of p^*; and an increase in rate z at which high quality agents produce low quality offspring lowers the equilibrium value of p^*.

There are of course various social and biological mechanisms that could lead to positive levels for either z or w. For example, the characteristics that confer high quality or low quality could be purely or substantially environmentally determined, or determined by cultural factors that are passed on through biased or horizontal rather than vertical cultural transmission (Cavalli-Sforza and Feldman 1981, Boyd and Richerson 1985). Or, the fitness benefits associated with prosocial signaling of high quality may be confined to some period of the life cycle, and be partially offset by disadvantageous effects at other ages.

6 Conclusion

Costly signaling of quality for purposes of mating, alliance formation, and warning off potential enemies has been reported in many species, including humans (Zahavi 1977a, Grafen 1990a, Maynard Smith 1991, Johnstone 1995, Wright 1999). We have proposed a multiplayer game-theoretic model of costly signaling and shown that under plausible parameter values, a class of signals that themselves contribute to group benefits may proliferate in a population when rare, and constitute evolutionary stable strategies once established. Costly signaling may thus provide a mechanism for the evolution of cooperative and other group beneficial practices capable of working independently of repeated interactions, positive assortment, and multilevel selection, though these latter factors may act to reinforce such evolution.

REFERENCES

Axelrod, Robert and William D. Hamilton, "The Evolution of Cooperation," *Science* 211 (1981):1390–1396.

Bergstrom, Carl T. and Michael Lachmann, "Signalling among Relatives. I. Is Signalling *Too* Costly?," *Phil. Trans. R. Soc. Lond. B* 352 (1997):609–617.

Bliege Bird, Rebecca, "Cooperation and Conflict: the Behavioral Ecology of the Sexual Division of Labor," *Evolutionary Anthropology* 8 (1999):65–75.

Bliege Bird, Rebecca L. and Douglas W. Bird, "Delayed Reciprocity and Tolerated Theft," *Current Anthropology* 38 (1997):49–78.

— , Eric A. Smith, and Douglas W. Bird, "The Hunting Handicap: Costly Signaling in Human Foraging Strategies," *Behavioral Ecology and Sociobiology (*in press).

Blurton Jones, Nicholas G., "Fitness Returns from Resources and the Outcome of Contests: Some Implications for Primatology and Anthropology," in J. G. Else and P. Lee (eds.) *Primate Ontogeny, Cognition and Social Behaviour: Proceedings of the 10th Congress of the International Primatology Society* (Cambridge: Cambridge University Press, 1986) pp. 393–406.

Boehm, Christopher, "Segmentary warfare and Management of Conflict: a Comparison of East African Chimpanzees and Patrilineal-patrilocal Humans," in A. Harcourt and Frans de Waal (eds.) *Coalitions and Alliances in Humans and Other Animals* (Oxford: Oxford University Press, 1992) pp. 137–173.

Boone, James L., "The Evolution of Magnanimity: When is it Better to Give than to Receive?," *Human Nature* 9 (1998):1–21.

Bowles, Samuel, "Individual Interactions, Group Conflicts, and the Evolution of Preferences," in Steven Durlauf and Peyton Young (eds.) *Social Dynamics* (Cambridge, MA: MIT Press, 2000) pp. ??–??

Boyd, Robert and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985).

— and — , "The Evolution of Cooperation," *Journal of Theoretical Biology* 132 (1988):337–356.

— and — , "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizeable Groups," *Ethology and Sociobiology* 113 (1992):171–195.

Brown, Charles R., Mary Bomberger Brown, and Martin L. Shaffer, "Food-sharing Signals Among Socially Foraging Cliff Swallows," *Animal Behavior* 42 (1991):551–565.

Brown, Jerram L., "Cooperation—A Biologist's Dilemma," *Advances in the Study of Behavior* 13 (1983):1–37.

Cashdan, Elizabeth A., "Coping with Risk: Reciprocity among the Basarwa of Northern Botswana," *Man* 20 (1985):454–474.

Cavalli-Sforza, Luigi L. and Marcus W. Feldman, *Cultural Transmission and Evolution* (Princeton, NJ: Princeton University Press, 1981).

Clutton-Brock, T. H. and G. A. Parker, "Punishment in Animal Societies," *Nature* 373 (1995):209–216.

Connor, Richard C., "The Benefits of Mutualism: a Conceptual Framework," *Biological Reviews* 70 (1995):427–457.

Dittus, W. P. J., "Toque Macaque Food Calls: Semantic Communication Concerning Food Distribution in the Environment," *Animal Behaviour* 32 (1984):470–477.

Dugatkin, Lee Alan, *Cooperation among Animals* (New York: Oxford University Press, 1997).

Frank, Steven A, "Mutual Policing and Repression of Competition in the Evolution of Cooperative Groups," *Nature* 377 (October 1995):520–522.

Getty, Thomas, "Handicap Signaling: When Fecundity and Viability do not Add Up," *Animal Behavior* 56 (1998):127–130.

Gintis, Herbert, *Game Theory Evolving* (Princeton, NJ: Princeton University Press, 2000).

— , "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology* 206 (2000):169–179.

Grafen, Alan, "Biological Signals as Handicaps," *Journal of Theoretical Biology* 144 (1990):517–546.

— , "Sexual Selection Unhandicapped by the Fisher Process," *Journal of Theoretical Biology* 144 (1990):473–516.

Gurven, Michael, Wesley Allen-Arave, Kim Hill, and A. Magdalena Hurtado, "'It's a Wonderful Life': Signaling Generosity Among the Ache of Paraguay," *Evolution and Human Behavior* 21 (2000):263–282.

Hardin, Russell, *Collective Action* (Baltimore: Johns Hopkins University Press, 1982).

Harsanyi, John C., "Games with Incomplete Information Played by Bayesian Players, Parts I, II, and III," *Behavioral Science* 14 (1967):159–182, 320–334, 486–502.

Hauser, Marc D. and Peter Marler, "Food-associated Calls in Rhesus Macaques (*Macaca mulatta*): II. Costs and Benefits of Call Production Suppression," *Behavioral Ecology* 4 (1993):206–212.

Hawkes, Kristen, "Sharing and Collective Action," in E. A. Smith and B. Winterhalder (eds.) *Evolutionary Ecology and Human Behavior* (New York: Aldine de Gruyter, 1992) pp. 269–300.

— , "Why Hunter-Gatherers Work: An Ancient Version of the Problem of Public Goods," *Current Anthropology* 34,4 (1993):341–361.

Heinrich, Bernd and John Marzluff, "Why Ravens Share," *American Scientist* 83 (1995):342–349.

Iwasa, Yoh and Andrew Pomiankowski, "Good Parent and Good Genes Models of Handicap Evolution," *Journal of Theoretical Biology* 200 (1999):97–109.

Johnstone, Rufus A., "Sexual Selection, Honest Advertisement and the Handicap Principle," *Biological Reviews* 70 (1995):1–65.

— , "The Evolution of Animal Signals," in John R. Krebs and Nicholas B. Davies (eds.) *Behavioral Ecology: An Evolutionary Approach (4th ed.)* (Oxford: Blackwell, 1997) pp. 155–178.

Kaplan, Hillard and Kim Hill, "Food Sharing among Ache Foragers: Tests of Explanatory Hypotheses," *Current Anthropology* 26,2 (1985):223–246.

— and — , "Hunting Ability and Reproductive Success among Male Ache Foragers: Preliminary Results," *Current Anthropology* 26,1 (1985):131–133.

— , — , and Magdalena Hurtado, "Fitness, Foraging and Food Sharing Among the Ache," in E. Cashdan (ed.) *Risk and Uncertainty in Tribal and Peasant Economies* (Boulder, CO: Westview Press, 1990) pp. 107–144.

Lachmann, Michael and Carl T. Bergstrom, "Signalling among Relatives, II. Beyond the Tower of Babel," *Theoretical Population Biology* 54 (1998):146–160.

Marlowe, Frank, “The Patriarch Hypothesis: an Alternative Explanation of Menopause,” *Human Nature* 11,1 (2000):27–42.

Maynard Smith, John, “Honest Signalling: The Philip Sidney Game,” *Animal Behavior* 42 (1991):1034–1035.

Michod, Richard E., “Cooperation and Conflict in the Evolution of Individuality. 1. The Multilevel Selection of the Organism,” *American Naturalist* 149 (April 1997):607–645.

Pusey, Anne E. and Craig Packer, “The Ecology of Relationships,” in John R. Krebs and Nicholas B. Davies (eds.) *Behavioral Ecology: An Evolutionary Approach*, 4th ed. (Oxford: Blackwell Science, 1997) pp. 254–283.

Roberts, Gilbert, “Competitive Altruism: From Reciprocity to the Handicap Principle,” *Proceedings of the Royal Society of London B* 265 (1998):427–431.

Schelling, Thomas C., *Choice and Consequence* (Cambridge, MA: Harvard University Press, 1978).

Selten, Reinhard, “A Note on Evolutionarily Stable Strategies in Asymmetric Animal Conflicts,” *Journal of Theoretical Biology* 84 (1980):93–101.

Siller, Steven, “The Epistatic Handicap Principle Does Work,” *Journal of Theoretical Biology* 191 (1998):141–161.

Smith, Eric Alden, “Risk and Uncertainty in the ‘Original Affluent Society’: Evolutionary Ecology of Resource Sharing and Land Tenure,” in T. Ingold, D. Riches, and J. Woodburn (eds.) *Hunters and Gatherers: History, Evolution, and Social Change* (Oxford: Berg, 1988) pp. 222–252.

— and Rebecca L. Bliege Bird, “Turtle Hunting and Tombstone Opening: Public Generosity as Costly Signaling,” *Evolution and Human Behavior* 21,4 (2000):245–261.

— and Robert Boyd, “Risk and Reciprocity: Hunter-gatherer Socioecology and the Problem of Collective Action,” in Elisabeth Cashdan (ed.) *Risk and Uncertainty in Tribal and Peasant Economies* (Boulder, CO: Westview Press, 1990) pp. 167–191.

Sober, Elliot and David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge, MA: Harvard University Press, 1998).

Sosis, Richard, “Costly Signaling and Torch Fishing on Ifaluk Atoll,” *Evolution and Human Behavior* 21 (2000).

Spence, A. Michael, “Job Market Signaling,” *Quarterly Journal of Economics* 90 (1973):225–243.

Stanford, Craig B., *The Hunting Apes* (Princeton, NJ: Princeton University Press, 1999).

— , J. Wallis, E. Mpongo, and J. Goodall, "Hunting Decisions in Wild Chimpanzees," *Behavior* 131 (1996):1–20.

Taylor, P. and L. Jonker, "Evolutionarily Stable Strategies and Game Dynamics," *Mathematical Biosciences* 40 (1978):145–156.

Trivers, R. L., "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46 (1971):35–57.

Wilson, David Sloan, "Structure Demes and the Evolution of Group-advantageous Traits," *American Naturalist* 111 (1977):157–185.

Winterhalder, Bruce, "Social Foraging and the Behavioral Ecology of Intragroup Resource Transfers," *Evolutionary Anthropology* 5,2 (1996):46–57.

Wright, Jonathan, "Altruism as Signal: Zahavi's Alternative to Kin Selection and Reciprocity," *Journal of Avian Biology* 30 (1999):108–115.

Zahavi, Amotz, "Mate Selection—A Selection for Handicap," *Journal of Theoretical Biology* 53 (1975):205–214.

— , "The Cost of Honesty (Further Remarks on the Handicap Principle)," *journal of Theoretical Biology* 67 (1977):603–605.

— , "Reliability in Communication Systems and the Evolution of Altruism," in B. Stonehouse and C. M. Perrins (eds.) *Evolutionary Ecology* (London: Macmillan Press, 1977) pp. 253–259.

— , "Arabian Babblers: the Quest for Social Status in a Cooperative Breeder," in P. B. Stacey and W. D. Koenig (eds.) *Cooperative Breeding in Birds: Long-term Studies of Ecology and Behavior* (Cambridge: Cambridge University Press, 1990) pp. 103–130.

— , "Altruism as Handicap: the Limitations of Kin Selection and Reciprocity," *Journal of Avian Biology* 26 (1995):1–3.

— and Avishay Zahavi, *The Handicap Principle: A Missing Piece of Darwin's Puzzle* (New York: Oxford University Press, 1997).

e\Papers\Eric Smith Paper\Costly Signaling and Reciprocity November 14, 2000