Disclosure and Information Transfer in Signaling Games

Justin P. Bruner*†

One of the major puzzles in evolutionary theory is how communication and information transfer are possible when the interests of those involved conflict. Perfect information transfer seems inevitable if there are physical constraints, which limit the signal repertoire of an individual, effectively making bluffing an impossibility. This, I argue, is incorrect. Unfakeable signals by no means guarantee information transfer. I demonstrate the existence of a so-called pooling equilibrium and discuss why the traditional argument for perfect information transfer (Frank's full disclosure principle) does not hold in all cases. Additionally, I demonstrate that deception is possible at equilibrium despite the fact that signals are impossible to fake.

1. Introduction. One of the major puzzles in evolutionary theory is how communication and information transfer are possible when the interests of those involved conflict. If the sender has incentive to deceive, then the receiver often does best to simply ignore her counterpart and communication breaks down. Recently, however, a variety of theorists have argued there are two basic mechanisms that can sustain high levels of information transfer (see Bradbury and Vehrencamp 1998; Maynard Smith and Harper 2003; Hurd and Enquist 2005, for a recent overview of the field). Simply put, communication is possible if there are substantial costs associated with

Received December 2014; revised March 2015.

*To contact the author, please write to: Australian National University, Canberra, Acton ACT 2602, Australia; e-mail: justin.bruner@anu.edu.au.

†Thanks to Kim Sterelny, Brian Skyrms, Simon Huttegger, Hannah Rubin, Luke Holman, Ben Fraser, and Chris Hunter-Lean for feedback and encouragement. I would also like to thank the participants of the New Zealand Association for Philosophy 2014 in Christchurch, the participants of the Social Dynamics Seminar at University of California, Irvine, and two anonymous referees. I am especially grateful for the comments of an anonymous referee who helped me better understand the relationship between indices and the notion of certification in economics.

Philosophy of Science, 82 (October 2015) pp. 649–666. 0031-8248/2015/8204-0006$10.00
Copyright 2015 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

sending a signal or, alternatively, if signals themselves are difficult, if not impossible, to fake. This first proposal, often referred to as costly signaling theory or the handicap principle, is viable because deception is often prohibitively costly when signals are differentially costly (see Zahavi 1975; Grafen 1990; see also Spence 1973, for an early model, albeit in economics). Likewise, perfect information transfer seems inevitable if there are physical constraints that limit the signal repertoire of an individual, effectively making bluffing an impossibility (Maynard Smith 1982; Frank 1988). While the handicap principle has been critically discussed and debated for decades now (see Grose 2011, for a good historical overview), little attention has been paid to unfakeable signals (also known as indices). This is in part because unlike the handicap principle, which has been called into question on both theoretical and empirical grounds (see Zollman 2013, for a good overview), there is a broad consensus that information transfer is guaranteed when signals cannot be faked. This, I will argue, is incorrect. Indices by no means guarantee information transfer. In this article, I demonstrate the existence of a so-called pooling equilibrium in which no information is transferred from sender to receiver. Furthermore, I show that under many conditions this uninformative equilibrium is the most likely outcome of an evolutionary model.

This article proceeds as follows. In section 2, I present the nature of indices and the traditional argument given for why unfakeable signals guarantee information transfer: the full disclosure principle. Section 3 is an overview of important concepts from game theory. In section 4, I construct a simple action-response game involving unfakeable signals, prove the existence of an uninformative pooling equilibrium, and register the evolutionary significance of this equilibrium. A continuous version of this model is briefly explored in section 6. Section 5 discusses the possibility of deception at equilibrium despite the fact that signals are impossible to fake. Section 7 concludes.

2. Signals, Costly Signals, and Indices. Unobservable or nonobvious qualities of individuals tend to play a key role in many strategic settings. As a result, organisms often have reason to deceive, exaggerate, or withhold information about these qualities from others. A stalked antelope may wish to exaggerate its evasiveness to a predator. Potential mates may have reason to overstate their health and underlying genetic quality. In these and other conflict-of-interest cases, signaling is the norm. But what ensures that signals are reliable? Recent work in signaling theory indicates that there are two ways to ensure honesty when the interests of sender and receiver do not completely overlap.[1] As mentioned above, the handicap

1. These two ways are not exhaustive. Other proposed mechanisms such as individual directed skepticism and repeated interactions can sustain high levels of honesty, al-

https://doi.org/10.1086/683016 Published online by Cambridge University Press

principle states that signals must be in some sense costly to produce.[2] In particular, if signals are more costly for low types, only high types will have incentive to send a signal, enabling the receiver to reliably differentiate high- from low-type senders.[3]

Alternatively, signals could be kept honest because signals are often difficult, if not impossible, to fake. Honesty is due to the fact that the signal is in some way causally related to the feature being signaled. Put a slightly different way, a signal is in some sense forced to be reliable "because of physiological or anatomical constraints on signal production" (Searcy and Nowicki 2005, 215). These so-called indices play a vital role in many contexts and are, for example, frequently used between potential mates as a means of indicating condition or health (Taylor, Hasson, and Clark 2000). Consider the often loud and deep croaks made by toads during mating season. Females prefer large toads, and experimental work indicates that larger toads are capable of producing deeper croaks due to a physiological constraint (Davies and Halliday 1978). Namely, large toads typically have a larger vocal apparatus, and it is the size of this apparatus that determines how deep of a croak the toad is capable of producing. Hence, croaks can be naturally interpreted as a reliable signal of size in light of these physiological constraints that restrict the types of calls a toad can muster.[4]

Yet, a natural complication arises that must be addressed. As Maynard Smith and Harper bluntly put it, if an index provides "accurate information about quality, why should a low-quality individual provide [this] information?" (2003, 45). Having a foolproof means of honestly conveying one's quality does not immediately entail that all have incentive to reveal their quality. Why would a small toad ever decide to emit its high-pitch croak?

---

though it is sensible to think these are just versions of costly signaling theory (Fraser 2012). Additionally, if there is not too great of a conflict of interest between sender and receiver, communication is also possible (Crawford and Sobel 1982).

2. This is a slight mischaracterization of the handicap principle. Cost is not necessary to ensure honesty. Instead, the marginal cost associated with deviating from one's "honest" signaling behavior ensures signal reliability.

3. As is common in the literature, we assume individuals are one of a number of different quality types. Depending on the context, quality can refer to fighting ability, evasiveness, or genetic quality. We also use the Kullback-Leibler entropy notion of information provided by Skyrms (2010) to discuss information transfer in signaling systems. Conditioning on a signal changes the probability of states (or in this case types).

4. There is a natural way of trying to fold indices into costly signaling theory (see Searcy and Nowicki 2005). Specifically, unfakeable signals can be seen as being minimally costly for those able to signal and infinitely costly to all others. There are also steep developmental costs associated with assembling the physical mechanisms used in the production of signals. In development, resources can be directed toward increasing the size of the vocal apparatus, allowing the toad a deeper croak than toads not employing this developmental gambit. Yet the reallocation of resources during development has real costs that can easily nullify any benefits the amphibian accrues from his low bellows.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

Unfortunately, Maynard Smith and Harper do not provide a compelling reason for receivers to behave in such a fashion.[5]

Frank (1988), in a discussion of his so-called full disclosure principle, picks up the slack.[6] Briefly, the full disclosure principle states that if signals cannot be faked, then when "some individuals benefit by revealing a favorable value of some trait, others will be forced to disclose their less favorable values" (104). This leads to a dynamic (described below) that results in all choosing to reveal their underlying quality. If signals are reliable due to physiological constraints, then perfect information transfer is unavoidable.

The logic behind the full disclosure principle is simple enough. Begin by considering a population of senders (toads), all of whom do not signal (i.e., croak). Is such an arrangement stable? No, argues Frank, because the largest toads will choose to disclose their type. Disclosure is in their interest because they are larger than the population average and thus prefer to distinguish themselves than be lumped together with the rest of the nondisclosers.[7] Yet once the largest toads select to signal, the second-largest toads have, for similar reasons, incentive to disclose their quality as well. Disclosure is again in their interest because they prefer correctly being identified by the receiver than being treated like an average member of the subpopulation that does not signal. This "unravelling process" continues until all but the smallest individuals disclose their type. Thus, receivers learn to interpret silence as an admission that one is of the "worst" type, and all but the worst types disclose their quality.

Economists have discussed similar strategic scenarios. George Akerlof, in his famous 'market for lemons' paper (1970), considers a marketplace in which multiple sellers and consumers interact. A merchant has some good on offer, but potential buyers cannot determine the underlying quality of the product. In the absence of a third-party tasked to certify the quality of the good, Akerlof shows that the market 'collapses'—sellers of high-quality merchandise exit the market, which in turn spurs merchants of slightly in-

5. Additionally, Maynard Smith and Harper question whether a formal model of indices can even be developed. Somewhat perplexingly, they mention that it is "hard to decide what assumption to incorporate in such a model. In particular, could [emitting] a signal be conditional on quality?" (2003, 47). As we see in sec. 4, it is possible to provide a helpful game-theoretic model of indices that takes many of their concerns into account.

6. While Robert Frank's full disclosure principle is widely known, the principle was anticipated by a number of different authors in both economics and biology. Dawkins and Krebs (1984) discuss why indices would lead to full information transfer. Before this, Grossman (1981) and Milgrom (1981) argue that full disclosure is unavoidable when agents have impossible-to-fake signals at their disposal.

7. This argument assumes that receivers will treat a nondiscloser as a typical or average nondiscloser. Frank's results hold (as long as there are no costs associated with signaling) even when this assumption is relaxed, as we see in sec. 4.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

ferior goods to leave, and so on. Yet, this so-called death spiral is avoidable. Grossman (1981) and Milgrom (1981) argue that if merchants have a cost-free means of certifying the quality of their goods, then all will eventually choose to disclose (through certification) their private information, and transactions will abound. Certification results in a similar unraveling process to that outlined by Frank above. This is somewhat unsurprising, because, put in the language of this article, certification provides merchants with impossible-to-fake signals (assuming the certification process is highly reliable): only merchants of high-quality goods will be able to certify that their good is, in fact, high quality. Unfortunately, this tight link between certification and indices has been by and large overlooked in both the biology and economics literature. In section 6 we explore a model that resembles some formal work in economics on the topic of certification.[8]

3. Signaling Games and Equilibria. Here we present the basic game-theoretic framework employed in sections 4 and 5. Consider a simple action-response game (see Lewis 1969; Hurd 1995, for canonical examples). This game consists of two individuals, a sender and a receiver. With some fixed probability, p, the sender is a high type, and with probability $1 - p$, the sender is a low type. The sender has the option of sending a signal to the receiver, and initially only the sender knows its type. Moreover, the sender can condition her behavior on type; that is, she can choose to only send the signal when a high type. The receiver observes only whether the signal is sent and not the underlying type of her counterpart. The receiver then must guess the type of the sender. The receiver can condition her guess on the receipt or nonreceipt of the signal, allowing for strategies such as "guess counterpart is a high type if signal, low type if otherwise." Since there are only two types (high and low) and two actions the sender can perform (send signal or not), there are a total of four sender strategies and four receiver strategies.

If the interests of sender and receiver completely overlap, this strategic situation is essentially a David Lewis signaling game. This game (displayed in fig. 1) was famously introduced by Lewis in part to respond to Quinean skepticism regarding the conventionality of meaning (Lewis 1969). In recent years, it has been the focal point for philosophers, social scientists, and biologists interested in the evolution of language, among other topics. In this game, receivers endeavor to correctly identify the type of the sender, and senders desire to be correctly identified by the receiver. There exist two stable signaling systems that enable the receiver to reliably determine the type of the sender. Senders can, for example, choose to only send the signal when a high type. Receivers will then identify their counterpart as a high type just in case they receive a signal. This separating equilibrium allows for full

8. I thank an anonymous referee for helping me clearly see these connections.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

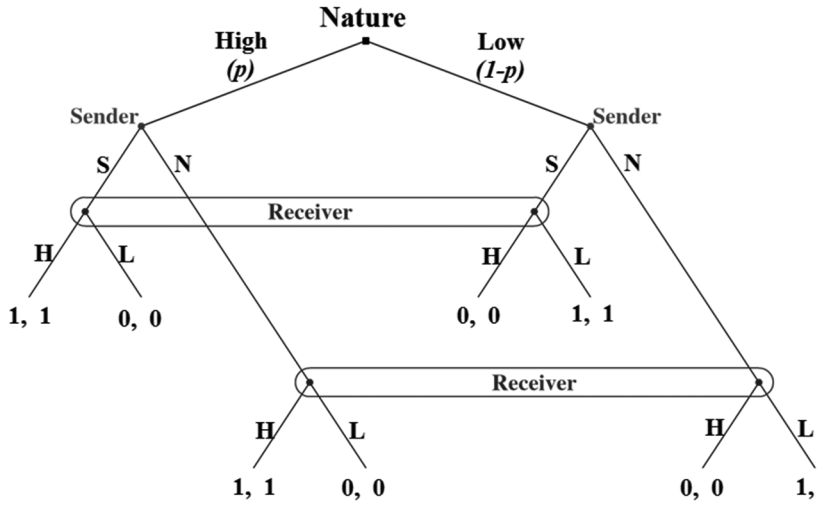


Figure 1. Extensive form of a David Lewis signaling game. Senders can either signal (S) or not signal (N), and receivers must attempt to determine whether the sender is of low or high quality. Nodes represent choices in the game. Nodes in the same oval are said to be in the same 'information set', meaning the agent cannot distinguish between these nodes when making a decision. Color version available as an online enhancement.

information transfer and is stable because neither sender nor receiver benefit from unilaterally changing their strategy. Early studies of the Lewis signaling game suggested that this separating equilibrium was the unique evolutionary outcome (Skyrms 1996), implying that perfect information transfer is guaranteed when common interest prevails. Yet this is too quick. There also exist uninformative equilibria or "pooling equilibria," in which, for example, either the sender sends the signal regardless of type or the receiver makes the same guess regardless of whether a signal was sent. Recent work in evolutionary game theory has shown that the pooling equilibrium is a real evolutionary possibility (see Huttegger 2007; Pawlowitsch 2008, for theoretical results; Bruner et al. 2014, for experimental results). Specifically, when nature is biased ($p \neq .5$), then an adaptive process such as the replicator dynamics (to be explained in sec. 4) will often result in the less desirable pooling equilibrium. Moreover, the pooling equilibrium is all the more likely as the strength of the bias increases. This is due to the fact that when nature is severely biased, the receiver often does best to simply treat her interactive partner as the more common type of sender, regardless of the signal she receives. In section 4, we observe that similar pooling outcomes are possible even when signals are impossible to fake.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

4. Indices and Information Transfer. Few formal models of indices exist in biology. Huttegger, Bruner, and Zollman (2015) explore a model in which signals are potentially costly and vary with respect to how difficult they are to send. To spell this out a bit more, both low and high types can attempt to send a signal, but low types are typically less successful at sending the signal than high types. In the extremal case, only high types can send the signal, and thus it is natural to classify this signal as an index. In the course of this section, we assume that both low and high types have their own type-specific signal that they alone can send. In other words, the low-type signal can be sent with probability 1 by low types but cannot be successfully sent by high types.[9]

This can all be nicely represented as an action-response game similar to those presented in the previous section. Once again with probability p the sender is a high type. The sender must then determine whether to send her type-specific signal. The sender desires to be identified as a high type, while the receiver simply wants to correctly identify the sender (i.e., there is a partial conflict of interest between sender and receiver). If the sender does not send a signal, then the receiver must determine whether to classify the sender as a high or low type. If the receiver is confronted by a type-specific signal, she correctly identifies her counterpart.[10]

The extensive form of this game is presented in figure 2.[11] Senders can choose to always signal (SS), only signal when a high type (SN), only signal when a low type (NS), or never signal (NN). Likewise, in the absence of a type-specific signal, the receiver may treat the sender as a low type (L) or a high type (H). When it is more likely a sender is a low type ($p < .5$), there are two fully informative equilibria: (SS, L) and (SN, L). In both cases the receiver is able to correctly determine the type of the sender because high types always emit their type-specific signal, and the receiver assumes a sender who fails to send a signal is a low type. When $p > .5$, (SS, L) and (SN, L) are also equilibria, as is (NN, H). In this latter case, senders never send their type-specific signal, and the receiver does best to guess the sender is a high type. No information is transferred between parties.

How is (NN, H) possibly stable? Low types obviously have no incentive to out themselves by sending their type-specific signal. Likewise, high types

9. It may appear odd that we assume that high types cannot send the low-type signal. This assumption does not change the qualitative results and helps simplify the analysis substantially.

10. Once again we can complicate the model by dropping this assumption. The qualitative results do not change, however.

11. Hurd and Enquist (2005) construct a similar signaling game but do not provide a game-theoretic or evolutionary analysis.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

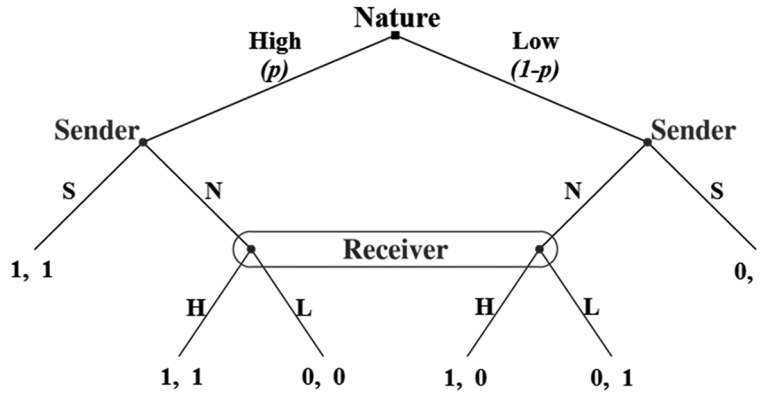


Figure 2. Extensive form of a signaling game in which both types have type-specific signals the receiver can recognize. Color version available as an online enhancement.

do not have incentive to signal. Note, however, that high types are by no means penalized if they send their type-specific signal—receivers will continue to identify them as a high type. Thus, it is clear that while (NN, H) is indeed an equilibrium, it is not a strict equilibrium: senders employing NN do equally well as those using SN. This means NN is only neutrally stable,[12] and a population consisting of senders that use NN can be invaded by senders that employ SN since both receive a payoff of 1 when pitted against a population of receivers all playing H. If the proportion of senders using SN passes a critical threshold, the receivers now do best to switch from H to L.[13] Once the receivers make this switch, senders using SN outcompete those employing NN.

We can provide further proof that (NN, H) is not a real evolutionary possibility by appealing to the replicator-mutator dynamics, a modified version of the more traditional replicator dynamics (Taylor and Jonker 1978; Hofbauer and Sigmund 1998). The replicator dynamics is a standard model of biological evolution. It assumes an infinite population and that individuals are randomly paired to play a predetermined game. Simply put, the replicator dynamics ensures strategies that outperform the population average increase in size in the next generation. While the simplest form of

12. We say a strategy is a neutrally stable strategy if there are no mutants who do strictly better than the natives but there do exist mutants who do just as well as the natives. See Hofbauer and Sigmund (1998) for an overview.

13. Specifically, this occurs when the proportion of senders using SN is greater than $(2p - 1)/p$.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

the replicator dynamics does not allow for mutations, we use the replicator-mutator dynamics (see Hofbauer and Huttegger 2008).[14] With some small probability, an individual switches (with uniform probability) from her current strategy to a new strategy. We find that in the presence of mutations, the population gravitates away from the (NN, H) equilibrium. Senders employing SN outperform those using NN because senders must occasionally interact with mutant receivers playing L. Thus, under reasonable conditions, it appears the full disclosure principle is vindicated—all but the lowest type reveal, and receivers assume senders who do not signal are low types.

How do these results change if we assume there is a small cost to sending a signal? Including such a cost is in many cases a sensible thing to do. First, theorists often distinguish between strategic costs and so-called efficacy costs (Guilford and Dawkins 1991). Strategic costs are those necessary to ensure honest communication (and thus typically refer to the type-specific costs stipulated by the handicap principle). Efficacy costs, however, refer to the costs necessary to ensure the signal is transmitted efficiently to the receiver. In complex environments, conveying information to others is often a serious task and may involve significant metabolic costs. For example, nightingales are known to signal for the majority of the night and can lose up to 10% of their body weight as a result. Additionally, there are other reasons to think that signaling in many contexts is costly. Calls may attract unwanted attention from predators (for more on the predation costs of signaling, see Pruden and Uetz 2004). Finally, there are real opportunity costs associated with signaling. Time spent croaking could instead be used to catch flies.

The introduction of efficacy costs indeed has drastic effects. Table 1 incorporates such a cost into our index-based signaling game.[15] If efficacy costs are great enough ($c > 1$), then high types do best to not signal, even if this means they will be classified by the receiver as a low type. Less extreme efficacy costs can also have important effects. In particular, if signaling is costly, (NN, H) becomes a strict Nash equilibrium. High types now have

14. In particular, we employ the discrete-time, two-population replicator-mutator dynamics. In this dynamic, we assume two populations of agents, senders and receivers, randomly paired to interact with out-group members. Let the composition of population 1 be represented by the vector $X = (x_1, \ldots, x_n)$, while the composition of population 2 is represented by the vector $Y = (y_1, \ldots, y_n)$. The fitness of x_i is simply $F_i = \sum_{j=1}^{n} f_i(y_j)x_i$, where $f_i(y_j)$ is x_i's payoff when paired with a counterpart using strategy j. How the composition of these two populations changes over time is governed by the following equation: the proportion of agents in population 1 using strategy i in the next generation is $(1 - \mu)x_i(F_i/\sum_{k=1}^{n} F_k x_k) + \mu/n$, where μ is the mutation rate.

15. We assume that efficacy costs are the same for low and high types. This of course may not be the case in nature.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

TABLE 1. SIGNALING GAME

	L	H
SS	$p - c, 1$	$p - c, 1$
SN	$p(1 - c), 1$	$1 - cp, p$
NS	$-(1 - p)c, 1 - p$	$p - (1 - p)c, 1$
NN	$0, 1 - p$	$1, p$

Note.—Normal form of the game presented in fig. 2 except there is a small cost, c, to sending a type-specific signal. Sender strategies are listed by row, receiver strategies by column.

reason to remain mute—by not sending the signal they recoup the benefits of being correctly identified as a high type without having to pay the extra cost. Thus, the inclusion of signal costs in the model ensures that the uninformative pooling equilibrium (NN, H) is evolutionarily significant: a population of senders all employing NN cannot be invaded by SN because these mutants unnecessarily pay an extra cost. Recall from above that in the absence of signal costs these pesky mutants ultimately lead to the demise of the (NN, H) equilibrium. Yet since these mutants can no longer gain a foothold in the population, the pooling equilibrium is evolutionarily stable.

We can further stress this point by appealing once again to the replicator-mutator dynamics. In particular, we see that for a wide range of parameter values a significant proportion of simulations head to the pooling equilibrium (see fig. 3).[16] How likely the system is to end up at the pooling equilibrium is determined by two values, the magnitude of the cost associated with signaling and the proportion of high types in the population. As the costs associated with signaling increases, the population is more likely to evolve to the uninformative pooling equilibrium. Likewise, as high types become more prevalent, information transfer becomes a more remote possibility.

One may question whether it is biologically realistic to assume that the majority of individuals are high types. In many contexts, however, this is a perfectly sensible assumption. Consider a case in which the underlying trait signaled is health. Male jumping spiders attempt to convey whether they are well fed or starving by presenting their abdomen to potential mates. Healthy spiders have plenty of white coloration on their abdomen, while the abdomen of a starving spider is primarily black (this signal is considered to be unfakeable; see Taylor et al. 2000). When times are good, healthy high-type spiders will vastly outnumber sickly low types. Thus, for traits such as health, it is sensible in many cases to assume high types will outnumber low types. Additionally, one may worry that the results of this section will not

16. The starting condition for each of these simulations was determined by randomly sampling the state space.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

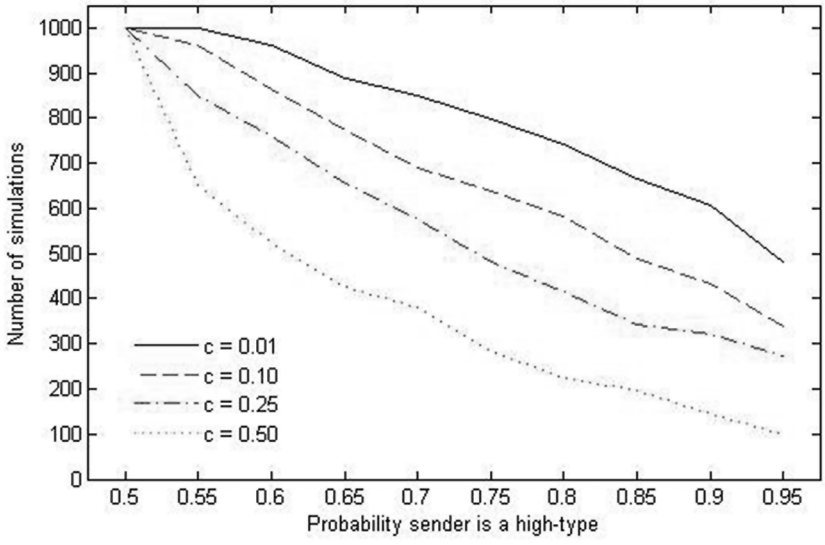


Figure 3. Number of simulations that result in perfect information transfer. Simulations not resulting in information transfer reach the uninformative (NN, H) equilibrium. Each simulation was run for 5,000 generations. Color version available as an online enhancement.

transfer to more complicated models, in which, for example, there is a continuum of quality types. This will be addressed in section 6. In the next section we discuss the possibility of deception.

5. Deception despite Unfakeable Signals. Just as in the Lewis signaling game, information transfer in our index signaling game is not guaranteed when nature is biased. Yet can we say something stronger than this? Specifically, are we licensed to say that deception at equilibrium is a real possibility, even when in the presence of indices? This hinges on what is meant by deception in this game-theoretic context. I restrict myself to one recent account of deception put forth by Brian Skyrms (2010) and note only in passing that there are a plethora of competing proposals in both the biological as well as the philosophical literature.[17] Briefly, a signal is said to be deceptive if it (i) contains misinformation and results in the receiver performing an act that is (ii) detrimental to the receiver but (iii) in the sender's interest. A

17. For criticisms of Skyrms's definition, see Godfrey-Smith (2011). Martinez (2015) provides a compelling discussion of Godfrey-Smith's reworking of Skyrms's account. In general, Skyrms's definition avoids many of the pitfalls of previous attempts at defining deception. See McWhirter (2015) for a comparison and appraisal.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

signal is said to carry misinformation if upon receiving a signal the probability of a state that does not occur goes up.

Consider a signaling game identical to the game of section 4, except there are now three levels of sender quality: low, medium, and high. Once again, senders can either send their type-specific signal or choose to stay mute. Receivers acquire a payoff of 1 for correctly identifying the sender, and senders receive a payoff of 1 if they are identified as a high type, 0.5 if identified as a medium type, and 0 if identified as a low type. If 20% of the population are low types, 40% are medium types, and the remaining 40% are high types, the following arrangement is stable. High types send their type-specific signal, while both medium and low types do not send their type-specific signal. Upon not receiving a signal, receivers assume their counterpart is a medium type. This arrangement is an equilibrium. Receivers do best to assume their counterpart is a medium type since medium types outnumber low types. Furthermore, medium types have no incentive to send their type-specific signal.

Where is the deception? When a low type sends the null signal (i.e., does not send a type-specific signal) it is treated as a medium type by the receiver. This, of course, is great news to the sender but detrimental to the receiver. Finally, since both low and medium types send the null signal, the signal contains misinformation. It raises the probability that the sender is a medium type (from 40% to 66.6%), when in fact the sender is a low type. Surprisingly enough, deception at equilibrium is possible even when signals cannot be faked.

6. The Continuous Case. The results of both section 4 and section 5 seem to be predicated on the relative frequency of types in the population. Moreover, the pooling equilibrium uncovered in section 4 was only possible when high types outnumbered low types. While I have argued that there are many traits of interest (such as health) for which it is sensible to assume high types are more prevalent, in what follows we briefly develop a more biologically realistic signaling game in which there is a continuum of sender quality types. While a full evolutionary analysis is too involved to include in this article, we nonetheless demonstrate that in this more complicated setting, full information transfer is not possible when there are costs associated with signaling. Moreover, we demonstrate that when there are more senders at the high end of the quality spectrum, fewer types will choose to signal at equilibrium. This is analogous to the main result of section 4.

We consider a continuum of sender quality types on the unit interval.[18] Overall, the results of this section (as well as the model explored) bear a

18. In other words, quality is now represented as a number from 0 to 1, where those closer to 1 are considered to be of high quality.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

close resemblance to prior work in economics by Viscusi (1978) as well as Jovanovic (1982). Once again, senders determine whether to keep quiet or send a type-specific signal, and senders desire to be identified as a higher type than they actually are. We assume for simplicity that when the sender issues her type-specific signal her payoff is just her quality level (i.e., a number on the unit interval). If the sender remains mute, her payoff is equal to her counterpart's guess (once again, a number on the unit interval). Receivers correctly identify senders who emit their type-specific signal (securing a payoff of 1). In the absence of a signal, receivers must guess the underlying type of their counterpart, and their payoff decreases as the absolute difference between their guess and the actual underlying type of the sender increases. This is a sensible assumption—it is reasonable to expect that the receiver's payoff decreases as her response becomes less appropriate for the underlying type of the sender. Specifically, we assume the receiver secures a payoff of $1 - \eta(x - g)^2$, where x is the actual quality of the sender, g is the receiver's guess, and $\eta > 0$ (see fig. 4).[19]

The sender's strategy is a number, n, on the unit interval and determines what we will refer to as the 'signaling threshold'. Only senders whose quality is above the signaling threshold signal. The strategy pair (n, g) is stable when three conditions hold. First, the highest quality type among the subset of quality types that do not signal must not have incentive to signal. Second, the lowest quality type among the subset of quality types that do signal must have incentive to signal. The only way both of these conditions can be simultaneously satisfied is for $n - c = g$ (where c is the cost associated with signaling). Finally, receivers must have no reason to adjust their strategy. In other words, given the distribution of types (discussed below), signaling threshold, and receiver payoff function, receivers do best to stick with their current guess. A formal statement of this last condition is provided in the appendix.

Finally, we consider the distribution of quality types in the population. We consider a simple class of probability density functions: $a + 2(1 - a)x$. When $a = 1$, this is just the uniform distribution on the unit interval. For $0 \leq a < 1$, "high types" (those whose underlying quality is close to 1) are more prevalent than "low types" (those closer to 0; see fig. 5). With all of this in place we can now determine the sender's signaling threshold at equilibrium. This calculation is contained in the appendix, and the results of this exercise are displayed in figure 6.

When there is no cost associated with signaling, then the only stable outcome is full disclosure ($n = 0$, $g = 0$). Yet even in this slightly more

19. This particular payoff function was chosen due to the fact that it allowed for easy analytic results. It is worth mentioning that the qualitative results are robust under choice of payoff function.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

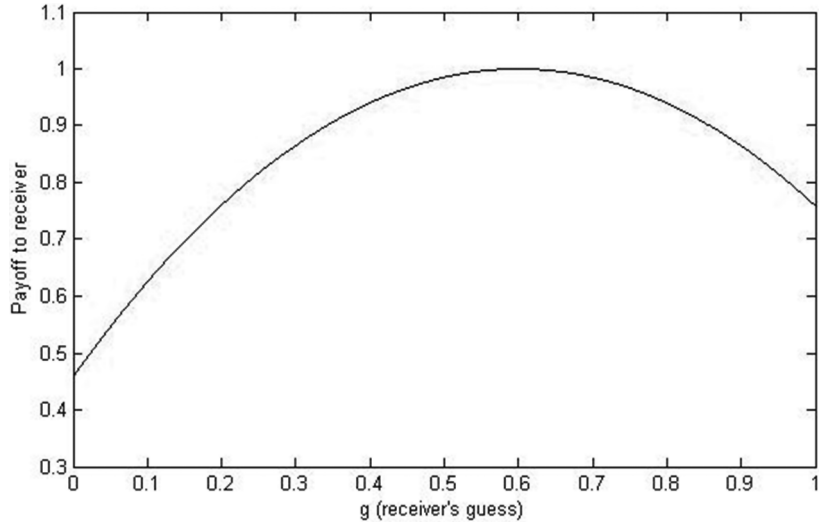


Figure 4. Receiver's payoff as a function of g when confronted by a sender of quality 0.6. Color version available as an online enhancement.

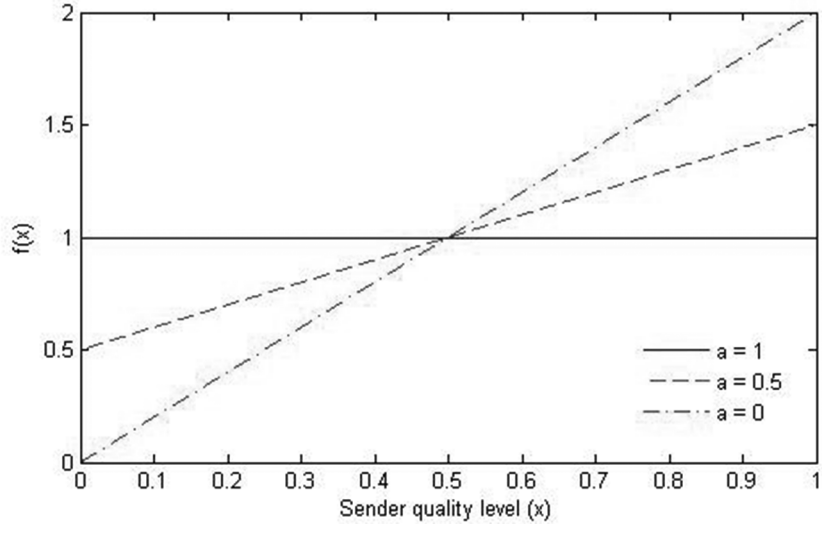


Figure 5. Distribution of sender quality types. Distribution skews to the left as a approaches 0. Color version available as an online enhancement.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

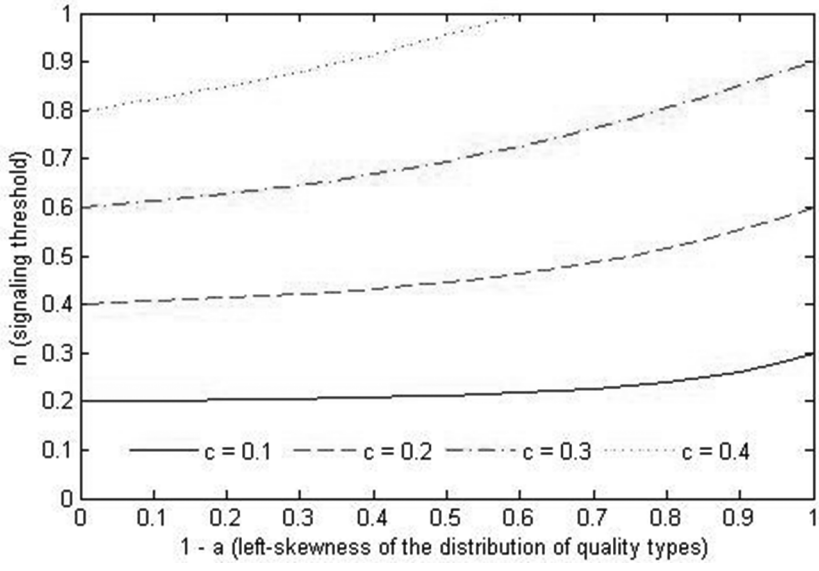


Figure 6. Signaling threshold at equilibrium as a function of $1 - a$. Color version available as an online enhancement.

complicated model, only minor signal costs can result in partial, or even no, information transfer. As figure 6 indicates, when the cost associated with sending the signal is 0.2, all senders with a quality level below 0.4 do not signal. As the distribution of types becomes skewed to the left, this signaling threshold increases (when $a = 0$, e.g., only those senders with a quality level above 0.6 signal). Moreover, whenever there is a cost attached to signaling, then at equilibrium there is a strictly positive signaling threshold, meaning that a subset of the population will not select to signal. Jovanovic (1982) proves this in the context of a slightly different model.

7. Conclusion. In this article we investigated whether unfakeable signals, or indices, invariably lead to perfect information transfer. There initially appeared to be much support for this conventional wisdom. Unlike fakeable signals, indices are reliable by design, and thus the issue of signal cost is completely avoided. Furthermore, Frank's full disclosure principle provided us with compelling reason to think that nearly all senders would have incentive to reveal their underlying type. Yet in the course of this article we developed a simple signaling game and demonstrated that the uninformative pooling equilibrium is evolutionarily stable when (i) high types are more prevalent than low types and (ii) there are minor costs associated with signaling. Both of these conditions will hold in many real-world scenarios.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

There are often, after all, significant efficacy costs associated with signaling due to the fact that signals can increase both an organism's energy expenditure as well as its visibility to predators (Ryan and Cummings 2005). Likewise, for many features (such as health) it is often sensible to assume that high types outnumber low types. Furthermore, the results of this article are rather robust—perfect information transfer was not guaranteed in the more biologically realistic 'continuous' model explored in section 6. The introduction of cost ensures that some subset of the population will not select to send their type-specific signal. This result was first uncovered by economists investigating certification in the context of Akerlof's pathbreaking work on asymmetric information. Yet these insights have not yet been appreciated by those working in animal communication. As we have demonstrated, these models from economics can easily be appropriated by biologists bent on studying the evolution of signaling. Furthermore, the evolutionary game-theoretic tools employed in this article can better clarify when pooling is evolutionarily significant, as well as the likelihood that this uninformative arrangement is to be the outcome of an evolutionary process.

Appendix

Let $x \in [0, 1]$ be the underlying quality of the sender and n the signaling threshold of the sender. If $x \leq n$, then the organism will not send its type-specific signal. Let $g \in [0, 1]$ be the guess of the receiver upon not receiving a signal. The fitness of a sender using strategy n is equal to either g if $x \leq n$ or x if $x > n$. The distribution of types is governed by the following probability density function, $p(x) = a + 2(1 - a)x$. The receivers' payoff is equal to 1 if they guess the type of their counterpart correctly. If they do not guess the type correctly, then their payoff is $1 - \eta(x - g)^2$, where g is their guess and x is the underlying quality of the sender.

The pair (n, g) is stable when $n - c = g$ and g maximizes $\int_0^n [(1 - (x - g)^2)(a + 2(1 - a)x)]\, dx$ (i.e., the expected payoff of the receiver who guesses g conditional on not receiving a signal). Taking the derivative of this integral with respect to g, we find that when g is equal to $n[a(4n - 3) - 4n]/[6a(n - 1) - 6n]$ the value of the above integral is maximized. We plug this value into the equation $n - c = g$ to determine the signaling threshold: $n - c = n[a(4n - 3) - 4n]/[6a(n - 1) - 6n]$. This yields a polynomial, and when $a = 1$ the root is $n = 2c$. When $a \neq 1$ the root is $n = -\left(\sqrt{3}\sqrt{12a^2c^2 - 4a^2c + 3a^2 - 24ac^2 + 4ac + 12c^2 + 6ac + 3a - 6c}\right)/4(a - 1)$. Senders of quality type at or below these values will refrain from signaling at equilibrium. Furthermore, it can be shown that for $c > 0$ the signaling threshold at equilibrium is strictly positive (i.e., $n > 0$). When $a = 1$, then at equilibrium $n = 2c > 0$. It can also be easily shown that

https://doi.org/10.1086/683016 Published online by Cambridge University Press

for $0 \leq a < 1, -\left(\sqrt{3}\sqrt{12a^2c^2 - 4a^2c + 3a^2 - 24ac^2 + 4ac + 12c^2} + 6ac + 3a - 6c\right)/4(a - 1) > 2c > 0$. Thus, whenever there is a cost attached to signaling, not all quality types will select to signal.

REFERENCES

Akerlof, G. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.

Bradbury, J. W., and S. L. Vehrencamp. 1998. *Principles of Animal Communication*. Sunderland, MA: Sinauer.

Bruner, J., C. O'Connor, H. Rubin, and S. Huttegger. 2014. "David Lewis in the Lab: Experimental Results on the Emergence of Meaning." *Synthese*. doi:10.1007/s11229-014-0535-x.

Crawford, V., and J. Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50:1431–51.

Davies, N., and T. Halliday. 1978. "Deep Croaks and Fighting Assessment in Toads *Bufo bufo*." *Nature* 274:683–85.

Dawkins, R., and J. Krebs. 1984. "Animal Signals: Mind-Reading and Manipulation." In *Behavioural Ecology: An Evolutionary Approach*, 2nd ed, ed. J. R. Krebs and N. B. Davies, 380–402. Sunderland, MA: Sinauer.

Frank, R. 1988. *Passions within Reason: The Strategic Role of Emotions*. New York: Norton.

Fraser, B. 2012. "Costly Signalling Theories: Beyond the Handicap Principle." *Biology and Philosophy* 27:263–78.

Godfrey-Smith, P. 2011. Review of *Signals: Evolution, Learning and Information*, by Brian Skyrms. *Mind* 120:1288–97.

Grafen, A. 1990. "Biological Signals as Handicaps." *Journal of Theoretical Biology* 144:517–46.

Grose, J. 2011. "Modelling the Fall and Rise of the Handicap Principle." *Biology and Philosophy* 26:677–96.

Grossman, S. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24 (3): 461–83.

Guilford, T., and M. Dawkins. 1991. "Receiver Psychology and the Evolution of Animal Signals." *Animal Behavior* 42:1–14.

Hofbauer, J., and S. Huttegger. 2008. "Feasibility of Communication in Binary Signaling Games." *Journal of Theoretical Biology* 254:843–49.

Hofbauer, J., and K. Sigmund. 1998. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

Hurd, P. 1995. "Communication in Discrete Action-Response Games." *Journal of Theoretical Biology* 174 (2): 217–22.

Hurd, P., and M. Enquist. 2005. "A Strategic Taxonomy of Biological Communication." *Animal Behavior* 70:1155–70.

Huttegger, S. 2007. "Evolution and the Explanation of Meaning." *Philosophy of Science* 74:1–27.

Huttegger, S., J. Bruner, and K. Zollman. 2015. "The Handicap Principle Is an Artifact." *Philosophy of Science* 82 (5), forthcoming.

Jovanovic, B. 1982. "Truthful Disclosure of Information." *Bell Journal of Economics* 13:36–44.

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Martinez, M. 2015. "Deception in Sender-Receiver Games." *Erkenntnis* 80:215–27.

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

Maynard Smith, J., and D. G. C. Harper. 2003. *Animal Signals*. Oxford: Oxford University Press.

McWhirter, G. 2015. "Deception and Formal Models of Communication." *British Journal for the Philosophy of Science*. doi:10.1093/bjps/axv001.

Milgrom, P. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics* 12:380–91.

Pawlowitsch, C. 2008. "Why Evolution Does Not Always Lead to an Optimal Signaling System." *Games and Economic Behavior* 63:203–26.

https://doi.org/10.1086/683016 Published online by Cambridge University Press

Pruden, A., and G. Uetz. 2004. "Assessment of Potential Predation Costs of Male Courtship Displays in Wolf Spiders Using Video Digitilization and Playback." *Journal of Insect Behavior* 17:67–80.

Ryan, M., and M. Cummings. 2005. "Animal Signals and the Overlooked Costs of Efficacy." *Evolution* 59:1160–61.

Searcy, W., and S. Nowicki. 2005. *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton, NJ: Princeton University Press.

Skyrms, B. 1996. *The Evolution of the Social Contract*. Cambridge: Cambridge University Press.

———. 2010. *Signals: Evolution, Learning and Information*. Oxford: Oxford University Press.

Spence, M. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87:355–74.

Taylor, P., O. Hasson, and D. Clark. 2000. "Body Postures and Patterns as Amplifiers of Physical Condition." *Proceedings of the Royal Society* B 267:917–22.

Taylor, P., and L. Jonker. 1978. "Evolutionarily Stable Strategies and Game Dynamics." *Mathematical Biosciences* 40:145–56.

Viscusi, W. 1978. "A Note on 'Lemons' Markets with Quality Certification." *Bell Journal of Economics* 9 (1): 277–79.

Zahavi, A. 1975. "Mate Selection: A Selection for a Handicap." *Journal of Theoretical Biology* 53:205–14.

Zollman, K. 2013. "Finding Alternatives to the Handicap Principle." *Biological Theory* 8:127–32.

https://doi.org/10.1086/683016 Published online by Cambridge University Press