The Handicap Principle Is an Artifact

Simon M. Huttegger, Justin P. Bruner, and Kevin J. S. Zollman*†

The handicap principle is one of the most influential ideas in evolutionary biology. It asserts that when there is conflict of interest in a signaling interaction signals must be costly in order to be reliable. While in evolutionary biology it is a common practice to distinguish between indexes and fakable signals, we argue this dichotomy is an artifact of existing popular signaling models. Once this distinction is abandoned, we show one cannot adequately understand signaling behavior by focusing solely on cost. Under our reframing, cost becomes one—and probably not the most important—of a collection of factors preventing deception.

1. Introduction. The handicap principle is one of the most influential ideas in evolutionary biology (Zahavi 1975; Grafen 1990; Maynard Smith 1991; Bergstrom and Lachmann 1997; Maynard Smith and Harper 2003; Searcy and Nowicki 2005). It asserts that when there is conflict of interest in a signaling interaction signals must be costly in order to be reliable. Such situations are relevant whenever an informed sender has an incentive to hide some information from an uninformed receiver. Examples of this can be found in models of sexual selection, in interactions between predator and prey, and in relationships between parents and offspring. If there are sufficiently high differential costs or differential benefits in the signaling interaction, then honest signaling becomes a feature of an evolutionarily stable state despite conflicts of interest. However, the handicap principle faces serious problems from both empirical (Borgia 1993; Haskell 1994; Caro et al. 1995; Chappell et al. 1995; Horn, Leonard, and Weary 1995; Gaunt et al. 1996; McCarty 1996; Silk, Kaldor, and Boyd 2000) and theoretical per-

Philosophy of Science, 82 (December 2015) pp. 997–1009. 0031-8248/2015/8205-0022\$10.00 Copyright 2015 by the Philosophy of Science Association. All rights reserved.

^{*}To contact the authors write to: Kevin J. S. Zollman, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213; email: kzollman@andrew.cmu.edu.

[†]The authors would like to thank Richard Scheines and audiences in Irvine and Chicago for helpful comments. This work was supported by National Science Foundation grant EF 1038456.

spectives (Bergstrom and Lachmann 1997; Huttegger and Zollman 2010; Zollman, Bergstrom, and Huttegger 2013). We show how the handicap principle is a limiting case of honest signaling, which can also be sustained by other mechanisms. This fact has gone unnoticed because in evolutionary biology it is a common practice to distinguish between cues, indexes, and fakable signals (Grafen 1990; Maynard Smith and Harper 2003; Hurd and Enquist 2005; Searcy and Nowicki 2005), where cues provide information but are not signals and indexes are signals that cannot be faked. We find that the dichotomy between indexes and fakable signals is an artifact of the existing signaling models. In our game-theoretic model, no such dichotomy is presupposed but arises naturally for extreme parameter settings, whereas in most cases other outcomes involving cheap honest signaling are shown to be evolutionarily significant. Our results suggest that one cannot adequately understand signaling behavior by focusing solely on cost. Under our reframing, cost becomes one—and probably not the most important—of a collection of factors preventing deception.

2. Action-Response Games and the Handicap Principle. The central situation of interest for biologists studying signaling is one in which there is some incentive for deception. When a predator approaches a prey, the prey has an interest in appearing uncatchable, and this incentive holds whether or not the prey is in fact uncatchable. The predator, however, would like to know whether the prey is in fact uncatchable—it would prefer to avoid pursuing prey that it cannot catch but would like to pursue prey it can catch.

Abstracting away from the particulars of this situation, one generates an interaction that is known as "partial common interest." One party is aware of some set of facts (either about itself or about something outside of itself). In some situations the aware individual has an incentive to honestly reveal some information, while in others there is an incentive for deception.

In classic discussions of signaling in biology, scholars distinguish between signals that can only be sent in some situations (that are "unfakable") and those that can be sent by all types of individuals (that can be "faked"). Although Zahavi (1975) referred to both fakable and unfakable signals as "handicaps" (cf. Grafen 1990), this distinction can now be found, under various names, throughout the biological discussion of signaling.² And often

- 1. Understanding concepts like "deception" in the context of non-intention-based communication is a tricky matter that we will leave to the side for the purposes of this article. What we mean by "some incentive for deception" will be made clear later in this section.
- 2. A signal that cannot be faked is called "an unbluffable signal" by Wiley (1983), "unambiguous" by Maynard Smith (1982), "a revealing handicap" by Maynard Smith (1985) and Grafen (1990), "an assessment signal" by Maynard Smith and Harper (1988) and Johnstone

these two are presented as competing hypotheses for the evolution of signals (Emlen et al. 2012; Shingleton and Frankino 2013).

Thompson Gazelles (*Eudorcas thomsonii*) provide an example of the effectiveness of nonfakable signals. These gazelles will jump high in the air when they observe a predator (a phenomenon known as stotting). It is thought that by doing so they reveal their ability to flee. Predators avoid pursuing those gazelles that jump higher and chase those who do not. It is believed that a slow gazelle will be unable to jump as high, and therefore the signal is unfakable. In such a situation honesty is not particularly surprising because an evolutionary constraint prevents dishonesty.³

But, when signals are available to all types then one must postulate some other mechanism for the maintenance of honesty—this motivated what is now called handicap or costly signaling theory. For example, male guppies (*Poecilia reticulata*) signal a number of features that make them good mates by coloring themselves red. Females respond by differentially mating with a male that is more red (Kodric-Brown 1985, 1989). It is widely believed that males of low quality could also color themselves red but do not. Here, honesty presents a mystery, and the handicap principle has been suggested as a solution to this problem.

The core structure of fakable signaling interactions can be captured by an *action-response game* (Lewis 1969; Hurd 1995). This game has two players, the "sender" and the "receiver," and captures the central role of information in biology. One organism, the sender, has access to some information that another organism, the receiver, cannot access but would benefit from knowing. This is modeled by supposing that the receiver would increase its fitness by conditioning its behavior on the information to which the sender has access. Depending on the situation, the sender may not benefit from providing the information to the receiver.

Figure 1 illustrates one example of this game. Nature begins with the choice at the center of the picture, determining whether the sender is a "high type" or a "low type." The types can be interpreted in any number of ways including the quality as a mate, the state of need of a child, or the ability to flee as prey. The sender can condition its behavior on its type and can choose whether to send a potentially costly signal. If it sends the signal, it pays a cost regardless of the behavior of the receiver. The cost paid depends on the type of the sender, c_L if the sender is a low type and c_H if the sender

^{(1998), &}quot;an index" by Maynard Smith and Harper (2003), Searcy and Nowicki (2005), and Számadó (2011), and "a performance signal" by Hurd and Enquist (2005).

^{3.} While stotting is often used as an example of an unfakable signal, Searcy and Nowicki (2005) suggest that it is better conceptualized as a developmental cost than an evolutionary constraint. Importantly, however, from the perspective of an adult gazelle, one that is slow does not have the option of jumping higher.

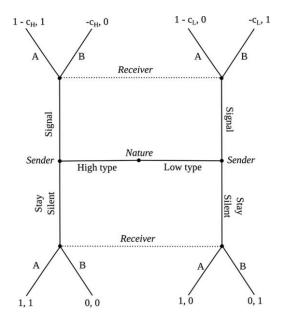


Figure 1. Extensive form representation of a "differential cost" model of the handicap principle. Adapted from Zollman et al. (2013).

is a high type. The receiver, who cannot condition its behavior on the sender's type but can condition its behavior on whether the signal was sent, takes one of two actions A or B. Action A can be interpreted as mating, provision for a child, or refraining from pursuit.

Regardless of type, the sender would prefer that the receiver take action A. When interacting with a high-type sender, the receiver also prefers to take action A. However, when interacting with a low-type sender, the receiver prefers action B. This fact is sufficient to categorize this game as a partial conflict-of-interest signaling game.

Using this game, one can easily illustrate the central concept of the handicap principle. If $1 > c_L$, then there is no equilibrium where the sender honestly reveals its type to the receiver.⁴ However, if $c_L > 1 > c_H$, there exists a separating equilibrium, where the sender reveals its type to the receiver. In this equilibrium, only the high-type sender sends the costly signal, and the receiver responds by taking action A. The receiver does well, securing a

4. It is worth noting that there is another equilibrium known as the hybrid equilibrium that exists when $1 > c_L > c_H$ (Huttegger and Zollman 2010; Wagner 2013; Zollman et al. 2013). Although there are interesting scientific issues surrounding this equilibrium, we will not discuss it further here.

payoff of 1, and the high-type sender receives a payoff of $1 - c_H > 0$. The low-type sender does not send the signal, and the receiver responds by taking action B. Again the receiver does well. The sender gets a payoff of 0, but because $c_L > 1$, it would receive a negative payoff by sending the signal in order to cause the receiver to take action A. So it will not change its strategy.⁵

Guppies color themselves red by allocating carotenoids for coloration that would otherwise be used for a variety of tasks including buffeting the immune system. If the guppy occupies an environment rich in carotenoids (is of "high quality") then the marginal cost of this reallocation is small. However, if the guppy does not have a large supply (is of "low quality"), the marginal cost of reallocating carotenoids to ornamentation is large.

This represents the central tenant of the handicap principle: in the face of conflict of interest, signals must be costly in order to sustain honest signaling. In this case, critically it must be that $c_L > 1$ if honest signaling is to be an equilibrium. For the last 20 years or so, this represents the dominant explanation for honest signaling (Pomiankowski and Iwasa 1998; Maynard Smith and Harper 2003; Searcy and Nowicki 2005; Grose 2011).

3. A Generalized Game. We propose analyzing the handicap principle in the context of a more general game, which we call the Pygmalion game. In our novel conflict-of-interest model, the sender can (again) be of one of two types: a high type or a low type. The high type has relative frequency p in the population. Conditional on its type, the sender can attempt to send a costly signal or not. If the sender attempts to send the signal, nature then determines whether the sender is successful. The probability that a high type succeeds is given by s_L . Critically, the sender pays a cost to attempt to signal regardless of whether she succeeds.

The receiver has no information about the sender's type and can only distinguish two outcomes, the outcome in which the sender succeeded in sending the costly signal and all other outcomes. It cannot tell the type of the sender, nor can it distinguish between the situation in which the sender attempted to signal but failed and the situation in which the sender did not attempt to signal.

Upon observing whether the signal was successfully sent, the receiver then will perform one of two actions. The receiver will take either action A or action B. Action A is best for both players when the sender is of the high type. When the sender is of the low type, the sender still prefers the receiver

5. Although the handicap principle was first stated by Zahavi (1975), the formal point described here was made first in biology by Grafen (1990). In economics, this had already been shown by Spence (1973).

to take action A, but the receiver would prefer to take action B. Again both types receive a payoff of 1 if A is taken (minus a cost for the sender if the signal is attempted). The extensive form of this game is pictured in figure 2.

Interpreted in the context of stotting, like in previous models, we suppose that the gazelle is either fast or slow. The gazelle can then attempt to jump or not. If the gazelle attempts to jump, it pays some cost (either energetic costs or lost time to flee). With some probability, s_H , the fast gazelle succeeds in jumping "high enough." With some other probability, s_L , the slow gazelle succeeds in jumping "high enough." The predator observes whether the gazelle jumped high enough and then decides to pursue on the basis of that observation.

In the context of coloration for guppies, the male guppy either occupies an environment rich in carotenoids or not. It can then allocate some of its carotenoids to coloration. It succeeds with probability s_H or s_L if it is in the

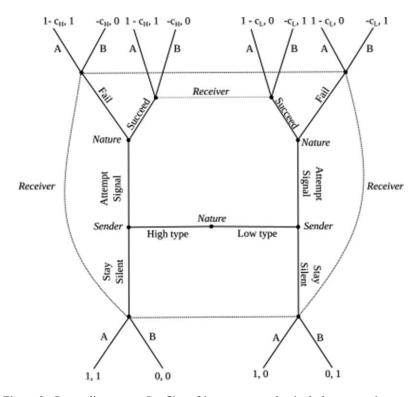


Figure 2. Pygmalion game. Conflict-of-interest game that includes uncertainty as to whether the signal will be successfully sent.

rich or poor environment, respectively. The female then observes whether the guppy is richly colored and decides whether to mate with the male on the basis of this observation.

3.1. Fakable and Unfakable Signals. What our game illustrates is that the difference between fakable and unfakable signals is one of degree not category. How one views signal cost and the mechanisms maintaining honesty changes throughout this continuum. In our general model, a separating equilibrium, where the type of sender is honestly communicated to the receiver, exists provided that the following conditions are satisfied:

$$s_H > c_H$$
 and $c_L > s_L$ and $p < \frac{1}{2 - s_H}$.

The first two conditions illustrate that the minimum cost necessary to sustain honest signaling is determined by the reliability with which the two types succeed in sending the signal.

The two modes of reliable signaling—handicaps and indexes—represent two extremes. When $s_H = s_L = 1$, we reproduce the classic models of costly signaling theory described in the previous section. When $s_H = 1$ and $s_L = 0$, then only the high type can send the signal and would be expected to do so even with no signal cost (when $c_H = c_L = 0$). In this case the signal is functioning as an unfakable signal. In the intermediate cases where $0 < s_L < s_H < 1$, the signal becomes more like an index, the lower s_L and the higher s_H is.

To illustrate how the traditional version of the handicap principle can be misleading, consider a situation appropriately modeled by our Pygmalion game. Suppose a researcher incorrectly believes that the situation fits the model of figure 1—the model of the traditional handicap principle. In an attempt to confirm the traditional handicap model, the researcher would endeavor to estimate the value of c_L . To do so, the researcher might find low-type individuals who display the signal or might experimentally manipulate low types by forcing them to signal and measure their fitness (or, usually, some proxy for fitness). This researcher might discover an apparent contradiction with the theory, namely, that $1 > c_L$.

Viewed from the perspective of the Pygmalion game, the researcher has not fully exhausted the space of possibilities, however. The researcher must also measure s_L . Under our hypothetical scenario it might be the case that $1 > c_L > s_L$, presenting no particular mystery from the perspective of the Pygmalion game.

Beyond the traditional separating equilibrium, there is a new equilibrium that does not have a counterpart in traditional signaling models. We call this a *pseudo-separating equilibrium*. Here both the high type and the low type

send the signal, while the receiver only chooses the favorable action upon receipt of the signal. It exists when

```
1. s_H > c_H, s_H \ge s_L, and s_L > c_L.

2. s_L/(s_H + s_L) .

3. If <math>s_H + s_L > 1, then p > (2s_L - 1)/(s_H + s_L - 1).

4. If s_H + s_L < 1, then p < (2s_L - 1)/(s_H + s_L - 1).
```

Since $s_L > c_L$, the separating equilibrium and the pseudo-separating equilibrium cannot both exist at the same time. Indeed, this equilibrium can exist when there is no signal cost $(c_L = 0)$. Furthermore, the pseudo-separating equilibrium can occur even when high types are rare, so long as low types are sufficiently bad at sending the signal.

Populations occupying this equilibrium will also appear to contradict the handicap principle because they may exhibit no signal cost whatsoever. This can occur even when s_L is significantly above zero. However, information is still communicated, and such cases will be difficult to classify into the categories of fakable or unfakable because they occupy a middle ground. Finally, like many signaling games, the Pygmalion game features "pooling equilibria," where neither type sends the signal and the receiver responds as best it can given its lack of information.

3.2. Evolutionary Analysis. The existence of equilibria cannot guarantee their evolutionary significance (Huttegger and Zollman 2013). To evaluate their evolutionary significance, we must consider an explicit evolutionary model that will allow us to determine whether the equilibria presented above are potential endpoints for an evolutionary process. Here we use the two-population replicator dynamics (Hofbauer and Sigmund 1998).

We generated 1,000 random parameterizations of the game where the separating equilibrium exists and another 1,000 random parameterizations of the game where the pseudo-separating equilibrium exists. For each parameterization, we estimated the proportion of populations that will evolve to the separating and pseudo-separating equilibrium respectively. These results are pictured in figures 3 and 4.

Figure 3 illustrates the relationship between the two equilibria and p, the probability with which the sender is a high-type sender. In both cases, one can see that populations are more likely to evolve to the separating or pseudo-separating equilibrium when the high type is relatively rare. This is sensible. When the high type is relatively common, the receiver does well by choosing A regardless of the sender's behavior. And, if this is the receiver's strategy, both types of senders would prefer to not pay a cost to send the signal.

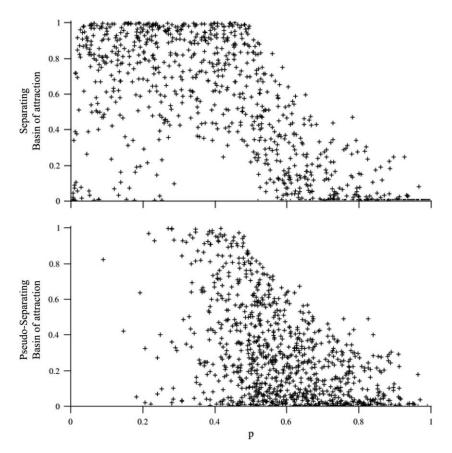


Figure 3. Relationship between p (the probability with which the sender is the high type) and the basins of attraction of the separating equilibrium (top) and pseudo-separating equilibrium (bottom).

Figure 4 illustrates the relationship between the two equilibria and the marginal benefit of signaling for the high-type receiver (i.e., $s_H - c_H$). The top plot only includes parameterizations where the high type is relatively rare, $p \le .5$. In both cases one can see that as the marginal benefit of signaling becomes higher for the high-type sender, one is more likely to evolve to an equilibrium where information is communicated.

Critically, this second result indicates another important caveat for the handicap principle. If one is observing signaling, one should expect the cost of signaling for high types to be relatively low compared to the probability of successful signaling. Without a relatively low cost, signaling is

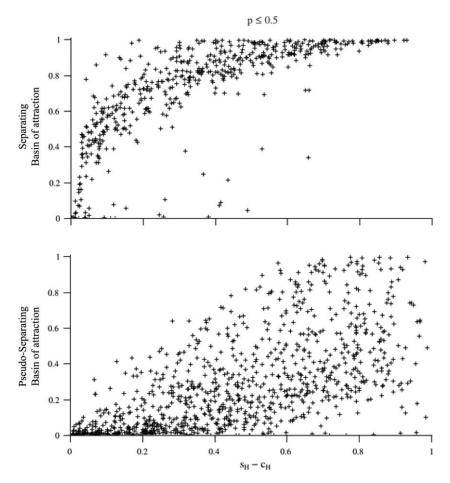


Figure 4. Relationship between the marginal benefit of signaling for the high type (i.e., $s_H - c_H$) and the basins of attraction for the separating equilibrium (*top*) and pseudo-separating equilibrium (*bottom*). Top plot contains only points where $p \le .5$.

unlikely to evolve, even if it is an equilibrium. Experimentalists who estimate signal costs should therefore be careful to estimate c_L as opposed to c_H or some combination of the two.

Overall, while the separating and pseudo-separating equilibria are far from certain to evolve in any situation in which they are equilibria, there are a significant number of cases where they are expected to evolve. As a result, we can conclude that they represent potentially important end points for evolution.

4. Conclusion. One important limitation of this model is that both the cost parameters and the ability to signal are exogenous; that is, from the perspective of the model they are constraints of evolution. Further research is needed to consider what would be the effect of allowing these to vary. This would be akin to studying the method by which selections settle on one of a large potential number of signals—what Maynard Smith and Harper (2003) call the "evolution of signal form."

One might argue that we have achieved the opposite of what we claimed to in the beginning. Rather than challenging the handicap principle, one might say we have shown it to have broader applicability—even to cases of index-like signaling. As a formal statement of our results, this is absolutely true. There is probably no canonical single theory that goes by the name "Handicap Principle" to which we could clearly point. For us, the handicap principle as a theory includes the traditional categorization of signaling interactions into the categories of "fakable" and "unfakable" signals. This categorization, we argue, places signal costs as central to the understanding of honesty (when signals are fakable) or as totally irrelevant (when signals are unfakable). The former situation represents what is usually provided as the central illustrative case for the handicap principle, while the latter case has been regarded as trivial in biological discussions of communication.

We argued that when one models the ability of a sender to fake a signal as a continuum, one develops a more nuanced understanding of signal costs that places it as one, but not the only, factor determining the stability of honest signaling. This we believe represents a significant departure from handicap signaling theory, where cost is posited as *the* stabilizing force.

REFERENCES

- Bergstrom, C. T., and M. Lachmann. 1997. "Signalling among Relatives." Pt. 1, "Is Costly Signalling Too Costly?" *Philosophical Transactions of the Royal Society of London* B 352: 609–17.
- Biernaskie, J. M., A. Grafen, and J. C. Perry. 2014. "The Evolution of Index Signals to Avoid the Cost of Dishonesty." *Proceedings of the Royal Society* B 281 (July). doi:10.1098/rspb.2014 0876
- Borgia, G. 1993. "The Cost of Display in the Non-resource-Based Mating System of the Satin Bowerbird." *American Naturalist* 141 (5): 729–43.
- Caro, T. M., L. Lombardo, A. W. Goldizen, and M. Kelly. 1995. "Tail-Flagging and Other Anti-predator Signals in White-Tailed Deer: New Data and Synthesis." *Behavioral Ecology* 6 (4): 442–50.
- Chappell, M. A., M. Zuk, T. H. Kwan, and T. S. Johnsen. 1995. "Energy Cost of an Avian Vocal Display: Crowing in Red Junglefowl." Animal Behaviour 49 (1): 254–56.
- 6. For models in which this is considered, see Holman (2012), Biernaskie, Grafen, and Perry (2014), and Kane and Zollman (2015).

- Emlen, D. J., I. A. Warren, A. Johns, I. Dworkin, and L. C. Lavine. 2012. "A Mechanism of Extreme Growth and Reliable Signaling in Sexually Selected Ornaments and Weapons." *Science* 337 (August): 860–65.
- Gaunt, A. S., T. L. Bucher, S. L. Gaunt, and L. F. Baptista. 1996. "Is Singing Costly?" Auk 113 (3): 718–21.
- Grafen, A. 1990. "Biological Signals as Handicaps." *Journal of Theoretical Biology* 144:517–46.
 Grose, J. 2011. "Modelling and the Fall and Rise of the Handicap Principle." *Biology and Philosophy* 26 (5): 1–20.
- Haskell, D. 1994. "Experimental Evidence That Nestling Begging Behaviour Incurs a Cost due to Nest Predation." Proceedings of the Royal Society of London B 257:161–64.
- Hofbauer, J., and K. Sigmund. 1998. Evolutionary Games and Population Dynamics. Cambridge: Cambridge University Press.
- Holman, L. 2012. "Costs and Constraints Conspire to Produce Honest Signaling: Insights from an Ant Queen Pheromone." Evolution 66 (7): 2094–2105.
- Horn, A. G., M. L. Leonard, and D. M. Weary. 1995. "Oxygen Consumption during Crowing by Roosters: Talk Is Cheap." Animal Behaviour 50:1171–75.
- Hurd, P. L. 1995. "Communication in Discrete Action-Response Games." Journal of Theoretical Biology 174 (2): 217–22.
- Hurd, P. L., and M. Enquist. 2005. "A Strategic Taxonomy of Biological Communication." Animal Behaviour 70 (5): 1155–70.
- Huttegger, S. M., and K. J. S. Zollman. 2010. "Dynamic Stability and Basins of Attraction in the Sir Philip Sidney Game." Proceedings of the Royal Society of London B 277:1915–22.
- 2013. "Methodology in Biological Game Theory." British Journal for the Philosophy of Science 64 (3): 637–58.
- Johnstone, R. A. 1998. "Game Theory and Communication." In Game Theory and Animal Behavior, ed. L. A. Dugatkin and H. K. Reeve, 94–117. New York: Oxford University Press.
- Kane, P., and K. J. S. Zollman. 2015. "An Evolutionary Comparison of the Handicap Principle and Hybrid Equilibrium Theories of Signaling." PLoS ONE 10 (9): e0137271. doi:10.1371/journal.pone.0137271.
- Kodric-Brown, A. 1985. "Female Preference and Sexual Selection for Male Coloration in the Guppy (*Poecilia reticulata*)." *Behavioral Ecology and Sociobiology* 17 (3): 199–205.
- 1989. "Dietary Carotenoids and Male Mating Success in the Guppy: An Environmental Component to Female Choice." *Behavioral Ecology and Sociobiology* 25 (6): 393–401.
- Lewis, D. 1969. Convention: A Philosophical Study. Cambridge, MA: Harvard University Press.Maynard Smith, J. 1982. Evolution and the Theory of Games. Cambridge: Cambridge University Press.
- ——. 1991. "Honest Signaling: The Philip Sidney Game." *Animal Behavior* 42:1034–35.
- Maynard Smith, J., and D. G. C. Harper. 1988. "The Evolution of Aggression: Can Selection Generate Variability?" *Philosophical Transactions of the Royal Society* B 319 (1196): 557–70.
 ——. 2003. *Animal Signals*. Oxford: Oxford University Press.
- McCarty, J. P. 1996. "The Energetic Cost of Begging in Nestling Passerines." *Auk* 113 (1): 178–88. Pomiankowski, A., and Y. Iwasa. 1998. "Handicap Signaling: Loud and True?" *Evolution* 52 (3): 928–32.
- Searcy, W. A., and S. Nowicki. 2005. The Evolution of Animal Communication. Princeton, NJ: Princeton University Press.
- Shingleton, A. W., and W. A. Frankino. 2013. "New Perspectives on the Evolution of Exaggerated Traits." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 35 (2): 100–107.
- Silk, J., E. Kaldor, and R. Boyd. 2000. "Cheap Talk When Interests Conflict." *Animal Behaviour* 59 (2): 423–32.
- Spence, M. 1973. "Job Market Signaling." Quarterly Journal of Economics 87 (3): 355-74.
- Számadó, S. 2011. "The Cost of Honesty and the Fallacy of the Handicap Principle." Animal Behaviour 81 (1): 3–10.

- Wagner, E. 2013. "The Dynamics of Costly Signaling." Games 4 (2): 163-81.
- Wiley, R. H. 1983. "The Evolution of Communication." In *Communication*, ed. T. Halliday and P. Slater, 156–89. Oxford: Blackwell Scientific.
- Zahavi, A. 1975. "Mate Selection: A Selection for a Handicap." *Journal of Theoretical Biology* 53:205–14.
- Zollman, K. J. S., C. T. Bergstrom, and S. M. Huttegger. 2013. "Between Cheap and Costly Signals: The Evolution of Partially Honest Communication." *Proceedings of the Royal Society* B 280 (1750): 20121878.