Responsiveness and Robustness in the David Lewis Signaling Game

Carl Brusse and Justin Bruner*†

We consider modifications to the standard David Lewis signaling game and relax a number of unrealistic implicit assumptions that are often built into the framework. In particular, we motivate and explore various asymmetries that exist between the sender and receiver roles. We find that endowing receivers with a more realistic set of responses significantly decreases the likelihood of signaling, while allowing for unequal selection pressure often has the opposite effect. We argue that the results of this article can also help make sense of a well-known evolutionary puzzle regarding the absence of an evolutionary arms race between sender and receiver in conflict-of-interest signaling games.

1. Signaling Games and Evolution. Common interest signaling games were introduced by David Lewis (1969) as a game theoretic framework that treats communicative conventions as solutions to coordination problems. In recent years, this has informed a growing body of work on the evolution of communication, incorporating signaling games into an evolutionary game theoretic approach to modeling the evolution of communication and cooperation in humans (Skyrms 1996, 2010).

As the basis for such work, Lewis signaling games are attractive in their intuitive simplicity and clear outcomes. They are coordination games between world-observing senders and action-making receivers using costless signals, in contrast to games where interests may differ and where costly signals are typically invoked. In the standard two-player, two-state, two-option Lewis signaling game (the '2 × 2 × 2 game'), the first agent (signaler) observes that the world is in one of two possible states and broadcasts one of

two possible signals that are observed by the second agent (receiver) who performs one of two possible actions. If the act chosen by the receiver matches the underlying state of the world, both agents receive a greater payoff than otherwise.

Most important, though, the game theoretic results are unequivocal. There exist two Nash equilibria that are, in Lewis's words, *signaling systems*, where senders condition (otherwise arbitrary) signaling behavior on the state of the world, and receivers act on those signals so as to secure the mutual payoff. The two systems only differ on which signal is associated with each state of the world.[1] Huttegger (2007) and Pawlowitsch (2008) have shown that under certain conditions a signaling system is guaranteed to emerge under the replicator dynamics, a standard model of evolution further discussed in section 4.

Of course the degree to which Lewis's approach makes sense is to some extent determined by the level of confidence we have in the interpretation and application of such idealized models to the more complex target systems (such as animal and human communication). Thus, one obvious worry is that by introducing more realistic features into the model, one may break or significantly dilute previous findings on the evolution of signaling.

Not surprisingly, then, recent work on Lewis signaling games has investigated the many ways in which such de-idealizations could occur and tested the robustness of signaling in the face of them. Some deviations from the standard Lewis signaling game include more and varied states of the world, the possibility of observational error or signal error, noisy signals, the introduction of a partial conflict of interest between senders and receivers, the reception of more than one signal, and so on. Many such concerns are dealt with favorably by Skyrms (2010) and in work by others. For example, Bruner et al. (2014) generalize beyond the $2 \times 2 \times 2$ case, while Godfrey-Smith and Martinez (2013) and Martinez and Godfrey-Smith (2016) mix signaling games of common interest and conflict of interest. One result (particularly important for our purposes) is that signaling systems are not guaranteed in the simple $2 \times 2 \times 2$ case when 'nature is biased'. In other words, when the probabilities of the world being in one state or the other are not equal, a pooling equilibrium in which no communication occurs between sender and receiver is evolutionarily significant (Huttegger 2007; Pawlowitsch 2008).

2. Symmetry Breaking. Our article will focus on the idealization that senders and receivers are equally responsive in strategic settings. Senders and receivers (in the evolutionary treatment of such games) are two popula-

1. There also exist a class of 'pooling equilibria' that involve the receiver performing an act unconditionally.

tions of highly abstract and constrained agency roles: all that signalers do upon observing the state of the world is send a signal, and the receivers choose to act as though the world is in one of the two possible sender-observable states. Of those two roles, it is the restriction on receivers that appears to be particularly unrealistic.

Imagine, for example, a forager sighting a prey animal at a location inaccessible to her, but close enough to be acquired by an allied conspecific (who cannot observe the animal). In this case, it is easy for the first forager to slip into the signaling role and execute it, whistling or gesturing to her counterpart. To play the receiver role, however, the second forager has to actually reorient his attention (to some degree) and attempt to engage in appropriate behavior for the world-state the first has observed (e.g., prey is to the east or to the west).

The Lewis signaling model by design is constrained such that the receiver's actions are limited to just those acts associated with the sender's observed world-states. It is of course sensible to begin inquiry with as simple of a model as possible and consider a limited range of responses to stimuli. However, our point is that it is more plausible to make these idealizations for signalers than for receivers. Signals are (by stipulation) cheap and easy to send, yet the actions available to the receiver are less plausibly interpreted as intrinsically cheap and free of opportunity cost.

In addition, the informational states drawn on by sender and receiver are also likely to be very different. Any real-life sender's observation of a world state will likely inform her motivations ('we should catch that animal') to dictate a fairly clear course of action ('try to direct the other agent's behavior'). But all the receiver gets is a whistle, gesture, or other signal that (by stipulation) has no preestablished meaning. The experience of observing a strategically relevant state of the world will typically be richer and more detailed than that of observing a strategically relevant artificial signal.

All this leads to two concerns. First, asymmetries in the strategic situations are likely to exist between senders and receivers. Receivers are likely to have locally reasonable options available to them other than those relevant to signaler-observed states of the world, and their responsiveness to the strategic situation is therefore less satisfactorily modeled by the strictly symmetric payoff structures of standard signaling games. Call this the *structural responsiveness* concern.

Second, given the likely differences in informational states, goal directness, workload, and opportunity cost between sender and receiver roles, we can expect the mechanisms (cognitive and otherwise) that instantiate them to differ as well, quantitatively and qualitatively. This implies that we should not expect the update responsiveness between sender and receiver to be equal either. Yet the working evolutionary assumption is that senders and receivers update their strategies in an identical manner, modeled using either

learning dynamics or replicator dynamics. Call this the *evolutionary responsiveness* concern.

3. Hedgehog Strategies and Update Asymmetry. The first of these concerns might sound like an argument for abandoning coordination games and moving toward 'conflict of interest' or 'partial conflict of interest' models. However, the issue is more specific than this.

The structural responsiveness concern provides parallel motivation to one of Sterelny's (2012) worries about Skyrms's (2010) use of the Lewis model. Sterelny asks whether the availability of 'third options' on the part of the receiver might undermine the evolution of signaling even when these third options are less valuable than the payoff for successful coordination. As part of a discussion of animal threat responses, he labels this a 'hedgehog' strategy, for it provides the agent with an action that pays off modestly regardless of the state of the world. To make this concrete, hedgehogs often roll into a ball in response to predators. This is a stark contrast to the more sophisticated behavior of vervets, who have specific responses to specific threats. Yet the optimal response a vervet takes to one threat—climb a tree when confronted by a leopard—may lead to total disaster when used in response to another threat, such as an eagle. Hedgehogs avoid such outcomes by 'hedging' unconditionally so as to secure a modest payoff. Translated to signaling games, such a gambit may, in many cases, be more attractive than attempting to respond optimally to a signal.[2]

Sterelny's 'hedgehog' strategy compliments the structural responsiveness concern: receivers (especially) might have other options of value that will stand in competition to those assumed in the standard signaling game. Something like these hedgehog strategies is a plausible departure from the idealization of the baseline Lewis signaling game and better captures the demandingness of the receiver role. The question is whether (as Sterelny suspects) including hedgehog strategies might undermine the evolutionary robustness of signaling systems.

Our second concern pertaining to evolutionary responsiveness parallels a well-known evolutionary hypothesis: the so-called Red Queen effect. In competitive relationships such as predator-prey or parasite-host, the Red Queen hypothesis states that species will be constantly adapting and evolving in response to one another just to "stay in the same place" (van Valen 1973). This should also be the case in competitive signaling situations—such as predator-

2. It is worth noting here that the 'hedgehog' strategy in this Lewis signaling game is in many ways analogous to the risk dominant 'hare' response in stag hunt games. Playing hare instead of stag allows the agent to avoid disaster but only guarantees the individual a mediocre payoff. Thus, the issues and trade-offs associated with the hedgehog strategy are general concerns not confined to just the Lewis signaling games. Thanks to Kim Sterelny for helping us better see this connection.

prey signaling systems or courtship displays among conspecifics. Signalers and receivers come to not just update their strategies but do so at faster or slower rates depending on the nature of the strategic encounter they are entwined in.[3]

It might seem that independent update rates should play no role in Lewis signaling games (along with games of common interest more generally). However, any realistic interpretation of the Lewis signaling game makes it plausible to consider asymmetry in evolutionary responsiveness as likely, if not the norm. First, as argued, the precise cognitive mechanisms and procedures employed by senders and receivers are likely to be different. Different systems (or systems used differently) will admit to different degrees of plasticity and evolvability and will have a different set of crosscutting tasks and utilities exerting distinct demands and pressures. Quick and easy signaling responses will have pathways of update and adaptation different from the (typically) more complex set of systems that appropriate receiver responses require. We should not assume that the evolution of sender and receiver strategies always proceeds at the same pace.

Finally, there is at least some evidence of a basic asymmetry between sender and receiver roles in the literature on great ape communication. For example, Hobaiter and Byrne (2014) stress the great sophistication and flexibility on the receiver side of Chimpanzee gestural communication, while Seyfarth and Cheney (2003) discuss how greater inferential sophistication on the receiver side is a feature of many primate communication systems. While these findings do not directly support the structural and evolutionary responsiveness concerns, they show that real-life sender and receiver strategies (in our near biological cousins at least) exhibit important differences, suggesting cognitive asymmetries compatible with those concerns.

In summary then, there is reason to consider two structural modifications to the Lewis signaling game as especially salient to the issue of responsiveness: the addition of 'hedgehog' strategies for receivers and differing rates of change in sender and receiver strategies.

4. The Model. The evolutionary model we use as the basis for our analysis is the pure strategy $2 \times 2 \times 2$ Lewis signaling game, with the two-population discrete-time replicator dynamics. Exact components of the model include two states of the world (L and R), a world-observing signaler with two pos-

---

3. An example of two groups adapting and evolving at different rates can be found in Richard Dawkins's discussion of his famous Life-Dinner principle (Dawkins and Krebs 1979). While we expect both predator and prey to adapt to each other, Dawkins claims the prey species will come to evolve at a faster rate than the predator species because of the different selection pressures exerted on both species. Failing to adapt quickly enough for the predator means going hungry for an extra day, while failing to adapt for the prey means death.

sible signals (V_1 and V_2), and a signal-observing receiver with two possible actions (A_L and A_R).

If the receiver's action matches the state of the world, then both signaler and receiver get a fixed positive success payoff; otherwise, their payoff is zero. Signalers and receivers both have four pure strategies available to them (see table 1).

For the evolutionary model, the proportions of the different strategies within sender and receiver populations are initially randomly generated. The fitness of each strategy at a time period t is determined by the composition of the opposing population and the payoff associated with each strategy pairing. The proportion of each strategy at play in the next time period $t + 1$ is determined by the standard discrete-time replicator dynamics. For the sender population this is

$$X_i(t + 1) = X_i(t)\frac{F_i^S}{F^S},$$

where X_i is the proportion of senders using the ith sender strategy at a given point in time, F_i^S is the fitness of that strategy, and F^S is the average sender strategy fitness. Likewise, for receivers

$$Y_j(t + 1) = Y_j(t)\frac{F_j^R}{F^R},$$

where Y_j is the proportion of receivers using the jth receiver strategy, F_j^R is the fitness of that strategy, and F^R is the average receiver strategy fitness. This is repeated until the populations settle into an evolutionarily stable arrangement. The update process is deterministic with no randomization or mutation.

5. Modifications and Results. We introduce two novel modifications to this model. First, we add a 'hedgehog' action A_H for the receiver. Second, we allow the rate of generational change of senders and receivers to

TABLE 1. SIGNALER AND RECEIVER STRATEGIES IN THE STANDARD $2 \times 2 \times 2$ COMMON INTEREST SIGNALING GAME

	Strategy Description
S1	Signal V_1 if L and signal V_2 if R
S2	Signal V_2 if L and signal V_1 if R
S3	Signal V_1 always
S4	Signal V_2 always
S5	Act A_L if V_1 and act A_R if V_2
S6	Act A_R if V_1 and act A_L if V_2
S7	Act A_L always
S8	Act A_R always

vary relative to one another. In addition, the bias of nature is varied, and we investigate the effects these three departures from the Lewis $2 \times 2 \times 2$ game have on the evolutionary significance of signaling systems.

5.1. Hedgehog Strategy.   Turning to our first modification, the receiver now has three possible actions upon observing the signal: A_L, A_R, and A_H. As before, a success payoff of 1 is received by both players in the case that the receiver plays A_L while the world is in state L or the receiver plays A_R while the world is in state R. A payoff of zero is received if A_L or A_R is played otherwise. A payoff of H is received unconditionally if the receiver plays A_H, where the value of H is between 0 and 1. The sender has four familiar pure strategies, whereas the receiver now has five (for simplicity we omit conditional strategies involving A_H).

To adapt the earlier forager story, we can imagine the sender and receiver as an egalitarian hunting party, and the game as a situation in which the sender remotely observes the location of a valuable prey animal (left or right) and calls out to the receiver. The receiver is initially unable to observe the prey but can choose to go left or go right (catching the prey if he goes in the matching direction) or alternatively to abandon the hunt in order to obtain a less valuable resource he does not need help from the sender to acquire (the hedgehog strategy). Varying the prior probability of the world is equivalent to it being in a situation in which it is systematically more likely that the prey is to the left or the right.

In the simple unbiased $2 \times 2 \times 2$ signaling game, one of the two signaling equilibria is guaranteed to be reached under the replicator dynamics (Huttegger 2007). In our notation, these equilibria are S1-S1 and S2-S2. Increasing the bias of the world (i.e., making L more probable than R or vice versa) will undermine this, with an increasing proportion of populations instead collapsing to a pooling equilibrium. This will occur when there are initially few conditional signaling strategies (such as S1) in the sender population. In such situations, receivers do best to simply perform the act that is most appropriate for the more likely state of the world. The incentive for senders to adopt a signaling system then disappears, and the community is locked into a pooling equilibrium.

Not surprisingly, we find a similar effect with the hedgehog strategy as values of H, the payoff for A_H, become significant. The hedgehog strategy provides the receiver with an additional unilateral response and is able to attract some proportion of initial populations away from the signaling equilibria when H is in excess of 0.5 (i.e., the average payoff for 'guessing'). This result, for an unbiased world, is illustrated in figure 1.[4]

---

4. Note that the exact range of this effect, including the point at which the effect becomes significant and the *y*-intercept, are artifacts of the number of world-states and strategies in the model and therefore not general.
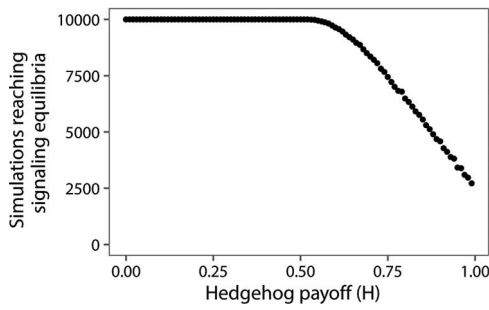
Figure 1. Effect of varying hedgehog payoff *H* on the number of simulations reaching signaling equilibria.

We observe a more surprising result when the bias and *H* are varied in combination. Figure 2 shows the results of varying bias for different values of *H*. As expected, when *H* is set to zero signaling is guaranteed when nature is unbiased and is less likely to emerge as bias increases. As before, when the hedgehog strategy pays significant dividends ($H > 0.5$) signaling is less likely when nature is unbiased. For these high values of *H*, however, increasing nature's bias does not immediately reduce the likelihood of signaling. In fact we observe a 'plateau' followed by an increase in signaling behavior, peaking at the point where bias is equal to the value of *H*, that is, at $P(L) = H$ and at $P(L) = 1 - H$. Within these bounds, all populations reach either a signaling equilibrium or a Hedgehog equilibrium (in which all receivers have adopted the hedgehog strategy), but for values of *H* out-
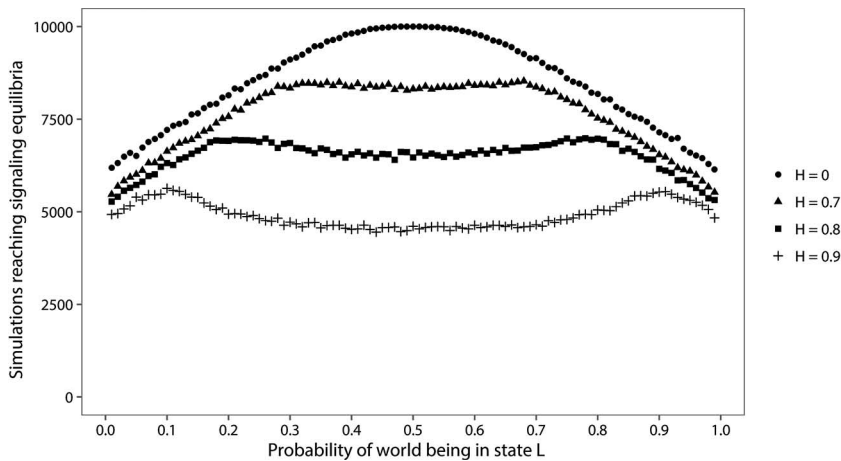

Figure 2. Effect of hedgehog strategy and bias of nature on number of simulations reaching signaling equilibria.

side of this interval traditional pooling instead occurs when the population fails to reach a signaling convention.

5.2. Generational Asymmetry.   We now turn to our second modification of the Lewis signaling framework in which we introduce a generational asymmetry. We introduced a 'slow-down factor' Z to the replicator dynamics in order to control the rate at which sender and receiver populations change over time. Composition of the sender and receiver populations is now governed by the following equations:

$$X_i(t + 1) = (1 - Z_S)X_i(t)\frac{F_i^S}{F^S} + X_i(t)Z_S,$$

$$Y_j(t + 1) = (1 - Z_R)Y_j(t)\frac{F_j^R}{F^R} + Y_j(t)Z_R.$$

Note that when both $Z_R = 0$ and $Z_S = 0$ the above equations are simply the standard replicator dynamics. However, the rate of change is reduced as these values are increased; for example, setting $Z_S = .5$ halves the rate of change for sender strategies. Setting Z_R to 1 (while holding Z_S fixed at 0) means the composition of the receiver population remains fixed, and only the sender population is allowed to evolve.

Introducing this generational asymmetry between senders and receivers has the effect of making signaling more likely when sender strategies evolve faster than receiver strategies. This is illustrated in figure 3, where, for var-
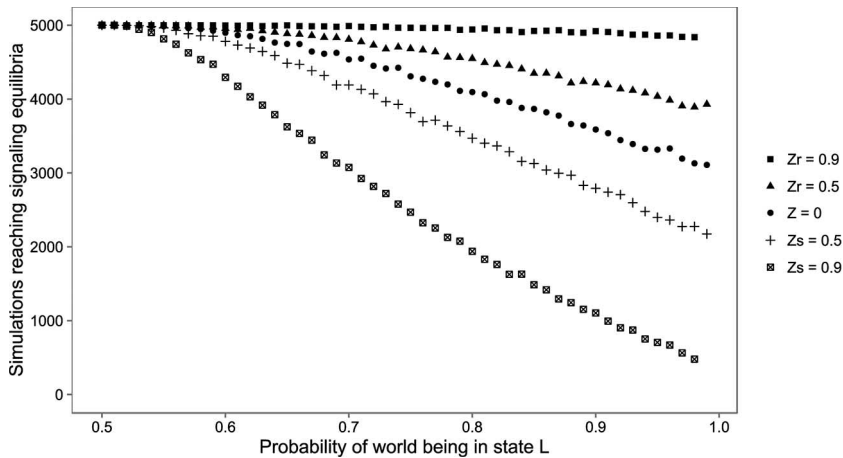

Figure 3. Effect of generational asymmetry and bias of nature on number of simulations reaching signaling equilibria.

ious levels of bias, senders (receivers) are slowed down to half and one-tenth speeds while receivers (senders) are left unaltered.

Slowing the evolution of the sender population leads to more pooling because, as before, receivers facing a sender population consisting of few conditional signalers will do best to simply perform that act that is best suited for the more likely state of the world. As figure 3 illustrates, this effect is more pronounced for high levels of bias. Slowing the evolution of the receiver population has the opposite effect. Senders now have time to adopt the best separating strategy given the initial mix of receiver strategies. Once this occurs, the receiver population slowly responds, and the end result is a robust signaling system. By a similar logic, it is easy to see that a faster-evolving sender population also mitigates against the effect of receiver hedgehog strategies.[5]

6. Discussion. We have explored a few well-motivated departures from the highly idealized and simple Lewis signaling game typically considered in the literature. As shown in section 4, breaking the symmetry between senders and receivers often significantly reduces the likelihood that a separating equilibrium emerges. For one, providing receivers with a safe third option that allows them to secure a decent payoff regardless of the state of the world significantly reduces the size of the basin of attraction of the separating equilibrium. Likewise, separating is a remote possibility when receivers outpace senders in the race to adapt.

However, the situation is less bleak when senders evolve at a faster pace than receivers.[6] Interestingly, many scholars in the animal communications literature have noted that such a response asymmetry holds between sender and receiver when the interests of the parties partially conflict. For instance, Owren, Rendall, and Ryan (2010) note that senders can easily adapt their signaling behavior while receivers for the most part have responses to the stimuli produced by senders that are more difficult to change. This has lead some

---

5. This also helps to explain the twin-peaked curves in fig. 2. In the region of the peaks the uptake rate of the fittest pooling strategy (either hedgehog or traditional) is reduced by the relatively high fitness of its near rival, as the average fitness F^R in the denominator of the replicator dynamics will be closer to that of the superior F_j^R strategy. For some populations (depending on proportions) this competition will allow more time for conditional sending strategies to tip the fitness balance toward conditional receiving strategies instead.

6. This effect is somewhat robust. For instance, Hofbauer and Huttegger (2008) find that under the selection-mutation dynamics signaling conventions are possible when the mutation rate of the receiver population exceeds the mutation rate of the sender population. Note that large mutation rates in effect slow down the rate of evolution of a given population. Thus, one way of interpreting their results is that signaling is likely to occur when the receiver population is relatively unresponsive to the sender population.

to the conclusion that signaling primarily involves the manipulation of receivers by senders.

Yet this leaves us with an evolutionary puzzle. If there is a partial conflict of interest between sender and receiver, what prevents receivers from increasing the speed at which they adapt to the behavior of the sender? In other words, what explains the absence of an evolutionary arms race between sender and receiver? Partial-conflict-of-interest settings are the exact circumstances to which we would expect the Red Queen hypothesis to apply. We believe the results of this article may form the basis of a novel explanation for this puzzling phenomena. Recall that when the interests of sender and receiver are perfectly aligned, it is actually in the interest of both parties for the sender population to 'take the lead' and evolve at the faster rate, as doing so ensures the community is more likely to hit upon a mutually beneficial signaling system.

Yet senders and receivers rarely find themselves engaged exclusively in either common or conflict-of-interest signaling games. As is well known by any parent, not all signaling interactions between relatives are free of conflict. Likewise, agents whose interests are typically thought to be partially opposed, such as two potential mates, may frequently engage in common interest signaling games in contexts unrelated to mating. The point here is that a variety of distinct strategic scenarios can hold between sender and receiver. There is no principled reason to think all interactions will involve the perfect alignment of interests or sizable conflict. Nor is it sensible to assume that communicative strategies will always be fine-grained enough for scenario-by-scenario optimization.

If this is correct, then when a sizable proportion of interactions between sender and receiver involve no or very low conflict of interest, the generational asymmetry result from the previous section may hold to some degree. Both sender and receiver profit when the sender population evolves at a faster rate than the receiver population. Receivers in this context do best to limit how responsive they are to senders so as to ensure the emergence of informative signaling systems in those cases in which their interests do overlap.

REFERENCES

Bruner, Justin, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger. 2014. "David Lewis in the Lab: Experimental Results on the Emergence of Meaning." *Synthese*, September, 1–19.

Dawkins, R., and J. R. Krebs. 1979. "Arms Races between and within Species." *Proceedings of the Royal Society of London* B 205 (1161): 489–511.

Godfrey-Smith, Peter, and Manolo Martínez. 2013. "Communication and Common Interest." *PLoS Computational Biology* 9 (11): e1003282.

Hobaiter, Catherine, and Richard W. Byrne. 2014. "The Meanings of Chimpanzee Gestures." *Current Biology* 24 (14): 1596–1600.

Hofbauer, Josef, and Simon M. Huttegger. 2008. "Feasibility of Communication in Binary Signaling Games." *Journal of Theoretical Biology* 254 (4): 843–49.

Huttegger, Simon M. 2007. "Evolution and the Explanation of Meaning." *Philosophy of Science* 74 (1): 1–27.

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

Martínez, Manolo, and Peter Godfrey-Smith. 2016. "Common Interest and Signaling Games: A Dynamic Analysis." *Philosophy of Science* 83 (3): 371–92.

Owren, Michael J., Drew Rendall, and Michael J. Ryan. 2010. "Redefining Animal Signaling: Influence versus Information in Communication." *Biology and Philosophy* 25 (5): 755–80.

Pawlowitsch, Christina. 2008. "Why Evolution Does Not Always Lead to an Optimal Signaling System." *Games and Economic Behavior* 63 (1): 203–26.

Seyfarth, Robert M., and Dorothy L. Cheney. 2003. "Signalers and Receivers in Animal Communication." *Annual Review of Psychology* 54 (1): 145–73.

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

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

Sterelny, Kim. 2012. "A Glass Half-Full: Brian Skyrms's Signals." *Economics and Philosophy* 28 (1): 73–86.

van Valen, L. 1973. "A New Evolutionary Law." *Evolutionary Theory* 1:1–30.