

Responsiveness and robustness in the David Lewis signalling game

Carl Brusse and Justin Bruner

October 28, 2016

Abstract

We consider modifications to the standard David Lewis signalling game and relax a number of unrealistic implicit assumptions that are often built into the framework. In particular, we explore realistic 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 signalling, while allowing for unequal selection pressure often has the opposite effect. We argue that the results of this paper 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 signalling games.

1 Signalling games and evolution

Common interest signalling games were introduced by David Lewis (Lewis, 1969) as part of a game theoretic framework which identified communicative conventions as the expected solutions to coordination problems. In recent years, this has informed a growing body of work on the evolution of communication, incorporating signalling games into an evolutionary game theoretic approach to modelling the evolution of communication and cooperation in humans (Skyrms, 2010; Skyrms, 1996).

As the basis for game theoretic modelling of such phenomena, David Lewis signalling games are attractive in their intuitive simplicity and clear outcomes. They are coordination games of common interest 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 David Lewis signalling game (hereafter the ‘2x2x2 game’), the first agent (signaller) observes that the world is in one of two possible states (state1 or state2) and broadcasts one of two possible signals (signal1 or signal2) which are observed by the second agent (receiver) who performs one of two possible actions (act1 or act2). If the acts match the state of the world (i.e. act1 if state1 or act2 if state2) then the players receive a greater payoff than otherwise.

Most importantly, though, the game theoretic results are unequivocal. There exist two Nash equilibria that are, in Lewis’s words, signalling systems where senders condition otherwise arbitrary signalling behaviour on the state of the world, and receivers act on those signals to secure the mutual payoff. The two

systems only differ on which signal gets to be associated with each state of the world¹. Huttegger (2007) and Pawlowitsch (2008) have shown that under certain conditions a signalling system is guaranteed to emerge under the replicator dynamics, a standard model of evolution to be discussed further in section 4.

Of course the degree to which Lewis' approach makes sense is the degree to which we have confidence in the interpretation and application of such a highly idealised model to the more complex target systems. The obvious worry is that by introducing more realistic features into the model one might break or significantly dilute previous findings on the evolution of signalling.

Not surprisingly, then, recent work on Lewis signalling games has investigated the many ways in which such de-idealizations could occur. Some deviations from the standard Lewis signalling game include: more and varied states of the world, the possibility of observational error or signal error, noisy signals, partial deviation in interest between senders and receivers, the reception of more than one signal, and so on. Many such concerns are dealt with favourably in Skyrms (2010), and in work by others. For example Bruner et al. (2014) generalizes beyond the 2x2x2 case and Godfrey-Smith and Martinez (2013) and Godfrey-Smith (2015) mix signalling games of common interest and conflict of interest. One complication of the Lewis signalling game (particularly important for our purposes) is that signalling systems are not guaranteed in the simple 2x2x2 case when the world is biased. In other words, when the probabilities of the world being in state1 or state2 are not equal, a pooling equilibrium in which no communication occurs between sender and receiver is evolutionarily possible.

2 Symmetry breaking

The focus here will be with the idealisation that sender and receiver are equally responsive in strategic settings. Senders and receivers (in the evolutionary treatment of such games) are two populations of highly abstract and constrained agency roles: all that signallers do on observing the state of the world is send a signal, and the receivers must act as though the world is in one or other of the sender-observable states. Of those two roles, it is the restriction on receivers which is the more problematic.

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 signalling role and execute it, whistling or gesturing to her counterpart. To play the receiver role, however, the second forager has to actually re-orient their attention (to some degree) and attempt to engage in appropriate behaviour for the world-state the first has observed (e.g. prey is to the east or to the west, etc.).

The Lewis signalling 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 signallers than

¹The other two possible outcomes of the game are 'pooling equilibrium', where the receiver plays act1 or act2 unconditionally.

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 their motivations (‘we should catch that animal’) to dictate a fairly clear course of action (‘try to direct the other agent’s behaviour’). But all the receiver gets is a whistle, gesture or other signal which (by stipulation) has no pre-established 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. Firstly, 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 signaller-observed states of the world, and their responsiveness to the strategic situation is therefore less satisfactorily modelled by the strictly symmetric payoff structures of standard signalling games. Call this the structural responsiveness concern.

Secondly, given the likely differences in informational states, goal-directness, workload and opportunity cost implications of sender and receiver roles, we can expect the mechanisms (cognitive and otherwise) which instantiate them to differ as well, quantitatively and qualitatively. This implies that we should not expect their update-responsiveness in any given game to be equal either. Yet the working evolutionary assumption is that senders and receivers update their strategies in an identical manner, modelled 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 Kim Sterelny’s (Sterelny, 2012) concerns about Skyrms (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 signalling 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 – taking an action which 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 behaviour 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 signalling games, such a gambit may, in many cases, be more attractive than attempting to respond optimally to a signal².

²It is worth noting here that the ‘hedgehog’ strategy in this Lewis signalling 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

This compliments the structural responsiveness concern: receivers (especially) might have other options of value which will stand in competition to those assumed in the standard signalling game. Something like these hedgehog strategies are plausible departures from the idealisation and should be expected on the part of the receiver given a realistic demandingness of the role. The question is whether (as Sterelny suspects) including hedgehog strategies might undermine the robustness of evolution toward signalling 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 signalling situations – such as predator-prey signalling systems or courtship displays among conspecifics. Signallers and receivers come to not just update their strategies, but to do so at faster or slower rates depending on the nature of the strategic encounter they are entwined in³.

It might seem that in David Lewis signalling games (as with games of common interest in general) the Red Queen effect should have no role to play. However any realistic interpretation of the Lewis signalling 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 will admit to different degrees of plasticity and evolvability – and will have a different set of cross-cutting tasks and utilities that will place their own demands upon them. Quick and easy signalling responses will have different pathways of update and adaptation than the (typically) more complex set of systems which appropriate receiver responses require.

The consideration of multiple use or adaptive reuse also makes the Red Queen hypothesis salient: it is wildly implausible that entirely separate cognitive systems would evolve to deal with competitive signalling situations and coordination-style situations. Cognitive structures which underpin sender or receiver behaviour will likely be subject to evolutionary pressures from competitive as well as cooperative situations, and the responsive nimbleness of sender and receiver strategies is therefore not guaranteed to be the same. 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 about how greater inferential sophistication on the receiver side is a feature of many primate communication systems. While these findings do

mediocre payoff. Thus the issues and trade-offs associated with the hedgehog strategy are general concerns not confined to just the Lewis signalling games. Thanks to [name redacted for review] for helping us better see this connection.

³An example of two groups adapting and evolving at different rates can be found in Richard Dawkin’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 due to 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.

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 signalling 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 a basis for our analysis is the pure-strategy 2x2x2 David Lewis signalling 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 signaller with two possible signals (V1 and V2), and a signal-observing receiver with two possible actions (AL and AR). If the receiver’s action matches the state of the world, then both signaller and receiver get a fixed positive success payoff, otherwise their payoff is zero. Signallers and receivers both have four pure strategies available to them (see table 1).

<i>S1</i>	Signal V_1 if L and signal V_2 if R
<i>S2</i>	Signal V_2 if L and signal V_1 if R
<i>S3</i>	Signal V_1 always
<i>S4</i>	Signal V_2 always
<i>S5</i>	Act A_L if V_1 and act A_R if V_2
<i>S6</i>	Act A_R if V_1 and act A_L if V_2
<i>S7</i>	Act A_L always
<i>S8</i>	Act A_R always

Table 1: Signaller and receiver strategies in the standard 2x2x2 common interest signalling game.

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:

$$X_i(t + 1) = X_i(t) \frac{F_i}{F_S}$$

where X_i is the i th sender strategy, F_i is the fitness of that strategy and F_S is the average sender strategy fitness. Likewise, for receivers:

$$Y_j(t + 1) = Y_j(t) \frac{F_j}{F_R}$$

where Y_j is the j th sender strategy, F_j 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 and no randomising or mutations are allowed.

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 vary relative to one other. In addition, the bias of nature is also varied, and we investigate the effects these three departures from the Skyrms/Lewis idealisation have on the evolutionary stability of signalling equilibria.

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 where 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 they go in the matching direction), or alternatively to abandon the hunt in order to obtain a less valuable resource they do 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 where it is systematically more likely that the prey is to the left or the right.

In the simple unbiased 2x2x2 signalling game, one of the two signalling equilibria is guaranteed to be reached under the replicator dynamics. In our notation, these equilibria are S1-R1 and S2-R2. 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 pooling equilibria. This will occur when there are initially few conditional signalling strategies 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 signalling system then disappears and the community is locked into a pooling equilibrium.

Not surprisingly, we found a similar effect with the hedgehog strategy as values of H, the payoff for A_H , becomes significant. The hedgehog strategy R5 is an additional unilateral response, and is able to draw some initial populations away from the signalling 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 1⁴.

⁴Note that the exact range of this effect, including the point at which the effect becomes significant and the y-intercept, are artefacts of the number of world-states and strategies in the model and therefore not general.

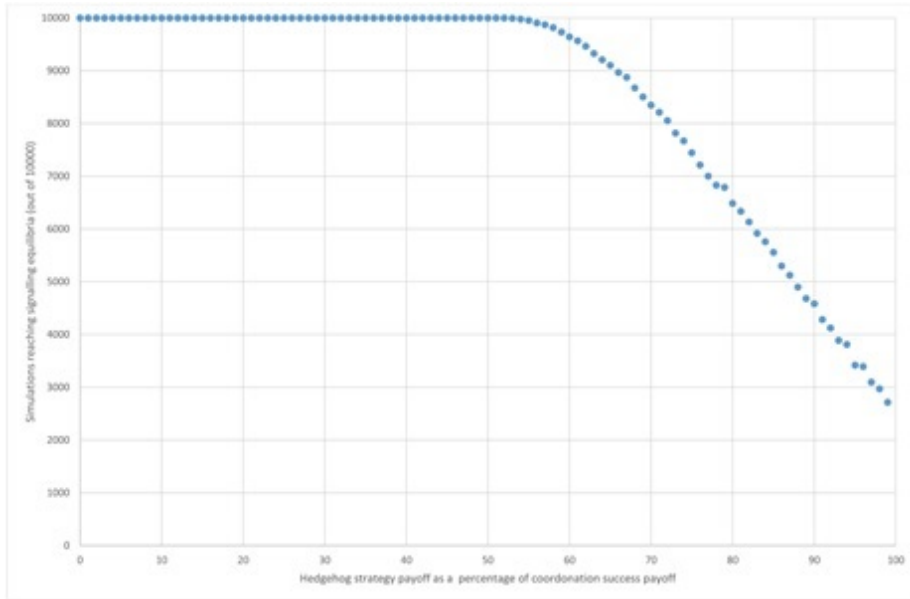


Figure 1: Effect of hedgehog payoff on proportion of signalling equilibria.

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 . The $H = 0$ curve has the expected n-shape, with perfect signalling being degraded as world-bias increases away from the mid-point of even bias between L and R . The inclusion of significant (i.e. $H \neq 0.5$) hedgehog payoffs decreases signalling at even bias. As nature becomes increasingly biased, however, the proportion of simulations that head to a signalling system does not go down. In fact we observe a ‘plateau’ followed by a gradual *increase* in the proportion signalling as nature becomes increasingly biased. However, once the bias becomes too extreme, the traditional pooling equilibrium becomes increasingly likely as the payoff associated with simply performing the appropriate act for the more likely state of the world approaches 1. This results in a steep decline in the proportion of simulations that result in signalling systems.

6 Generational asymmetry

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

$$X_i(t+1) = (1 - Z_S)X_i(t) \frac{F_i}{F_S} + X_i(t)Z_S$$

$$Y_j(t+1) = (1 - Z_R)Y_j(t) \frac{F_j}{F_R} + Y_j(t)Z_R$$

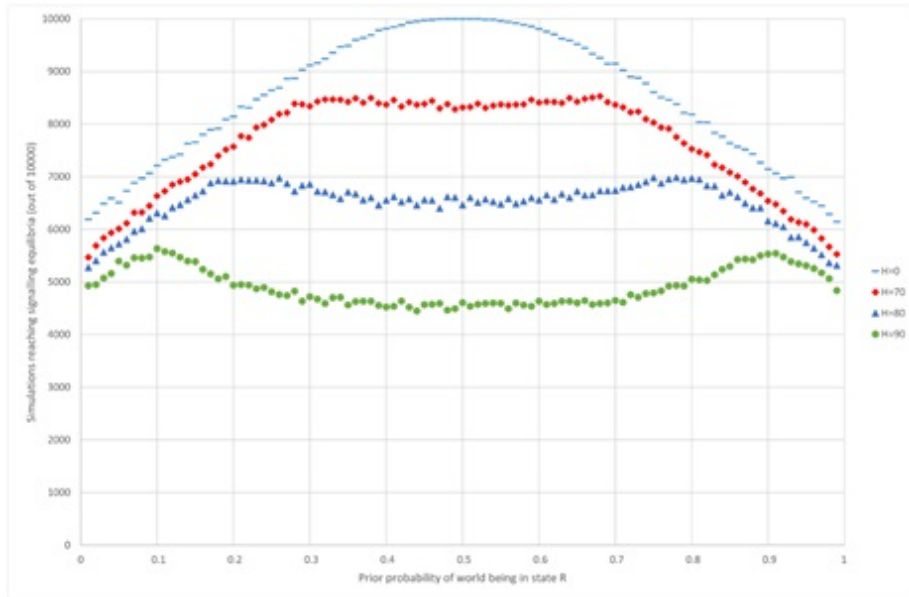


Figure 2: Effect of hedgehog strategy and bias of nature on proportion of signalling equilibria.

Note that when both Z_R and Z_S are zero there is no deviation from the standard replicator dynamics. Rates of changes are slowed as their values increase; for example setting $Z_S = .5$ halves the rate of change for sender strategies. Z_R (alone) being set to 1 means that the composition of the receiver population would not change over time, and only the sender population would evolve.

The result of introducing this generational asymmetry between senders and receivers is that signalling is more likely when sender strategies evolve faster than receiver strategies. This is illustrated in figure 3, where senders (Z_S) and receivers (Z_R) are slowed down to half and one-tenth speeds (with the other population unaltered) as the bias of nature is varied.

Slowing the evolution of the sender population leads to more pooling because, as before, receivers facing a sender population whose conditional signalling is low will begin to gravitate to the act that matches the more likely state of the world (and the threshold for ‘low’ is higher at higher bias). This evolutionary trajectory only reverses if conditional signalling increases rapidly enough to tip the fitness balance toward its matching conditional response, before that response is overpowered. Thus signalling becomes quite a remote possibility when bias is high and senders are slow, occurring in less than 10% of simulations for some parameter values. Slowing the evolutionary responsiveness of the receiver population evolves has the opposite effect – as senders will have time to adopt the best separating strategy given the mix of receiver strategies, and the receiver population slowly adjusts and a robust signalling system establishes. By a similar logic, it is easy to see that a quickly evolving sender population also mitigates against the effect of hedgehog strategies.

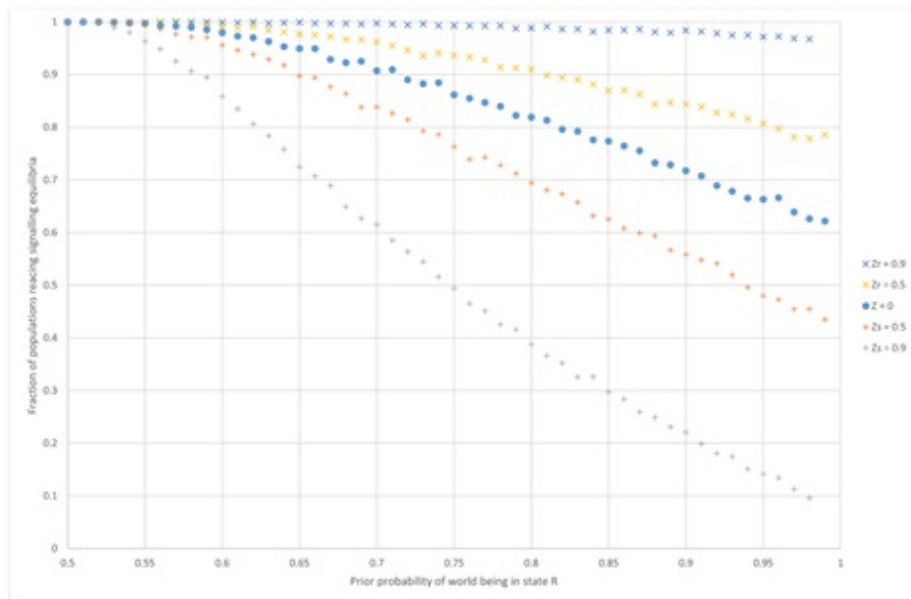


Figure 3: Effect of generational asymmetry and bias of nature on proportion of signalling equilibria.

7 Discussion

We have explored a few well-motivated departures from the highly idealized and simple Lewis signalling 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 which 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 interaction between hedgehog payoffs and bias shows that signalling-undermining effects are not strictly additive. Likewise, the situation is much less bleak when senders evolve at a faster pace than receivers. Interestingly, many scholars in the animal communications literature have noted a similar response asymmetry between sender and receiver in conflict of interest and partial conflict of interest signalling games. For instance, Owren, Rendall, and Ryan (2010) note that senders can easily adapt their signalling behaviour while receivers for the most part have responses to the stimuli produced by senders that are more difficult to change. Thus some have taken to think of signalling as primarily involving the manipulation of receivers by senders.

But this leaves us with an evolutionary puzzle. If there is a conflict of interest between sender and receiver, then what prevents receivers from increasing the speed at which they adapt to the behaviour of the senders? In other words, what explains the absence of an evolutionary arms race between sender and receiver? These are the exact circumstances we would expect the red queen hypothesis to apply. We believe the results of this paper may form the basis of

a novel explanation for this puzzling phenomena. 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 signalling system. When the interests of sender and receiver significantly diverge, however, we would expect this not to be the case since both parties now have reason to adapt at a faster pace than the other.

Yet individuals who routinely interact rarely find themselves playing either common interest or conflict of interest signalling games exclusively. As is well known by any parent, not all signalling 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 signalling games in contexts unrelated to mating. The point is that a variety of strategic scenarios can hold between sender and receiver, and there is no principled reason to think all interactions will involve perfect alignment or sizable conflict. If so, then a proportion of signalling interactions between sender and receiver may involve no conflict, a partial conflict, or a full conflict of interest. When the proportion of no or low conflict signalling games is significant, the generational asymmetry result from the previous section may hold to some degree. Both sender and receiver will then profit from the sender population evolving at a faster rate than the receiver population, and receivers do best to limit how responsive they are to senders so as to ensure the emergence of informative signalling systems when their interests do overlap. Thus, while it may appear puzzling as to why a receiver is not more responsive when her interests diverge from that of the sender, this confusion might be resolved when the interaction is put into context.

The robustness analysis considered in this paper has in some sense shown how fragile the evolution of signalling can be. Slightly altering the framework in a sensible fashion leads to significantly different results. While many variants of the baseline Lewis signalling game have been explored by philosophers in recent years, more work is required in order to better assess the prospect of signalling in realistic environments.

8 Acknowledgements

We thank Kim Sterelny, Ron Planer and the audiences at the Sydney-ANU Philosophy of Biology Workshop and the 2016 Meeting of the Philosophy of Science Association.

9 Bibliography

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. doi:10.1007/s11229-014-0535-x.

Dawkins, R., and J. R. Krebs. 1979. “Arms Races between and within Species.” *Proceedings of the Royal Society of London B: Biological Sciences* 205 (1161): 489–511. doi:10.1098/rspb.1979.0081.

- Godfrey-Smith, Peter. 1996. *Complexity and the Function of Mind in Nature*. Cambridge; New York: Cambridge University Press.
- Godfrey-Smith, Peter, and Manolo Martínez. 2013. “Communication and Common Interest.” *PLoS Comput Biol* 9 (11): e1003282. doi:10.1371/journal.pcbi.1003282.
- Hobaiter, Catherine, and Richard W. Byrne. 2014. “The Meanings of Chimpanzee Gestures.” *Current Biology* 24 (14): 1596–1600. doi:10.1016/j.cub.2014.05.066.
- 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.
- Martinez, Manolo, and Peter Godfrey-Smith. 2015. “Common Interest and Signaling Games: A Dynamic Analysis.” <http://petergodfreysmith.com/wp-content/uploads/2013/06/Martinez-GS-paper2-Dynamic-Preprint.pdf>.
- 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. doi:10.1007/s10539-010-9224-4.
- Pawlowitsch, Christina. 2008. “Why Evolution Does Not Always Lead to an Optimal Signaling System.” *Games and Economic Behavior* 63 (1): 203–26. doi:10.1016/j.geb.2007.08.009.
- Seyfarth, Robert M., and Dorothy L. Cheney. 2003. “Signalers and Receivers in Animal Communication.” *Annual Review of Psychology* 54 (1): 145–73. doi:10.1146/annurev.psych.54.101601.145121.
- Skyrms, Brian. 1996. *Evolution of the Social Contract*. Cambridge University Press. ———. 2010. *Signals: Evolution, Learning, and Information*. Oxford; New York: Oxford University Press.
- Sterelny, Kim. 2012. “A Glass Half-Full: Brian Skyrms’s Signals.” *Economics and Philosophy* 28 (01): 73–86. doi:10.1017/S0266267112000120.
- Van Valen, Leigh. 1973. “A New Evolutionary Law.” *Evolutionary Theory* 1 (1-30). <http://tmtfree.hd.free.fr/albums/files/TMTisFree/Documents/Biology/A>