Cost, expenditure and vulnerability

Justin P. Bruner[1] · **Carl Brusse[1]** · **David Kalkman[1]**

Abstract The handicap principle (HP) stipulates that signal reliability can be maintained if signals are costly to produce. Yet empirical biologists are typically unable to directly measure evolutionary costs, and instead appeal to expenditure (the time, energy and resources associated with signaling behavior) as a sensible proxy. However the link between expenditure and cost is not always as straightforward as proponents of HP assume. We consider signaling interactions where whether the expenditure associated with signaling is converted into an evolutionary cost is in some sense dependent on the behavior of the intended recipient of the signal. We illustrate this with a few empirical examples and demonstrate that on this alternative expenditure to cost mapping the traditional predictions of HP no longer hold. Instead of full information transfer, a partially informative communication system like those uncovered by Wagner (Games 4(2):163–181, 2013) and Zollman et al. (Proc R Soc B 20121878, 2012) is possible.

Keywords Handicap principle · Evolutionary biology · Game theory

Introduction

One of the central problems in evolutionary biology concerns the stability of honest communication in contexts where the interests of those involved do not entirely overlap. Communication appears to be a remote possibility due to the fact that the

✉ Justin P. Bruner
Justin.bruner@anu.edu.au

Carl Brusse
carl.brusse@anu.edu.au

David Kalkman
david.kalkman@anu.edu.au

[1] Australian National University, Canberra, Australia

sender often has incentive to deceive and, given this behaviour, the receiver typically does best to ignore her counterpart. Yet despite these misaligned interests, honest communication frequently occurs, taking place between potential mates, predator and prey, as well as parent and offspring. What accounts for such high levels of honesty?

The handicap principle (HP) is one plausible explanation of honesty (Grafen 1990; Zahavi 1975). Briefly, the HP stipulates that signal reliability can often be maintained if signals are costly to produce. Importantly, a signal is said to be costly if emitting the signal in some way reduces the fitness of the sender. The introduction of such costs can ensure that deception, while still possible, is no longer profitable. Yet despite its warm reception, a variety of empirical issues have threatened to discredit HP. Namely, experimentalists have had difficulty registering the existence of such signal costs (see Zollman 2012 for an overview). This is further complicated by the fact that empirical biologists are typically unable to directly measure cost, and instead look to what is often referred to as expenditure (the time, energy and resources associated with signaling behaviour). How expenditure translates into cost is a rather complex issue, and for this reason many are hesitant to take the existence of expenditure as evidence confirming HP (Kotiaho 2001).

Indeed, the distinction between expenditure and cost is something both experimentalists and theorists should be mindful of. In the course of this paper we demonstrate that the game-theoretic predictions regarding the likelihood of information transfer and honesty hinges on the particular way expenditure maps to cost. Specifically, we consider signaling interactions where whether the time, energy and resources associated with signaling convert into an evolutionary cost depends on the behaviour of the intended recipient of the signal. There is ample empirical evidence that such *vulnerability handicaps* (as they are sometimes referred to in the literature) exist in a variety of strategic settings. We construct a simple partial conflict of interest signaling game and demonstrate that if the costs associated with signaling in part depend on the response of the receiver, the high levels of honesty predicted by the handicap principle is no longer attainable. This is not to say that communication breaks down. In fact, a partially honest equilibrium similar to that uncovered in a series of recent papers (see Wagner 2013; Zollman et al. 2012; Huttegger and Zollman 2010) remains. We register the evolutionary significance of this partially honest equilibrium and determine that under a wide range of parameters this arrangement is the most likely outcome of a simple evolutionary model. *Overall, this analysis highlights the crucial role the overlooked relationship between expenditure and cost plays in the study of animal communication.*

This paper will proceed as follows. "Cost, expenditure and signals" section discusses expenditure, cost and their relationship to HP, as well as briefly surveys a number of attempts to vindicate HP by measuring both expenditure and cost. Vulnerability handicaps are discussed in "Vulnerability signals" section, and it is in this section that we introduce a number of empirical cases which motivate the novel partial conflict of interest signaling game considered in "A new signaling game" section. An analysis of this new signaling game is provided in "A new signaling game" and "Evolutionary significance" sections. In "Vulnerability signals versus intrinsically costly signals" section we attempt to determine when vulnerability

handicaps or traditional costly signals will be utilized in cases where senders have both signal types at their disposal. "Discussion" section concludes.

Cost, expenditure and signals

Unobservable features of organisms tend to play an important role in many strategic interactions. As a result, individuals often have incentive to exaggerate or withhold information regarding these unobservable features. For instance, potential mates have reason to overstate their underlying genetic quality. Yet honest communication is often the norm in these partial conflict of interest cases.[1] What then explains the sustained reliability of signals?

One major advance in biological theory was the development of the so-called handicap principle (HP). Briefly put, HP stipulates that honest communication in partial conflict of interest scenarios can often be sustained if signals are differentially costly. To see this, consider a simple case in which two types of senders, those of low and those of high genetic quality, both attempt to convince a potential mate they are high quality. If signals are differentially costly and low-types must pay a greater cost than high-types to send the signal, then it is possible that only high-types will have incentive to signal, providing receivers with a means of distinguishing between the two kinds of senders.[2]

For instance, the extravagantly elongated and decorated tail of the male peacock is a classic example of a (mating) signal kept honest by the differential cost of signaling. The cost of developing, maintaining and displaying such an array falls more heavily on low quality males than high quality males, such that it is in the adaptive interests of only the latter to signal (Searcy and Nowicki 2005). This logic underlying HP has been widely accepted and discussed in the literature and we examine the specifics in more detail with the help of a simple formal game-theoretic model in "A new signaling game" section.

One aspect of HP worth highlighting is that signals only count as costly if, holding fixed the response of the receiver, they lower the *biological fitness* of the target organism. Counting costs (and benefits) in this fashion is necessary because HP is an *evolutionary* explanation of honesty. Thus signals are said to be costly if they reduce either the survival probability or future reproductive success of the target organism. Not surprisingly, biologists interested in testing HP have had great difficulty measuring the costs associated with signaling behaviour. As a result, empirical biologists have instead turned to expenditure—the time, energy and resources associated with signaling—as a sensible proxy for evolutionary cost (Ryan 1988; Pretwich 1994).

---

[1] We follow the animal communications literature here and use the phrase 'honest communication' to refer to the scenario in which receivers can reliably infer the unobservable feature of interest based on the sender's signaling behavior. In general, the notion of reliability, honesty and deception is a tricky matter when dealing with non-intention based communication. We do not address these issues in the course of this paper.

[2] This basic story can easily be generalized to more complicated cases in which, for instance, there are a continuum of sender types. See Grafen (1990) for instance.

While signaling displays often do require significant time and energy on the part of the sender, the exact connection between expenditure and cost has only recently been investigated. Kotiaho (2000) discusses a variety of empirical studies which seem to be on track. For instance, Moeller and de Lope (1994) examine the fitness costs associated with barn swallow mating signals, in which males develop elongated tails that are ostensibly an indicator of genetic quality. It is clear that these long tails require the agents to exert more effort to move about and require more energy to maintain. Moller and de Lope find that this expenditure associated with tail length does appear to translate into a real fitness cost for the sender.[3]

Yet the way in which expenditure translates into evolutionary cost is of interest to theoreticians as well as empirical biologists. In particular, we contend that a clearer distinction between expenditure and cost will sharpen our understanding of the conditions under which honest communication is possible. To see this, note that in the above examples it is assumed that there is a rather simple expenditure to cost mapping that is independent of the behaviour of the intended recipient of the signal. According to this mapping, expenditure is converted into cost at some rate, with greater expenditure resulting in greater signal cost. Yet as we will demonstrate in the next section, there are a variety of instances in which the simple expenditure to cost mapping just described is not appropriate. In particular, there are many scenarios in which whether or not the expenditure associated with signaling becomes a fitness cost to the sender *depends in some way on the response of the receiver.* Such cases problematize the receiver-independent expenditure-to-cost mapping that is tacitly assumed by many in the animal communications literature. Moreover, as we argue in "A new signaling game" and "Evolutionary significance" section, this alternative mapping leads to drastically different predictions regarding the likelihood of honest communication.

Vulnerability signals

Thus far we have considered signals that are costly regardless of the behaviour of the intended recipient of the signal. Following Searcy and Nowicki, we refer to such displays as 'receiver-independent' and distinguish them from so-called 'receiver-dependent' signals in which the cost associated with a display is in part determined by the response of the receiver (Searcy and Nowicki 2005).

In what follows, we focus on a particular type of receiver-dependent signal, so-called vulnerability signals.[4] Vulnerability signals function to place the sender in a position of disadvantage such that the sender is at increased risk of harm should the receiver respond in a particular fashion. We illustrate this below by drawing on two examples of predator–prey signaling systems from the literature.[5] Note that both

---

[3] Of course this is not a vindication of HP for it must be further shown that the costs are differential—i.e., that low-quality males displaying the signal incur a higher fitness penalty than high-quality males.

[4] In addition to vulnerability signals, so-called retaliatory signals or retaliatory handicaps are also examples of a receiver-dependent signal. See Molles and Vehrencamp (2001).

[5] In general, those working in animal communication tend to ignore vulnerability signals. We think they are a bit more prevalent than previously thought (as evidenced by the examples provided in this section)

scenarios conform to the same general pattern. Namely, the energy, time and resources associated with signaling (i.e., expenditure) temporarily places the sender at great risk should the receiver respond in a particular fashion.

The 'pushup' display given by the lizard *Anolis cristatellus* is one such example. Here, upon spotting a predator, the lizard signals by becoming stationary and moving its body up and down in a vertical plane by flexing and extending its legs. Its natural predator, the snake *Alsophis portoricensis*, often responds to the signal by stopping its approach and observing the lizard before deciding whether to continue its pursuit. If the snake does select to engage, the lizard can often escape death by struggling violently after capture, dragging its pursuer along as it attempts to flee.

Since interactions between lizard and snake often result in a long, drawn-out grappling match, it is not surprising that a lizard's ability to escape from a predator after capture correlates with its endurance capacity (Bennett and Huey 1990). Thus engaging in a physically demanding activity (such as pushups) that reduces the lizard's endurance capacity before a potential struggle places the organism at risk. Physically exerting oneself *before* the scuffle decreases the likelihood of escape.[6] Note, however, that this signaling display only negatively impacts the fitness of the sender when the receiver decides to give chase. In other words, the display on its own does not appear to be costly.[7] Engaging in the display does, however, ensure the organism is vulnerable (by in this case, burning through its energy reserves), thereby increasing the chance of death should the snake aggress. In other words, the expenditure associated with the display is translated into a real fitness cost only in those cases where the predator attacks the lizard.

Moreover, the evidence suggests pushup displays are in fact communicative (and moreover, partially honest). Leal (1999) finds a positive correlation between the number of pushups a lizard gives in the presence of a snake and its endurance capacity, suggesting that pushups are an honest indication of evasiveness. Furthermore, it has also been determined that lizards engaged in the pushup display are significantly less likely to be attacked than those not engaged in the display. Taken together, these findings suggest that the pushup display is best seen as a signal of the underlying evasiveness of the signaller.

Other canonical predator–prey signaling systems arguably have similar features to that of the *Anolis cristatellus* pushup display. Consider the stotting behaviour of the gazelle. Upon spotting a predator, gazelles often make prodigious vertical leaps into the air. Such stotting displays do not appear to help the gazelle outrun the predator. In fact, as is the case with the pushup display, the time and energy required to produce the display may reduce the likelihood of evasion. Yet the display on its own is not costly—it is only when stotting is meet with a particular response from

Footnote 5 continued

and think they also highlight important assumptions typically made at the core of costly signaling theory (such as the assumed straight-forward relationship between expenditure and cost).

[6] Brandt (2003) found endurance of lizards fell below baseline levels immediately after a signaling display, meaning the display is in fact hampering the lizard's ability to effectively fend off predators.

[7] It is worth noting here that we do not have any definitive proof that the push display on its own is not costly. From all descriptions of the display we've encountered in the literature, however, we feel safe asserting that the production of the signal alone does not reduce the fitness of the sender.

the receiver that the signaling display negatively impacts the expected fitness of the sender.

As mentioned earlier, both of the above cases conform to the same basic pattern, namely, that the expenditure associated with the display places the sender in a worse position should they have to engage with or flee from the predator.[8] Both cases involve a more complicated expenditure to cost mapping than is typically assumed in the literature on animal communication. Whether the energy, time and resources associated with signaling is converted into a fitness cost is in large part mediated by the receiver. Finally, in both cases it appears these vulnerability signals result in (to some degree) honest communication. Thus it appears vulnerability signals are in fact capable of generating some level of honesty between parties when the interests of those involved conflict. We now construct a simple game-theoretic model of these interactions and illustrate that the game-theoretic predictions on this alternative expenditure to cost mapping deviate from previous work on the evolution of honest communication.

A new signaling game

We introduce our new signaling game by way of contrast with standard costless and costly signaling games.

The familiar extended form of a two-player signaling game with variable receiver-independent signal costs and partial conflict of interest is illustrated in Fig. 1. Nature assigns player one, the *sender,* to one of two types (high or low). Player one has two actions available, to signal (S) or not signal (N) and four possible pure strategies, summarised in Table 1. Player two, the *receiver*, can observe player one's action but not their type and has two actions available, which we call "treat as high" (H) or "treat as low" (L). The receiver's four possible strategies are also included in Table 1.

The partial conflict of interest between sender and receiver means that both prefer the receiver treat the sender as a high type when the sender is in fact a high type, but disagree as to what the best response is when the sender is a low type. When the receiver-independent, signal cost for high types (c_h^i) and low types (c_l^i) are both set to 0 then talk is cheap and honest communication (i.e. 'separating equilibria', in which senders conditionalize their behaviour on the state of the world and receivers conditionalize on the behaviour of the sender, to the advantage of both) cannot be stabilised. Differential signal costs can prevent this collapse to dishonesty by making it prohibitively costly for low types to send the signal, allowing for the classic separating equilibrium in which senders signal just in case they are a high type and receivers only treat as a high type upon receipt of a signal. The separating equilibrium is possible if c_l^i is greater than 1, while c_h^i is less than 1. As mentioned in

---

[8] Note that as has been acknowledged in the literature, expenditure is not the only means of placing individuals in a vulnerable position. Consider the predator–prey signal sent by the guppy poecilia reticulate, often referred to as a 'predator approach' signal. In this case the guppy, upon spotting a predator, moves toward the predator. This behavior is taken in the literature as communicative and can be seen as an instance of a vulnerability signal (that does not involve any expenditure).
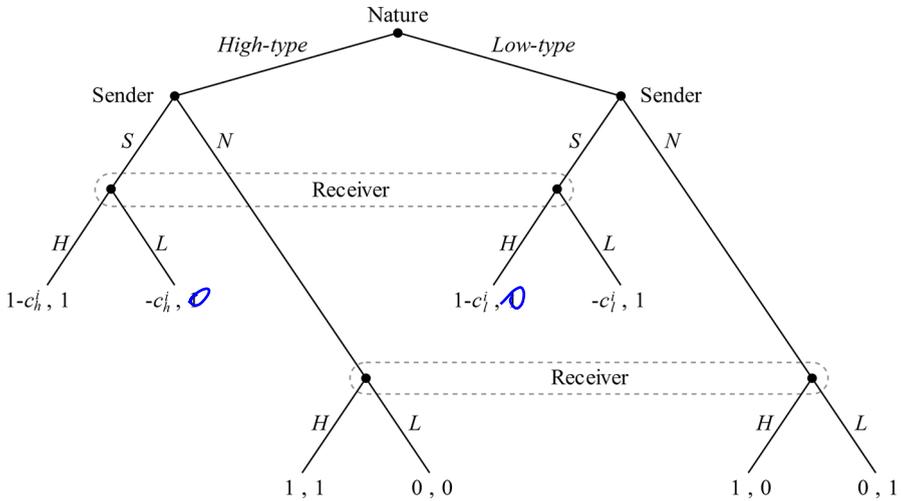
Fig. 1 A signal-response game with receiver-independent costs

Table 1 All possible pure strategies in the signal-response games pictured in Fig. 1

Strategy	Description
S1	Signal if high type, do not signal if low type
S2	Signal always
S3	Never signal
S4	Signal if low type, do not signal if high type
R1	Treat as high type if signal, treat as low type if no signal
R2	Always treat as low type
R3	Always treat as high type
R4	Treat as low type if signal, treat as high type if no signal

the previous sections, differentially costly signals can therefore allow honest communication despite the conflict of interest. Furthermore, as is common in the literature, signals are assumed to be receiver-independent, meaning the cost associated with signaling is "imposed regardless of whether or how receivers respond" (Searcy and Nowicki 2005: 15). We refer to such signals as either receiver-independent signals or intrinsically costly signals, as what determines the fitness cost associated is internal to the sender.[9]

Note that this classic separating solution considers pure strategies only. Zollman et al. (2012) demonstrate that when c_l is less than 1, there exists a mixed strategy hybrid equilibrium which allows for partially honest communication. This hybrid

---

[9] Calling signals that are costly regardless of the response of the intended recipient of the signal 'intrinsic' is a bit of a misnomer because the cost associated with signalling could be due to a third-party eavesdropper (in which case costs are neither intrinsic to the sender, nor due to the response of the intended recipient).

equilibrium (hereafter the *intrinsic cost hybrid*) exists when the sender mixes between strategy S_1 (signal only if high type) and S_2 (always signal), while the receiver mixes between R_1 (only treat as high type if signal) and R_2 (always treat as low type). In other words, the signaller always signals if a high type and *sometimes* signals if a low type, while the receiver sometimes treats as a low type if a signal is sent, but always treats as a low type in the absence of a signal. So while the receiver cannot reliably determine the underlying type of the sender, signals nonetheless carry some information about the underlying type of the sender. The intrinsic cost hybrid is illustrated in Fig. 2.

We now introduce a novel signaling game, the so-called 'vulnerability game' displayed in Fig. 3. While similar to the costly signaling game discussed above, the vulnerability game departs from the costly signaling game in one important respect. Namely, unlike the costly signaling game, senders do not always incur a cost upon sending a signal. Whether signals negatively impact the fitness of the sender depends on the response of the receiver. In particular, senders incur a cost upon emitting a signal if the receiver treats the sender as a low type. Relating this back to our example from the previous section involving the lizard *Anolis cristatellus*, the push-up display given in the presence of a snake only negatively impacts the evolutionary fitness of the lizard if the predator decides to give chase (i.e., the predator treats the sender as a low type or un-evasive type they can easily capture). The display requires both time and energy to perform, but the expenditure associated with the display is not significant enough to on its own result in a fitness cost for the sender. Only when the sender must flee from its predator is the expenditure associated with signaling translated into a fitness cost.

Note that just as in the traditional receiver-independent signaling game displayed in Fig. 1, low- and high-type senders face a slightly different cost structure from each other in the vulnerability signaling game. If low-types are pursued after engaging in the signaling display, they incur a higher cost (c_l^v) than their high-type counterpart would have had they signalled and been likewise pursued (c_h^v). In other words, while signaling places both types of senders in a vulnerable position, the signaling display ensures low-types are especially vulnerable.

It is easy to see that in the vulnerability game, there is no classic separating equilibrium (S_1, R_1) in which only high-type senders engage in a signaling display and receivers classify only those sending the signal as a high-type. Separating arrangements are not possible because the sender does best to employ S_2 (always signal) if the receiver employs R_1 (treat as high if signal, treat as low if no signal). However, some level of information transfer is nonetheless possible in the vulnerability game. Just as in the intrinsic costs signaling game, there is a partially informative 'hybrid' equilibrium in the vulnerability game. At this *vulnerability hybrid equilibrium*, senders mix between S_1 and S_2, always signaling when a high type and sometimes signaling when a low type. Receivers, on the other hand, mix between R_1 and R_2, always interpreting the absence of a signal as admission one is of low quality, and sometimes treating those who signal as high quality.

It is easy to show that if x is the prior probability a sender is a high type, then at the vulnerability hybrid equilibrium the sender utilizes S_1 with probability $\alpha_v = \frac{x}{1-x}$
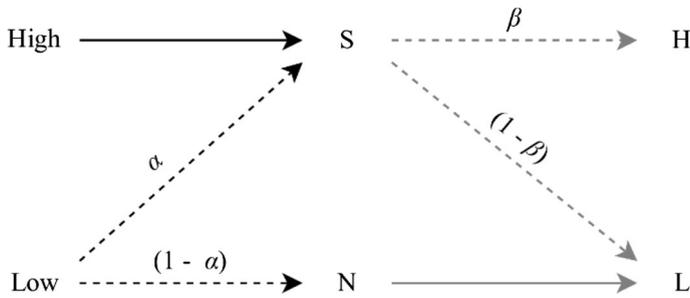
Fig. 2 Illustration of sender (*black*) and receiver (*grey*) strategies at the hybrid equilibrium
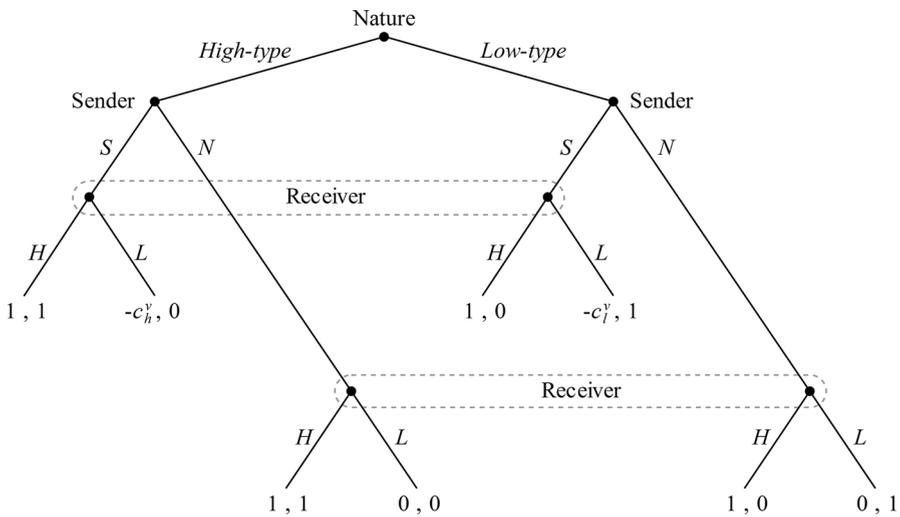

Fig. 3 A signal-response game with receiver-dependent signal costs

and S_2 with probability $1 - \alpha_v$. Likewise, the chance the receiver performs R_1 at the vulnerability hybrid is $\beta_v = \frac{c_l^v}{1+c_l^v}$ and R2 is performed with probability $1 - \beta_v$. Finally, the condition $c_h^v \leq c_l^v$ must hold for the vulnerability hybrid to be stable. In other words, a signaling display must make low types more 'vulnerable' than high types performing the same display. Table 2 compares the conditions for the vulnerability hybrid equilibrium to those for the intrinsic cost hybrid outlined in Zollman et al.

Note that the relationship between c_l and the probability with which receivers utilize R_1 (β_c and β_v) is the key difference between the intrinsic cost hybrid and the vulnerability hybrid. At the intrinsic cost hybrid, c_l^i is equal to β_c, meaning c_l^i cannot exceed 1. However, at the vulnerability hybrid c_l^v can take on rather extreme values, approaching infinity as β_v approaches 1. In other words, our vulnerability signaling game can potentially represent *extremely risky behaviour as communicative*, so long as receivers very rarely treat those sending the signal as a low type.

Table 2 Conditions for hybrid signaling equilibrium in the costly signaling and vulnerable signaling games

Conditions for intrinsic cost hybrid (Zollman et al.)	Conditions for differential vulnerability hybrid
$\alpha_c = \frac{x}{1-x}$	$\alpha_v = \frac{x}{1-x}$
$c_l^i = \beta_c$	$c_l^v = \frac{\beta_v}{1-\beta_v}$
$c_h^i \leq c_l^i$	$c_h^v \leq c_l^v$

Finally, a pooling equilibrium (in which senders all engage in the same signaling behaviour) is always possible in the vulnerability signaling game. In this case, neither low nor high type senders signal, and receivers treat senders as low types if x < .5 and high types if x > .5. Figure 4 illustrates the comparative locations of equilibria for both the vulnerability and standard costly signaling game.

Evolutionary significance

We now turn our attention to the evolutionary significance of the vulnerability hybrid. The fact an arrangement is a Nash equilibrium in no way guarantees evolutionarily significance, and in this section we determine in what sense the vulnerability hybrid is stable.
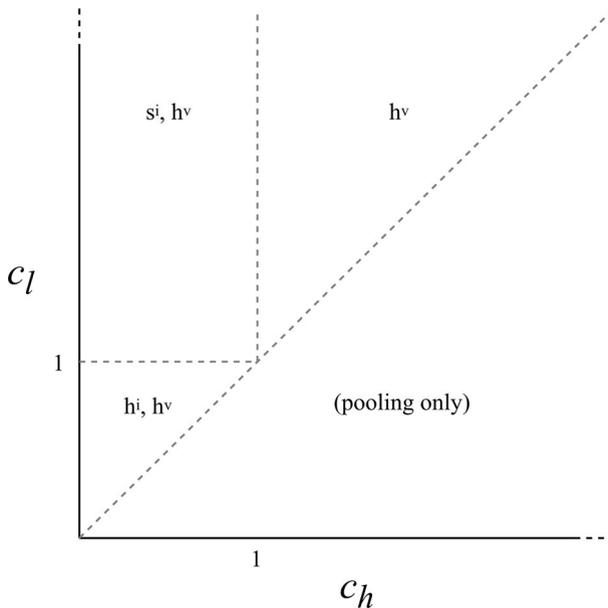
Fig. 4 An illustration of the location of separating (s) and hybrid (h) equilibria for the costly signaling and vulnerability signaling game. Note that in both games pooling equilibria exist for all combinations of parameter values

To precisely determine the importance of the vulnerability hybrid we appeal to the two-population replicator dynamics. The replicator dynamics is a common tool in both evolutionary biology as well as the social sciences, and can be interpreted as a model of either cultural or biological evolution (Weibull 1995). The replicator dynamics provides us with a means of tracking the strategic behaviour of a community over time. As a rule of thumb, strategies that do better than the population average increase in size, while those that do worse than the population average decline in number. The two-population version of the replicator dynamics is commonly used to study the evolution of signaling behaviour (a population of senders is tasked to repeatedly interact with a second population of receivers).

Our analysis of the stability properties of the vulnerability hybrid under the replicator dynamics mirrors the analysis of Huttegger and Zollman (2010), Zollman et al. (2012) and Wagner (2013) on the intrinsic cost hybrid (where, recall, low-type senders mix between emitting the signal and not emitting the signal and receivers, upon receipt of the signal, mix between treating the sender as a high-type and low-type). We discuss the details of our mathematical analysis in the "Appendix" and in the remainder of this section unpack our two central results. The first pertains to the behaviour of the community on the so-called 'hybrid plane.' The hybrid plane is the region of strategy space consisting of just those strategies that constitute the vulnerability hybrid equilibrium (i.e., S_1, S_2, R_1 and R_2). We find that on this hybrid plane there are closed orbits circling the vulnerability hybrid equilibrium.

This cycling is in some sense unsurprising. Recall that in the vulnerability game signals are costly only when a sender is called on their bluff, so to speak, and treated like a low type. Thus receivers 'police' the use of the signal by treating those sending the signal as low-types in order to ensure that the majority of individuals emitting the signal are in fact high types. Yet whether it pays to police as a receiver crucially depends on the proportion of senders misusing the signal by sending it when a low type (if only high types signal, then receivers do best to treat those sending the signal as a high type). Likewise, whether one does best to send the signal as a low type hinges on the proportion of receivers engaged in policing the use of the signal (i.e., the proportion of receivers who utilize the strategy 'always treat as low type.').

Taken together, these incentives result in a cyclic dynamic that has been observed in a variety of other contexts, from models of peer-policing in science (Bruner 2013) to corruption cycles in politics (Bicchieri and Duffy 1997). Furthermore, in the "Appendix" we show that these closed orbits exist when the proportion of high types is less than ½, which is coincidentally a necessary condition for the existence of the vulnerability hybrid equilibrium. Hence, whenever the hybrid equilibrium exists there will be closed orbits around it on the hybrid plane under the replicator dynamics.

Although it is unclear whether we should expect to see cycling in real world communities, our next finding indicates that we have good reason to think that many real world populations may gravitate toward and remain on the hybrid plane. To determine this we show that there is evolutionary pressure ensuring that populations on the hybrid plane near the vulnerability hybrid equilibrium remain on the plane (see the "Appendix" once again). In other words, mutant strategies such as R_3 or S_4
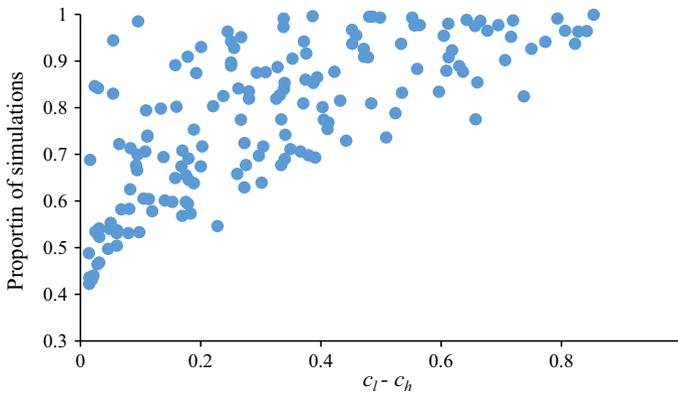
Fig. 5 Proportion of simulations that result in the vulnerability hybrid equilibrium for randomly generated parameters as a function of the difference in signal cost (i.e., $c_l^v - c_h^v$). Populations not at the hybrid end up at the pooling equilibrium

do particularly poorly against a population near the hybrid equilibrium and thus go to extinction when rare. Thus states near the hybrid plane will converge to the plane and once on the plane, begin to cycle around the vulnerability hybrid equilibrium. In this sense the vulnerability hybrid equilibrium is evolutionarily significant and, as we will now discuss, a sizable portion of the strategy space will in fact converge to the plane.

Figure 5 showcases the results of a simulation that establish the size of the basin of attraction of the hybrid plane and pooling equilibrium (where high- and low-type senders behave similarly) in the vulnerability signaling game. Using the discrete-time two-population replicator dynamics we track the trajectory of 1000 randomly generated initial population states for a randomly determined set of parameters. Two things are worth noting. In general, the likelihood of partial information transfer is exceptionally high—the community often makes it to the hybrid plane and avoids the pooling equilibrium. In fact, for the vast majority of parameter values we examined, well over 50% of the strategy space goes to the hybrid plane. Second, as the value of $c_l^v - c_h^v$ (the x-axis of Fig. 5) increases it becomes all the more likely that the hybrid plane is reached.[10] In other words, when a signaling display makes low types exceptionally vulnerable relative to high types, partially honest communication is likely to emerge.

The case of the *Anolis* lizard discussed in "Vulnerability signals" section is arguably an example of the hybrid equilibrium. As we argued, the *Anolis* lizard's push-up displays are best conceived of as vulnerability signals since they do not appear to on their own reduce the fitness of the sender. It is only when the predator decides to give chase that the expenditure associated with the display negatively impacts the lizard's chance of survival (and thus imposes a real evolutionary cost on

---

[10] Parameter values were selected randomly via the following procedure. The value of c_l was determined by a draw from the uniform distribution on the unit interval. c_h^v was the determined by a draw from the uniform distribution spanning from zero to the value of c_l^v. The probability senders are high-types was determined by a draw from the uniform distribution spanning from zero to 0.5.

the sender). Leal (1999) has shown that while high-type lizards (those with high endurance capacity) on average produce significantly more push-ups than low-types, some low-type lizards nonetheless engage in the signaling behaviour. Likewise, snakes, upon observing a lizard not engaged in the display, almost always attack. Snakes are significantly less likely to pursue, however, when the lizard in question produces the display. This nicely corresponds to the vulnerability hybrid equilibrium.

Finally, it is worth noting that prior work on vulnerability handicaps by and large focus on rather complicated strategic scenarios in which agents compete for some contested resource (see, for instance, Enquist et al. 1985; Waas 1991a, b). In these circumstances, an agent takes on both the role of sender and receiver and must determine whether to escalate or flee from potentially violent encounter on the basis of their underlying type, their signaling behaviour and the signaling behaviour of their counterpart. Thus, while Adams and Mesterton-Gibson (1995) develop a game-theoretic model of these scenarios and find that partial honesty is possible at equilibrium, one may be hesitant to extract general conclusions about the likelihood of honest communication from this complex set-up. Our simpler vulnerability signaling game, on the other hand, corresponds to a variety of real world examples and allows for a better understanding of vulnerability handicaps. Additionally, past work on vulnerability has relied solely upon so-called ESS (referring to the notion of an 'evolutionarily stable strategy') or 'static' methodology. In this framework, researchers determine the strategies or set of strategies that withstand invasion by mutants. As Huttegger and Zollman (2013) have appreciated, this approach, while useful in many ways, overlooks evolutionarily significant outcomes such as the vulnerability hybrid equilibrium discovered in this paper. Furthermore, ESS methodology does not provide the theorist with a means of determining how likely a particular stable arrangement is to emerge. Dynamic game-theoretic tools, such as the replicator dynamics and models of individual and social learning, are necessary to accomplish this task.

Vulnerability signals versus intrinsically costly signals

As shown in the previous section vulnerability signals as well as intrinsically costly signals can result in partially honest communication when the interests of sender and receiver are not perfectly aligned. Yet what explains the uptake of vulnerability signals in certain strategic scenarios and intrinsically costly signals in others? The question of when a particular *type* of signal is adopted over alternative signal types has received scant attention in the literature despite the fact that a variety of high-profile biologists have emphasized its importance.[11]

Lachmann et al. (2001) and Fraser (2012) briefly address this question, and in particular discuss when evolution will select for the use of conventional signals (i.e., intrinsically costless signals, such as language) over intrinsically costly signals. Lachmann et al. contend that conventional signals are likely to be used in those cases

---

[11] See for instance, chapter 5 ('The evolution of signal form') in Maynard-Smith and Harper (2003).

where receivers can easily verify 'the integrity of the message' emitted by the sender. While it is difficult to verify the underlying genetic quality of a potential mate, for instance, it is possible to verify whether a conspecific is lying about the presence of berries on the other side of the hill—all one has to do is simply go check. Thus while signal costs in the former case are crucial, signal costs in the latter setting appear superfluous, and for this reason we'd expect conventional signals to be the norm.

We turn our attention away from conventional signals and instead focus on when vulnerability as opposed to intrinsically costly or receiver-independent signals will be utilized. To do this we informally consider below a setting in which within a community some senders utilize intrinsically costly signals, while others rely upon vulnerability signals. Note that this assumes for simplicity that a particular sender will only rely upon one type of signal.[12] With this idealization in hand, we now informally outline general conditions for when vulnerability signals are likely to be utilized.[13]

We first determine the average fitness for sender and receiver at the vulnerability hybrid equilibrium and the traditional intrinsically costly separating equilibrium (recall, at the separating equilibrium only high-type senders signal and only those who signal are treated as a high-type by the receiver). All else held constant, we should expect that the intrinsically costly signals used in the separating equilibrium are more likely to be utilized when both sender and receiver prefer the separating equilibrium to the vulnerability hybrid equilibrium.[14] Note that this isn't to say that the separating equilibrium is more likely than the vulnerability hybrid equilibrium when both sender and receiver prefer the separating equilibrium. Instead, we claim that the separating equilibrium is more likely to emerge when both agents prefer it to the vulnerability hybrid than the case in which one, or both, of the parties rank the separating equilibrium below the vulnerability hybrid.[15] By the same token, vulnerability signals are more likely when both parties have higher fitness at the vulnerability hybrid equilibrium. While this rule of thumb is of course a poor substitute for a thorough evolutionary game-theoretic analysis, it provides us with some insight as to when vulnerability and intrinsically costly signals are utilized.[16]

---

[12] We think this is a somewhat reasonable starting assumption since we know of no case in which both intrinsically costly signals and vulnerability signals are utilized in the same strategic context.

[13] A formal model investigating when certain types of signals will be taken up is mathematically challenging and thus will not be explored in the course of this paper.

[14] We compare the vulnerability hybrid to the separating equilibrium and not the intrinsically costly hybrid equilibrium because both sender and receiver do better at the separating equilibrium than at the intrinsically costly hybrid.

[15] It is clear that the traditional separating equilibrium could be more likely to emerge (i.e., have a larger basin of attraction) than the vulnerability hybrid equilibrium despite the fact that both parties prefer the vulnerability hybrid to the separating equilibrium. For instance, in the so-called Stag Hunt game, all prefer the stag hunting equilibrium to the hare hunting equilibrium, yet the hare hunting equilibrium is more likely under the replicator dynamics due to the inherent risk attached to hunting stag (see Skyrms 2004 for more on this point).

[16] Kane and Zollman (2015) construct a model to investigate when different types of signals will go to fixation, but do not consider the vulnerability signals we've addressed in this paper. Holman (2012) considers a similar question in a slightly different mathematical framework and Huttegger, Bruner and Zollman (2015) explore 'hybrid' signals—signals which have properties common to two or more traditional signal types (for instance, signals that are both intrinsically costly *and* result in the sender being placed in a vulnerable position).

At the vulnerability hybrid equilibrium, the sender attains an average payoff of $x(c_l^v - c_h^v)/(1 + c_l^v)$. When signals are intrinsically costly, senders receive on average $x(1 - c_h^i)$ at the separating equilibrium, where c_h^i is the intrinsic cost associated with sending the signal as a high type and c_l^v and c_h^v are the costs low- and high-type incur if they signal and are treated as a low-type in the vulnerability signaling game. Receivers have a payoff of $1 - x$ at the vulnerability hybrid and a payoff of 1 at the separating equilibrium. Thus, while receivers always prefer the separating equilibrium to the vulnerability hybrid (as their biological fitness is higher at the separating equilibrium), senders sometimes prefer the vulnerability hybrid to the separating equilibrium (as their fitness is sometimes maximized at the vulnerability hybrid). When c_h^i is less than $\frac{1 + c_h^v}{1 + c_l^v}$, both parties prefer the separating equilibrium and as a result intrinsically costly signals are likely to be utilized (see Fig. 6). When $c_h^i > \frac{1 + c_h^v}{1 + c_l^v}$, however, senders do best at the vulnerability hybrid while receivers still prefer the separating equilibrium. How is this conflict resolved? Which group, sender or receiver, gets their preferred equilibrium? We only offer a few speculative remarks in this section and note that a more thorough analysis is necessary but outside the scope of this paper.
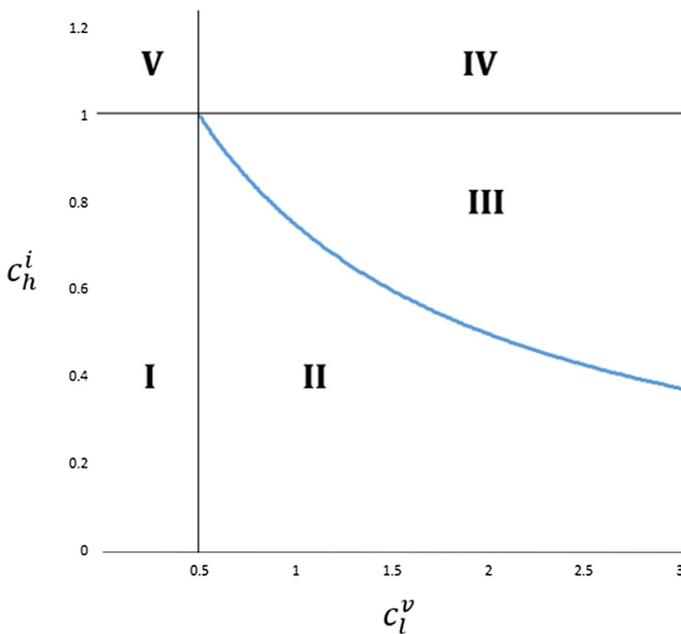
Fig. 6 Conditions under which the separating and vulnerability hybrid exist as a function of c_l^v and c_h^i (c_h^v is held fixed at 0.5). In region *I* only the separating equilibrium is possible. In region *II* both separating and vulnerability hybrid equilibria are possible, and both senders and receivers prefer the separating to the vulnerability hybrid. Region *III* is similar to region *II* except senders prefer the vulnerability hybrid to the separating equilibrium. In region *IV* only the vulnerability hybrid is possible, and in region V neither the separating nor vulnerability hybrid are possible (i.e., the pooling equilibrium is the unique evolutionary outcome)

Note that in many coevolving processes the group (or species) that evolves more rapidly is often at a distinct advantage. Adapting at a faster rate than one's counterpart typically gives one the upper hand, and instances of this so-called Red Queen effect abound, occurring between parasite and host as well as predator and prey. Hence we'd expect the senders' preferred equilibrium (the vulnerability hybrid) to attain when senders evolve at a faster rate than receivers. Likewise, when receivers adapt more rapidly than senders the separating equilibrium should win out. Recall that many of our motivating examples discussed in "Vulnerability signals" section involve communication between predator and prey. In these cases it is sensible to assume that senders (the prey) will outpace receivers (predators) in the evolutionary race due to a 'built-in imbalance between predator and prey with respect to the penalty of failure' (Dawkins and Krebs 1979). For the prey, one false move may result in death, while failure for the predator simply means going hungry for a longer period of time. In other words, members of the prey species are under stronger selection pressure than members of the predator population. Furthermore, predators are typically larger, and hence have smaller populations, meaning that selection is less efficient, and generational turnovers occur less frequently. This gives us some reason to believe vulnerability signals are likely to emerge—since senders are more responsive than receivers, the sender population is more likely to drive the evolutionary process, tilting the scales in the direction of their preferred outcome (the vulnerability hybrid).

Discussion

This paper began with a pervasive problem confronting empirical biologists bent on vindicating costly signaling theory. Since the *cost* in costly signaling theory refers to a reduction of some component of fitness, it is not enough for experimentalists to simply register that the production of a signal is metabolically draining or time consuming. For the signal to be considered a *costly* signal, this expenditure must somehow reduce the fitness of the sender. We argue that the exact way expenditure maps to cost is of importance to theorists and experimentalists alike, and demonstrate that alternative expenditure to cost mappings yield outcomes that differ wildly from the traditional predictions of costly signaling theory.

Specifically, if the cost of the signal hinges on the behaviour of the intended recipient of the signal (as when the signal is a vulnerability handicap, outlined above), then full information transfer is no longer possible. Instead, a partially honest equilibrium remains, in which low-types sometimes emit the signal and receivers sometimes treat their counterpart as a high-type upon receipt of the signal. We have shown how this arrangement is evolutionarily stable and is a likely outcome of a simple evolutionary model. Furthermore, on one popular definition of deception, it is clear that deception is possible at equilibrium when signals are receiver-dependent. Specifically, according to Skyrms (2010) a signal is said to be deceptive if it: (1) increases the probability of a state that does not obtain and results in the receiver taking behaviour that is (2) to the detriment of the receiver but (3)

benefits the sender.[17] It is clear that all three conditions can be satisfied when low-types engage in the signaling display.[18]

Appendix

In this brief appendix we establish the evolutionary significance of the hybrid equilibrium. We begin by first considering a game consisting of just those strategies involved in the hybrid equilibrium. The payoff matrix for the sender, and receiver, are listed below.

$$S = \begin{pmatrix} x & -xc_h^v \\ 1 & -xc_h^v - (1-x)c_l^v \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & x \\ 1-x & 1-x \end{pmatrix}$$

We alter both matrices by subtracting different constants from each column. This results in the following (equivalent) game:

$$S' = \begin{pmatrix} 0 & (1-x)c_l^v \\ 1-x & 0 \end{pmatrix}$$

$$R' = \begin{pmatrix} 0 & 2x-1 \\ -x & 0 \end{pmatrix}$$
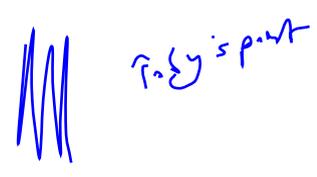
Under certain parameter values, there exists a mixed Nash equilibrium that corresponds to the hybrid equilibrium. We now attempt to determine the stability of the hybrid equilibrium under the replicator dynamics. Hofbauer and Sigmund (1998) prove that for there to be closed orbits around an equilibrium in the two-population replicator dynamics, two things most hold. First, all non-zero entries in S' must be of the same sign, and, likewise, all non-zero entries in R' must be of the same sign. Secondly, non-zero entries in S' must be of the same sign as the non-zero entries in R'. Clearly, the non-zero entries in S' are all positive, since p is a probability restricted between zero and one. The non-zero entries in R' are all negative when $x < .5$. Yet recall that the hybrid equilibrium only exists when $x < .5$, meaning that whenever the hybrid exists there are closed orbits circling the equilibrium on the plane consisting of all four strategies involved in the hybrid equilibrium.

---

[17] For criticism of Skyrms' notion of deception, see Martinez (2015) and McWhirter (2015).

[18] In particular, when the receiver responds to the signal by treating the sender as a high-type, (2) and (3) will hold. It is clear that (1) will always hold since all high-types signal but only a fraction of low-types signal, meaning the probability of a high-type conditional on receipt of the signal is greater than the unconditional probability of the sender being a high-type (which must hold if (1) is to be satisfied).

What of states that are not on this plane? Will they head to the plane and begin cycling around the hybrid? To determine this we examine small perturbations and consider how the remaining strategies would do when playing against the hybrid equilibrium. The payoff to senders at the hybrid equilibrium is:

$$x\left(\frac{c_h^v}{1+c_h^v} - \left(1 - \frac{c_h^v}{1+c_h^v}\right)(c_l^v)\right) + (1-x)\left(\frac{x}{1-x}\right)\left(\frac{c_h^v}{1+c_h^v} - \left(1 - \frac{c_h^v}{1+c_h^v}\right)(c_h^v)\right)$$
$$= \frac{x(c_h^v - c_l^v)}{1+c_h^v}$$

The payoff to receivers at the hybrid equilibrium is: $1 - x$. We now compare this to the payoff of the remaining two sender strategies (never send; send if low, do not send if high) against the hybrid. Both yield a payoff of zero, which is less than $\frac{x(c_h^v - c_l^v)}{1+c_h^v}$. The payoff of the remaining two receiver strategies (always classify as high; classify as high if no signal, low if signal) is p. Since a condition of the hybrid is that $x < .5$, it is clear that $1 - x > x$, and thus whenever the hybrid does exist states near it will be pulled to the plane. Just as Zollman et al. conclude, the hybrid equilibrium in our model is strongly stable with respect to states not on the plane, and weakly stable with respect to states on the plane.

References

Adams E, Mesterton-Gibson M (1995) The cost of threat displays and the stability of deceptive communication. J Theor Biol 175(4):405–421

Bennett A, Huey R (1990) Studying the evolution of physiological performance. Oxf Surv Evolut Biol 7:251–284

Bicchieri C, Duffy J (1997) Corruption cycles. Polit Stud 45(3):477–495

Brandt Y (2003) Lizard threat display handicaps endurance. Proc R Soc B 270(1519):1061–1068

Bruner J (2013) Policing epistemic communities. Episteme 10(4):403–416

Dawkins R, Krebs JR (1979) Arms races between and within species. Proc R Soc Lond Biol Soc 205(1161):489–511

Enquist M, Plane E, Roed J (1985) Aggressive Communication in fulmars (*Folmarus glacialis*) competing for good. Anim Behav 33:1007–1020

Fraser B (2012) Costly signaling theories: beyond the handicap principle. Biol Philos 27:263–278

Grafen A (1990) Biological signals as handicaps. J Theor Biol 144(4):517–546

Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. 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

Huttegger S, Zollman K (2010) Dynamic stability and basins of attraction in the Sir Philip Sidney game. Proc R Soc B 277:1915–1922

Huttegger S, Zollman K (2013) Methodology in biological game theory. Br J Philos Sci 64(3):637–658

Huttegger S, Bruner J, Zollman K (2015) The handicap principle is an artifact. Philos Sci 82(5):997–1009

Kane P, Zollman K (2015) An evolutionary comparison of the handicap principle and hybrid equilibrium theories of signaling. PloS ONE 10(9):e0137271

Kotiaho J (2000) Testing the assumptions of conditional handicap theory: costs and condition dependence of a sexually selected trait. Behav Ecol Sociobiol 48(3):188–194

Kotiaho J (2001) Costs of sexual traits: a mismatch between theoretical considerations and empirical evidence. Biol Rev 76(3):365–376

Lachmann M, Szamado S, Bergstrom C (2001) Cost and conflict in animal signals and human language. Proc Natl Acad Sci USA 98(23):13189–13194

Leal M (1999) Honst Signaling during prey-predator interactions in the lizard *Anolis cristatellus*. Anim Behav 58:521–526

Martinez M (2015) Deception in sender–receiver games. Erkenntnis 80:215–227

Maynard-Smith J, Harper D (2003) Animal signals. Oxford University Press, Oxford

McWhirter G (2015) Behavioral deception in formal models of communication. Br J Philos Sci. doi:10.1093/bjps/axv001

Moeller A, De Lope F (1994) Differential costs of a secondary sexual character: an experimental test of the handicap principle. Evolution 48:1676–1683

Molles L, Vehrencamp S (2001) Songbird cheaters pay a retaliation cost: evidence for auditory conventional signals. Proc R Soc Lond B Biol Sci 268(1480):2013–2019

Pretwich K (1994) The energetics of acoustic signaling in anurans and insects. Am Zool 34:625–643

Ryan M (1988) Energy, calling, and selectin. Am Zool 28:885–898

Searcy W, Nowicki S (2005) The evolution of animal communication: reliability and deception in signaling systems. Princeton University Press, Princeton

Skyrms B (2004) The stag hunt and the evolution of social structure. Cambridge University Press

Skyrms B (2010) Signals: evolution, learning and information. Oxford University Press, Oxford

Waas J (1991a) Do little blue penguins signal their intentions during aggressive interactions with strangers? Anim Behav 41(3):375–382

Waas J (1991b) The risks and benefits of signaling aggressive motivation: a study of cave-dwelling little blue penguins. Behav Ecol Sociobiol 29(2):139–146

Wagner E (2013) The dynamics of costly signaling. Games 4(2):163–181

Weibull J (1995) Evolutionary game theory. MIT Press, Cambridge

Zahavi A (1975) Mate selection—a selection for a handicap. J Theor Biol 53(1):205–214

Zollman K (2012) Finding alternatives to handicap theory. Biol Theory 8(2):127–132

Zollman KJS, Bergstrom CT, Huttegger SM (2012) Between cheap and costly signals: the evolution of partially honest communication. Proc R Soc B doi:10.1098/rspb.2012.1878