Honest signalling made simple.

Jacob Chisausky[A], Carl Bergstrom[B], Kevin Zollman[C] and Graeme Ruxton[D]

A: Ludwig Maximilian University of Munich; B: University of Washington, Seattle; C: Carnegie Mellon University; D: University of St Andrews

Corresponding Author: Jacob Chisausky, jacob.chisausky@evobio.eu

Abstract

Honest communication is a common phenomenon in animal behaviour, and is frequently explained by appeal to the so-called handicap hypothesis by which signal costs deter dishonesty. However, the handicap models commonly used to explain honest signalling have has been subject to several lines of criticism in recent literature. This trend may have led researchers outside of the field of animal communication to believe that the handicap hypothesis is logically invalid. That is untrue. Here we explain the handicap hypotheses and explore these critiques in non-mathematical terms in order to more widely disseminate a contemporary understanding of the theory of honest signalling. We address what is currently known and unknown in signalling theory, and suggest directions for future research.

Introduction

There are many situations in which one organism, which we will call the *receiver*, is interested in traits of another, which we call the *signaller*, that are not discernible. These might relate to quality as a potential mate, aggressive intent, or energetic state, to name only a few examples. If the signaller can gain advantage by influencing the behaviour of the receiver, then it may be selected to modify how it stimulates the receiver's senses. That is, it may be selected to signal to the receiver. However, the receiver is only expected to heed the signal if on average it benefits from doing so. In most such situations, there will not be perfect alignment of interests of both parties. Hence there is a tension: the receiver should only be expected to heed the signal if it conveys some useful information to it, but the signaller will only send an informative signal if doing so is in its own interest (Dawkins & Krebs 1978) Since the origins of the field, researchers in animal behaviour have sought to understand when animal signals are reliable, and what mechanisms are responsible for such honest signalling.

Mathematical analyses of signalling are complicated by the fact that the utility of signalling behaviour depends on the behaviour of signal receivers, and vice-versa. Fortunately, game theory is well-suited for this sort of strategic analysis, and has been the most popular avenue of theoretical research in animal communication. Several elements of animal communication are included in signalling games, reflecting the fundamental components of signalling in nature:

1. Signallers vary in some quality of interest to receivers. Signallers have access to this information, but receivers do not.
2. Signallers may select a strength or intensity with which to signal. (This may include choosing to send no signal at all; the key point is that some variation in signal form is possible)
3. Receivers respond with a given strength of response. If communication is occurring, then this response is conditioned on the received signal.
4. Receivers derive a benefit depending on the signaller's quality of interest and their response.
5. Signallers derive a benefit depending on the receiver's response and their own quality of interest.

While details very widely between models, this general outline reconstructs the most fundamental aspects of signalling. Signals and responses may be discrete or continuous. Discrete models typically allow only two possible values for qualities, signal strengths, and response strength. Continuous models, while analytically more complex, better approximate most biological systems by allowing continuously graded traits and actions. Because of this, and because discrete models have been thoroughly treated elsewhere (Maynard Smith & Harper 2003; Hurd 1996; Számadó 1999), we focus on continuous models in this paper. This abstraction of animal communication is general and can apply in some way to most or all signalling systems in nature. Three examples which we will use repeatedly throughout this paper are as follows.

Mate choice: Peacocks, the signallers, vary in genetic quality. Peahens, the receivers, are selected to mate with males of the highest genetic quality. Male tails can have many eyespots (the signal); the strength of the signal can be characterised as reflecting the number of eyespots. Males all are selected to receive the strongest response from females, i.e., receive a mating.

Nestling begging: Bird nestlings, the signallers, vary in hunger. Their parents, the receivers, are selected to feed the hungrier nestlings. Nestlings beg (the signal); the strength of the signal is the amount of time spend begging. Nestlings generally benefit from receiving more food regardless of hunger level, though consideration of inclusive fitness may change these payoffs.

Aposematism: Poisonous frogs, the signallers, vary in toxicity. A predator, the receiver, is selected to eat the nontoxic frogs and avoid the toxic frogs. Frogs produce aposematic coloration (the signal); signal strength is brightness of the aposematic display. All frogs benefit from being avoided by predators.

It is important to note that not all interactions mediated by a structure or behaviour qualify as signalling systems by our definition above. Sexual selection, through a Fisherian runaway-like process, can result in the evolution of elaborate, costly structures which signal no unobservable quality of interest to receivers (Prum 2010). Similarly, sensory exploitation may result in signal-like structures. For instance, an anglerfish lure mimics the prey of another fish, and may cause that fish to approach the lure. Here the anglerfish manipulates the senses of its prey, but the prey has not been selected to respond to this signal, rather it has not been sufficiently selected to avoid it because anglergfish are rare in comparison to the similar stimuli of the victim's food.

Discussions of signalling can become hampered by varying terminology (Maynard Smith & Harper 2003; Penn & Számadó 2020) and to our knowledge, a general and self-descriptive set of terms for the above aspects of signalling games has yet to coalesce in the literature. We will briefly present one here for the purposes of this paper. In the examples we treat, and indeed in a large fraction of the signalling literature, all signallers in a population prefer to elicit the same response from receivers (e.g. receive mating, receive maximal food, avoid predation). This is the *signaller-preferred response*. Receivers benefit from giving the signaller-preferred response to signallers of a specific quality (e.g. highest quality males, hungry nestlings, toxic prey); we call these signallers *consonant* because these signallers have the same interest as receivers[1]. There are other signallers to whom the receivers should not give the signaller-preferred response; we call these *dissonant* because these signallers have diverging interests from the receiver. The terms consonant and dissonant can consistently be applied to signalling systems that differ in details, and their use highlights the fact that it is the alignment of signaller and receiver preferences that shapes a signalling equilibrium, rather than 'low' or 'high' quality per se.

The most prevalent question in the theory of animal communication is the problem of honesty. A signalling system is honest if signallers signal with strength proportional to their quality in a way that allows receivers to use signal strength to infer signaller quality. If receivers determine their response based on signal strength, and signallers prefer a stronger response, then signallers will be incentivised to 'exaggerate' their quality by sending a signal that is stronger and thus elicit a more preferrable response. For instance, a low-quality peacock could produce a tail with many eyespots, a well-fed nestling could beg, and a non-

---

[1] 'High-quality' is a term often used for this quality of signaller, but this language can be confusing in contexts outside of mate choice, as it defines hungry nestlings and toxic frogs as high-quality

toxic frog could produce aposematic coloration. These are ‘deceitful’ signals, and given the apparent benefits to sending such, the absence (or near-absence) of deceit requires evolutionary explanation.

The handicap hypothesis

The most well-known explanation for maintaining honest signalling in animal communication is the handicap hypothesis. First proposed by Zahavi (1975) — albeit in substantially different form than the modern interpretation — the handicap hypothesis became widely accepted by biologists after publication of mathematical models by Grafen (1990a, Grafen 1990b, see Penn & Számadó 2020 for a historical review of the adoption of the handicap hypothesis). Definitions of the handicap hypothesis vary by author and year and finding a single definition of the handicap hypothesis is notoriously difficult (Penn & Számadó 2020). The general idea, however, is as follows:

If signals are costly to produce and more costly for less consonant signallers, then signallers will produce costly signals and the strength of signals will be honest indicators of the signallers' qualities. More consonant individuals will produce stronger signals than less-consonant individuals.[2]

An example of a handicap signalling equilibrium is shown in Figure 1A. At equilibrium, signallers of a given quality pay a cost to signal, and receive a benefit from receivers that infer signaller quality from signal strength and respond accordingly. The cost that signallers pay is given by a *signal cost function:* the relationship that determines how the cost of sending a signal depends on the signaller quality and the signal strength. Such functions are illustrated as red curves C_{con} and C_{dis} in Figure 1A. The *benefit function* represents the payoff that a signaller receives at equilibrium from sending a signal of a given strength. At equilibrium, each signaller selects a signal intensity that maximizes benefit minus cost. In continuous models, a necessary condition for this equilibrium is that the marginal cost of signalling (i.e., the slope of the cost function) equals the marginal benefit (i.e., the slope of the benefit function).[3] At equilibrium, signallers of different qualities have different optimal signals and so receivers can discriminate among signaller types.

Costs and benefits are the fundamental ingredients of costly signalling models, and must be mapped onto real biological properties to be meaningful to biological inquiry. In the models, signal cost is intended to capture the deleterious consequences of sending a signal on an individual's viability or fecundity; this may be due to energetic expenditure, resource use, increased predation risk, reduced movement speed, etc. Signal costs may accrue from different sources; for instance, a bird call may both increase predation risk and incur energetic costs. The strength at which a signal is minimally costly (often but not always 0) represents the optimal value of that trait when receiver responses are not considered. Signal benefits reflect an increase to either viability or reproductive success as a result of receiver responses. Models of begging focus on viability benefits; models of mate choice on reproductive success.

Because the handicap hypothesis requires that signal costs vary with signaller quality, it is a case of a **differential costs** model. Specifically, it is a differential costs model in which at equilibrium signallers pay costs to signal. While models of the handicap hypothesis can be shown to support honest signalling in models of both discrete (Maynard Smith & Harper 2003; Hurd 1996; Számadó 1999) and continuous signals (Grafen 1990a; Grafen 1990b; Lachmann et al. 2001; Biernaskie, Grafen & Perry 2014), decades of theoretical developments have uncovered a variety of nuances and corrections to the handicap hypothesis

---

[2] This definition excludes differential benefits models (discussed below) which some authors consider to be cases of the handicap hypothesis. As there is no definite definition of the handicap hypothesis, this distinction is somewhat arbitrary, and we have elected to consider only differential costs models as the handicap hypothesis.

[3] Assuming continuous and continuously differentiable functions - discussed more below.

as it was first conceived. One of the most important of these is that differential costs models do not actually require signal costs to be paid at equilibrium.

Minimal cost signalling

Early treatments of the handicap hypothesis concluded that signals must be costly for all signallers at an honest signalling equilibrium[4] (Grafen 1990a, Maynard Smith & Harper 2003). This has now been shown to be incorrect. Differential costs can support honest signalling where signallers pay no or very little costs at equilibrium. The cost of signals at equilibrium depends on the shape of the signal cost function, and signal cost functions exist in which signalling honestly is arbitrarily cheap yet deception is still disfavoured (e.g., Figure 1B) [5] (Lachmann et al. 2001). This is possible because what prevents cheating in a signalling system is the potential cost of deception, rather than signal costs as paid (Számadó 2011, Lachmann et al. 2001, Hurd 1995). If a signaller would pay more in costs to send an out of equilibrium (deceitful[6]) signal than it would gain from deceiving a receiver, then dishonest signalling will not be favoured. Because the potential cost of deception is what drives honesty, the cost any signaller is actually paying to send the honest signal is irrelevant: the honest signal can be free. We call this sort of cost-free or very low-cost signalling equilibrium **minimal cost signalling.** Differential costs for different signaller types are still required, but the shape of signal cost functions allow low or no costs to be paid.

One biological mechanism that allows this type of cost-free honest signalling involves signal costs that are socially imposed (Lachmann et al 2001; Maynard Smith & Harper 2003) [7]. If dishonest signallers are punished by other individuals but no one signals dishonestly, signalling can be cost-free for all signallers. Cheating is deterred by potential cost of being punished. Such a mechanism has been documented in *Polistes dominulus* paper wasps. Members of this species have unique facial markings, such that an individual's fighting ability is indicated by the brokenness of spots on their face (Tibbetts & Lindsay 2008). 'Cheaters' with broken facial patterns but low fighting ability pay a large cost in increased aggression from conspecifics. Production costs of these markings are likely very low, suggesting that the only signal costs are socially imposed (Tibbetts & Dale 2004). Punishment might also take the form of ignoring the signaller's future signals in such cases (Rich and Zollman 2016).

---

[4] 'Equilibrium' refers to a signalling system that has settled into a stable state where each signaller and receiver is using the best strategy available to them. This means that each signaller sends a signal that provides it the best payoff based on receiver behaviour; sending a weaker or stronger signal would reduce payoff due to a decrease in benefit minus cost. At an honest signalling equilibrium, each quality of signaller sends a different signal strength from other qualities, allowing receivers to correctly determine the quality of a signaller from the signal sent and respond accordingly.

[5] If the signal cost functions are not constrained to be continuously differentiable, signal cost can even be zero at equilibrium signalling when there is an explicit kink between the region of zero signal cost and that of rapidly increasing signal costs (Lachmann et al. 2001).

[6] By 'deceit', we mean sending a signal which causes a receiver to be unable to accurately recover a signaller's quality from the strength of their signal. If all signallers signal more strongly than their naturally selected optimum and receivers can use signal strength to determine signaller quality, then the signals are honest. If a signaller exaggerates their signal beyond this point, and a receiver thus identifies their quality as higher than it really is, then that signal is deceitful.

[7] Some authors (Maynard Smith & Harper 2003) don't consider cost-free signals to be handicaps, though we believe this distinction is arbitrary. There is no intrinsic requirement on cost functions that allows only socially imposed signal costs to support cost-free signalling, and a signal cost function produced by social punishment, production cost, or a mix of the two generate signalling equilibria in exactly the same way (Fraser 2012).

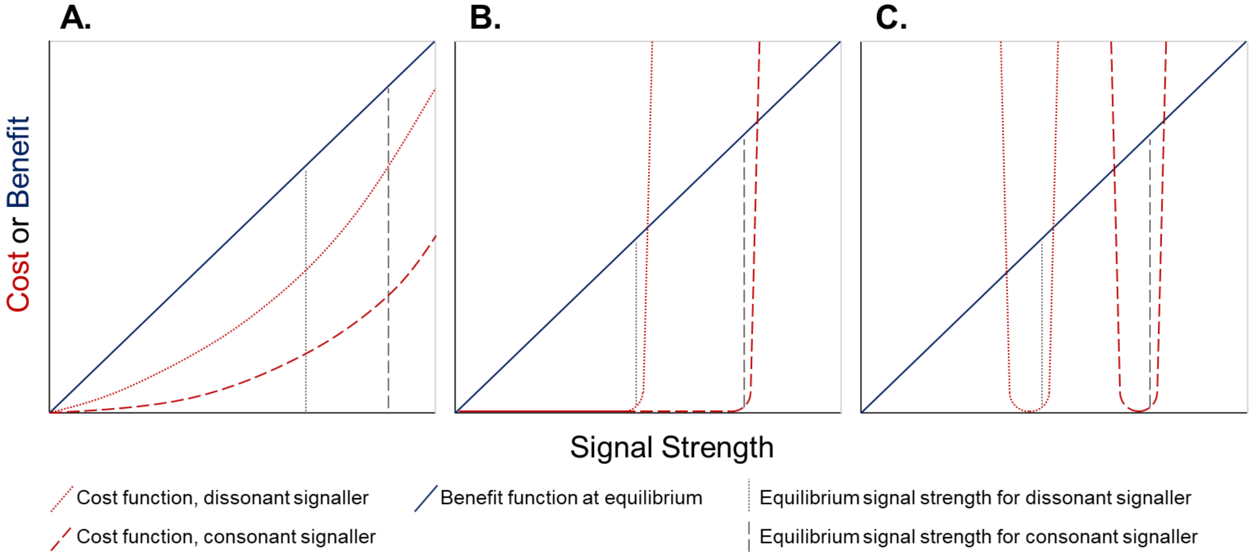


Figure 1: Three cases of honest signalling equilibria under differential costs. The red curves show signal cost functions for an exemplar consonant (dashed) and dissonant (dotted) signaller. The blue lines show benefit functions at equilibrium. The grey lines show equilibrium signal strength for the two depicted signallers; these are signal intensities that maximize the difference between benefit and cost (and at which marginal costs equal marginal benefits). Panel A shows a traditional handicap hypothesis, where costs are paid at equilibrium. Panels B and C show cases where costs are very low at equilibrium. Panel C shows an index-like signal, in which any signal away from equilibrium reduces fitness by either increasing costs more than benefits (a stronger signal) or reducing benefits more than costs (a weaker signal).

Cost-free differential costs models do not invalidate the handicap hypothesis (although they do invalidate the handicap *principle*; discussed below). Rather, minimal cost signals and 'traditional' handicap signals (where all individuals pay strategic costs) are both varieties of the differential cost model. Signal costs do not need to be paid at equilibrium for a differential costs model to work, but they may be; costs at equilibrium do not modify the rationale of the differential costs model and are instead just the outcomes of different signal cost functions. If the cost of a signal rises sharply at the point where it would be dishonest, signalling can be cheap or free (as in Figure 1B). If the cost of a signal rises more smoothly, as in Figure 1A, signalling will be costly for all individuals. We call the latter of these situations handicap signals, and the former minimal cost signals. However, this suggests a false dichotomy between handicap signals and cost-free signals. Whether signal costs are paid at equilibrium (handicap signals) or not is a question of degree, not qualitative difference. Differential cost models operate in the same way regardless of the amount of strategic costs paid at equilibrium.

Grafen's (1990) model predicted a cost of signalling paid at equilibrium. This is due to the specific cost function used in the model, that viability is quality raised to the power of signal strength (Lachmann et al. 2001). A model that used different cost functions could allow cost-free signalling without altering the fundamental logic at play.

Differential benefits models

While the traditional handicap hypothesis requires that costs vary with signaller quality, a similar solution to honest signalling exists where costs are uniform but benefits differ according to signaller type. The **differential benefits model** shows that signalling can be honest if more consonant individuals receive greater benefits from receiving a response of a given strength than do less consonant individuals (Johnstone 1997; Godfray 1991; Maynard Smith 1991). Costs may be paid at equilibrium, but because more consonant signallers receive a greater benefit from a given response, they can more effectively compensate for the cost. Thus, more consonant individuals can afford to signal more strongly and still receive net benefit, and so signal strength is proportional to signaller quality. Grafen's (1990a) influential model assumed that signal benefits are uniform, which is why Grafen asserted that costs must be differential for a signalling equilibrium to exist (Getty 2006).

Begging behaviour by nesting birds has been characterized as a differential benefits model (Godfray 1991). Imagine that signalling costs were equal for all nestlings irrespective of condition, as would be the case if signalling were energetically cheap but expensive in terms of predation risk. Given that hungry nestlings receive a greater benefit from a given quantity of food than do satiated nestlings, a satiated bird that begs too much will pay more cost than the benefit gained from the additional food it receives, and so deception is unprofitable.

Indices (unfakable signals)

The final mechanism we will discuss which can support honest signalling under a conflict of interest (i.e., when some signallers are dissonant) is that of indices, or unfakable signals. Indices are signals for which a physiological constraint exists between signal magnitude and some aspect of a signaller's physiology which is of interest to the receiver. For example, if a spider signals its weight to a rival by shaking the web, and the weight of the spider determines how much the web shakes, web-shaking is an honest signal of the spider's weight since all spiders – regardless of weight – should be selected to shake the web as vigorously as they can (Maynard Smith & Harper 2003). When signal and underlying quality are physiologically constrained in this manner, no strategic costs are necessary to ensure honesty.

While indices have typically been considered a distinct means for cost-free honest signalling to evolve, a clear distinction between fakable and unfakable signals is not straightforward (Huttegger, Bruner & Zollman 2015; Biernaskie, Grafen & Perry 2014). Indices are typically signals that are indirectly correlated with the quality of interest to a receiver: e.g., a red deer bull's roar depth correlates with neck length, neck length correlates with body size, which in turn correlates with fighting ability, so roar depth is viewed as an index of fighting ability (Maynard Smith & Harper 2003). However, because the relationship between signal (sound production in the upper respiratory system) and quality of interest (fighting ability) is indirect, there is no intrinsic reason why these cannot be faked. Rather, most purported indices probably are fakable, though the cost of doing so (signalling dishonestly) would be so high that it would not pay. For instance, a deer bull could feasibly develop a proportionally longer neck to decrease the pitch of its roar without this larger neck improving fighting ability, signalling that its fighting ability is better than it actually is. However, the cost of this to other aspects of fitness (unrelated to influencing competitors) is probably so great that the benefit from dishonestly signalling would be far exceeded by these costs. In this way, physiological constraints on signal size can often be recast in terms of developmental (or other) costs (Searcy & Nowicki 2005). Thus, it is the potential cost of cheating that prevents cheating in this purported index, exactly as for costly signalling.

If a signal is truly unfakable then the logic of indices correctly applies. The only class of signals for which there is an intrinsic reason for unfakability are those which directly reveal the underlying quality. For instance, if a receiver is interested in a bird's agility, an intricate dance requiring a great amount of agility would be a true index of agility. It would not, however, be a true index of foraging ability, though we may be tempted to call it such; while forging ability may be correlated with agility, a very agile bird may have (for example) sacrificed investment in eyesight for investment in agility and consequently be a very poor forager.

Conditions on cost and benefit functions

Much of the goal of this paper is to highlight the diversity of different cost and benefit functions can support honest signalling, with the details of the equilibrium (costs actually paid) determined by the shape of those functions. Most models of costly signalling, however, restrict the diversity of cost and benefit functions in way that excludes many realistic scenarios. Typically, signals of any strength are assumed to be equally or less costly for more consonant signallers, and benefits are assumed to be equal or greater for more consonant signallers. While this is valid for most biological systems, some important exceptions may exist. For instance, a hungry (consonant) begging nestling may pay higher costs to signal (due to poor condition), but receive much greater benefits.

Another standard assumption in costly signalling models is monotonicity of cost and benefit functions. This may be invalidated in nature more often: for instance, barn swallow (*Hirundo rustica*) tail streamers are used as a signal in mate choice by females, and the naturally selected (aerodynamic) optimum for an individual is of an intermediate length (and increases with male quality) (Evans & Thomas 1997; Bro-Jørgensen, Johnstone & Evans 2007). The cost function of these tails is thus characterized by a monotonic cost function, where intermediate tails are the cheapest and increasing or decreasing signal strength increases signal cost. Figure 1C depicts a signalling system of this sort. Interestingly, in nature tail streamers are longer than the aerodynamic optimum; and this is predicted by signalling theory. Signals at the strength where the marginal benefit equals the marginal cost will be, when cost functions are smooth, stronger than the naturally selected optimum (Lachmann & Bergstrom 1999). This is visualized by the way that equilibrium signal strengths in Figures 1B-C are 'pulled up' the cost function a small amount by the slope of the benefit function.

This non-monotonicity especially applies to index-like signals, many of which may be costly to modify in either direction because physiological considerations cause an intermediate value of a signal to be optimal. If the costs of increasing or decreasing a signal's strength rise very steeply from the naturally selected optimum, that signal may seem to be an index though costs and benefits, rather than genuine unfakability, are factors that ensure honesty.

Putting it All Together

The differential cost and differential benefits models presented thus far simplify affairs by assuming that costs differ across signaller qualities and benefits are uniform, or vice versa. More elaborate models allow both costs and benefits to differ. Where a conflict of interest is present, differential costs and differential benefits may be combined to enable honest signalling in cases where consonant signallers either pay more cost to signal (if greater benefits outweigh the extra cost) or receive lower benefits from responses (if lower costs outweigh the lower benefit).

Several authors have tried to generalize a form of 'conversion efficiency' in which consonant signallers must be more efficient at converting signal costs into benefits than dissonant signallers[8], though we show that this notion is inadequate as an explanation for honest signalling (Appendix). We find that the

---

[8] Getty (2006) writes "[…] higher quality signallers must be more efficient at converting advertising into fitness." Számadó (1999) writes: "[…] the signal should be more costly for low-quality individuals not in itself but in relation to the whole benefit of the interaction." In these papers, 'high-quality' can be read as 'consonant'. Maynard Smith and Harper (2003) write: "The ratio of the cost of a signal to the benefit received is lower for individuals giving stronger signals."

following three principles capture the requirements for honest signalling to be evolutionarily stable, applicable under a less strict set of assumptions than are required by other models (that is, costs and benefits need not be monotonic, consonant signallers can pay higher costs or receive lower benefits).

1) Signallers will be selected to maximise their fitness, which is influenced both by the costs of sending signals and by the benefits that ensure as a consequence of receiver responses.

2) If we can describe cost and benefit curves as two functions of signal strength (for continuous models), then at equilibrium we would expect fitness to be maximised (i.e. benefits less costs to be maximised) at the signal strength where these two curves have the same slope[9]. This emphasises that the key to understanding an equilibrium is the marginal decrease in fitness associated with all marginal changes in signal strength near the equilibrium point.

3) If the fitness-maximising signal strength is different for signallers that differ in the quality of interest to the receiver, then an honest signalling equilibrium is supported. This does not require that more consonant signallers pay lower costs or receive higher benefits from signals of any strength (Figure 1C). Indeed, one can readily sketch signal cost functions that support honest signalling yet for which consonant signallers have neither a favourable cost:benefit ratio nor a higher conversion efficiency (Appendix).

Our framework for characterising signalling unifies all of the above models (differential costs and benefits, non-monotonic cost or benefit functions, and index-like costly signals) into one simple heuristic useful to theoreticians and empiricists alike. It also shows that a suite of assumptions included (though not well-justified) in most signalling models are not required for honest signalling equilibria.

Common interest between signaller and receiver

Honest signalling evolves more straightforwardly when common interest between signaller and receiver obviates any incentive for dishonesty (in other words, when all signallers are consonant). If signaller and receiver rank possible outcomes in the same order of preference, then a signaller does not gain by misrepresenting its quality to the receiver and so signalling is honest (Maynard Smith & Harper 2003). Because no additional enforcement of honesty is required, minimally costly signals can support honesty. Common interest may occur in systems such as mate recognition and when a high degree of relatedness exists between signaller and receiver (Bergstrom & Lachmann 1998; Bruner & Rubin 2020).

Signallers and receives may also share partial common interest. This will have the effect of modifying a signaller's benefit function: while consideration of direct fitness only may produce a benefit function that increases monotonically with response strength, inclusive fitness may produce a non-monotone benefit function, in which a very strong response gives less benefit to a signaller's inclusive fitness than does a weaker response. For instance, a begging nestling receiving too much food from a parent will have a lower inclusive fitness through depriving its siblings of food; the benefit of a strong response is lower than the benefit of a weaker response. Since the optimal response strength for each signaller is reduced under partial common interest, this will have the effect of reducing the strength at which each individual needs to signal at equilibrium.

The handicap hypothesis is not universal (and so the handicap *principle* is invalid)

---

[9] Cost and benefit slopes can be equal at local minima, local maxima, or saddles. If cost and benefit slopes are equal at multiple points, or if cost and benefit functions not are continuous and continually differentiable, the point with the maximum benefit minus cost is the equilibrium signal strength. .

One mistake by early proponents of the handicap hypothesis was to overextend the theory and claim that all examples of honest signalling in nature that involve costly signals must be handicap signals[10] (Grafen 1990a; Zahavi & Zahavi 1999). This assertion of generality for the handicap hypothesis is the handicap *principle*, a notion which is incorrect (Penn and Számadó 2020). In fact, as we have discussed above, the handicap hypothesis is just one of many ways by which signals can theoretically be honest and evolutionarily stable, alongside cost-free differential cost models, differential benefits models, true indices, and systems with both differential costs and differential benefits. Rejecting the handicap *principle* does not, however, invalidate the handicap *hypothesis* for signalling systems in which it may be operating.

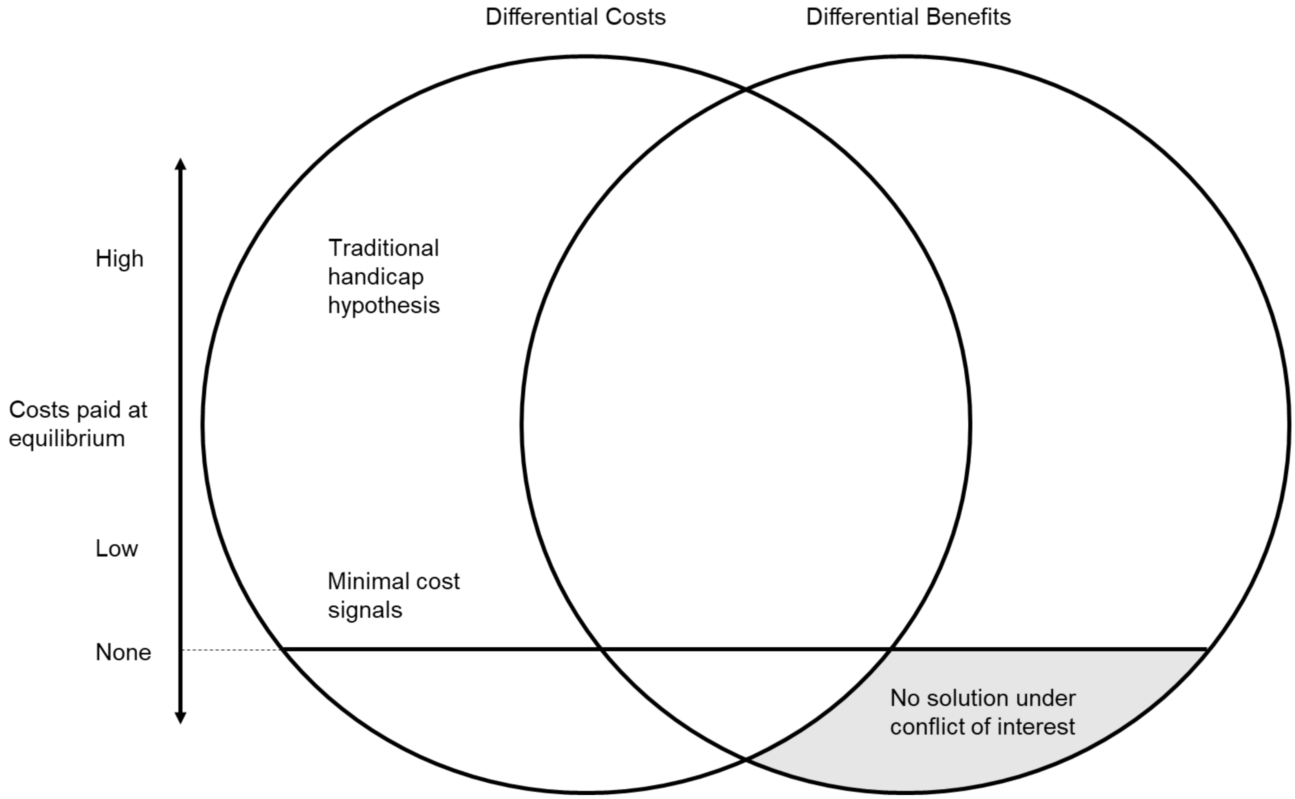


Figure 2: Different ways to support honest signalling. This shows the set of conditions under which honest signalling is explained by the unified costly signalling model where signals are fakable (i.e., not true indices) and a conflict of interest exists. The y-axis shows the extent of signal costs paid at equilibrium, emphasizing the fact that cost-free or low-cost signals and expensive (traditional handicap hypothesis) signals are different ends of a gradient rather than qualitatively different signalling systems, determined by the shape of signal cost functions. Only when costs are uniformly free are true indices or no conflict of interest required to explain honest signalling.

Recent Criticisms of the Handicap Hypothesis

A series of recent papers have suggested that the handicap principle is invalid. These papers correctly note the overgeneralization of the handicap hypothesis into the handicap principle, point out

---

[10] at least under a conflict of interest and for signals which are potentially fakable.

oversights in early models of the handicap principle, and describe the various alternative mechanisms for honest signalling as we have addressed above. However, they could easily leave a casual reader with the impression that the handicap *hypotheses* is invalid and has no place in signalling theory.

Following Penn and Számadó (2020), we consider it crucial to distinguish between the handicap *hypothesis*, which is a viable explanation for honest signalling, and the handicap *principle*, which is false and readily disproven mathematically and empirically. Numerous critiques of many aspects of costly signalling theory throughout decades don't invalidate the idea that differential costs can promote honest signalling with costs paid at equilibrium, which we see as the basic idea of the handicap hypothesis. Rather they offer alternatives; which of these possible alternative routes to signal honesty are most commonplace in nature cannot be resolved by logical reasoning alone – it requires application of empirical understanding of the biologies of different signalling systems.

Conclusion: a modern understanding of costly signalling

The theory of costly signalling has undergone a turbulent history, much of it centred around the primacy of the handicap hypothesis (Penn & Számadó 2020). We show that the handicap hypothesis is theoretically valid, though it should be considered as just one of several mechanisms to support honest signalling. Specifically, the handicap hypothesis is a differential costs model in which substantial costs are paid at equilibrium; this is fundamentally no different from differential costs models in which no or minimal signal costs are paid at equilibrium. Rather, the shape of a system's cost and benefit functions will determine the degree of costs actually paid at equilibrium. We also argue that most purported instances of indices, similarly, can also be subsumed by a more general framework in which equilibrium signal strengths are determined by costs and benefits.

Which biological situations encapsulated in these various theoretical models are more common in nature now falls to empirical researchers to determine. It is likely that some communication types or taxa tend to rely on different mechanisms for honest signalling, and determining which of the above mechanisms evolves in different scenarios and why — these questions should motivate the next generation of theoretical work on animal communication. The relative importance of different subsets is essentially an empirical question. We advocate a shift in this research area away from refining theory to exploring what existing theories best fit the signalling that is actually going on in the natural world around us. Once we know that, there will be a new and interesting challenge for theoreticians: to explain any unevenness in the distribution of natural systems across the spectrum of possible situations that should theoretically lead to signal honesty. To facilitate this way forward we need to present theories in as straightforward a fashion as possible to facilitate dialogue with empiricists – we hope this paper is a modest but valuable step in that direction.

References

Bergstrom, C. T., & Lachmann, M. (1998). Signalling among relatives. III. Talk is cheap. *Proceedings of the National Academy of Sciences*, *95*(9), 5100-5105.

Biernaskie, J. M., Grafen, A., & Perry, J. C. (2014). The evolution of index signals to avoid the cost of dishonesty. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1790), 20140876.

Bro-Jørgensen, J., Johnstone, R. A., & Evans, M. R. (2007). Uninformative exaggeration of male sexual ornaments in barn swallows. *Current Biology*, *17*(10), 850-855.

Bruner, J. P., & Rubin, H. (2020). Inclusive fitness and the problem of honest communication. *The British Journal for the Philosophy of Science*.

Dawkins, R. & Krebs, J. R. (1978). Animal signals: information or manipulation? In *Behavioural Ecology. An Evolutionary Approach* (eds J. R. Krebs and N. B. Davies), pp. 282–309. Blackwell, Oxford.

Evans, M. R., & Thomas, A. L. R. (1997). Testing the functional significance of tail streamers. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1379), 211-217.

Fraser, B. (2012). Costly signalling theories: beyond the handicap principle. *Biology & Philosophy*, *27*, 263-278.

Johnstone, R. A. (1997). The evolution of animal signals. Behavioural ecology: An evolutionary approach, 155178.

Getty, T. (2006). Sexually selected signals are not similar to sports handicaps. *Trends in Ecology & Evolution*, *21*(2), 83-88.

Godfray, H. C. J. (1991). Signalling of need by offspring to their parents. *Nature*, *352*(6333), 328-330.

Grafen, A. (1990a). Biological signals as handicaps. *Journal of theoretical biology*, *144*(4), 517-546.

Grafen, A. (1990b). Sexual selection unhandicapped by the Fisher process. *Journal of theoretical biology*, *144*(4), 473-516.

Hurd, P. L. (1995). Communication in discrete action-response games. *Journal of Theoretical Biology*, *174*(2), 217-222.

Huttegger, S. M., Bruner, J. P., & Zollman, K. J. (2015). The handicap principle is an artifact. *Philosophy of Science*, *82*(5), 997-1009.

Lachmann, M., & Bergstrom, C. T. (1999). When honest signals must be costly. *Santa Fe Institute, SFI working paper*, *99*, 08-059.

Lachmann, M., Számadó, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences*, *98*(23), 13189-13194.

Maynard Smith, J. (1991). Honest signalling: the Philip Sidney game. *Animal behaviour*.

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

Prum, R. O. (2010). The Lande–Kirkpatrick mechanism is the null model of evolution by intersexual selection: implications for meaning, honesty, and design in intersexual signals. *Evolution, 64*(11), 3085-3100.

Rich, P., & Zollman, K. J. (2016). Honesty through repeated interactions. *Journal of theoretical biology*, *395*, 238-244.

Searcy, W. A., & Nowicki, S. (2005). *The Evolution of Animal Communication: Reliability and Deception in Signalling Systems: Reliability and Deception in Signalling Systems*. Princeton University Press.

Számadó, S. (1999). The validity of the handicap principle in discrete action–response games. *Journal of theoretical biology*, *198*(4), 593-602.

Tibbetts, E. A., & Dale, J. (2004). A socially enforced signal of quality in a paper wasp. *Nature*, *432*(7014), 218-222.

Tibbetts, E. A., & Lindsay, R. (2008). Visual signals of status and rival assessment in Polistes dominulus paper wasps. *Biology Letters*, *4*(3), 237-239.

Zahavi, A., & Zahavi, A. (1999). *The handicap principle: A missing piece of Darwin's puzzle*. Oxford University Press.

Zollman, K. J., Bergstrom, C. T., & Huttegger, S. M. (2013). Between cheap and costly signals: the evolution of partially honest communication. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1750), 20121878.

Appendix

Conversion efficiency

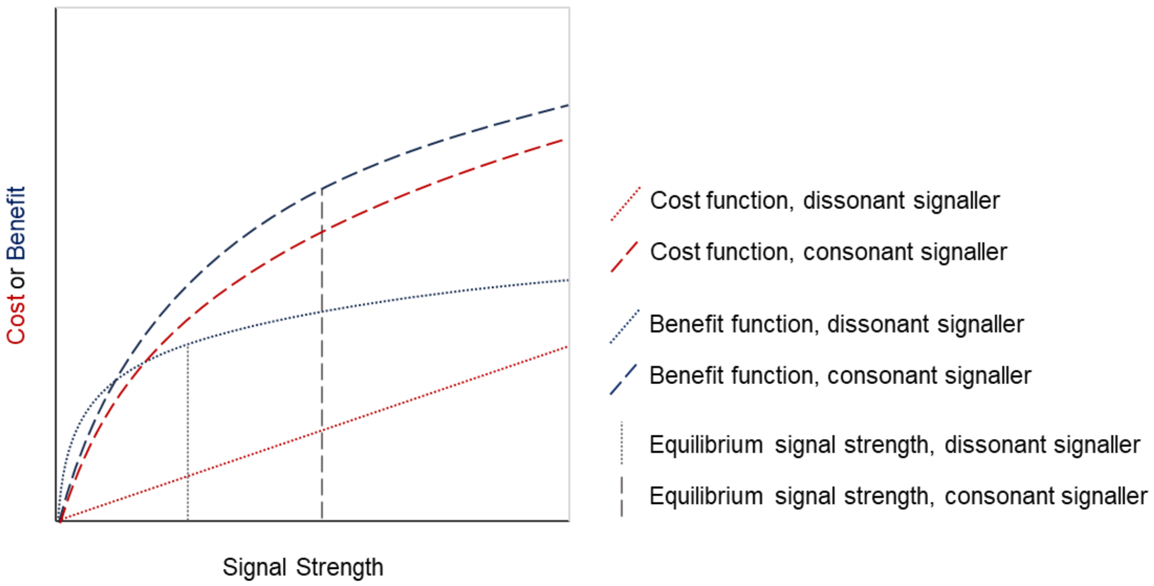
Several authors (Getty 2006, Számadó 1999; Maynard Smith & Harper 2003) have made claims about consonant signallers having higher "conversion efficiency", benefit-cost differences, and the likes in honest signalling models. While these quantities are never defined formally, candidates for conversation efficiency would be ratios such as $r_1 = b/c$ or $r_2 = (b - c)/c$, and cost-benefit differences would be terms such as $d = b-c$. The issue of how to make the comparisons also arises: are these claims about how r or d for a consonant signaller compares to r or d for a dissonant signaller with (case 1) each type sending its own particular equilibrium signal, or (case 2) both types sending the same signal? The aim of this appendix is to offer some justification for cases in which something like these conversion efficiency claims might hold.

We discuss here a continuous signalling model. Let $c_i(x)$ be the cost to type i of sending signal x and $b_i(x)$ be the benefit to type i of receiving the equilibrium response to signal x. Typically, for minimal hassle we assume that for all i and all x, cost and benefit are twice continuously differentiable, monotone increasing, and that the second derivatives don't change sign. That is, for all x in the domain of feasible signal intensities,

$$c'_i(x) > 0 \text{ and } c''_i(x) \geq 0 \quad (1)$$
$$b'_i(x) > 0 \text{ and } b''_i(x) \leq 0 \quad (2)$$

Notice that these are conditions that compare cost and benefit to the same signaller i for different signal intensities x.

If we restrict ourselves to only these two assumptions on signal form, it is easy to produce examples of signalling equilibria which violate the claim that consonant signallers have a more favourable conversion efficiencies than dissonant signallers. Figure S1A gives an example in which consonant signallers pay greater costs to signal, but also receive greater benefits. This is biologically reasonable, for instance a nestling bird which is very hungry and receives a greater benefit from a unit of food, but also pays higher costs to signal due to low energetic reserves. Under these cost and benefit functions, the consonant signallers receives a lower d, r_1, and r_2 than the dissonant signaller not only at each signaller's respective equilibrium signal strength (case 1) but across all signal strengths depicted on the graph > 0.

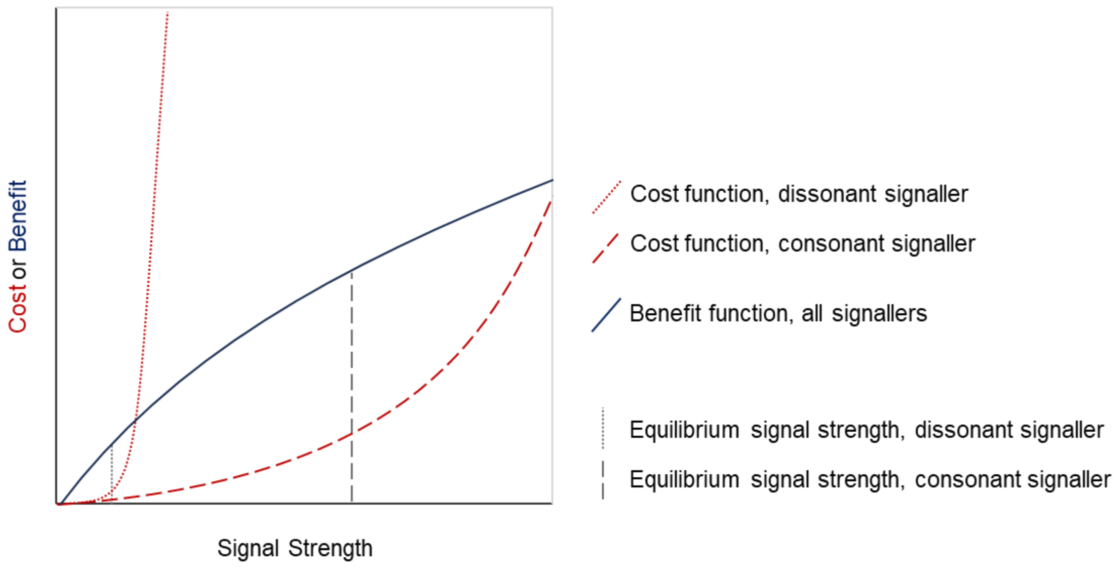


Supplementary Figure 1: An example of a signalling equilibrium in which consonant signallers have lower conversion efficiencies than dissonant signallers. The red and blue curves show signal cost and benefit functions (respectively) for an exemplar consonant (dashed) and dissonant (dotted) signaller. The grey lines show equilibrium signal strength for the two depicted signallers. The equations used in this plot are: consonant benefits: 4*ln(.9x+1) – 0.14x; consonant cost: 3*ln(x+1); dissonant benefits: .9*ln(15x+1); dissonant costs: x/3. The domain [0,10] is plotted.

Similar counterexamples can be drawn in which consonant signaller pay lower costs but receive lower benefits, receive lower benefits while paying the same cost, or receive the same benefits while paying higher costs, and have lower d, r_1, and r_2. This shows that, if cost and benefit functions are free to vary, claims about conversion efficiency do not hold. However, this example violates two typical yet often unwritten and unjustified other assumptions in signalling models: consonant signallers are restricted to signal more cheaply (in a differential costs model), receive larger benefits (in a differential benefits model), or both for any intensity of signal x. Where the subscript c indicates a consonant signaller and d indicates a dissonant signaller, the assumption is that

$$c_c(x) \leq c_d(x) \quad (3)$$
$$b_c(x) \leq b_d(x) \quad (4)$$

Here we analyse conversion efficiency assuming conditions (3) and (4) as well as (1) and (2). It is easily shown that it is not universally true that consonant signallers have to have a higher conversion efficiency at its equilibrium signal strength compared to a dissonant signaller at its equilibrium signal (case 1). Figure S2 gives such an example.



Supplementary Figure 2: An example in which the consonant signaller has a lower conversion efficiency *b/c* at its equilibrium signal choice than does the dissonant signaler at its equilibrium signal choice.

We use $r_1 = b/c$ as a measure of conversion efficiency, though the same logic applies to $r_2 = (b\text{-}c)/c$. The basic intuition is that the dissonant signaller can pay an arbitrarily low signal cost at equilibrium while the consonant signaler pays a cost bounded away from zero at equilibrium.

What is true is that for any particular signal intensity, the consonant signaler has a higher conversion efficiency than the dissonant signaler (case 2). This follows directly from (3) and (4), given strict inequality at at least one of the two comparisons. Let $r_{i|j}$ be the conversation efficiency of a signaller of quality *i* who signals at the equilibrium intensity of a signaller of quality *j*. So for example, the consonant signaller has higher conversion efficiency at its equilibrium signal than the dissonant signaller has at the consonant signaller's equilibrium signal when:

$$r_{C|C} = b_C(C)/c_C(C) \geq b_C(C)/c_D(C) \geq b_D(C)/c_D(C) = r_{D|C}$$

where at least one of the two internal inequalities is strict. The consonant signaller also has higher conversation efficiency at the dissonant signaller's equilibrium signal level than the dissonant signaler does when:

$$r_{C|D} = b_C(D)/c_C(D) \geq b_C(D)/c_D(D) \geq b_D(D)/c_D(D) = r_{D|D}$$

again with one of the internal inequalities strict. The first inequality holds due to (3) and (4), at least one of which is strict, and the second holds by the definition of equilibrium for a consonant signaler. It also follows trivially from (3) and (4) that the consonant signaler has a higher benefit-cost difference at its equilibrium signalling level, and also at the dissonant signaller's equilibrium signalling level.

Benefit-cost difference

If we assume (3) and (4) it is, however, true that the consonant signaler has a bigger benefit-cost difference *d* at its equilibrium than the dissonant signaler does at its equilibrium signal level. Where $d_{i|j}$is the benefit-cost difference for type *i* signaler using type *j*'s equilibrium signal,

$$d_{D|D} = b_D(D) - c_D(D) < b_C(D) - c_C(D) \leq b_C(C) - c_C(C) = d_{C|C}$$

Conclusion: Even with the assumption that consonant signallers can signal more clearly across the board (3), or receive greater benefits across the board (4), it is not true that consonant signallers have higher conversion efficiencies than dissonant signallers at their respective equilibria (case 1). It is true, however, that consonant signallers have higher benefit-cost differences than dissonant signallers at their respective equilibria. Consonant signallers also have higher conversion efficiencies when compared at the same signal strength (case 2).

While the model here applies to a model with discrete qualities but continuous signals, it is readily extended to a model of continuous qualities, by noting that the same comparisons can be made between any two signallers, a relatively more consonant one and a relatively more dissonant one. All that is required is the corresponding extension of assumptions (3) and (4).