The trade-offs of honest and dishonest signals

Szabolcs Számadó[1,2,*], István Zachar[3,4,5] & Dustin J. Penn[6]

December 15, 2023

[1] Department of Sociology and Communication, Budapest University of Technology and Economics, Egry J. u. 1. H-1111 Budapest, Hungary

[2] CSS-RECENS “Lendület” Research Group, HUN-REN Centre for Social Science, Tóth Kálmán u. 4., H-1097 Budapest, Hungary

[3] Institute of Evolution, HUN-REN Centre for Ecological Research, Konkoly-Thege Miklós út 29-33., H-1121 Budapest, Hungary

[4] Parmenides Foundation, Centre for the Conceptual Foundation of Science, Kirchplatz 1., D-82049 Pullach im Isartal, Germany

[5] Department of Plant Systematics, Ecology and Theoretical Biology, Eötvös Loránd University, Pázmány P. sétány 1/C, H-1117 Budapest, Hungary

[6] Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine, Vienna, Savoynestrasse 1a, 1160 Vienna, Austria

[*] Corresponding author: szamado.szabolcs@tk.mta.hu

Abstract

Explaining the evolution of honest versus dishonest signals under conflicts of interest has long posed a major challenge, but several recent developments should spur renewed interest in this problem. First, the Handicap Principle, which maintains that signals must be costly to be honest, has been refuted and the model that claimed to validate this idea has been shown to have been misinterpreted. Second, more recent theoretical models demonstrate that signal honesty can be maintained by condition-dependent signalling trade-offs rather than costs. Third, we propose that signalling trade-offs may provide a general theory of honest signalling. According to signalling trade-off theory, signallers that differ in quality face different trade-offs at the honest equilibrium and therefore they are bound to invest differently. Such differential trade-offs, or lack of, can explain honest versus dishonest signals according to both experiments and models. Signalling trade-offs are found in every example of honest communication in nature under conflict of interest. Moreover, signalling trade-offs couple various fitness components, including both short-term investments into long-term fitness benefits, providing the necessary link between proximate and evolutionary explanations. Furthermore, trade-offs can also help bridge biological and economic theories of honest communication, which have developed independently in parallel for decades.

The evolution of honest signals

Explaining why humans and other species communicate honestly has puzzled scientists and layman for centuries. Communication allows individuals to exchange information and influence each other's behaviour (Maynard Smith and Harper 1995, 2003, Searcy and Nowicki 2006). It is central to mediating inter- and intra-specific (social and sexual) interactions and is essential for the organization of human societies. The problem is that information can be unreliable, especially when there are conflicts of interest between senders and receivers and deception offers advantages (DePaulo et al. 1996, Christy and Rittschof 2011, Brown et al. 2012, Fischbacher and Föllmi-Heusi 2013, Dalziell and Welbergen 2016, Casewell et al. 2017, Wang et al. 2017, Fujisawa et al. 2020). Signals provide information, but they function to influence and persuade, rather than inform receivers per se (Dawkins and Krebs 1978), and they need not be honest – not even on average (Számadó 2000, 2008, 2017). Honest signalling theory aims to determine how honesty evolves and how it is maintained by natural selection, despite the constant risk of liars. Here, we (i) summarize why the most common explanation for honesty, the Handicap Principle (HP), is erroneous and can be fully rejected; (ii) argue why tradeoffs are central to analysing the evolution of honest signals; and (iii) explain how signalling tradeoffs provide a causal connection necessary to integrate proximate and ultimate explanations for both honest and dishonest signals. For terminology, see **Glossary**.

The Handicap Principle is dead

Research on honest signals for three decades has been dominated by Zahavi's (Zahavi 1975) Handicap Principle (HP) for three decades (reviewed in (Penn and Számadó 2020)). Its core tenet holds that '... in order to be effective, signals have to be reliable; in order to be reliable, signals have to be costly' (Zahavi and Zahavi 1997 p. XIV). It predicts that signals are wasteful, as well as costly, because 'by wasting one proves conclusively that one has enough assets to waste and more. The investment – the waste itself – is just what makes the advertisement reliable' (Zahavi 1981, 1987, Zahavi and Zahavi 1997 p. 229). The HP was initially criticised for being illogical and non-Darwinian, as it suggests that honest signals evolve because and not despite of their costs (Dawkins 1976). In response to critics, Zahavi clarified his theory and proposed that signals are different from other traits because they evolve under a non-Darwinian process of 'signal selection' that favours waste rather than efficiency (Zahavi

1981, 1987, Zahavi and Zahavi 1997). Moreover, he asserted that his idea provides a general principle, the Handicap Principle, to explain honest signals. The HP remained controversial, but then it gained widespread acceptance after Grafen (Grafen 1990) reportedly validated Zahavi’s verbal arguments for the HP with his 'strategic signalling handicap' model. Grafen concluded that ‘If we see a character which does signal quality, then it must be a handicap. The handicap principle lies at the heart of evolutionary signalling, and must therefore play a major role in our understanding of it’ (Grafen 1990 p. 521). Grafen's conclusions were widely accepted, even by previous critics (Dawkins 1990, Godfray 1991, Maynard Smith 1991, Hammerstein and Hagen 2005), and the HP subsequently became the most-cited and most common textbook explanation for honest signalling.

The HP was also embraced in the social sciences, including evolutionary anthropology (Bliege Bird et al. 2001, Hawkes and Bliege Bird 2002, McAndrew 2021) and evolutionary psychology (Miller 2011), though often re-labelled as 'costly signalling theory'. Although some have become critical of the HP (Bliege Bird et al. 2018, Barker et al. 2019, Stibbard-Hawkes 2019), this idea remains widely cited.

Yet, both the handicap hypothesis and the broader HP can be rejected for theoretical and empirical reasons, which will help to redress the enormous confusion that this idea has generated.

First, the theoretical arguments for the HP have been completely refuted. Contrary to the HP, signals need not be costly to be honest at the honest equilibrium (Hurd 1995, Számadó 1999, Lachmann et al. 2001, Bergstrom et al. 2002), and signal cost itself (paid at the equilibrium) cannot maintain honesty at the evolutionary equilibrium (Lachmann et al. 2001, Számadó et al. 2023). That is, signal cost paid at the equilibrium a.k.a. handicap is neither necessary nor sufficient condition of honest signalling; see fallacy of the HP (Számadó 2011). The problem, it turns out, is that Grafen has misinterpreted his strategic signalling model as a model of the HP (Grafen 1990), while it is not (Penn and Számadó 2020). In his model, poor-quality signallers pay a higher fitness cost for signalling compared to high-quality males (differential signalling costs). More precisely, the marginal cost for low-quality males to produce a high intensity signal (cheating) is greater than the marginal benefit gained by cheating at the honest equilibrium (Hurd 1995, Számadó 1999, Lachmann et al. 2001, Számadó et al. 2023). Males honestly signal their quality in this model, not because their signals are wasteful or costly per se, but because there is an inescapable cost for deception (and more specifically, a net marginal fitness cost for cheating). Another version of this model shows that selection also favours honest signals if they confer differential fitness benefits (Godfray 1991, Getty 1998a, Godfray and Johnstone 2000). This model demonstrates that differential signal costs is not necessary for honesty at the evolutionary equilibrium. Yet, these theoretical models and other hypotheses were originally misinterpreted as versions of the HP (Penn and Számadó 2020) (see Box 1). Since the existing theoretical models contradict the handicap hypothesis, as well as the broader claims of the HP, this idea can be rejected on theoretical grounds alone.

Second, the empirical evidence supporting the HP has been misinterpreted and the anomalous findings have been ignored (Penn and Számadó 2020)**.** Despite decades of empirical research on the HP, there is no support for any of its predictions. For example, there is no evidence that signal costs predict their reliability, or that signals are wasteful and more costly than they need to be (i.e., 'strategic costs' sensu (Maynard Smith and Harper 1995, 2003), as predicted by the HP (Számadó 2011, Penn and Számadó 2020). Numerous studies have failed to find support for ‘costly signals’ (Borgia 1993, 1996, McCullough and Emlen 2013, Guimarães et al. 2017), including offspring begging calls (McCarty 1996), see (Moreno-Rueda 2006) for review), and on the contrary, many results directly contradict the core tenet of the HP. Even the peacock’s elaborate plumage, flagship example of the HP, has been shown to increase rather than decrease efficiency and flight performance (Askew 2014, Thavarajah et al. 2016). It has been demonstrated that signals can be cheap and efficient (rather than costly and wasteful) and yet still honest (Számadó et al. 2019, 2022, 2023). **Many results have been interpreted as supporting the HP, but they confused correlation with causation.** Signals have often been found that are both honest and costly; but such results do not show that signals are honest because they are costly (also see the Affirming the consequent (Penn and Számadó 2020)). Thus, there is no empirical evidence to support the HP, and conducting further experiments would be futile as there are no models of the handicap hypothesis to test.

Third, the HP has generated enormous confusion, especially due to muddling proximate and ultimate levels of explanation (Penn and Számadó 2020). Beginning with Zahavi (Zahavi 1975), researchers have often confused proximate mechanisms proposed to explain honest signals with evolutionary or 'ultimate' explanations for their selective maintenance. Both types of explanations are necessary, but it is crucial to avoid confusing them. Signals with metabolic and other proximate costs were labelled as ‘handicaps,’ which resulted in a gamut of various ‘handicaps,’ including immunocompetence handicaps, revealing handicaps, vulnerability handicaps, and performance handicaps (Folstad and Karter 1992, Vehrencamp 2000, Hurd and Enquist 2005). Labelling traits as ‘handicaps’ provided no explanatory power, and adopting the illogical handicap paradigm and equating proximate with ultimate costs contributed to the confusion. Metabolic and other proximate costs of signalling have been used as proxies for fitness, which also led to fruitless debates over how high such costs must be to support the HP (Kotiaho 2001, Searcy and Nowicki 2006). To avoid confusion, signals having metabolic or other physiological costs at the proximate level are better labelled as *investments* or *expenditures*, rather than *costs* or *handicaps*, and they cannot be assumed that they have fitness costs (handicaps), especially since increased investment may instead enhance their fitness benefits. There are many additional reasons why the HP has generated enormous conceptual confusion and should therefore be rejected.

Rejecting the Handicap Principle means that decades of research on honest signalling needs to be re-evaluated (Box 1). Interestingly, all of the other hypotheses that have been proposed to explain honest signals were originally mistakenly interpreted as versions of the HP, and also like Grafen's model (Grafen 1990), these proposals all predict that honesty is selectively maintained by signalling tradeoffs (Box 2). Next, we explain how a general theory of honest signalling based on tradeoffs might be possible and could help to better integrate proximate and evolutionary analyses.

Signalling trade-offs: towards a unified theoretical framework for explaining honest and dishonest signalling

Here we propose a novel, unifying theoretical approach for studying honest signalling under conflict of interest based on signalling trade-offs (Box 2). While early theoretical models in biology emphasised the role of equilibrium costs of signals (Grafen 1990, Godfray 1991), subsequent models show that honesty is instead maintained by signalling trade-offs (Számadó et al. 2023). That is, honesty is enforced by condition-dependent trade-offs between the marginal cost and marginal benefits of honest versus dishonest signalling behaviour (Hurd 1995, Számadó 1999, Lachmann et al. 2001, Bergstrom et al. 2002) (see (Számadó 2011, Higham 2014) for reviews). A recent experimental test of a signalling game with human subjects demonstrated that condition-dependent trade-offs can account for honesty, whereas equilibrium signal costs cannot (Számadó et al. 2022): honesty was observed in all treatments with a condition-dependent signalling trade-off *regardless of the cost of equilibrium signals* (costly, cost-free or beneficial). Even signals providing immediate benefits can maintain honest signalling, as long as there is a condition-dependent trade-off (Számadó et al. 2022, 2023). Consequently, traits with condition-dependent trade-offs need not be harmful, wasteful or even costly for honest individuals at the equilibrium. In other words, honesty is maintained by the fact that different types of signallers face different trade-offs (see Figure 1). The *differential investment* (or proximate cost) is the result of such differences in trade-offs.

These trade-offs might be obvious in some contexts; e.g., a tradeoff between mating success versus survival (Darwin 1859, Poulton 1890, White et al. 2022) or between present versus future offspring, and yet obscure in others (e.g., in gossip, reputation systems, advertisement, PR, and modern social media). Despite their obscurity, signalling tradeoffs potentially govern honesty in humans and non-human species, and both in the biological and the economical contexts: every single signalling model that has an honest equilibrium also requires a signalling trade-off and thus, identifying and understanding these tradeoffs is central to explaining honest and also dishonest signals.

Deceptive signals have been too often overlooked in honest signalling theory, although even some proponents of the HP acknowledged it, cheating has been too often depicted as a rare, marginal behaviour (Johnstone and Grafen 1993, Maynard Smith and Harper 2003, Searcy and Nowicki 2006). Zahavi never addressed deception, other than to assert that signals should be honest and cheat-proof (Zahavi and Zahavi 1997). Yet, dishonest signals do exist and they also require a Darwinian explanation. Examples of dishonest signals abound in nature, including Batesian mimicry (Joron 2009), sexual mimicry ((Dominey 1980, Gross and Charnov 1980)) or predatory mimicry ((Lloyd 1965)). We propose that condition-dependent signalling tradeoffs can potentially explain dishonest as well has honest signals.

Understanding the adaptive function of honest and dishonest signals – like all traits requires determining their costs and benefits, thus signalling tradeoffs provides a unified framework for studying signal efficiency and the existence of both honest and dishonest signals. Whereas honesty is maintained by the fact that signallers of different qualities (condition) have different tradeoffs, dishonesty will evolve when different signallers have the same or very similar trade-offs (see (Számadó et al. 2022) *trade-off treatment*). If edible and toxic butterfly species face the same tradeoff when producing a given wing pattern then both can produce it, one being dishonest; if both males and females face the same tradeoff when producing a given morphological feature (advantageous to both in some context) then both can produce it, etc. In other words, dishonesty is the expected outcome when signalling tradeoffs are not depending on condition.

Finally, signalling tradeoffs clearly address how proximate investments are translated to ultimate fitness benefits: life-history tradeoffs explicitly provide the fitness return for a given investment (see Figure 1). Note, that at the honest equilibrium, signallers of different qualities obtain different returns from the same investment, resulting in *differential investment*, i.e., honest signalling. Practically, tradeoffs define the '*exchange rate*' of the invested 'currency' of individuals into fitness gains. High quality individuals enjoy a more favourable exchange rate (better tradeoffs) than low quality ones at the honest equilibrium (see (Getty 1998b) 'efficiency principle'), thus they can invest more. In other words, the same investment by a high-quality signaller yields a larger fitness increase than by a low-quality signaller (see Figure 2).

Concluding remarks

Explaining the evolution of honest and dishonest signals presents a challenging problem. Unfortunately, the Handicap Principle has been a serious setback by generating confusion for many reasons. First, equilibrium signal cost (a.k.a. handicap) is neither necessary nor sufficient condition of honest signalling (Számadó 1999, 2011, Lachmann et al. 2001, Számadó et al. 2023). Second, the HP misinterprets classic sexual selection tradeoff and life-history models as 'costly signalling' models. All the models that claimed support for the HP (Grafen 1990, Godfray 1991); see Box 1) are based on life-history tradeoffs, whereas none show the necessity of equilibrium cost. Third, the HP confused correlation with causation. There are honest and costly signals in nature (costly in terms of proximate investments), yet these signals are not honest because they are costly but because trade-offs guarantee that ***honest individuals are the most efficient*** (Getty 1998b, Számadó 1999, Lachmann et al. 2001, Számadó et al. 2022, 2023). Lastly, the HP confused proximate investment with ultimate fitness cost. Proximate investment (time, energy, resources, etc.) is a necessity of life; when this investment is returned as increased reproduction, it is never a waste, nor a handicap, not even fitness cost. Animals (and humans) invest in the short-term to accrue fitness benefits on the long run.

There is a need for a new unified theory of honest and dishonest signalling, now that both the Handicap Principle and the handicap hypothesis can be rejected. Signalling tradeoffs offer an opportunity for unification under Darwinian principles. The presence or absence of condition-dependent life-history tradeoffs can account for both honest and dishonest signals. Signalling tradeoffs readily translate short-term investments into long-term fitness benefits, thus providing the exchange rate between proximate investments and fitness benefits. Every signalling model that has an honest equilibrium under conflict of interest also assumes a signalling tradeoff. We predict that such condition-dependent tradeoffs are to be found in every honest communication system in nature as well.

Outstanding questions

- Re-evaluate the past three decades of research on honest signals and replace the handicap paradigm with an evolutionary framework based on life-history evolution and integrating proximate and evolutionary analyses (Penn and Számadó 2020, Számadó et al. 2023).
- Identify the costs, benefits and signalling tradeoffs in different contexts in nature.
- Re-classify the empirical literature based on the life-history tradeoffs present in the given context.
- Understand and classify the various life-history tradeoffs that were proposed in signalling games.
- Provide predictions based on the different tradeoffs.
- Test predictions in vivo and in vitro; using animal models in the laboratory.
- Integrate research on proximate mechanisms of traits with their evolutionary consequences by focusing on how tradeoffs at the proximate level influence fitness and vice versa (Box 1) (Garland et al. 2022).

Boxes and figures

Box 1. All viable hypotheses for honest signalling involve tradeoffs

There are four main hypotheses proposed to explain honest signalling, which are logical and Darwinian, despite that they were originally mistakenly classified as versions of the HP (Penn and Számadó 2020). First, strategic signalling models, although originally misinterpreted as handicap models (Dawkins 1990, Grafen 1990, Godfray 1991), explain honesty through tradeoffs rather than costs per se. Grafen's strategic signalling model (Grafen 1990) is a life-history model in which males are predicted to allocate investment into mating signals, depending upon their quality or condition, where the optimum investment is defined by tradeoffs between survival and reproduction (Box 1). The first version of this model was used to investigate signalling games with information asymmetries (Spence 1973). A related model shows how offspring begging calls can be honest indicators of hunger (Godfray 1991) due to a life-history tradeoff between the survival of current versus future offspring. Second, honesty can be due to inescapable constraints that make signals impossible or difficult for individuals to fake, but such 'index signals' were also originally misinterpreted as 'revealing handicaps' (Maynard Smith and Harper 1995, 2003). These two types of hypotheses are not mutually exclusive alternatives, as index signals provide a proximate rather than an evolutionary explanation (Holman 2012, Biernaskie et al. 2014, 2018), and proximate constraints can become selective tradeoffs in evolutionary models. Third, specific proximate mechanisms have been proposed to explain how honest signal expression is regulated through endocrine signals that allocate investment into sexual signalling versus other traits (e.g., immunocompetence handicap hypothesis (Folstad and Karter 1992), oxidative handicap hypothesis (Alonso-Alvarez et al. 2006, 2008). These ideas were also originally confused with the HP, but they are logical and consistent with life-history tradeoffs (Wedekind and Folstad 1994), even if the precise mechanisms are controversial and unresolved (Roberts et al. 2004). Fourth, signals can be honest due to social punishment or other risks inherent in social interactions (Enquist 1985, Számadó 2003, 2008, 2011). These signals and the models describing them were also misinterpreted as handicaps or 'vulnerability handicaps' (Grafen 1990, Zahavi and Zahavi 1997, Vehrencamp 2000, Hurd and Enquist 2005). Yet all these models describe a viability versus resource acquisition trade-off, where strong (high quality) individuals are better off by investing in resource acquisition, whereas weak individuals are better off by investing in viability. These seemingly different hypotheses for honest signals are all based on tradeoffs at different levels of biological organisation, and the challenge is to determine whether a more general theory can be found that connects various proximate and ultimate tradeoffs.

Box 2. Trade-offs are central to evolutionary theory

Tradeoffs are a type of constraint that prevent the simultaneous maximization of two or more traits, and they are central to evolutionary theory (Roff 1993, Reznick et al. 2000, Roff and Fairbairn 2007), e.g., life history theory (survival versus reproduction (Stearns 1989, 1992)), sex allocation theory (Charnov 1982), evolution of virulence (Anderson and May 1982), adaptive behaviour (Houston et al. 2023), and aging (growth and reproductive versus repair (Kirkwood 1977)). Theoretical analyses of tradeoffs were first formally investigated in economics (e.g., see opportunity costs, analysed with production-possibility frontier graphs and the Pareto front, which involves the optimization of multiple functions). Sexual selection theory since Darwin has predicted that secondary signals evolve under fitness tradeoffs, i.e., they enhance mating success (reproductive benefits), but can also attract predators (viability costs). This is a classic example of a selective tradeoff, and there is much empirical evidence for such sexual signalling tradeoffs (e.g., see (Heinen-Kay et al. 2015); and review by (White et al. 2022)). Secondary sexual traits generally provide reliable indictors of condition (condition-dependence) (Dougherty 2021), and the challenge is to explain why (Fromhage and Henshaw 2022). Honest signalling theory predicts that sexual signal expression has differential fitness cost/benefits depending upon condition (Grafen 1990). However, most research has been distracted with testing the handicap predictions, and surprisingly few studies have attempted to test the actual model (e.g., (Møller et al. 1994, Kotiaho 2001, Dinh and Patek 2023) – and these erroneously concluded to test and support the HP. Secondary sexual signals, as well as any life-history trait, are expected to have fitness tradeoffs (Darwin 1859, White et al. 2022), and yet, sexual signal expression is usually positively correlated with longevity (Jennions et al. 2001). Van Noordwijk and de Jong pointed out that when there is individual heterogeneity in quality or condition due to acquisition or allocation of resources, then this could mask actual tradeoffs and even generate positive correlations between reproduction and longevity (Noordwijk and Jong 1986, Jong 1993, Laskowski et al. 2021). They suggested that energy allocation tradeoffs occur when individual variation in resource allocation is greater than variation in resource acquisition (Y model), and their prediction holds even when relaxing the assumption that resource allocation is independent from acquisition (Descamps et al. 2016). Therefore, more empirical studies are needed to test the predictions of signalling theory, and our aim here is to help clarify the theoretical predictions.

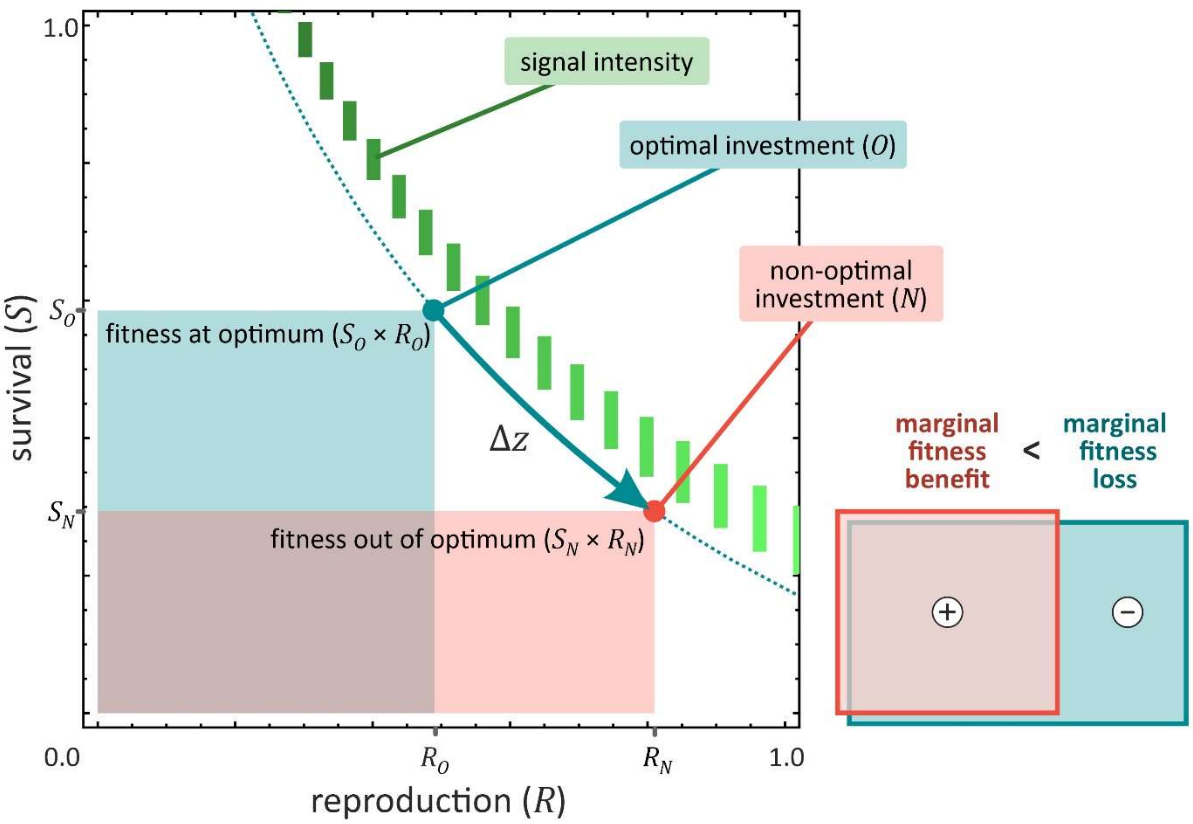

Figure 1. Relation between proximate investment and ultimate benefits as specified by a life-history trade-off, modified from Grafen (Grafen 1990)*. The dotted curve shows a life-history trade-off of a Signaller between reproduction success R and survival probability S. Reproductive success depends on the immediate (e.g., energetic) cost of signalling: this is the proximate investment, that increases along the trade-off curve as the signal intensity increases from left to the right (green bars). Ultimate fitness (dependent on signal intensity and Signaller quality) is measured by the actual area of the rectangles corresponding to the position on the trade-off curve (fitness is the multiple of R and S). When moving out of the local optimum (from blue dot to red dot) by increasing signal intensity by Δz (green bars), fitness is decreased, not because of increasing signal cost but because the marginal cost paid (the area lost from the blue fitness rectangle, depicted as the blue box on the right) surpasses the marginal benefit gained (the area gained by the pink fitness rectangle, depicted as the pink box on the right). The same would happen if the Signaller decreased signal intensity (i.e., moving to the left on the curve; fitness rectangle not depicted), proving that fitness optimum does not directly depend on signal cost (contrary to the main tenet of the HP).*

Box 3. Life-history trade-offs under the hood of seminal models of honest signalling

Like Grafen's model, all seminal models of so-called 'costly signalling' rely on life-history and other tradeoffs, often hidden beneath complex assumptions. As a result, none of them supports the claim that honesty requires equilibrium costs (Lachmann et al. 2001, Számadó et al. 2023). Grafen assumed sexual signals are condition-dependent and that signallers face a trade-off between investing into signals for reproduction versus survival (Grafen 1990), Godfray assumed a trade-off between the survival of current versus future offspring (Godfray 1991) (see (Számadó et al. 2019) for details). Honest signalling in these models is always maintained exclusively by these condition-dependent trade-offs instead of any equilibrium cost (Számadó et al. 2023). Figure 2 shows that Signallers of different qualities face different life-history trade-offs in Grafen's model (Grafen 1990). Due to the different trade-offs, the simultaneous optimization of Signaller and Receiver strategies is ensured.

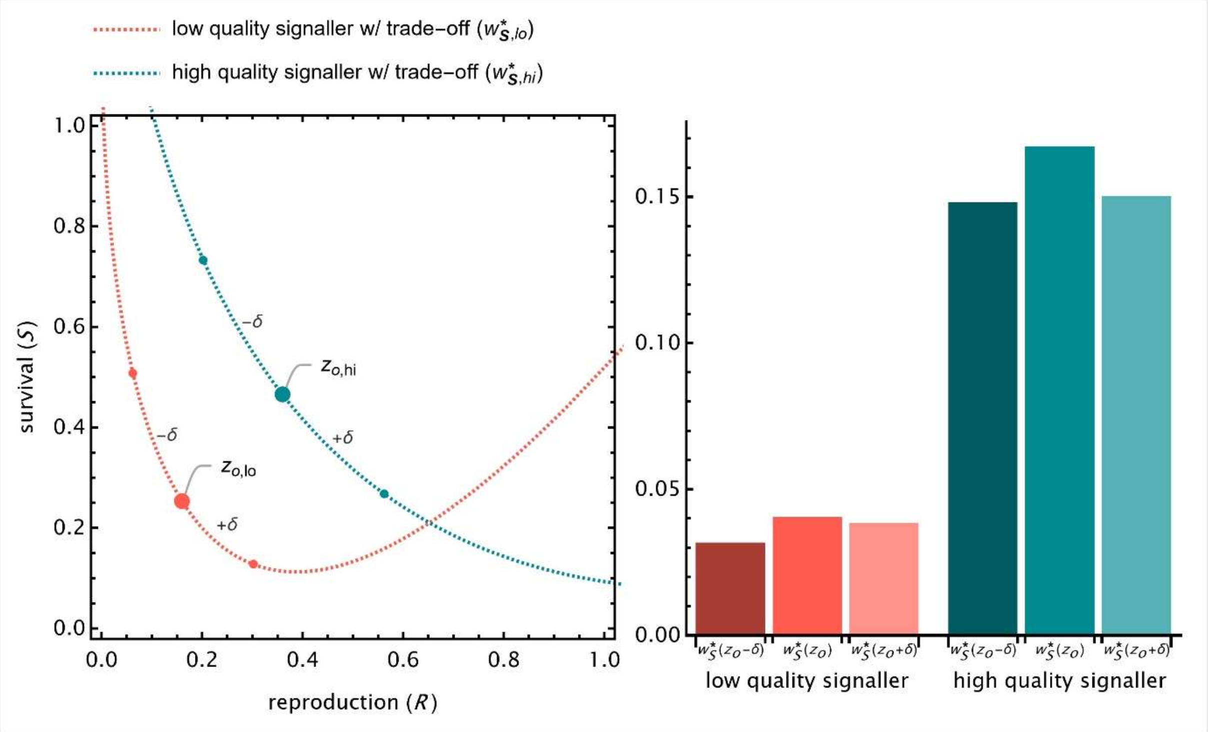

Figure 2. Life-history trade-offs explain honesty in Grafen's model (Grafen 1990). *The blue and red curves show the fitness of high- and low-quality Signallers, respectively, under tradeoff between survival probability* S *and reproductive success* R. *Large dots indicate the fitness optima for the different signallers where fitness is maximized and the trait value resolves the double-optimization problem of Signaller and Receiver. In the optimum, both Signaller and Receiver are at equilibrium, and the Receiver shares exactly what the Signaller asks for,* z_o. *Note, that the optimum is at different* z *values for low- and high-quality Signallers, red and teal dots respectively, due to their different tradeoffs (i.e.* $z_{o,lo}$ *does not equal* $z_{o,hi}$*). Small dots show out-of-optimum values, when the signal is changed by increasing intensity (to request a different amount of resource) with* $-\delta$ *(left) or* $+\delta$ *(right). The same* δ *in signal intensity yields different fitness changes for low- and high-quality Signallers. Bars on the right show the actual Signaller fitness values at the optima (middle) and out of optima (flanking) for low-quality (red) and high-quality (teal) individuals. Any deviation from the optimum trait value results in sub-optimal fitness. This marginal decrease of fitness when leaving the optimum is what maintains honesty at the equilibrium. There is no ultimate cost paid at the equilibrium as both reproduction and survival is a positive contribution to fitness, though there could be proximate investments (see Figure 1). Note, that the same* δ *in* z *yields a larger fitness change for a high-quality than for a low-quality Signaller.*

Glossary

Affirming the consequent (fallacy of the converse): A logical fallacy due to inverting a conditional statement so that the conclusion becomes the premise. For example, if a theoretical model uses costs to explain honest signals, then erroneously concluding that it is a model of the HP. For example: Jill got sunburnt in Egypt; seeing Jack has sunburns too, it is fallacious to conclude that he too was in Egypt.

Benefit: Positive effects on individual survival or reproductive success.

Cheating: Producing dishonest or deceptive signals (see dishonest signal), and the functional definition neither requires nor implies intentionality.

Condition-dependence: Phenotypic plasticity in the expression of signals or other traits that are constrained because depend upon an individual's condition (quality), i.e., nutritional state, health, etc. Such plasticity can be adaptive or maladaptive.

Cost: Negative effects on individual survival or reproductive success (functional definition), but also used for proximate 'costs' (suppressive effects on trait expression and energetic/metabolic investment into traits), even if such effects are functional and provide fitness benefits.

Costly signalling theory: Hypotheses and theoretical models explaining the selective maintenance of honest signals by differential signalling costs, and often mistakenly interpreted as models of the HP (see (Penn and Számadó 2020)).

Dishonest signal: A signal that is not reliably correlated with an attribute of a signaller (e.g., quality or identify), and functions to mislead receivers; intentionality is not necessary or implied (e.g., see Dennet's intentional stance (Dennett 1989)).

Double-optimization problem: Situations in which a sender and receiver's signalling strategy vary and depend upon each other's strategy, and therefore, requires finding the optima of both parties.

Equilibrium cost: Fitness cost of signalling at the evolutionary equilibrium.

Evolutionarily stable strategy (ESS): A strategy (morphology, behaviour, or both) that once fixed cannot be outcompeted and replaced by an alternative strategy invading a population; an equilibrium refinement of the Nash equilibrium, which it defines a stable condition at the evolutionary equilibrium (Maynard Smith 1982). In other words, it does not pay to deviate and adopt another strategy on the evolutionary timescale.

Fitness: Various measures of individual (or genetic) survival and reproductive success. It is not a measure of physical fitness and, contrary to what is often assumed, signals cannot provide direct indicators of fitness (since individual fitness cannot be known until after death).

Handicap: Originally, referred to a signal with a survival cost (at the evolutionary equilibrium), but then it became used refer to hypothetical signals that are honest and evolve because they are costly (*Zahavian handicap*). The term *cost* is usually interpreted to include either or both proximate and ultimate costs. We recommend avoiding this confusing term.

Handicap Principle (HP): Widely used as a synonym for Zahavi's handicap hypothesis, which predicts that signals are wasteful as well as costly, and that they are maintained because signal costs enforce honesty. It also refers to Zahavi's broader claim that this hypothesis provides a general principle to explain honest signals.

Honest signal: A signal that is consistently correlated with an attribute of the signaller, such as their condition, status, identity, or resources (versus dishonest signal, see above) and does not imply intentionality (see signal).

Hybrid equilibrium: An evolutionary equilibrium in which low quality signallers use a mixed strategy giving both low- and high-quality signals and receivers respond with a mixed strategy as well: with probability p, they respond positively to signals and with probability $1-p$ they reject them. Such mixed equilibria are expected to evolve when the marginal cost of producing a high-quality signal for low-quality individuals is not high enough to support a separating equilibrium (Zollman et al. 2013).

Life history theory: Analytical framework used to study the evolution of mechanisms and strategies to allocate investment into survival (growth and repair) versus reproduction (number and quality of offspring). A key postulate is that adaptive allocation mechanisms evolve under certain tradeoffs and other constraints that limit evolutionary trajectories.

Marginal cost: The cost to increase the intensity or expression of a signal, which can be represented and approximated by the rate of change of the cost function, $C(x)$, defined as the derivative of the cost function $C'(x)$.

Potential cost of cheating: The potential fitness cost paid by dishonest signallers (cheaters) if they produce a dishonest signal. It is the marginal cost of cheating. This cost is not paid by honest signallers, or when all individuals are at the honest equilibrium.

Proximate cost: Individual investment of time, energy, or any other resources (such as pigments) into the expression or function of a trait. Not to be confused with ultimate fitness cost.

Quality: See condition.

Realized cost: See equilibrium cost.

Reliable signal: See honesty.

Signal: Any behaviour or structure produced (by a sender) that functions to transmit information and alter the behaviour of other organisms (receivers) in a way that benefits the sender. In contrast, *cues* transmit inadvertently information, which may benefit receivers but not senders.

Signal cost: Ultimate fitness costs of a signal. It may be different for different levels of quality of Signallers. This concept is often confused with the energetic and other proximate costs of signalling, but they are not equivalent (e.g., signals that require a high investment may have large energetic costs but can still evolve because they have low fitness costs or none).

Signalling trade-off: Type of constraint that prevent the simultaneous maximization of two or more traits in the context of signalling interactions, e.g. mating success versus survival, attracting mates versus attracting parasites and predators (Darwin 1859, Poulton 1890).

Trade-off: When the expression of one trait cannot increase without also decreasing the expression or function of another trait (proximate definition), or when a trait enhances one fitness component, but reduces another ('selective tradeoff'; evolutionary definition). The term is also used to refer to a situation in which a decision or strategic choice requires weighing the available options to find the optimal solution to allocate time, energy, or other limited resources.

Ultimate cost: Harmful effects on individual survival or reproductive success (fitness cost).

Funding

The authors acknowledge support from the Hungarian Research Fund under grant numbers NKFIH #144625 (SS), from the János Bolyai Research Scholarship of the Hungarian Academy of Sciences #BO/00570/22/8 (IZ), ÚNKP-23-5 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research Development and Innovation Fund #ÚNKP-23-5-ELTE-117 (IZ), from the Austrian Science Fund (FWF) (P28141-B25 and P36446-B) and Human Science Program (RGP000/2020) (DJP). For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

Competing interests

The authors declare that they have no competing interests.

References

Alonso-Alvarez, C., S. Bertrand, B. Faivre, O. Chastel, and G. Sorci. 2006. Testosterone and oxidative stress: the oxidation handicap hypothesis. Proceedings of the Royal Society B: Biological Sciences 274:819–825.

Alonso-Alvarez, C., L. Pérez-Rodríguez, R. Mateo, O. Chastel, and J. Viñuela. 2008. The oxidation handicap hypothesis and the carotenoid allocation trade-off. Journal of Evolutionary Biology 21:1789–1797.

Anderson, R. M., and R. M. May. 1982. Coevolution of hosts and parasites. Parasitology 85:411–426.

Askew, G. N. 2014. The elaborate plumage in peacocks is not such a drag. Journal of Experimental Biology 217:3237–3241.

Barker, J. L., E. A. Power, S. Heap, M. Puurtinen, and R. Sosis. 2019. Content, cost, and context: A framework for understanding human signaling systems. Evolutionary Anthropology: Issues, News, and Reviews 28:86–99.

Bergstrom, C. T., S. Számadó, and M. Lachmann. 2002. Separating equilibria in continuous signalling games. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357:1595–1606.

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

Biernaskie, J. M., J. C. Perry, and A. Grafen. 2018. A general model of biological signals, from cues to handicaps. Evolution Letters 2:201–209.

Bliege Bird, R., E. Ready, and E. A. Power. 2018. The social significance of subtle signals. Nature Human Behaviour 2:452–457.

Bliege Bird, R., E. Smith, and D. W. Bird. 2001. The hunting handicap: costly signaling in human foraging strategies. Behavioral Ecology and Sociobiology 50:9–19.

Borgia, G. 1993. The cost of display in the non-resource-based mating system of the satin bowerbird. The American Naturalist 141:729–743.

Borgia, G. 1996. Satin bowerbird displays are not extremely costly. Animal Behaviour 52:648–650.

Brown, C., M. P. Garwood, and J. E. Williamson. 2012. It pays to cheat: tactical deception in a cephalopod social signalling system. Biology Letters 8:729–732.

Casewell, N. R., J. C. Visser, K. Baumann, J. Dobson, H. Han, S. Kuruppu, M. Morgan, A. Romilio, V. Weisbecker, K. Mardon, S. A. Ali, J. Debono, I. Koludarov, I. Que, G. C. Bird, G. M. Cooke, A. Nouwens, W. C. Hodgson, S. C. Wagstaff, K. L. Cheney, I. Vetter, L. van der Weerd, M. K. Richardson, and B. G. Fry. 2017. The evolution of fangs, venom, and mimicry systems in blenny fishes. Current Biology 27:1184–1191.

Charnov, E. L. 1982. The theory of sex allocation. . Princeton University Press, Princeton, NJ.

Christy, J. H., and D. Rittschof. 2011. Deception in visual and chemical communication in crustaceans. Pages 313–333 *in* T. Breithaupt and M. Thiel, editors. Chemical Communication in Crustaceans. . Springer, New York, NY.

Dalziell, A. H., and J. A. Welbergen. 2016. Mimicry for all modalities. Ecology Letters 19:609–619.

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. Page 502. . John Murray, London.

Dawkins, R. 1976. The selfish gene. Page 224. . Oxford University Press, New York.

Dawkins, R. 1990. The selfish gene. . Oxford University Press.

Dawkins, R., and J. R. Krebs. 1978. Animal signals: Information or manipulation? Pages 282–309 *in* J. R. Krebs and N. B. Davies, editors. Behavioural ecology: An evolutionary approach.

Dennett, D. C. 1989. The intentional stance. . MIT Press, Cambridge, Mass.

DePaulo, B. M., D. A. Kashy, S. E. Kirkendol, M. M. Wyer, and J. A. Epstein. 1996. Lying in everyday life. Journal of Personality and Social Psychology 70:979–995.

Descamps, S., J.-M. Gaillard, S. Hamel, and N. G. Yoccoz. 2016. When relative allocation depends on total resource acquisition: implication for the analysis of trade-offs. Journal of Evolutionary Biology 29:1860–1866.

Dinh, J. P., and S. N. Patek. 2023. Weapon performance and contest assessment strategies of the cavitating snaps in snapping shrimp. Functional Ecology 37:327–342.

Dominey, W. J. 1980. Female mimicry in male bluegill sunfish—a genetic polymorphism? Nature 284:546–548.

Dougherty, L. R. 2021. Meta-analysis reveals that animal sexual signalling behaviour is honest and resource based. Nature Ecology & Evolution 5:688–699.

Enquist, M. 1985. Communication during aggressive interactions with particular reference to variation in choice of behaviour. Animal Behaviour 33:1152–1161.

Fischbacher, U., and F. Föllmi-Heusi. 2013. Lies in disguise-an experimental study on cheating. Journal of the European Economic Association 11:525–547.

Folstad, I., and A. J. Karter. 1992. Parasites, bright males, and the immunocompetence handicap. The American Naturalist 139:603–622.

Fromhage, L., and J. M. Henshaw. 2022. The balance model of honest sexual signaling. Evolution 76:445–454.

Fujisawa, M., Y. Sakai, and T. Kuwamura. 2020. The false cleanerfish relies on aggressive mimicry to bite fish fins when benthic foods are scarce in their local habitat. Scientific Reports 10:8652.

Garland, T., C. J. Downs, and A. R. Ives. 2022. Trade-offs (and constraints) in organismal biology. Physiological and Biochemical Zoology 95:82–112.

Getty, T. 1998a. Handicap signalling: when fecundity and viability do not add up. Animal Behaviour 56:127–130.

Getty, T. 1998b. Reliable signalling need not be a handicap. Animal Behaviour 56:253–255.

Godfray, H. C. J. 1991. Signalling of need by offspring to their parents. Nature 352:328–330.

Godfray, H. C. J., and R. A. Johnstone. 2000. Begging and bleating: the evolution of parent–offspring signalling. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 355:1581–1591.

Grafen, A. 1990. Biological signals as handicaps. Journal of Theoretical Biology 144:517–546.

Gross, M. R., and E. L. Charnov. 1980. Alternative male life histories in bluegill sunfish. Proceedings of the National Academy of Sciences 77:6937–6940.

Guimarães, M., R. Munguía-Steyer, P. F. Doherty, and R. J. Sawaya. 2017. No survival costs for sexually selected traits in a polygynous non-territorial lizard. Biological Journal of the Linnean Society 122:614–626.

Hammerstein, P., and E. H. Hagen. 2005. The second wave of evolutionary economics in biology. Trends in Ecology & Evolution 20:604–609.

Hawkes, K., and R. Bliege Bird. 2002. Showing off, handicap signaling, and the evolution of men's work. Evolutionary Anthropology: Issues, News, and Reviews 11:58–67.

Heinen-Kay, J. L., K. E. Morris, N. A. Ryan, S. L. Byerley, R. E. Venezia, M. N. Peterson, and R. B. Langerhans. 2015. A trade-off between natural and sexual selection underlies diversification of a sexual signal. Behavioral Ecology 26:533–542.

Higham, J. P. 2014. How does honest costly signaling work? Behavioral Ecology 25:8–11.

Holman, L. 2012. Costs and constraints conspire to produce honest signaling: insights from an ant queen pheromone. Evolution 66:2094–2105.

Houston, A. I., L. Fromhage, and J. M. McNamara. 2023. A general framework for modelling trade-offs in adaptive behaviour. Biological Reviews n/a.

Hurd, P. L. 1995. Communication in discrete action-response games. Journal of Theoretical Biology 174:217–222.

Hurd, P. L., and M. Enquist. 2005. A strategic taxonomy of biological communication. Animal Behaviour 70:1155–1170.

Jennions, M. D., A. P. Moller, and M. Petrie. 2001. Sexually selected traits and adult survival: A meta-analysis. The Quarterly Review of Biology 76:3–36.

Johnstone, R. A., and A. Grafen. 1993. Dishonesty and the handicap principle. Animal Behaviour 46:759–764.

Jong, G. de. 1993. Covariances between traits deriving from successive allocations of a resource. Functional Ecology 7:75–83.

Joron, M. 2009. Mimicry. Pages 633–643 *in* V. H. Resh and R. T. Cardé, editors. Encyclopedia of Insect. . Academic Press, San Diego.

Kirkwood, T. B. L. 1977. Evolution of ageing. Nature 270:301–304.

Kotiaho, J. S. 2001. Costs of sexual traits: a mismatch between theoretical considerations and empirical evidence. Biological Reviews 76:365–376.

Lachmann, M., S. Számadó, and C. T. Bergstrom. 2001. Cost and conflict in animal signals and human language. Proceedings of the National Academy of Sciences 98:13189–13194.

Laskowski, K. L., M. Moiron, and P. T. Niemelä. 2021. Integrating behavior in life-history theory: Allocation versus acquisition? Trends in Ecology & Evolution 36:132–138.

Lloyd, J. E. 1965. Aggressive mimicry in |Photuris : firefly femmes fatales. Science 149:653–654.

Maynard Smith, J. 1982. Evolution and the theory of games. . Cambridge University Press.

Maynard Smith, J. 1991. Honest signalling: the Philip Sidney game. Animal Behaviour 42:1034–1035.

Maynard Smith, J., and D. Harper. 2003. Animal signals. Page 166. . Oxford University Press, Oxford.

Maynard Smith, J., and D. G. C. Harper. 1995. Animal signals: Models and terminology. Journal of Theoretical Biology 177:305–311.

McAndrew, F. T. 2021. Costly signaling theory. Pages 1525–1532 *in* T. K. Shackelford and V. A. Weekes-Shackelford, editors. Encyclopedia of Evolutionary Psychological Science. . Springer.

McCarty, J. P. 1996. The energetic cost of begging in nestling passerines. The Auk 113:178–188.

McCullough, E. L., and D. J. Emlen. 2013. Evaluating the costs of a sexually selected weapon: big horns at a small price. Animal Behaviour 86:977–985.

Miller, G. F. 2011. The mating mind: How sexual choice shaped the evolution of human nature. Page 528. . Knopf Doubleday Publishing Group.

Møller, A. P., F. de Lope, J. Moreno, G. González, and J. J. Pérez. 1994. Ectoparasites and host energetics: house martin bugs and house martin nestlings. Oecologia 98:263–268.

Moreno-Rueda, G. 2006. Is there empirical evidence for the cost of begging? Journal of Ethology 25:215–222.

Noordwijk, A. J. van, and G. de Jong. 1986. Acquisition and allocation of resources: Their influence on variation in life history tactics. The American Naturalist 128:137–142.

Penn, D. J., and S. Számadó. 2020. The Handicap Principle: How an erroneous hypothesis became a scientific principle. Biological Reviews 95:267–290.

Poulton, E. B. 1890. The colours of animals. . Paul, Trench, Trübner & Co., London, UK.

Reznick, D., L. Nunney, and A. Tessier. 2000. Big houses, big cars, superfleas and the costs of reproduction. Trends in Ecology & Evolution 15:421–425.

Roberts, M. L., K. L. Buchanan, and M. R. Evans. 2004. Testing the immunocompetence handicap hypothesis: a review of the evidence. Animal Behaviour 68:227–239.

Roff, D. 1993. Evolution of life histories: Theory and analysis. . Springer US.

Roff, D. A., and D. J. Fairbairn. 2007. The evolution of trade-offs: Where are we? Journal of Evolutionary Biology 20:433–447.

Searcy, W. A., and S. Nowicki. 2006. Signal interception and the use of soft song in aggressive interactions. Ethology 112:865–872.

Spence, M. 1973. Job market signaling. The Quarterly Journal of Economics 87:355.

Stearns, S. C. 1989. Trade-offs in life-history evolution. Functional Ecology 3:259.

Stearns, S. C. 1992. The evolution of life histories. Page 249. . Oxford University Press, Oxford.

Stibbard-Hawkes, D. N. E. 2019. Costly signaling and the handicap principle in hunter-gatherer research: A critical review. Evolutionary Anthropology: Issues, News, and Reviews 28:144–157.

Számadó, S. 1999. The validity of the handicap principle in discrete action–response games. Journal of Theoretical Biology 198:593–602.

Számadó, S. 2000. Cheating as a mixed strategy in a simple model of aggressive communication. Animal Behaviour 59:221–230.

Számadó, S. 2003. Threat displays are not handicaps. Journal of Theoretical Biology 221:327–348.

Számadó, S. 2008. How threat displays work: species-specific fighting techniques, weaponry and proximity risk. Animal Behaviour 76:1455–1463.

Számadó, S. 2011. The cost of honesty and the fallacy of the handicap principle. Animal Behaviour 81:3–10.

Számadó, S. 2017. When honesty and cheating pay off: the evolution of honest and dishonest equilibria in a conventional signalling game. BMC Evolutionary Biology 17:270.

Számadó, S., D. Czégel, and I. Zachar. 2019. One problem, too many solutions: How costly is honest signalling of need? PLoS ONE 14:e0208443.

Számadó, S., F. Samu, and K. Takács. 2022. Condition-dependent trade-offs maintain honest signalling. Royal Society Open Science 9.

Számadó, S., I. Zachar, D. Czégel, and D. J. Penn. 2023. Honesty in signalling games is maintained by trade-offs rather than costs. BMC Biology 21.

Thavarajah, N. K., P. G. Tickle, R. L. Nudds, and J. R. Codd. 2016. The peacock train does not handicap cursorial locomotor performance. Scientific Reports 6.

Vehrencamp, S. L. 2000. Handicap, index, and conventional signal elements of bird song. Pages 277–300 *in* Y. Espmark, T. Amundsen, and G. Rosenquist, editors. Animal signals: signalling and signal design in animal communication. . Tapir Academic Press, Trondheim.

Wang, M.-Y., V. Vasas, L. Chittka, and S.-H. Yen. 2017. Sheep in wolf's clothing: multicomponent traits enhance the success of mimicry in spider-mimicking moths. Animal Behaviour 127:219–224.

Wedekind, C., and I. Folstad. 1994. Adaptive or nonadaptive immunosuppression by sex hormones? The American Naturalist 143:936–938.

White, T. E., T. Latty, and K. D. L. Umbers. 2022. The exploitation of sexual signals by predators: a meta-analysis. Proceedings of the Royal Society B: Biological Sciences 289:20220444.

Zahavi, A. 1975. Mate selection—A selection for a handicap. Journal of Theoretical Biology 53:205–214.

Zahavi, A. 1981. Natural selection, sexual selection andthe selection of signals. Pages 133–138 *in* G. G. E. Reveal and J. L. Scudder, editors. Evolution Today. Proceedings of the Second International Congress of Systematics and Evolution. . Carnegie Mellon, Pittsburg.

Zahavi, A. 1987. The theory of signal selection and some of its implications. International Symposium of Biological Evolution. . Adtiatici Editrici.

Zahavi, A., and A. Zahavi. 1997. The handicap principle: a missing piece of Darwin's puzzle. Page 304. . Oxford University Press, New York, NY.

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