

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Forum

Why does costly signalling evolve? Challenges with testing the handicap hypothesis

Szabolcs Számádó^a, Dustin J. Penn^{b,*}^a MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group and, Department of Plant Systematics, Ecology and Theoretical Biology, L. Eötvös University, Budapest, Hungary^b Konrad Lorenz Institute of Ethology, Department of Integrative Biology and Evolution, University of Veterinary Medicine, Vienna, Austria

ARTICLE INFO

Article history:

Received 26 March 2015

Initial acceptance 29 April 2015

Final acceptance 26 May 2015

Available online 20 October 2015

MS. number: 15-00259R

Keywords:

action-response games

animal communication

handicap principle

receiver psychology

strategic costs

Zahavi's handicap hypothesis (Grafen, 1990; Zahavi, 1975; Zahavi & Zahavi, 1997) is a popular explanation for the evolution of honest and costly signalling. The general idea is that individuals honestly signal their quality because signalling is costly and therefore low-quality individuals cannot afford to produce dishonest signals. However, this hypothesis is controversial for several reasons. (1) Zahavi suggested that selection favours the evolution of honest signalling because (and not despite) of their costs, and he made the radical suggestion that when it comes to the evolution of signalling, natural selection favours waste rather than efficiency. (2) Zahavi argued that this idea is a general principle, not merely a hypothesis, which explains honest signalling in most or all contexts. (3) There are several versions of the handicap hypothesis, but attempts to provide theoretical support have largely failed. The main exception is a model proposed by Grafen (1990), which has become widely accepted among behavioural ecologists; however, his conclusions have been challenged (Bergstrom, Számádó, & Lachmann, 2002; Getty, 1998, 2006; Hurd, 1995; Lachmann, Számádó, & Bergstrom, 2001; Számádó, 1999, 2000, 2011). (4) There have been many attempts to empirically test the handicap

hypothesis, but there is no consensus regarding how it might be tested (Kotiaho, 2001).

Despite these difficulties, Polnaszek and Stephens (2014) recently conducted a study with trained blue jays, *Cyanocitta cristata*, to experimentally test the handicap hypothesis. They concluded that their findings provide the first experimental evidence that signal costs enforce honesty, and they interpreted their results to support the handicap principle. This experiment is unusually clever and insightful, and the findings provide important implications for honest signalling and receiver psychology (Guilford & Dawkins, 1991). However, we raise several caveats about the theoretical background, interpretations and conclusions of the study, and we explain why this study and other attempts to test the handicap hypothesis will be problematic as long as there is not a clear theoretical model to test.

THE JAY TRAINING EXPERIMENT

In this experiment, pairs of blue jays occupying adjacent cages were trained to play a communication game in which one bird, the sender, could choose to hop onto one of two perches, which could be used as a signal about the state of the environment, and the receiver responded by selecting a perch on the same or opposite side of the sender, depending upon the signal it perceived (Polnaszek & Stephens, 2014). The sender could choose to send an honest or dishonest signal about the environment, depending on whether one of the two red lights in the signaller's cage (visible only for the signaller) were turned on or off indicating the state for the given trial as either true or false. The birds were experimentally rewarded depending on their choices and they were tested under two conditions. In the incentives-aligned treatment, there was mutual interest between signaller and receiver, as both birds were rewarded for choosing a response that corresponded to the state of the environment. In the incentives-opposed treatment, there was a conflict of interest, as the signaller was interested in selecting the signal state regardless of the state of the environment, whereas the receiver was only rewarded if the response corresponded to the state of the environment. The authors also experimentally manipulated the cost of signalling by forcing the sender to take loops of

* Correspondence: D. J. Penn, Konrad Lorenz Institute of Ethology, Department of Integrative Biology and Evolution, University of Veterinary Medicine, Vienna, Savoyenstrasse 1a, A-1160 Vienna, Austria.

E-mail address: Dustin.Penn@vetmeduni.ac.at (D. J. Penn).

shuttle flights between a third perch and its current position before it could use the signalling or the nonsignalling perch. The authors showed that when there was no conflict (incentives-aligned treatment), the jays produced honest signals, and increasing cost of the signals had no effect on honesty. However, when they increased the conflict (incentives-opposed treatments), increasing the signalling costs affected their honesty: when the costs of signalling were low, they were often dishonest (not corresponding to the state of the environment), whereas when the costs of signalling increased, the jays produced more honest signals. The study also showed that the receivers followed or trusted the signals more often when they were reliable. The authors concluded that their study provides the first experimental evidence demonstrating that signal costs stabilize honesty, and they imply that this finding confirms the handicap principle.

ZAHAVI'S HANDICAP PRINCIPLE

Rather than supporting Zahavi's handicap principle (Zahavi & Zahavi, 1997), the findings in this study contradict this proposal. The costs of signalling stabilized honesty, but only when there was a conflict of interest between signaller and receivers. To our knowledge, this study provides the first experimental evidence that signals need not be costly to be honest under shared interests, and that signal cost has no effect on honesty under such conditions. This result is theoretically expected, but it contradicts suggestions that the handicap hypothesis is a general principle that explains honest signalling (with and without conflicts of interest; Zahavi & Zahavi, 1997). Also, Zahavi assumed that honest signals must be perceptibly costly or wasteful, since this is the only way to demonstrate honesty, and yet the birds' shuttle flights (the costs that maintained honesty) could not be seen by the receivers. There are other restricted versions of the handicap hypothesis, but as we explain next, these models were not supported either.

HANDICAPS AS STRATEGIC COSTS

The jay study was also interpreted to support a version of the handicap hypothesis proposed by Maynard Smith and Harper (1995), which views handicaps as strategic costs of signalling, and Polnaszek and Stephens (2014, p. 2) defined handicaps accordingly, i.e. 'any signal whose reliability is ensured by costs that exceed the minimal cost necessary to make the signal'. All signals have production or efficacy costs, which are necessary for a trait to transmit information or influence the behaviour of conspecifics, and the Maynard Smith and Harper (1995) version crucially predicts that they have additional strategic costs (the cost component that maintains honesty under conflict of interests). A cricket's song is costly to produce to reach females from afar (production costs), but the question is whether the males' songs are more costly than they need to be to reach female receivers. Do gazelles jump higher than they need to jump to signal their health to predators when stotting? No one has proposed how to measure such strategic costs, and the jay experiment did not attempt to distinguish strategic versus efficacy costs of signalling, which is the basis for this definition of handicaps.

THE STRATEGIC HANDICAP HYPOTHESIS

Polnaszek and Stephens (2014, p. 6) also cited Grafen's (1990) strategic handicap hypothesis as the 'authoritative mathematical statement of the handicap principle'; however, criticisms of his model (Getty, 1998, 2006) and conclusions (Hurd, 1995; Lachmann et al., 2001; Számádó, 1999, 2011) were too lightly brushed off. Grafen's (1990) main results were that (1) signals are honest, (2)

signals are costly and (3) signals are costlier for worse signallers, and yet these conditions have all been challenged by later models and empirical results (see Számádó, 2011 for a review). Signals need not be honest, not even on average, to evolve (Számádó, 2000). Honest signals need not be costly even under conflicts of interest (Bergstrom et al., 2002; Hurd, 1995; Lachmann et al., 2001; Számádó, 1999) and honest costly signals need not be costlier for poor-quality signallers (Getty, 1998, 2006).

It is also unclear how the jay experiment provides evidence or a test of Grafen's strategic handicap model. The versions of the model proposed by Grafen (1990) and Zahavi and Zahavi (1997) assume that the costs of signalling that enforce honesty are a strategic choice (where individuals can choose their level of investment) rather than an unavoidable constraint imposed on the signallers, for example high-quality signallers could use low-intensity signals but they 'choose' not to and vice versa. However, in the jay experiment costs of shuttle flights were artificially forced on the signallers: the birds could not use the signalling perch before paying the full cost of the signal. In addition, an experimental test requires showing that the marginal cost of producing the same signal is greater for low- than high-quality individuals, but this hypothesis was not tested for two reasons. First, the quality or condition of the birds was not known or examined, and quality was only mimicked by imposing two different conditions ('true' versus 'false') on the jays, which were signalled by red lights. This implementation is irrelevant to the jays' ability to bear the cost of signalling. Second, the model in the jay study is a differential benefit model (like the Sir Philip Sydney game, Maynard Smith, 1991), rather than a differential cost model (Grafen, 1990). The costs imposed on the signallers were the same in the two different conditions, and thus, by definition, there cannot be any difference in the marginal costs.

ACTION-RESPONSE GAME VERSUS HANDICAP MODEL

The authors constructed a simple model to derive the conditions of honesty for the jay experiment, and they cited Grafen's model (1990) as the 'authoritative cost condition' (Polnaszek & Stephens, 2014, p. 3) of honesty. However, the authors' model is an example of an action-response game (Hurd, 1995; Számádó, 1999) rather than a handicap model, and the conditions of honesty that can be derived from these games are different (see Appendix). The results of action-response games show that honest signals need not be costly not even under conflict of interest for high-quality signallers (Bergstrom et al., 2002; Hurd, 1995; Lachmann et al., 2001; Számádó, 1999), contrary to previous authors' claims (Grafen, 1990; Maynard Smith & Harper, 1995; Zahavi & Zahavi, 1997), assuming that signal costs vary as a function of quality. The explanation is that it is not the cost paid by 'high-quality', i.e. true condition, signallers at the equilibrium that maintains honesty, but the potential cost of cheating for 'low-quality signallers', i.e. false condition (Hurd, 1995; Számádó, 1999). This potential cost of cheating will be paid at the equilibrium for high-quality signallers only if there is a constraint linking the signal cost paid by low-quality signallers to the cost paid by high-quality signallers.

In terms of the jay experiment, if the experimenters impose a cost only on the 'false' condition, the system still remains honest and individuals under the 'true' condition (i.e., 'high-quality' individuals) do not have to pay a cost at the equilibrium. Consequently, if individuals pay a cost under the 'true' condition, then it is only because the constraint imposed by the experimenters was chosen that way (i.e. they implemented a differential benefit model). Therefore, results of the experiment cannot be used as evidence in favour of the necessity of such cost (as assumed by the handicap models), as it only reflects the choice made by the experimenters.

UNAVOIDABLE SIGNAL COSTS MAY ENFORCE HONESTY, BUT ARE SUCH INDICES HANDICAPS?

The findings in the jay experiment are more consistent with another explanation for the evolution of honest signalling called the 'index hypothesis' (Maynard Smith & Harper, 1995; Maynard Smith & Harper, 2004). This hypothesis assumes that honesty is enforced due to physical, developmental or physiological constraints that cannot be cheated, rather than additional costs that evolve on top of the (efficacy) costs required to produce a minimal signal. Because the costs of signalling were experimentally manipulated, as an unavoidable constraint, the findings are more consistent with the index hypothesis than the handicap hypothesis (Grafen, 1990; Maynard Smith & Harper, 1995). The index hypothesis is not controversial, but it is not considered to be a version of the handicap hypothesis, and classifying it as such would require redefining the handicap hypothesis.

TESTING EVOLUTIONARY HYPOTHESES WITH ARTIFICIAL LEARNING EXPERIMENTS?

Finally, we raise additional caveats about using such learning experiments for testing the handicap hypothesis, or any other ideas about the evolution of animal signals, i.e. adaptive behaviours, morphology or other phenotypic features that function to influence the behaviour of receivers (Maynard Smith & Harper, 1995; Searcy & Nowicki, 2005). The experimental design was set up to copy the structure of general action–response games, yet the elements of this game (the state of nature, the action used as a signal, the cost and the benefit) were all artificial (i.e. red light, perch hopping, flying loops and food pellets). It is unclear whether the signal in this study (perch hopping) functions as a signal in jays or other species. Moreover, it is unclear how such learning experiments can directly test hypotheses about the evolution of animal signals. Polnaszek & Stephens (2014, p. 6) acknowledged that their approach was a 'fairly drastic departure' and 'radically different, from the traditions of 'costly signalling' research'. To justify their methods, the authors pointed out that there are similarities between learning and evolutionary and new approaches are needed to test the handicap hypothesis. We agree with all of these points, but it is still unclear how this learning experiment can be extrapolated to test an evolutionary hypothesis. The unstated assumption is that if experimentally increasing the costs of signalling results in honest signals when animals are trained to produce a signal, then selection will favour the evolution of such costs as a mechanism to enforce honesty. It remains unresolved how the costs of signalling evolve, and whether any proximate rewards for honesty that might occur in nature will provide enough fitness benefits to overcome the costs of signalling. We agree that such learning studies provide a valuable tool that allows one to experimentally manipulate variables that would otherwise be difficult or impossible to test, but they are more akin to so-called 'proof of concept' studies than empirical tests of the handicap hypothesis.

CONCLUSIONS

We raised these caveats regarding the theoretical background, interpretations and conclusions of the study by Polnaszek and Stephens (2014) to emphasize the problems with the handicap hypothesis and the challenges with testing this idea. Future studies should consider the theoretical objections with the handicap hypothesis, or provide more convincing justifications for why these critiques can be ignored. The critics of the handicap hypothesis do not question the potential role of signal costs in maintaining honesty, and on the contrary, they classify their models as part of

the 'costly signalling' paradigm. No one has shown how selection can possibly favour costly signals because of their costs (contrary to Zahavi, costly signals can only evolve despite, not because, of their costs), and the jay experiment falls short of providing such evidence. Future efforts to test the handicap hypothesis defined as strategic signalling costs (Maynard Smith & Harper, 1995) should be aware that distinguishing strategic from efficacy costs may not be possible even in principle. For example, if the information being transmitted by a sender and evaluated by a receiver are the costs of the signal, as Zahavi proposed, then all of the signalling costs are strategic. We suggest that the jay study provides evidence that uncheatable constraints can enforce honesty (index signal hypothesis; Maynard Smith & Harper, 2004), but studies are needed to find an explanation for the evolution of such constraints (Biernaskie, Grafen, & Perry, 2014).

Despite our concerns, we commend the authors on their clever and innovative approach to studying animal signals. Showing that signal costs enforce honesty is an important step, and we suggest that similar experiments have great potential to provide insights into the underlying proximate mechanisms that control receiver psychology (Guilford & Dawkins, 1991). Studies are needed to determine how costly signalling evolves, and whether costly signals function to enforce honesty (i.e. do low-quality individuals pay a higher marginal cost or receive more benefits than high-quality signalers?). Finally, it would be especially helpful if future studies would identify inconsistencies, as well as the support for the handicap hypothesis.

Acknowledgments

We thank Sarah Zala for helpful comments. The research was supported by the Hungarian Scientific Research Fund, OTKA, grant No. K108974 and Austrian Science Fund (FWF P 24711–B21). Author contributions: S.S. and D.J.P. conceived the idea and wrote the article.

References

- Bergstrom, C. T., Számádó, S., & Lachmann, M. (2002). Separating equilibria in continuous signalling games. *Philosophical Transaction of the Royal Society of London B*, 357, 1595–1606. <http://dx.doi.org/10.1098/rstb.2002.1068>.
- 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, 20140876. <http://dx.doi.org/10.1098/rspb.2014.0876>.
- Getty, T. (1998). Handicap signalling: when fecundity and viability do not add up. *Animal Behaviour*, 56, 127–130. <http://dx.doi.org/10.1006/anbe.1998.0744>.
- Getty, T. (2006). Sexually selected signals are not similar to sports handicaps. *Trends in Ecology & Evolution*, 21, 83–88. <http://dx.doi.org/10.1016/j.tree.2005.10.016>.
- Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, 144, 517–546.
- Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, 42, 1–14. [http://dx.doi.org/10.1016/S0003-3472\(05\)80600-1](http://dx.doi.org/10.1016/S0003-3472(05)80600-1).
- Hurd, P. L. (1995). Communication in discrete action–response games. *Journal of Theoretical Biology*, 174, 217–222. <http://dx.doi.org/10.1006/jtbi.1995.0093>.
- Kotiaho, J. S. (2001). Costs of sexual traits: a mismatch between theoretical considerations and empirical evidence. *Biological Reviews of the Cambridge Philosophical Society*, 76, 365–376. <http://dx.doi.org/10.1017/S1464793101005711>.
- Lachmann, M., Számádó, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13189–13194. <http://dx.doi.org/10.1073/pnas.231216498>.
- Maynard Smith, J. (1991). Honest signalling: the Philip Sidney game. *Animal Behaviour*, 42, 1034–1035. [http://dx.doi.org/10.1016/S0003-3472\(05\)80161-7](http://dx.doi.org/10.1016/S0003-3472(05)80161-7).
- Maynard Smith, J., & Harper, D. G. C. (1995). Animal signals: models and terminology. *Journal of Theoretical Biology*, 177, 305–311. <http://dx.doi.org/10.1006/jtbi.1995.0248>.
- Maynard Smith, J., & Harper, D. (2004). *Animal signals*. New York, U.K.: Oxford University Press.
- Polnaszek, T. J., & Stephens, D. W. (2014). Why not lie? Costs enforce honesty in an experimental signalling game. *Proceedings of the Royal Society B: Biological Sciences*, 281. <http://dx.doi.org/10.1098/Rspb.2013.2457>, 20132457.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication. Reliability and deception in signaling systems*. Princeton, NJ: Princeton University Press.

Számadó, S. (1999). The validity of the handicap principle in discrete action-response games. *Journal of Theoretical Biology*, 198, 593–602. <http://dx.doi.org/10.1006/jtbi.1999.0935>.

Számadó, S. (2000). Cheating as a mixed strategy in a simple model of aggressive communication. *Animal Behaviour*, 59, 221–230. <http://dx.doi.org/10.1006/anbe.1999.1293>.

Számadó, S. (2011). The cost of honesty and the fallacy of the handicap principle. *Animal Behaviour*, 81, 3–10. <http://dx.doi.org/10.1016/j.anbehav.2010.08.022>.

Zahavi, A. (1975). Mate selection – A selection for a Handicap. *Journal of Theoretical Biology*, 53, 205–214. [http://dx.doi.org/10.1016/0022-5193\(75\)90111-3](http://dx.doi.org/10.1016/0022-5193(75)90111-3).

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: a missing piece of Darwin's puzzle*. New York, NY: Oxford University Press.

Appendix

Conditions of honesty in action-response games

Table A1 gives the variables of a general action-response game (Hurd, 1995; Számadó, 1999), Table A2 gives the values of these variables according to the model by Polnaszek and Stephens (2014) and Table A3 gives the conditions of honesty (Hurd, 1995; Számadó, 1999) and the values used in the model by Polnaszek and Stephens (2014) substituted into these conditions. One can see that assuming $r = 0$ we get $a - b < c$, the condition derived in Polnaszek and Stephens' (2014) article. In contrast, Table A4 shows Grafen's conditions (Grafen, 1990; Maynard Smith & Harper, 1995) and the corresponding values according to the current game. Polnaszek and Stephens (2014) provide a different set of conditions, which is not surprising as Grafen's conditions do not describe the conditions of honesty in action-response games (Hurd, 1995; Számadó, 1999).

Table A1
Parameters and notations of the action-response game (Hurd, 1995; Számadó, 1999)

Parameter	Description
W	Value of the receiver's response for the receiver
V	Value of the receiver's response for the signaller
C	Cost of signalling
r	Degree of relatedness
H	High
L	Low
U	Up
D	Down
$V_h = V(H,U) - V(H,D)$	Difference in the value of the receiver's responses for high-quality signallers
$V_l = V(L,U) - V(L,D)$	Difference in the value of the receiver's responses for low-quality signallers
$W_h = W(H,U) - W(H,D)$	Difference in the value of the receiver's responses for receivers with high-quality signallers
$W_l = W(L,U) - W(L,D)$	Difference in the value of the receiver's responses for receivers with low-quality signallers
$C_h = C(H,S) - C(H,N)$	Difference in the cost of signals for high-quality signallers
$C_l = C(L,S) - C(L,N)$	Difference in the cost of signals for low-quality signallers

Table A2
Parametrization of the action-response game according to Polnaszek and Stephens (2014)

Cost–benefit functions	Parametrization in the Polnaszek and Stephens (2014) model
$V(H,U)$	1
$V(H,D)$	0
$V(L,U)$	a
$V(L,D)$	b
$W(H,U)$	1
$W(H,D)$	0
$W(L,U)$	0
$W(L,D)$	1
$C(H,S)$	c
$C(H,N)$	0
$C(L,S)$	c
$C(L,N)$	0
$V_h = V(H,U) - V(H,D)$	1
$V_l = V(L,U) - V(L,D)$	$a - b$
$W_h = W(H,U) - W(H,D)$	1
$W_l = W(L,U) - W(L,D)$	-1
$C_l = C_h = C(H,S) - C(H,N)$	c

Table A3
Conditions of honesty according to action-response games (Hurd, 1995; Számadó, 1999)

Conditions of honesty (Hurd, 1995; Számadó, 1999)	Conditions of honesty in the Polnaszek and Stephens (2014) model
$W_l + rV_l < 0$	$-1 + r(a - b) < 0$
$W_h + rV_h > 0$	$1 + r > 0$
$V_l + rW_l < C_l$	$a - b - r < c$
$V_h + rW_h > C_h$	$1 + r > c$
$V_l + rW_l > 0$	$a - b - r > 0$
$V_h + rW_h > 0$	$1 + r > 0$
Assuming $r = 0$ according to Polnaszek and Stephens (2014) model	
$W_l < 0$	$-1 < 0$
$W_h > 0$	$1 > 0$
$V_l < C_l$	$a - b < c$
$V_h > C_h$	$1 > c$
$V_l > 0$	$a - b > 0$
$V_h > 0$	$1 > 0$

Table A4
Conditions of honesty according to Grafen (Grafen, 1990; Maynard Smith & Harper, 1995)

Grafen's conditions	Grafen's conditions using the values of the Polnaszek and Stephens (2014) model
$C > 0$	$c > 0$
$C_h / V_h < C_l / V_l$	$c / 1 < c / (a - b)$ which results in: $a - b < 1$