



CrossMark  
click for updates

## Research

**Cite this article:** Polnaszek TJ, Stephens DW.

2014 Why not lie? Costs enforce honesty  
in an experimental signalling game.

*Proc. R. Soc. B* **281**: 20132457.

<http://dx.doi.org/10.1098/rspb.2013.2457>

Received: 19 September 2013

Accepted: 17 October 2013

### Subject Areas:

behaviour

### Keywords:

communication, honesty, signal costs,  
handicap principle, game theory

### Author for correspondence:

Timothy J. Polnaszek

e-mail: [polna005@umn.edu](mailto:polna005@umn.edu)

# Why not lie? Costs enforce honesty in an experimental signalling game

Timothy J. Polnaszek and David W. Stephens

Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN 55108, USA

Communication depends on reliability. Yet, the existence of stable honest signalling presents an evolutionary puzzle. Why should animals signal honestly in the face of a conflict of interest? While students of animal signalling have offered several theoretical answers to this puzzle, the most widely studied model, commonly called the ‘handicap principle’, postulates that the costs of signals stabilize honesty. This model is the motivating force behind an enormous research enterprise that explores signal costs—whether they are physiological, immunological, neural, developmental or caloric. While there can be no question that many signals are costly, we lack definitive experimental evidence demonstrating that costs stabilize honesty. This study presents a laboratory signalling game using blue jays (*Cyanocitta cristata*) that provides, to our knowledge, the first experimental evidence showing honesty persists when costs are high and disappears when costs are low.

## 1. Introduction

### (a) Honesty and the handicap principle

Why are signallers honest even though lying often produces higher benefits? The problem of honest (and dishonest) communication is important for many disciplines. In economics, for example, understanding the conditions in which we can expect agents to exchange honest information represents a fundamental problem in the analysis of markets [1]. One could, moreover, make similar statements about nearly any field in the social and behavioural sciences including anthropology, behavioural ecology, cognitive science, linguistics, political science, psychology, sociology and others. Although other models of honest signalling exist (reviewed in [2]), Zahavi’s handicap principle [3,4] holds a central position in explaining the evolution of honest signalling. Handicap signals are generally defined as signals whose reliability ‘is maintained by signal production costs’ [5]. The handicap principle’s focus on signal costs has produced a diverse and energetic research programme that explores the mechanisms and nature of signal costs. This research programme has considered many forms of signalling (e.g. mate choice [6], predator deterrence [7] and even the practice of religion [8]), and many ways in which signals can be costly including the developmental costs of pigments [9], oxidative stress [10], reduced immunocompetence [11] and heightened risk of predation [12].

### (b) History and criticisms

Notwithstanding the enormous efforts devoted to the study of signal costs, the absence of definitive experimental evidence of the role of costs in stabilizing honesty has made this an increasingly contentious research programme (see [5,13,14] for detailed historical overviews). Critics of the handicap principle have questioned the basic approach, often emphasizing the value of alternative ways to maintain honesty (summarized in [2]). Some authors advance the claim that studies of costly signalling have simply measured the wrong costs [15–17]. Other critics even question the value of demonstrating that signals have costs [2,17], because the mere existence of signal costs is necessary but far from sufficient to substantiate their role in the maintenance of honesty [15]. Overall, critics offer a varied list of complaints about the literature on signal cost and cumulatively they underscore the urgency of the need for innovative and

rigorous new approaches to test the claims of handicap models [18–20]—models whose current influence clearly outweighs their own empirical support [17,21,22].

In the face of these critiques, direct experimental evidence on the role of costs in maintaining honesty seems critically important. Evidence demonstrating the link between cost and honesty would narrow the gap between theory and empirical research. Yet, this represents a challenging experimental problem for at least two reasons. First, it is difficult to meaningfully manipulate the costs of natural signals, such as colourful tails or stridulation. Second, it can be difficult to characterize natural signals as honest or dishonest because we are seldom certain of the underlying states that animals signal about [23,24].

This paper presents a direct experimental test of the role of costs in stabilizing honesty using a laboratory simulation of signalling that solves these problems. As a first step, we develop a simple model of ‘handicap signalling’. Here, we consider ‘handicaps’ in the general sense, where a handicap is any signal whose reliability is ensured by costs that exceed the minimal costs necessary to make the signal [13]. Our model is a variant of Maynard Smith’s Sir Philip Sidney game [25], modified to suit our experimental situation.

### (c) Model and predictions

We imagine that some aspect of the environment exists in two possible *states* that we call ‘true’ and ‘false.’ We assume that the state is true with probability  $p$ , and we restrict our attention to the case where  $0.5 \leq p < 1$ . For any given play of this game, the signaller knows the state but the receiver does not. The receiver must choose between two alternative *actions* that we call ‘accept’ and ‘reject’. The receiver does best by matching its action to the state of the environment; specifically, ‘accepting’ is best when the state is true and ‘rejecting’ is best when the state is false (table 1A). Table 1B shows how the receiver’s actions interact with the state to determine the signaller’s benefits: when the state is true, the accept action is in the best interest of both players. There can be a conflict of interest, however, if the state is false. The receiver always benefits from rejecting a false state, but the signaller’s best interests are determined by the parameters  $a$  and  $b$ . If  $b > a$ , we have a mutualism where the signaller’s and receiver’s incentives are aligned; if, however,  $b < a$  we have a case where the incentives of the two individuals are opposed (such that the signaller benefits if the receiver ‘accepts’, but rejecting is best for the receiver). Finally, we imagine that the signaller can choose between two actions that we call ‘signal’ (which indicates the true state) and ‘no signal’ (which indicates the false state). Note that while the terms used here are convenient, formally these are arbitrary labels for three general types of entities: states (here: true versus false); receiver actions (here: accept versus reject) and possible signals (here: signal versus no). Table 1C shows the costs paid by the signaller to emit these two signal types. For simplicity, we assume cost is independent of environmental state. Table 2 shows a simple game matrix derived from these assumptions. We consider two signaller strategies. An honest strategist signals when the state is true and gives no signal when the state is false. A dishonest strategist signals in both true and false states. We also consider two receiver strategies. A follow strategist accepts when a signal is observed and rejects when no signal is

**Table 1.** Pay-off matrices for our experimental game. (A) Receiver’s pay-offs: the receiver always does best by matching its behaviour to the state of the environment. The receiver obtains 1 unit of food (three pellets in the experiment) if it chooses the accept action in the true state, or chooses the reject action in the false state; otherwise it obtains nothing. (B) Signaller’s pay-offs: if the state is true, the signaller obtains 1 unit of food (three pellets) when the receiver chooses accept. If the state is false however, the variables  $a$  and  $b$  determine whether the signaller’s incentives are aligned with the receiver’s. When  $a > b$ , the signaller does best when the receiver ‘accepts’ regardless of which state applies. When  $a < b$ , therefore, the signaller and receiver incentives are opposed when the state is false. In the reverse situation, when  $b > a$ , the signaller and receiver incentives are always aligned. In our *incentives-aligned* condition  $a = 0.33$  and  $b = 1.0$ , while in our *incentives opposed*  $a = 1.0$  and  $b = 0.33$ . (C) Signalling costs: if the signaller chooses to signal, it must pay a cost ( $c$ ) determined by the current treatment (where  $c = 0, 1, 4$  or  $7$  shuttle flights). Models predict stable honest signalling when  $(a - b) < c$ . In order to make correct comparisons, we assume that the parameters  $a, b$  and  $c$  are normalized to the value of a single unit of food.

		state	
		true	false
(A) receiver’s pay-offs			
receiver action	accept	1	0
	reject	0	1
(B) signaller’s pay-offs			
receiver action	accept	1	$a$
	reject	0	$b$
(C) signalling costs			
signal emitted	signal	$c$	$c$
	no signal	0	0

**Table 2.** Game matrix based on the model we outlined. (Here, we assume  $(p > 0.5)$  and set  $A = p(1 - c)$ . The strategies honest/follow are stable under two conditions: (i) when  $b > a$ , this is when signaller and receiver incentives are aligned and (ii) when  $c > (a - b)$ , which means the cost of signalling is greater than the benefit gained from deceiving the receiver.)

		signaller	
		honest	dishonest
receiver	follow	$A + b(1 - p)$ 1	$A + (1 - p)(a - c)$ $p$
	ignore	$A + a(1 - p)$ $p$	$A + (1 - p)(a - c)$ $p$

observed. An ignore strategist ignores the signal and adopts the strategy that is best on average, which is to always accept because  $p \geq 0.5$ . To determine the stability of the honest/follow equilibrium, we compare the signaller’s pay-off in the honest/follow cell (upper left) to the signaller’s pay-off in the dishonest/follow cell (upper right), and we find that stability

requires  $c > a - b$ . The handicap principle emerges here because honest signalling can only be stable when the cost of *dishonestly* signalling true ( $c$ ) exceeds the difference  $a - b$ , which measures the conflict of interest between signaller and receiver. In this case, we can infer the effect of signal cost on honesty directly from the game matrix; we can also derive this condition by applying Grafen's authoritative cost condition ([26], pp. 519–520). Following Grafen, our model assumes that costs and benefits accumulate additively; while multiplicative costs and benefits are important for some applications of the handicap principle [16], additive combinations are appropriate for our experimental situation. In general, we expect honesty whenever the costs exceed the benefits of dishonesty and in our model these are parametrized by  $c$  and the difference  $a - b$ , respectively.

#### (d) Experimental approach

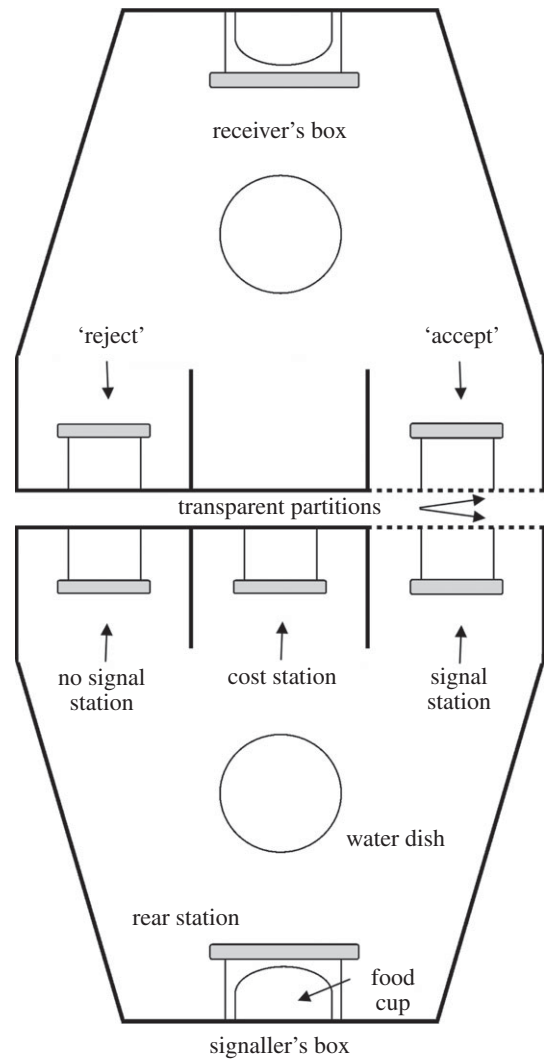
We developed a laboratory signalling game that gave us direct control over the theoretically significant variables of our model. Using this approach, we tested our model in a factorial experiment which used four levels of signal cost and two distinct pay-off structures: *incentives aligned*, in which signaller and receiver share a common pay-off structure and *incentives opposed*, in which the signaller sometimes benefits when the receiver acts against its own best interests (table 1). Our model predicts that signallers will signal honestly in the incentives-aligned condition regardless of cost, but that signallers will only signal honestly when costs are high in the incentives-opposed condition.

Our approach is unconventional in at least two respects. While our discussion will address these issues in detail, they merit some mention here because they represent a fairly drastic departure from typical studies of handicap signalling. First, we have studied the role of costs in the maintenance of honesty in a game involving food reward rather than mate selection. While the handicap principle is most widely applied to sexual signalling, the effects of cost are clearly applicable to honesty in a much broader sense as both Grafen [26] and Zahavi [4] pointed out. Second, the game theoretical equilibria studied here are established and maintained by learning rather than fixed genetics. While we cannot know the extent to which our findings apply to other situations, our approach, with its dependence on learning, gives us access to variables that are very difficult to manipulate and even measure in other situations.

## 2. Material and methods

### (a) Procedural overview

We tested our model using a procedure in which pairs of blue jays (*Cyanocitta cristata*) assumed the roles of signaller and receiver in an experimental game played for food rewards (see figure 1 for apparatus and within-trial details). The basic components of our signalling game, as in the model above, are: (i) two possible states of the 'environment' (designated 'true' and 'false'), (ii) two signal states ('signal' or 'no signal'), and (iii) two possible receiver responses ('accept' or 'reject'). To create the experimental incentive structures (*aligned* or *opposed*), we manipulated the pay-offs (in food pellets) associated with each outcome of the game. We also directly controlled the level of cost, paid in 'shuttle flights' before the signaller can access the signalling option. Finally, we note that the game allows a clear operational definition of honesty (see below), because we can observe situations in which the signaller indicates an action that is not in the receiver's best interest.



**Figure 1.** Apparatus. Blue jays occupied side-by-side Skinner boxes with one jay assigned to the role of signaller and another assigned to the role of receiver. Each box was fitted with perch stations, which could record a subject's presence or absence. The signaller and receiver used these perches to play a signalling game in a sequence of trials during the course of the day. At the beginning of each trial, a computer determined the 'state of environment' to be true or false (table 1). The state in a given trial was unknown to the receiver but revealed privately to the signaller via a system of cue lights. Signallers provided positional signals to the receiver by choosing either the 'signal' or 'no signal' perch station. In response, the receiver must choose between one of two stations at the front of its chamber. If the state was true, the receiver could only obtain food by hopping onto the 'accept' perch; and if the state was false, the receiver could only obtain food by hopping onto the 'reject' perch (table 1A). Table 1B shows how the receiver's action determined the pay-offs received by the signaller. To implement signal costs, the signaller was required to hop back and forth between the cost and rear stations  $c$  times (table 1C). The signaller was always free to choose the 'no signal' station without cost.

### (b) Subjects, housing and experimental apparatus

We randomly selected 14 adult blue jays from our colony of individuals. All subjects had previous experience in other experimental signalling games. To minimize effects from previous experience, we grouped subjects into random pairs and randomly assigned roles (as signaller or receiver). During the experiment, we kept each member of a pair in separate but adjacent Skinner boxes. A transparent partition allowed signallers and receivers to interact, such that the signaller could provide a positional signal by hopping in front of the partition (figure 1). We kept subjects in their

Skinner boxes for 23 h d<sup>-1</sup>. The 1 h provided time for daily sanitation of boxes, as well as health and weight checks. We kept subjects on a 12 L:12 D cycle and provided water ad libitum. The experiment maintained signallers and receivers on a closed economy, where each individual earned all its food from playing the signalling game, with a few exceptions. If an individual earned less than 6 g d<sup>-1</sup>, we provided food up to this minimum amount. In addition, we closely monitored the overall health of the subjects, providing appropriate veterinary care as required.

### (c) Treatments, trials and blocks

Table 1 shows the economic consequences associated with all possible combinations of signaller and receiver behaviour. Our experimental conditions were factorial combinations of the four levels of signal cost ( $c$  in table 1C; 0, 1, 4 or 7 shuttle flights required to 'signal') and two arrangements of incentives (incentives aligned and incentives opposed). These arrangements correspond to values of the variables  $a$  and  $b$  in table 1B. In our *incentives-aligned* condition  $a = 0.33$  (one food pellet) and  $b = 1.0$  (three food pellets), while in our *incentives opposed*  $a = 1.0$  and  $b = 0.33$ . Whereas we cannot quantitatively predict the value of shuttle flights in terms of food pellets, these experimental conditions allowed us to qualitatively test our model and document the effects of increasing signal costs.

We presented forced (i.e. no-choice) and free trials in blocks of 36 trials. Each block started with four forced trials, followed by 32 free trials. We scheduled eight types of forced trials, accounting for each permutation of state, signaller action and receiver response. As only four types of forced trials were sampled per block, we restricted randomization of the appearance of forced trials such that each type occurred once within a set of two blocks of trials. We completely randomized the order of each set of eight forced trials. We also randomized the current state for each free trial, but restricted such that 16 true and 16 false trials occurred per block.

### (d) Within-trial procedures

#### (i) Free trials

An intertrial interval (ITI) of 120 s preceded every free trial. After the ITI expired, a light illuminated at the rear of the signaller's box. The signaller moved to the rear perch, away from the shared partition, in order to begin a new trial. One of two red state lights illuminated, indicating the state for the given trial as either true or false. The signaller then could choose to either hop to the 'no signal' perch (on one side, in front of an opaque partition) or pay the cost of signalling. Note, when the cost of signalling was zero, the signaller could immediately choose between the 'no signal' perch and the 'signal' perch (in front of the transparent window into the receiver's box). If signalling was costly, the signaller indicated its choice to pay the cost of signalling by hopping to the cost perch. The trial was aborted if the signaller used the signalling perch before paying signal costs. Therefore, to activate the signalling perch, the signaller needed to complete  $c$  shuttle flights between the cost perch and the rear perch (where  $c = 0, 1, 4$  or  $7$ , depending on the current treatment). Each shuttle flight consisted of one circuit between the cost perch and the rear perch. After completing the shuttle flights, an orange LED indicated the signaller could access the signalling perch. Once the signaller either (i) completed paying the signal cost or (ii) chose the 'no signal' perch, a light illuminated at the rear of the receiver's box. The receiver then proceeded to the back of the box (away from the signaller and its box) if not already at its rear perch station. Once at the rear perch, orange LEDs above each response perch illuminated and the receiver was free to choose either the 'accept' perch (in front of the window) or the 'reject' perch (in front of the

opaque partition). After signaller action and receiver response, both individuals received food pellet rewards, the amount determined by the current treatment, the 'state of the environment' on given trial (true or false) and the receiver response. We manipulated the shared incentives between subjects by changing the pellet pay-off for combinations of state and receiver response as shown in table 1. The free trial was aborted if the pair did not complete it within 7 min, and the pair repeated the same type of free trial after an ITI of 120 s.

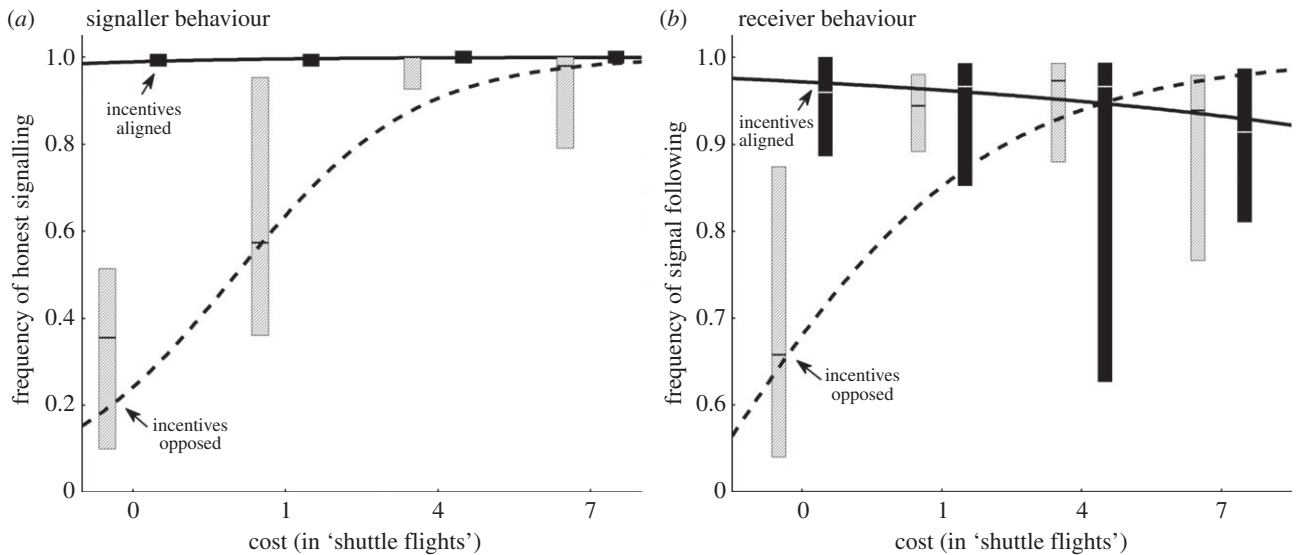
#### (ii) Forced trials

Forced trials proceeded similar to free trials, except signallers and receivers were forced into particular actions and responses. Two possible states (true or false), two signaller actions (signal or no) and two receiver responses (accept or reject) create eight possible combinations. In order to force a particular action or response, only the required perch was activated (e.g. on a forced accept trial, only hopping onto the accept perch allowed the receiver to progress through the trial). If a forced trial was not completed within 7 min, the trial was aborted and the subject pair restarted the same type of forced trial. This ensured all subjects experienced each combination of events at least once every 72 trials. Otherwise, the pay-offs and timing of events within forced trials were identical to free trials.

### (e) Termination criteria

Our experiment tested the stability of honest signals. The rationale for this is that game theoretical models predict the stability, but not necessarily the development of equilibrium states. To achieve these stability tests, we preceded each test condition (that is, each cost by incentive structure combination) with a mutualism treatment (similar to our *incentives-aligned* treatment, such that signaller and receiver both obtained food only for correct receiver responses). All birds experienced this pretreatment condition until subjects completed at least 300 trials and reached a signalling/signal-following equilibrium. We used a signalling criterion of more than 0.80 reliability of signaller action (i.e. 'signal' when state is true and 'no signal' when false). For signal following, the receiver response needed to match signaller action at a frequency of more than 0.80. We assessed these criteria at the end of each experiment day, requiring both values for either (i) the entire day or (ii) the last three blocks of trials (96 free trials total) to exceed the critical value of 0.80. After signallers were consistently honest (matching signal state to environmental state) and receivers were consistently following the signaller's action in this mutualism pretreatment, we changed the parameters of the signalling game to test the stability of communication in one of the treatment conditions specified in our factorial design. All subject pairs experienced the full set of treatments in a randomized order. Each pair of subjects experienced experimental treatment parameters for 900 free trials, and we used the final 300 in each treatment condition for our statistical analysis. In the results to follow, we first consider the observed behaviour of signallers, and then the behaviour of receivers.

While we have already defined the honest and dishonest strategists within the context of our model and experimental game, we need to recognize specific actions as either honest or dishonest to analyse our data. We adopt, therefore, the following operational definition of honest and dishonest signals. Our definition has two components. First, the receiver must have a history of responding to signal  $S$  with action  $A$ . If this condition is satisfied, we say that the signaller is acting *dishonestly* if it emits  $S$  when action  $A$  is in the best interests of the signaller but not in the best interests of the receiver. Similarly, we say that the signaller is acting *honestly* if it emits  $S$  when action  $A$  is in the best interests of the receiver. Note that this definition can be applied directly and objectively within our experiment. Our mutualism pretreatment establishes



**Figure 2.** The observed level of honest signalling (*a*) and signal following (*b*) for each of eight different treatments. Box plots summarize the results for each treatment (medians plus or minus upper and lower quartiles), and accompanying curves show the best-fitting logistic regressions. (*a*) The *y*-axis shows the relative frequency with which the signaller honestly chose the 'no signal' action when 'reject' was the receiver's best option. (*b*) Identical to (*a*) except that the *y*-axis shows the relative frequency with which the receiver followed the signaller's action (e.g. 'reject' if 'no signal'; note, the *y*-axis scale differs from (*a*)). For both panels, the *x*-axis shows the four levels of signal cost ( $c = 0, 1, 4$  or  $7$ ), measured in 'shuttle flights'. The diagonally hatched boxes and accompanying dashed curve show our *incentives-opposed* treatments. The black boxes and accompanying solid curve show our *incentives-aligned* treatments.

the history of receiver responses, and we recognize economic conflicts of interest because we have the control of the pay-off structure of the game.

### 3. Results

#### (a) Signaller behaviour

Figure 2*a* illustrates the significant interaction of cost and incentive structure on signaller honesty (repeated measures ANOVA,  $F_{3,18} = 14.022$ ,  $p = 0.00006$ ). We observe close to zero dishonesty in the *incentives-aligned* treatment. Moreover, changing the cost of signals has virtually no effect of the frequency of dishonest signals when incentives are aligned. The data tell a different story, however, when incentives are opposed. When the costs of signalling are low, signallers are commonly dishonest in the sense that they signal an action that is not in the receiver's best interest. When the costs of signalling increase, we observe a dramatic decrease in the frequency of these dishonest signals.

#### (b) Receiver behaviour

Of course, we also predict that the receiver's behaviour should be sensitive to the reliability of the signaller's signals. Receivers respond to the behaviour of signallers as we expect, that is they follow signals more closely when they are reliable (figure 2*b*; repeated measures ANOVA,  $F_{3,18} = 4.9113$ ,  $p = 0.0115$ ). Specifically, receivers dramatically decreased their signal-following behaviour in our 'no cost' treatment where signallers often signalled unreliably.

### 4. Discussion

#### (a) Significance of results

Models of handicap signalling [3,26] hypothesize that signal costs stabilize honesty. Following accepted standards of

scientific inference, experimental manipulation of costs should provide the strongest possible evidence that costs do, indeed, stabilize honest signalling. Our results provide, to our knowledge, the first experimental evidence of this type. Specifically, our experiment shows that in the presence of an incentive to cheat, subjects signalled honestly when costs were high but signalled dishonestly when costs were low; providing direct experimental support for the handicap principle's most basic causal hypothesis. Yet, our result also illustrates a simple but important limitation of the typical approach to studying handicaps, because costs did not affect signal honesty when signaller and receiver incentives were mutualistically aligned. Models of signalling have been clear about the unimportance of signal cost in mutualisms [25], but empirical studies of signal costs often proceed without first confirming that a conflict of interest exists.

In the shorthand of the signalling literature, investigators commonly emphasize the importance of differential costs or state-dependent costs in the control of honesty. A reader with this terminology in mind may feel that the cost structure shown in table 1C does not satisfy the requirements for handicap signalling, because the costs are the same in both the true and false states. It is more accurate to say, however, that handicap signalling requires 'differential marginal' costs, because this phrase captures the comparison of differences implicit in Grafen's cost condition. In our situation, this comparison of differences requires that the difference in pay-off between the honest and dishonest strategists in the bad state (the cost of the dishonest strategy in the bad state) must be greater than the same difference in the good state (the cost of the dishonesty strategy in the good state). Recall, however, that a dishonest strategist signals correctly in the good state (because to do otherwise would be against the dishonest individual's best interests). This means that the dishonest and honest strategists behave the same way in the good state, so the difference in their expected pay-offs (and hence the state-dependent cost of the dishonest strategy) is zero. The difference in marginal costs required by the

handicap principle, therefore, depends on whether the 'cost of dishonesty in the bad state' is greater than zero, which is true when  $c > a - b$  (which we can also directly observe from the game matrix, table 2).

The approach adopted here is different, even radically different, from the traditions of 'costly signalling' research. While we will discuss the possible weaknesses of our approach in the following paragraphs, we remark that our approach also has many advantages. It allows the manipulation of many theoretically important variables and it lends itself to straightforward operational definitions of honesty and dishonesty. Overall, we argue that it offers a valuable new tool in the study of signalling games.

## (b) Results in context

Models of handicap signalling have motivated the publication of many studies of signal cost. These papers are diverse and ingenious. They include, for example, efforts to characterize the energy and predation costs of begging (reviewed in [20]); and sophisticated studies emphasizing the developmental costs and signal value of carotenoid pigments [27–29]. There can be no doubt that signals have costs that take many forms (e.g. production costs, developmental costs and receiver-imposed costs) and arise via a wide range of mechanisms (increased predation risk, energy, immunological, genetic etc.). Yet, we have a growing list of reviews that express their dissatisfaction with the signal cost literature [2,14,17–20,30]. The litany of complaints includes: (i) the simple existence of signal costs tells us little about signal honesty; (ii) absolute costs of signalling are often small; and (iii) very few papers have measured the difference in marginal costs that is central to all handicap models. These critics unanimously stress the need for new approaches that rigorously make the connection between costs and honesty.

Most reviewers agree, though, that there are several examples that come close to demonstrating the hypothesized effects of cost—such as that of barn swallow tail length (discussed in [5,14]). Research shows that experimental tail elongations alter survival rates differently for individuals of different quality (inferred from original tail length) [31]—which is consistent with the differential marginal costs that handicap models require. Additional evidence demonstrates that elongated tails are costly and that male tail length is related to aspects of male quality [32–35]. Overall, this system provides a substantial body of evidence showing the ways in which this signalling system is consistent with honesty stabilized by cost. However, not all are convinced [36], and the evidence falls short of an experimental demonstration that increased costs promote honesty. Our study stands out because we directly control the level of signal cost in our experimental system and it allows clear operational definitions of what constitutes honest and dishonest signals. The approach allows us to address the concerns of some critics and measure honesty at different levels of signal cost.

## (c) Limitations and further questions

### (i) Does the handicap model apply outside of mate choice?

The handicap principle leads two lives. For some investigators, it is 'the handicap model of sexual selection' [28,37,38]. This seems natural to some because Zahavi [3] originally proposed it in this context, because within the

field of animal behaviour it is most commonly discussed in this context, and most models of handicap signalling (such as Grafen [26] and Getty [16]) develop this example. For others, it is a more general claim about the forces that stabilize honesty. For example, both Grafen [26] and Zahavi [4] took pains to point out the applicability of their ideas to problems beyond mate selection. Clearly, neither Grafen nor Zahavi felt that handicap signalling was limited to sexual advertisement. Similarly, students of handicap signalling have focused on signals that reveal information about the signaller—male quality, ability to escape or hunger [30,39,40]. Again this is more important historically than conceptually. The claim that a low-quality signaller pays a larger 'marginal cost' to signal helps make the idea of marginal cost more intuitive. As we remarked in the introduction, it is straightforward to show that the conditions required for honesty in our simple experimental game are algebraically identical to a discrete version of Grafen's cost condition, which is widely regarded as the authoritative mathematical statement of the handicap principle. The algebra of the handicap principle does not depend on the example of mate choice signalling or the idea that signals necessarily indicate signaller quality. In general, the hypothesis holds that a dishonest strategist must pay larger marginal costs, but the dishonest individual does not have to be lying about its own abilities, reproductive or otherwise. Moreover, the idea that models and tests of handicap signalling ought to be restricted to special situations seems counterproductive. Surely, behavioural ecologists want to study broadly applicable models, such as the general determinants of honesty, that have relevance to questions in the behavioural sciences that go far beyond mate choice.

### (ii) Learning?

Traditional studies of signalling often emphasize genetically determined traits—like tail length—in natural or semi-natural situations. Readers from this tradition may be uncomfortable with studies of signalling equilibria that depend on relatively arbitrary learned behaviour. While this sentiment is understandable, there are several reasons to think that this novel approach can provide useful insights. First, the game theoretical equilibria predicted by models of handicap signalling do not depend on the mechanism underlying their stability: Thorndike's law of effect or natural selection against a 'cheater gene' are equally admissible theoretically. Moreover, learning is a well-documented and well-understood biological mechanism [41–43], and this is surely why Maynard Smith devoted considerable theoretical attention to the role of learning in behavioural equilibrium ([44], ch. 5); reminding us that both 'learning and genetic change may be involved' (p. 77) in the establishment of behavioural equilibria (and see [45–47]). Second, as the quotation of Maynard Smith suggests, it is simply incorrect to think of learning and natural selection as incompatible opponents. Learning, similar to hormones and sensory receptors, is an evolved feature of animals [48]. Moreover, learning is clearly an important mechanism in animal social interactions including signalling. Finally, several recent reviews [2,14,18,30] have emphasized the difficulties faced by traditional approaches and called for fresh approaches to the study of signals and signal costs. Our studies will provide one new line of evidence that represents such a fresh approach. Our ultimate contribution will depend on how this new line of evidence interacts with and informs more traditional studies.

## (d) Final summary

Our results represent the first experimental test of the causal relationship between signal costs and honesty. While recent criticisms have characterized the efforts to connect cost and honesty as misguided or incomplete, our results suggest a position that lies somewhere between the critics and the supporters of the 'signals as costly handicaps' research programme. The experimental signalling game defines signal costs and other game parameters, and allows this research to begin addressing the concerns raised by 'signal cost' detractors. Our results support the role of costs in some situations, but they also place the onus on students of signal costs to clearly demonstrate that signaller's and receiver's incentives are indeed opposed. That is, the costliness of signals itself does not imply the role of signal cost in honesty; mutualisms or other mechanisms can also drive honesty. However, as we demonstrate, costs do enforce honesty in some cases. While studies of animal signals probe ever

deeper into the physiological and genetic mechanisms that make signals costly or condition dependent, these studies seem increasingly distant from the problem of honesty. Our approach represents an effort to reconnect these studies with the question that motivated them: how and when do costs affect signal reliability?

We conducted this research with the approval of the Institutional Animal Care and Use Committee, University of Minnesota (protocol no. 1109A04421)

**Acknowledgement.** T.J.P. developed the project and implemented the experiment with consistent advice from and in consultation with D.W.S. The two authors cowrote the manuscript.

**Data accessibility.** All data are freely available upon request, please direct requests to T.J.P. at polna005@umn.edu.

**Funding statement.** This work was supported by NSF grant no. IOS-077221, and a grant from the Dayton Fund of the Bell Museum of Natural History. T.J.P. received support from the Ecology, Evolution and Behavior Department, University of Minnesota.

## References

- Spence M. 1973 Job market signaling. *Q. J. Econ.* **87**, 355–374. (doi:10.2307/1882010)
- Számádó S. 2011 The cost of honesty and the fallacy of the handicap principle. *Anim. Behav.* **81**, 3–10. (doi:10.1016/j.anbehav.2010.08.022)
- Zahavi A. 1975 Mate selection: a selection for a handicap. *J. Theor. Biol.* **53**, 205–214. (doi:10.1016/0022-5193(75)90111-3)
- Zahavi A. 1977 The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.* **67**, 603–605. (doi:10.1016/0022-5193(77)90061-3)
- Bradbury JW, Vehrencamp SL. 2011 *Principles of animal communication*, 2nd edn. Sunderland, MA: Sinauer Associates, Inc.
- Johnstone RA. 1995 Sexual selection, honest advertisement and the handicap principle: reviewing the evidence. *Biol. Rev.* **70**, 1–65. (doi:10.1111/j.1469-185X.1995.tb01439.x)
- Vega-Redondo F, Hasson O. 1993 A game-theoretic model of predator–prey signaling. *J. Theor. Biol.* **162**, 309–319. (doi:10.1006/jtbi.1993.1089)
- Bird RB, Smith EA. 2005 Signaling theory, strategic interaction, and symbolic capital. *Curr. Anthropol.* **46**, 221–248. (doi:10.1086/427115)
- Huggins KA, Navara KJ, Mendonça MT, Hill GE. 2010 Detrimental effects of carotenoid pigments: the dark side of bright coloration. *Naturwissenschaften* **97**, 637–644. (doi:10.1007/s00114-010-0679-6)
- Noguera JC, Morales J, Pérez C, Velando A. 2010 On the oxidative cost of begging: antioxidants enhance vocalizations in gull chicks. *Behav. Ecol.* **21**, 479–484. (doi:10.1093/beheco/arq005)
- Moreno-Rueda G. 2010 An immunological cost of begging in house sparrow nestlings. *Proc. R. Soc. B* **277**, 2083–2088. (doi:10.1098/rspb.2010.0109)
- Ibáñez-Álamo J, Arco L, Soler M. 2012 Experimental evidence for a predation cost of begging using active nests and real chicks. *J. Ornithol.* **153**, 801–807. (doi:10.1007/s10336-011-0797-8)
- Maynard Smith J, Harper D. 2003 *Animal signals*. Oxford, UK: Oxford University Press.
- Searcy WA, Nowicki S. 2005 *The evolution of animal communication: reliability and deception in signaling systems*. Princeton, NJ: Princeton University Press.
- Lachmann M, Szamado S, Bergstrom CT. 2001 Cost and conflict in animal signals and human language. *Proc. Natl Acad. Sci. USA* **98**, 13 189–13 194. (doi:10.1073/pnas.231216498)
- Getty T. 2006 Sexually selected signals are not similar to sports handicaps. *Trends Ecol. Evol.* **21**, 83–88. (doi:10.1016/j.tree.2005.10.016)
- Grose J. 2011 Modelling and the fall and rise of the handicap principle. *Biol. Phil.* **26**, 677–696. (doi:10.1007/s10539-011-9275-1)
- Kotiaho JS. 2001 Costs of sexual traits: a mismatch between theoretical considerations and empirical evidence. *Biol. Rev.* **76**, 365–376. (doi:10.1017/S1464793101005711)
- Roberts ML, Buchanan KL, Evans MR. 2004 Testing the immunocompetence handicap hypothesis: a review of the evidence. *Anim. Behav.* **68**, 227–239. (doi:10.1016/j.anbehav.2004.05.001)
- Moreno-Rueda G. 2007 Is there empirical evidence for the cost of begging? *J. Ethol.* **25**, 215–222. (doi:10.1007/s10164-006-0020-1)
- Johnstone RA. 1997 The evolution of animal signals. In *Behavioural ecology: an evolutionary approach* (eds JR Krebs, NB Davies), pp. 155–178. Oxford, UK: Blackwell.
- Enquist M, Hurd PL, Ghirlanda S. 2010 Signaling. In *Evolutionary behavioral ecology* (eds DF Westneat, CW Fox), pp. 266–284. New York, NY: Oxford University Press.
- Lotem A. 1993 Secondary sexual ornaments as signals: the handicap approach and three potential problems. *Etologia* **3**, 209–218.
- Kokko H, Brooks R, Jennions MD, Morley J. 2003 The evolution of mate choice and mating biases. *Proc. R. Soc. Lond. B* **270**, 653–664. (doi:10.1098/rspb.2002.2235)
- Maynard Smith J. 1991 Honest signalling: the Philip Sidney game. *Anim. Behav.* **42**, 1034–1035. (doi:10.1016/S0003-3472(05)80161-7)
- Grafen A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546. (doi:10.1016/S0022-5193(05)80088-8)
- Hill GE. 1991 Plumage coloration is a sexually selected indicator of male quality. *Nature* **350**, 337–339. (doi:10.1038/350337a0)
- McGraw KJ, Hill GE, Parker RS. 2005 The physiological costs of being colourful: nutritional control of carotenoid utilization in the American goldfinch, *Carduelis tristis*. *Anim. Behav.* **69**, 653–660. (doi:10.1016/j.anbehav.2004.05.018)
- Fitze PS, Tschirren B, Gasparini J, Richner H. 2007 Carotenoid-based plumage colors and immune function: is there a trade-off for rare carotenoids? *Am. Nat.* **169**, S137–S144. (doi:10.1086/510094)
- Kilner R, Johnstone RA. 1997 Begging the question: are offspring solicitation behaviours signals of need? *Trends Ecol. Evol.* **12**, 11–15. (doi:10.1016/S0169-5347(96)10061-6)
- Møller AP, de Lope F. 1994 Differential costs of a secondary sexual character: an experimental test of the handicap principle. *Evolution* **48**, 1676–1683. (doi:10.2307/2410256)
- Møller AP. 1989 Viability costs of male tail ornaments in a swallow. *Nature* **339**, 132–135. (doi:10.1038/339132a0)
- Møller AP. 1994 Male ornament size as a reliable cue to enhanced offspring viability in the barn swallow. *Proc. Natl Acad. Sci. USA* **91**, 6929–6932. (doi:10.1073/pnas.91.15.6929)
- Saino N, Møller AP. 1996 Sexual ornamentation and immunocompetence in the barn swallow. *Behav. Ecol.* **7**, 227–232. (doi:10.1093/beheco/7.2.227)

35. Møller AP, de Lope F, Caballero JML. 1995 Foraging costs of a tail ornament: experimental evidence from two populations of barn swallows *Hirundo rustica* with different degrees of sexual size dimorphism. *Behav. Ecol. Sociobiol.* **37**, 289–295. (doi:10.1007/BF00174132)
36. Saunders S. 2009 Costly signalling: a work in progress. *Biol. Phil.* **24**, 405–416. (doi:10.1007/s10539-007-9100-z)
37. Matyjasiak P, Jablonski PG, Olejniczak I, Boniecki PI, Lee S-D. 1999 Foraging cost of a long tail ornament: an experiment with sand martin females. *Ethology* **105**, 521–530. (doi:10.1046/j.1439-0310.1999.00422.x)
38. Hedrick A. 2005 Environmental condition-dependent effects on a heritable, preferred male trait. *Anim. Behav.* **70**, 1121–1124. (doi:10.1016/j.anbehav.2005.02.011)
39. Rivero A, Alatalo RV, Kotiaho JS, Mappes J, Parri S. 2000 Acoustic signalling in a wolf spider: can signal characteristics predict male quality? *Anim. Behav.* **60**, 187–194. (doi:10.1006/anbe.2000.1452)
40. Caro TM. 1995 Pursuit-deterrence revisited. *Trends Ecol. Evol.* **10**, 500–503. (doi:10.1016/S0169-5347(00)89207-1)
41. Thorndike EL. 1911 *Animal intelligence: experimental studies*. New Brunswick, NJ: Transaction Publishers.
42. Domjan M. 1998 *The principles of learning and behavior*, 4th edn. Belmont, CA: Wadsworth.
43. Kandel ER. 2000 Cellular mechanisms of learning and the biological basis of individuality. In *Principles of neural science* (eds ER Kandel, JH Schwartz, TM Jessell), pp. 1247–1279. New York, NY: McGraw-Hill.
44. Maynard Smith J. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
45. Giraldeau L-A, Dubois F. 2008 Social foraging and the study of exploitative behavior. *Adv. Study Behav.* **38**, 59–104. (doi:10.1016/S0065-3454(08)00002-8)
46. Morand-Ferron J, Giraldeau L-A. 2010 Learning behaviorally stable solutions to producer–scrounger games. *Behav. Ecol.* **21**, 343–348. (doi:10.1093/beheco/arp195)
47. Fawcett TW, Hamblin S, Giraldeau L-A. 2013 Exposing the behavioral gambit: the evolution of learning and decision rules. *Behav. Ecol.* **24**, 2–11. (doi:10.1093/beheco/ars085)
48. Dunlap AS, Stephens DW. 2009 Components of change in the evolution of learning and unlearned preference. *Proc. R. Soc. B* **276**, 3201–3208. (doi:10.1098/rspb.2009.0602)