Journal of Economic Behavior & Organization 106 (2014) 281–297

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

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

Signs and (counter)signals of trustworthiness☆

Diego Gambetta[a,b], Áron Székely[c,d,*]

[a] European University Institute, Via dei Roccettini 9, San Domenico, Firenze, Italy
[b] Nuffield College, New Road, Oxford OX1 1NF, UK
[c] University of Oxford, Department of Sociology, Manor Road, Oxford OX1 3UQ, UK
[d] Institute of Cognitive Sciences and Technologies, CNR, Via San Martino della Battaglia 44, Rome, Italy

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form 17 July 2014
Accepted 20 July 2014
Available online 28 July 2014

JEL classification:
D83
C72
C91

Keywords:
Information
Costly signals
Generosity
Trust
Laboratory experiments

ABSTRACT

When we become aware that our past actions carry information about qualities that we possess or lack, which others use to decide how to deal with us, are we unconcerned, content to rely on what we have done, or do we take action to alter this information? We study this question experimentally using generosity as a sign and a signal of trustworthiness, and a trust game. Subjects play a dictator game unaware that later they will play a trust game and that their level of generosity in the dictator game will be revealed to trusters, with some inaccuracy, before trusters decide whether to trust or not. Once made aware of what follows, trustees have the option to play a second dictator game, from which their choice will be accurately conveyed to trusters in addition to their decision in the initial game. Consistent with 'countersignalling theory', those who, in the first dictator game, were either miserly or generous do not play the second dictator game, resigned or content with the information conveyed by their past actions. Those neither miserly nor generous in the first dictator game, an intermediate generous group, are likeliest to use the second dictator game; many of them for the purpose of signalling, so that they are not confused with the miserly.

1. Introduction

We investigate a common case in which people realise that some past action of theirs, which they performed for motives other than informing others of their qualities, becomes known to others in situations in which this could have consequences for them.

Going about their lives, people act in myriad ways, good and bad. Many of these actions are trivial and go unnoticed, forever lost. But some actions are perceived by other people, other actions are recorded on media whether agents are aware of it or not, others still leave enduring marks on them, perceivable well after the event that produced them occurred. The information potential contained in these past actions can remain dormant. However, out of this baggage with which everyone travels, people often extract *signs* of other people's qualities, good or bad, and use this information in deciding how to deal with them.

☆ We thank Ozan Aksoy, Dominic Burbidge, Giorgia Gon, Dan Honig, Wojtek Przepiorka, Dingeman Wiertz, Nan Zhang, and two anonymous referees for their helpful comments. This work was supported by the Economic and Social Research Council [Grant number ES/I022066/1]. Data are available upon request from the authors.

* Corresponding author at: Institute of Cognitive Sciences and Technologies, CNR, Via San Martino della Battaglia 44, Rome, Italy. Tel.: +39 06 44362370x1; fax: +39 06445 952 43.
E-mail addresses: diego.gambetta@eui.eu (D. Gambetta), aron.szekely@sociology.ox.ac.uk (Á. Székely).

http://dx.doi.org/10.1016/j.jebo.2014.07.009
0167-2681/

Our question is: what do people do when they realise that this is happening? Under what conditions do they take further action or do nothing about it? If the sign is of a quality of theirs that can help them in the situation in which they find themselves, should they still invest resources in signalling to further strengthen the receiver's belief that they truly have that quality? If an inmate enters jail preceded by the information that he has received a 20-year sentence for armed robbery, does he need to, as it were, flex his muscles publicly and vigorously to ensure that other inmates will leave him alone, or can he afford to sit back and just bank on that indirect sign of his toughness?[1] And if the sign is of a quality that could be detrimental, should people engage in some new deed to make up for it? Can an inmate rumoured to be a 'snitch' do anything to persuade other inmates that it is all a big misunderstanding, and that he really is a most loyal fellow, or would further actions make his situation even worse?

2. Theoretical background

Observable features of an agent that are displayed *intentionally* by that agent for the purpose of changing the beliefs of others about a quality of that agent are called signals. Like signals, signs carry information about people's qualities, but unlike signals signs are not produced with the intention of informing anyone. Signs, in our context, are perceivable features of an agent which *unintentionally* convey information about qualities of that agent (Gambetta, 2009, p. 170). They are by-products of our actions rather than being produced by us in order to send a particular message. On the one hand, the lack of intention to inform makes signs a source of reliable information—for instance, small gestures of generosity when it is clear that they are not carried out instrumentally are signs of genuine generosity (Gambetta and Przepiorka, 2014). On the other hand, the information that signs convey tends to be *noisy*: while raising the probability that an agent has or lacks a certain quality, they often fall short of providing full reassurance. This is either because some 'noise' in the transmission affects their reception or because the actions that created the signs are imperfectly linked to the quality under scrutiny.

Signalling theory—arguably the best instrument we have to understand intentional communication when agents' interests do not fully overlap and information is asymmetric—is silent about signs.[2] In the theory-stylised landscape, the receiver can only use the signalling action taken (or avoided) by the sender to update his beliefs about that sender. Until the sender takes an action, the receiver has to rely on prior beliefs about the base-rate distribution of types to identify the sender.[3,4] In real life, however, this simplification—especially outside of anonymous markets—is almost always violated: before receiving the signal, receivers already know something relevant about the qualities of the sender they are interacting with.

An extension of signalling theory, known as 'countersignalling theory' (CS) (Feltovich et al., 2002), can help us to deal with these richer situations. This theory discards the assumption of a single source of information and admits, realistically, three conditions:

- the existence of prior information independent of signalling but dependent on a sender's quality—sometimes called 'exogenous' information—which the receiver has about the sender and which the sender knows the receiver has;
- the noisiness of the exogenous information[5]; and
- the existence of three levels of sender quality, or types, instead of the customary two types—countersignalling theory includes not only those who have or do not have the quality of concern, known in the theory jargon as 'high types' and 'low types' respectively, but also those who fall somewhere in between, the 'medium types'.

When these features are introduced, the main finding[6] of CS theory is that in certain conditions medium types will signal more intensely than both low types and, more surprisingly, high types too.[7]

---

[1] What, exactly, committing armed robbery conveys depends to some extent on the context. In America, it may be less convincing as a sign of toughness than in Britain. Anticipating that civilians will be armed, robbers in America arm themselves too, making armed robbery the standard form of robbery. In Britain, armed robbery counts as an extreme action that only those strongly committed to the possibility of using violence take.

[2] We refer here only to the signalling literature developed in economics. The theory in biology differentiates between signals and signs—which are called cues (Diggle et al., 2007, p. 1242; Scott-Phillips, 2008).

[3] In some economic renditions of the theory, agents are conceived as strategic to the hilt, ready to always take into account the signalling value of anything they do before they do it, so nothing is ever a sign. This may be true in some extreme cases in which the peril of mutual aggression is permanently present as among Mafiosi or inmates. But this seems hardly the case for most ordinary people in most daily activities.

[4] In his seminal paper (1973) and subsequent book (1974) Michael Spence touches on this issue. In his paper, his second model pp. 368–374 (and Model 3a in his 1974 book), addresses the impact that an uninformative—evenly distributed according to quality—non-alterable characteristic (which he calls an *index*) has on signalling equilibria. Also, in Model 3b in his book (pp. 38–46), he looks at the effect that an index has when it is linked to signalling costs. However, Spence's indexes are different to what we are concerned with here for they are not correlated with quality.

[5] Specifically, four features typify the exogenous information included in Feltovich et al.'s model. First, senders know the probability distribution that their exogenous information can take, but not its specific realised outcome that receivers observe. Second, receivers observe the specific point outcome of the exogenous information that is drawn from the sender's probability distribution. Third, the probability of sending a higher level of exogenous information correlates positively with quality. Fourth, senders have no control over it: it is exogenous to sender choice.

[6] Although there are multiple equilibria as with other signalling games, the countersignalling equilibrium is robust to common out of equilibrium belief refinements, namely the Intuitive Criterion, D1, and D2.

[7] Like signalling theory, CS theory assumes that there is a negative relationship between quality and signalling cost, and that as the signalling level increases the signalling cost increases at an increasing rate. In other words, cost is decreasing in quality and convex in signalling level. The cost of signalling is assumed to satisfy the single-crossing property, thus the marginal cost to signalling is lower for higher type senders.

This is how, intuitively, the reasoning goes. Low types carry prior information that makes their type partially apparent to receivers. The best they can reasonably hope for, given the noise of their exogenous information, is to be mistaken for medium types, while they know that they are very unlikely to be mistaken for high types. But what for low types is a dream for medium types is a nightmare, so the prediction is that

- *medium types will signal vigorously (s^*) investing resources that ensure that they will distinguish themselves from the low types.*[8] Expecting this
- *low types would then find it too costly to signal with such an intensity as to be thought of as medium types, hence signalling for them is not worth it (s_0).*

The decision of whether to signal or not to signal for low types is a consequence of two contrasting pressures: low types who emit the signal s^*, would be sometimes mistaken, to their advantage, for medium types. However, taking into account the exogenous information that already allows receivers to gauge them as low types, a substantial proportion of the time, their gains from signalling s^* would not be enough to offset their costs.

Medium types, in contrast, by sending a signal s^*, distinguish themselves from those below, bringing themselves benefits. Yet, signalling has its downside for medium types: by sending s^* they also distinguish themselves from those high types who choose not to signal (s_0). The consequences of signalling are double-edged for mediums types.

At the other extreme, high types display exogenous information that gives receivers a strong hunch that they are in fact high types. The worst that could reasonably happen to them is to be mistaken for medium types, while they can be confident that they will not be mistaken for low types. For high types too, signalling vigorously may seem like the obvious course of action to pursue but, under certain conditions, it would be an unwise one: if the high types did that they would behave as, and be mistaken for, medium types. Enter countersignalling: a provocative result of the model is that

- *high types separate themselves from medium types by not signalling and pooling with low types (s_0).*

The lack of investment in signalling itself has an informative effect for receivers and is not just a saving of resources achieved by not signalling. Doing nothing, being 'laid back', turns out to be informative: high types fully separate themselves from medium types for, given what is already known about them, they can afford to signal by not signalling. Their silence confirms their type.

But why, one may wonder, do high types not raise their signal intensity to separate themselves from medium types given that they could afford it? This question, addressed by Feltovich et al. (2002) in their founding paper, can be summarised as follows. First, there are situations in which increasing the signal intensity is simply impossible and CS becomes the only separating course of action open to high types: this occurs when the signal intensity is capped upward (p. 640).[9] Next, there is a more complex reason. When high types countersignal, they send the same signal as low types—they 'pool' in their actions—but are still separated from low types by their noisy exogenous information, which allows them to be identified correctly most of the times. If the exogenous information sufficiently differentiates between high and low types, the gain in payoffs that could potentially be obtained by high types who signal is too small as they are already correctly recognised often enough. With a certain payoff distribution the cost of the minimum signal that high types would have to send to make it unaffordable for any medium type to mimic would offset the additional benefit (pp. 636–639). This means that high types prefer pooling with low types, instead of trying to out-signal medium types.

CS theory can make sense of two facts suggested by casual observation and some evidence. First, that people whose quality is open to question go out of their way to convince others that they truly possess the relevant quality—*nouveau riche* display opulent appurtenances,[10] inmates or school children with feeble reputations behave more aggressively to signal their toughness.[11] Second, the converse observation, that high types who have what it takes seem blasé about it, and do nothing to persuade others of their qualities, even playing them down—as people often like to repeat, Steve Jobs was a college drop-out.[12] One could conjecture that high types do not try hard to persuade anyone that they are skilful, clever, wealthy, tough or trustworthy simply because it is perfectly clear that this is how they are. Their past actions speak for them loud and clear. High types' abstention from signalling could thus be motivated by a desire to economise on resources rather

[8] s represents a signal in the set $S \subset \mathbb{R}_+$ and $s^* > s_0$.

[9] For example, if one can only pass or fail an exam. Medium types can make an effort and pass, and so can high types of course. But if high types passed, they would not be distinguishable from medium types. Failing for them can be the optimal (counter)signalling option for, coupled with the exogenous information, it ensures that they are not confused with low types, and that therefore, *a fortiori*, they can only be high types (remember that medium types are forced to try and pass the exam for their exogenous information is not strong enough to sufficiently distinguish them from low types).

[10] In *La Distinction*, Pierre Bourdieu writes that 'Where the petit bourgeois or nouveau riche "overdoes it", betraying his own insecurity, bourgeois discretion signals its presence by a sort of ostentatious discretion, sobriety and understatement, a refusal of everything which is "showy", "flashy" and pretentious, and which devalues itself by the very intention of distinction.' (1984, p. 249).

[11] For inmates see James (2003, p. 17); Sykes (1958, pp. 104–105); for high school children see the longitudinal study by Faris and Felmlee (2011, p. 60).

[12] Oddly, Steve Jobs' alma mater—Reed College—out of which he dropped, also behaves nonchalantly about its quality; it is an intensely academic liberal arts college that became the first U.S. university to refuse to participate in the college ranking system.

than to send information through a non-signal.[13] But countersignalling theory makes a subtler claim: its counter-intuitive results speak to those situations in which the exogenous information is not rock solid and could be thrown into doubt: then, the theory predicts, it is the very lack of signalling that fully reassures receivers.

CS theory to our knowledge has been investigated experimentally before only by Feltovich et al. (2002). Student-subjects played the role of senders and a computer played the role of employer-receivers. Senders were, at the start of each experimental interaction, randomly assigned a productivity type (high, medium or low). They were then asked to choose whether to invest some money into obtaining a good grade or not investing money and getting a bad grade; here 'grade' formed the binary signal. Obtaining a bad grade in a test was costless for all types, while the cost of a good grade correlated inversely with type. Additionally, and outside the control of senders, they each automatically sent a 'test score' which could be a pass or fail, and which functioned as their exogenous information. The computer-employer was programmed to employ all senders and pay them an amount equal to their expected productivity, conditional on both grades (signal) and scores (exogenous information). The subjects were not told about the countersignalling equilibrium, although they knew that they were being judged according to their expected productivity. Feltovich et al. allocated subjects to one of two groups in a between-subjects design. One was a control group in which the test scores (the exogenous information) of senders did not depend on type (they assigned a 0.5 probability of emitting a pass or fail to all types)—making the exogenous information useless to evaluate productivity potential. The other was the countersignalling treatment group in which the exogenous information was informative, for the probability of emitting a pass correlated positively with type: high types always emitted a pass, medium types sent a pass half of the time and a fail the other half of the time, and low types always produced a fail.

Their results proved partially satisfactory: they found that (i) high types chose to obtain good grades less often in the countersignalling treatment group than in the other group—as predicted by theory; however, they also found that (ii) medium types too chose to invest in getting good grades less often in the countersignalling treatment group, meaning that a comparison between high type and medium type signalling behaviour (a key part of the theory) in the countersignalling treatment led to a non-significant difference.[14]

3. Design and hypotheses

Instead of academic skills and productivity, for our experiment we chose generosity and trustworthiness as the qualities of concern. We exploit the fact that generosity is correlated with trustworthiness and can thus be a sign and a signal of the latter (Ashraf et al., 2006, p. 202; Gambetta and Przepiorka, 2014).

Our experiment consists of three parts.[15] First, subjects participate in a *veiled* dictator game (DG) through which we expect them to reveal the level of their generosity, and which will be used as their noisy sign—'veiled' because they are not informed of what will follow in parts two and three. Second, they are informed that they will take part in a trust game (TG) in which what they chose in the veiled DG will be revealed, albeit somewhat 'noisily' (uncertainty is added to the information) to trusters. Third, before playing the TG they are given a chance to revise what they gave in the veiled DG. Here they can choose to send a signal to trusters to strengthen or override what they did previously. Both their veiled and unveiled giving is transmitted to trusters in the TG. Those who choose to revise participate in a repeat of the DG, this time unveiled; those who choose not to revise, whether because they are content or resigned, move directly to the TG.

We employ a DG with a non-binary discrete choice-space in order to enable subjects to generate their sign, on the basis of which we then categorise their type, as high, medium or low.[16] We use instead a dichotomous TG, which is complex enough to capture the information that we need but simple enough to be readily interpreted.

With a few exceptions (Fehrler and Przepiorka, 2013; Gambetta and Przepiorka, 2014), in studies of signalling (including Feltovich et al.'s experiment), experimenters randomly assign participants their type and corresponding payoffs. This exogenous attribution of type relies on an unrealistic ability to switch type at will, on a mental plasticity which is not in everyone's gift: subjects who in the real-world belong to one type may be assigned a different type in the lab, and while some subjects may adapt to the incentive structure provided by the experimenter, in other cases their imprinted habits may

[13] This explanation rests on the assumption that the exogenous information is noisier for low and medium types than it is for high types. That is, low types may be mistaken for medium types and vice versa, but high types are unlikely to be mistaken for anything but high types. Then medium types may have an incentive to separate themselves from low types by signalling and high types would not have an incentive to signal, saving resources and accounting for the inverted U-shaped curve pattern that CS can explain.

[14] They report a p-value of 'about 0.12' for a χ^2 test and a p-value of 0.0643 for a robust rank-order test.

[15] Subjects were told at the start of our experiment that it consists of three parts. What we did not tell them at the time was what parts two and three consisted of. A number of other papers use a similar design. Schniter et al. (2013) reveal an end trust game to their subjects only after these make their initial choices; Ellingsen et al. (2010) first elicit their subjects' beliefs and only after tell them that their beliefs will be sent to others, and Binmore et al. (1985) reveal the existence of a second round only after their subjects finish a first round of bargaining (reported in Thaler, 1988, p. 199). See Ellingsen et al.'s (2010, p. 96) discussion about why such a design arguably does not violate the norm of non-deception in experiments.

[16] We could have adopted a binary sign structure as Feltovich et al. (2002) did. In this case, we would a priori decide that a given amount would be shown to receivers as 'generous' all of the times, a lower amount would be shown as generous half of the times, and a still lower amount would never be shown as generous. So the sign the receivers observe would not be the actual amount sent in the DG, but an indicator of the probability that a sender is of a given type. We decided against using this structure because we preferred to preserve a closer link between the action of the sender and the content of the sign in order to keep the experiment as simple as possible.

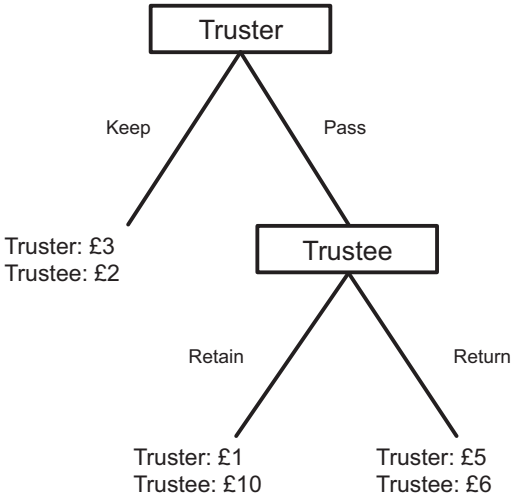

Fig. 1. Our trust game. A truster has to believe that the probability of the trustee playing return is greater than or equal to 0.5 before choosing to pass.

determine their lab behaviour. In order to avoid this confounder, we sort participants out according to their *real* generosity as manifested in the veiled DG; in other words, we allocate subjects their 'true' type.[17]

3.1. Veiled dictator game

Subjects are given an endowment of £8 and asked to decide how much of it to allocate to an anonymous passive subject, who is given no endowment, and how much of it to keep for themselves (for screenshots of all the stages see 'Experimental Instructions' in Supplementary Data). Passive subjects were randomly allocated participants whose only role was to receive money given to them by trustees in the unveiled and veiled dictator game. After the two stages, passive subjects left the experiment and did not play the Trust Game. This ensured that the decisions of trusters was only affected by the sign and signal the trustees generate.[18,19]

3.2. Unveiling

When making their initial DG choice, subjects are 'veiled', that is not informed of the further stages of the experiment. Thus, their decision can be regarded as a natural sign of their generosity. Subsequently, subjects are unveiled and given instructions for the rest of the experiment.

They are then given the option to revise their initial giving. If they choose to revise, they repeat the DG, but this time knowing what is to come later. If they choose not to revise they move directly to the final stage that is the TG. Trustees who choose to revise (*revisionists*) are randomly paid for *either* their initial or revised giving and trustees who choose not to revise (*non-revisionists*) are paid only for their initial choice.[20]

3.3. Trust game

The final stage of the experiment is a dichotomous TG. Here, each subject first plays as a trustee and then as a truster. Trusters decide whether to keep their endowment or pass it to an anonymous trustee, and trustees choose either to retain or return any money sent to them by a truster (Fig. 1). Only if trusters believe that trustees will return does it make sense for them to pass, otherwise they should keep, making it important for trusters to discriminate between trustees who will

[17] We did not tell subjects what proportion of the participants were high, medium, or low type. In this respect, in our experiment subjects do not have common priors, contravening an assumption of signalling theory.

[18] Subjects are shown eight one-pound coins on their part of the screen that they can allocate to a passive subject in £1 increments from a minimum of £0 to a maximum of £8. To give money to a passive subject, active subjects click on a coin on their part of the screen that transfers it to the passive subject's side of the screen. To reverse this, subjects click on a coin on the passive subject's screen section. They can change their minds as many times as they wish before clicking on the 'Continue' button.

[19] We used two sets of control questions to ensure that all subjects understood the experiment. One set of control questions was asked before subjects participated in the veiled dictator game. The other set was asked after the unveiling and before subjects decided whether to revise or not. After each of these, the experimenter verbally went over the correct answers to the questions and gave a brief explanation.

[20] This corresponds to reality: being naturally generous incurs a monetary cost to the individual, then not revising—continuing on with one's natural behaviour and being generous to others—also re-incurs the same monetary cost.

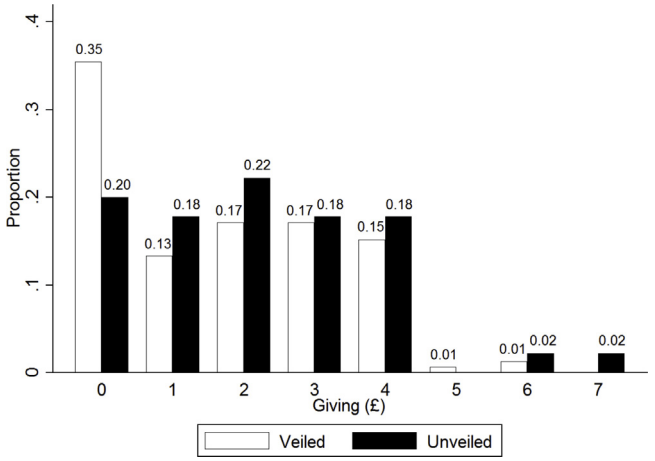

Fig. 2. The amounts given by trustees in the veiled and unveiled dictator games.

return and trustees who will retain. Trustees have a strong incentive to be trusted because this gets them an additional £4 if they are trustworthy in the TG or £8 if they are not, relative to not being trusted.

When making their choice, trusters are given noisy information about the trustees' initial giving in the form of six overturned blue cards: they are told that only four of the cards hide the true amount the trustees gave, while one card covers a value that is £1 higher than the true value and one card covers a value that is £1 lower.[21] Trusters are allowed to turn over only one card and see the value underneath. This injects some noise into the sign, which affects both sides, for trustees too do not know for sure what card trusters drew. Noisiness is a key part of the countersignalling mechanism as well as being a feature commonly present in signs.[22]

Trusters are then shown whether trustees chose to revise or not, and if they did a single red card appears displaying the exact value a revisionist trustee decided to give. If a trustee chose not to revise then the red card is absent.

We elicited trusters' choices by asking them how they would respond to different combinations of trustees' veiled and unveiled giving that were displayed to them sequentially.[23] Some of these combinations were chosen by trustees; others were never picked by trustees in the experiment and remained hypothetical combinations. Not all theoretically possible veiled–unveiled combinations were shown to trusters, but, to avoid a list that could prove tiresome for subjects, only a sub-set that seemed more interesting for us (for details on the combinations that we chose see Table 2 in Section 4). Trusters were not told which combinations were hypothetical and which ones were actually chosen by trustees. Trusters were not given feedback between their responses to the different veiled-unveiled combinations, making our implementation a partial version of the strategy method. Trustees were asked to specify whether they would, in the event of being trusted, choose to return an equitable amount or keep most of the resources. In one randomly chosen round each truster was matched with a trustee and shown that trustee's veiled–unveiled combination. Payment for the TG was decided for both based on their decisions in this round (neither was aware which round would determine their payments).

The giving in the unveiled DG is designed to offer trustees an opportunity to send an exact signal of their generosity. For this signal to be reliable, it must be more affordable for a trustworthy trustee than for an untrustworthy trustee. There are two reasons why this can be the case. First, the expected cost of giving in the unveiled condition is higher for trustees who were less generous in the veiled DG—and are likelier to be untrustworthy—because trusters are less likely to bestow trust upon them, precluding the possibility for them to obtain either of the two higher payoffs of £6 and £10. Second, for truly generous individuals (who are more likely to be trustworthy) giving money, up to a point at least, is not a cost; it is only a cost for the selfish. There is however an opposing argument, in that an untrustworthy trustee may find it more affordable to give because he anticipates a higher gain to being trusted than a trustworthy trustee: the former will keep most of the money, £10, while the latter will split it, keeping only £6.

If we vary only the probability that a trustee expects to be trusted, then a trustee who intends to return in the TG has to believe that the probability that a truster sends him money is more than twice as high as the probability that a truster

[21] For those who give £0 naturally, one of the cards covers £1 and the rest cover £0. Consequently, there is less noise around the sign of someone who gave £0 than around the sign of someone who gave more. The upper limit of the sign (£8) was never reached, so the asymmetry is irrelevant at the top.

[22] As the correlation between generosity and trustworthiness is imperfect, some noise is already objectively present. However, we cannot assume that subjects know about this or that, if they do, they know how strong the link really is; for instance, trusters may believe that the link between meanness and untrustworthiness is near-perfect, and trustees may believe that that is what trusters believe. So we introduced this additional uncertainty, which aims to make trustees more insecure of their position with respect to trusters, and trusters less confident about drawing conclusions from observing intermediate natural generosity.

[23] The order of veiled-unveiled combinations was randomised across sessions. So trusters who took part in the same session observed the veiled-unveiled combinations in the same sequence, and trusters who took part in different sessions saw the combinations in a different order.

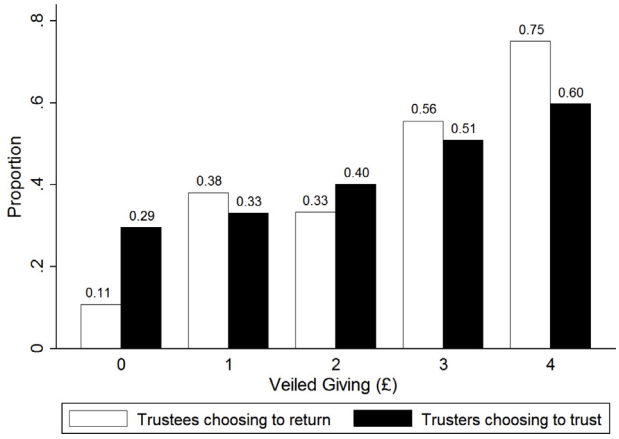

Fig. 3. The proportion of trusting trusters and trustworthy trustees by trustees' veiled giving.

sends him money if he were ungenerous, before he can expect to gain by spending more on signalling. For instance, a trustee intending to be untrustworthy and who believes that he will be trusted with a probability of 0.2 if he signals, and never otherwise, can spend up to £1.6 on signalling ($10 \times 0.2 + 2 \times 0.8 - 2$). If he intends to be trustworthy, he can spend the same amount on signalling if he believes that he will be trusted with a probability of 0.4 ($6 \times 0.4 + 2 \times 0.6 - 2$). By itself, this condition is fulfilled in a number of cases. Including unequal signalling costs into the calculation makes the condition easier to fulfil, as it decreases the difference in the probabilities necessary to allow a trustworthy trustee to signal more: e.g. if every £1 given by an untrustworthy trustee is twice as costly as for a trustworthy trustee, then a generous trustee only has to believe that he is slightly more likely than an ungenerous trustee to be sent money before he can afford more in signalling.

At the end of the experiment we also administered a short questionnaire on the motivation of subjects for making their choices.[24]

We decided against having a control group with no communication between participants. Other studies of the binary-choice trust game (e.g. Bracht and Feltovich, 2009; Ermisch et al., 2009) reveal that about 40–60% of trusters trust strangers about whom they have no information. These studies include one carried out only six months before ours in the same laboratory, drawing subjects from the same pool and offering similar payoffs (Gambetta and Przepiorka, 2014), and we will take the level they found, 46%, as our benchmark. This spared us from having to run a control group. We programmed and ran the experiment using *z*-Tree (Fischbacher, 2007).

3.4. Hypotheses

We expect trustees' giving in the veiled DG to be positively correlated with their trustworthiness in the TG (H1a); and we expect trusters to realise that this is the case, and thus more of them to trust those who gave more in the veiled DG (H1b).

The other hypotheses are inspired by the countersignalling model described in Section 2.[25] We expect trustees who gave intermediate amounts in the veiled DG to be the most uneasy about their generosity not being recognised by trusters, so they should be more likely to revise to make sure that trusters see their generosity precisely. By contrast, naturally low givers and naturally high givers will be less likely to revise: the former resign themselves to not being trusted and the latter rely on their generosity showing (H2).

Trustees who give an intermediate amount in the veiled DG and choose to revise should do so to persuade trusters to trust them—by revising and either increasing their giving or re-giving the same amount they gave naturally (H3).

More trusters will send money to naturally low generosity trustees who revise and increase their giving than to naturally low generosity trustees who do not revise (by signalling they look like medium trustees some of the time). Trusters respond to signals from low generosity trustees (H4a). However, low types, 'marked' by their sign, will have to invest more in signalling than medium types to achieve the same level of trust (H4b).

More trusters will choose to trust medium generosity trustees who revise and increase their giving than medium types who do not revise (H5).

Trusters will *not* treat high type trustees who revise and re-give the same amount differently to high type trustees who do not revise; revising and re-giving the same amount will be irrelevant. If the sign generated naturally by high types suffices

[24] The questionnaires were carried out at the end of the experiment while subjects' payments were being prepared—by which point they may have been tired and losing concentration—and they were not incentivised monetarily to complete them. Both of these points imply that the answers subjects wrote down regarding their decisions will be a 'lower bound' estimate of their reasoning and decision process.

[25] This holds assuming that our subjects' behaviours conform to the model's countersignalling equilibrium predictions. Similar to Powell and Wilson's experiment on Hobbesian jungles (2008) we do not solve our experiment for equilibria.

for them to be separated often enough from those below them, then supplementing it by re-giving the same amount in the form of a signal—which is a more precise channel of information—will not improve their probability of being trusted as they are already recognised as high types. Eliminating noise that we injected in our experiment from the information will not gain high types benefits (H6a). If the sign of high generosity trustees is sufficiently separating, then trusters will also treat high type trustees who revise and *increase* their giving the same as high types trustees who do not revise because they are already recognised as generous, making signalling wasteful for them (H6b).

4. Results

We collected a total of 158 independent decisions by trustees in the veiled DG.[26] The modal giving was £0, although the majority of trustees (63%) gave between £1 and £4 (Fig. 2). Our distribution is broadly similar in both mean (21.3%) and spread to that typically found in experimental dictator games (Belot et al., 2010; Engel, 2011, p. 589).[27,28,29,30]

Comparing the distribution of unveiled giving by revisionists, we find that, relative to veiled decisions, a smaller proportion of revisionists choose to give £0 (20% vs. 35%) and larger proportions choose to give between £1 and £4 (76% vs. 63%). Here too, however, few trustees give £4 (18%).

H1a. We expect trustees' giving in the veiled DG to be positively correlated with their trustworthiness in the TG.

As expected, trustees who are naturally more generous are also more trustworthy ($r = 0.47$)[31] (Fig. 3). For every one-pound increase in the amount given naturally (relative to £0), the odds of being trustworthy more than double ($\exp(b) = 2.22, z = 5.22, p < 0.001$).[32]

H1b. We expect trusters to realise that H1a is the case, and thus more of them to trust those who gave more in the veiled DG.

Trusters recognise that natural giving is a strong sign of trustworthiness and act accordingly, more of them bestowing trust on naturally generous trustees: the odds of being trusted increases by 40% for every pound increase trustees naturally give (Fig. 3) ($\exp(b) = 1.40, z = 9.04, p < 0.001$).[33]

H2. We expect trustees who gave intermediate amounts in the veiled DG to be more likely to revise.

Although only 28% of trustees chose to revise their giving (43 out of 155), those among them who gave an intermediate amount in the veiled DG were the most likely to revise as our hypothesis predicts (Fig. 4). Among trustees initially giving £1, £2 or £3, respectively 38%, 41% and 37%, chose to revise. In contrast only 18% and 17% of trustees who initially gave a low (£0) or a high (£4) amount respectively revised. In a logistic regression containing both a linear and a quadratic term for veiled generosity, we find a positive and significant linear coefficient ($p = 0.002$) and a significant and negative quadratic coefficient ($p = 0.002$). This implies an inverted U-shaped relationship between the amount given when veiled and the probability of revising.

These results reinforce a way of conceptualising subjects' types which rests on clear cut behavioural foundations, and which accurately capture our three types: those perfectly self-interested who give nothing (*low types*); those somewhat other-regarding who give something but less than half of their endowment (*medium types*); and the altruists who give half (*high types*). If we re-analyse the data using these three groups we find, on the basis of a logistic regression, that medium

[26] Only three subjects gave more than 50% of their endowment in the veiled DG—one gave £5 and two gave £6. We remove them from our analysis for a combination of reasons. While those who give up to 50% (£4 in our case) are common in dictator games, those who give more are outliers (Engel, 2011, p. 589). Some authors consider this a sufficient reason to remove subjects from their analysis of DGs as it suggests that they failed to understand the experiment (Pradel et al., 2009; Schotter et al., 1996). Our three outliers display additional indicators in support of this inference. One subject who gave £6 in the veiled DG and revised upwards went on to exhibit incoherent behaviour in the TG, and the other two outliers wrote incorrect or contradictory answers in their ex-post questionnaire. We could have alternatively combined these outliers into a group with trustees who gave £4. Our results are only slightly affected if we do this.

[27] Our experiment was carried out at the Centre for Experimental Social Science, Nuffield College, Oxford.

[28] We collected data on 158 active subjects and had 35 passive subjects from a total of 7 experimental sessions run over three days from 31/10/12 to 2/10/12. Subjects were graduate and undergraduate students who had not taken part in Gambetta and Przepiorka's (2014) experiment.

[29] Our trustees had a mean giving of 21.3%, which was slightly lower than that reported by Engel, at 28.4%, in a recent meta-review of 129 papers using dictator games published between 1992 and 2010. We also find more people give £1–3 than that reported by Engel, and no spike in the distribution at 50%.

[30] Belot et al. look at the differences between student and non-student subjects at CESS and finds that students are less other regarding than non-students. As we used students for our experiment, it makes sense that we find a lower mean giving.

[31] This result includes both revisionist and non-revisionist natural generosity, removing the former group strengthens the association.

[32] All logistic regressions include control variables for age, sex, and experimental session. The full regression outputs can be found in Appendix A.

[33] The logistic regressions used to test Hypotheses 1b, 4a, 5, 6a and 6b have S.E.s clustered according to subject to account for the multiple observations per subject.

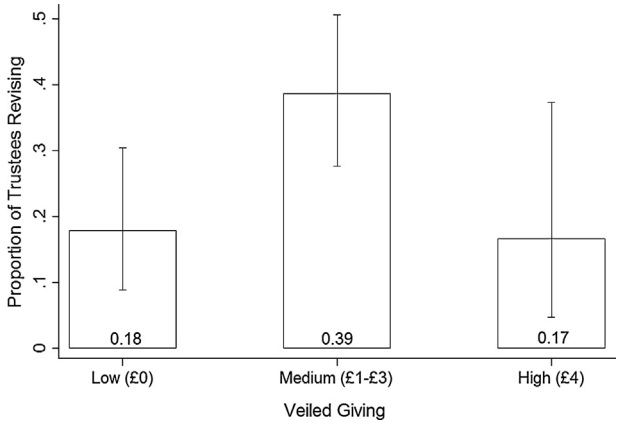

Fig. 4. The proportion of subjects revising by their veiled giving. Error bars indicate 95% confidence intervals.

types revise more than low types ($p = 0.007$) and more than high types ($p = 0.037$). Thus we find broad support for Hypothesis 2.[34,35]

H3. Trustees who gave an intermediate amount in the veiled DG and choose to revise should do so to persuade trusters to trust them.

A slight majority, 52% (15/29) of intermediate type revisionists revise and re-give the same amount again or revise and increase their giving. A decision to revise and re-give the same amount as naturally given we consider as a signal because it removes noise around the giving of trustees; it is a costless signal that purifies the naturally given sign.[36] We also interpret the action of revising and increasing as a signal, this time a costly one. The other 48% of intermediate type revisionists decreased the amount they gave upon revision implying an unexpected—unexpected for they became *less* other-regarding after unveiling—short-term selfish motivation to revise.

When we compare the actions of intermediate types to the actions of low and high types, we find that 100% of low types (10/10) who revised and 25% (1/4) of the high types who revised did so with a signalling motive. From this angle, relative to low types, medium types do not seem to be more likely to revise for a signalling purpose—although they are still more likely to revise than high types. However this interpretation of our results does not consider the different choice that low and medium types face: the former cannot decrease their investment, while the latter can. This makes a straight comparison between revisionists' decisions problematic.

A better way to view our results is to take into account both the proportion of types who choose to revise and the proportion within each type who revise in order to signal. When we do this, we find that 20% (15/75) of the intermediate types revise and increase their giving or re-give the same amount again relative to what they gave in the veiled DG. Of the low generosity trustees, 18% (10/56) revise and increase or re-give the same amount again; and of the high generosity trustees just one individual revises and gives the same amount (1/24, 4%) while none of them increase. Using one-sided Fisher's test for equality of proportions, we find that medium types do not significantly differ from low types ($p = 0.47$), but significantly differ from high types at the 10% level ($p = 0.056$). Intermediate generosity individuals appear to signal at similar levels to low types, but more than high types (Fig. 5 and Table 1).

We analyse 2450 decisions by trusters, which gives us an average of 77 observations for each veiled-unveiled combination that we are interested in (Table 2).[37] We obtained more results for trusters than trustees because when eliciting truster

[34] Schniter et al. (2013) report experimental results that mesh well with ours. Their subjects took part in a trust game. Before trusters sent the trustees any money, the trustees had the opportunity to make a non-binding promise to trusters about how much they will return if they are trusted. Once trusters made their investment decision, and the trustees decided how much to return (and whether to honour their promise), it was revealed to participants that they will take part in a second trust game still matched with the same partner. However, before this new trust game began, trustees had another chance to send a non-binding promise. Schniter et al. find that trustees who broke their initial promise subsequently increase how much they promise to send in the second trust game by a greater amount than those who had initially kept their promise (p. 249) and promise-breakers chose to send longer messages than promise-keepers (p. 250); both actions are interpreted as an effort to assuage the concerns of trusters who could see trustees' previous untrustworthy behaviour.

[35] As the trusting behaviour of trusters may act as a proxy for their beliefs about how they will be treated as trustees, in one version of this regression we included a proxy for this: mean trusting behaviour. This had no effect on our findings. In addition, we find a strong link between the trustworthiness of trustees and their trusting behaviour as trusters ($r = 0.31$, $p = 0.001$).

[36] The questionnaire responses of the revisionists who re-gave the same amount support this interpretation e.g. 'It meant that people could actually see what I had choosen [sic]', 'More opportunity for person to see how much I sent'.

[37] We do not analyse all trusters' decisions that we collected. Some trustees chose combinations which were not included in the partial strategy method set of veiled-unveiled combinations which we had planned to elicit, and we do not use trusters' decisions in response to those (even though we had to show them to trusters to work out their payments). We also cap our analysis of trusters' decisions in response to combinations including only a maximum

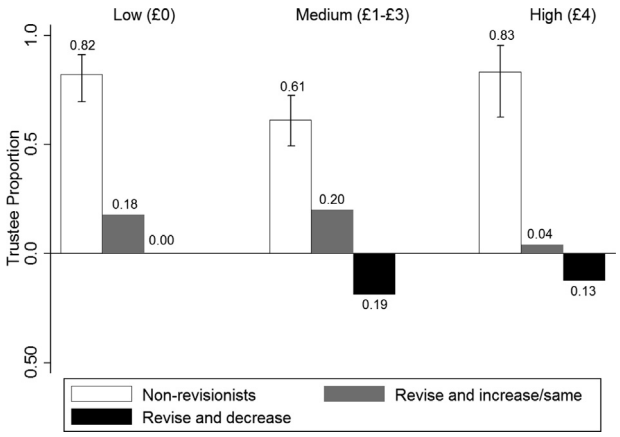

Fig. 5. Proportion of revisionist and non-revisionist trustees by their natural generosity. Error bars indicate 95% confidence intervals.

Table 1
Subjects' actions according to their type. The number of subjects is enclosed in parentheses, and row proportions are shown above.

Type	Action				Total
	Not revise	Increase	Same	Decrease	
Low	0.82	0.18	0.00	0.00	1.00
	(46)	(10)	(0)	(0)	(56)
Medium	0.61	0.16	0.04	0.19	1.00
	(46)	(12)	(3)	(14)	(75)
High	0.83	0.00	0.04	0.13	1.00
	(20)	(0)	(1)	(3)	(24)
Total	0.72	0.14	0.03	0.11	1.00
	(112)	(22)	(4)	(17)	(155)

Table 2
The veiled and unveiled giving combinations that we showed to trusters and subsequently analyse. A dash '–' indicates that we did not collect data for this combination.

Unveiled giving (£)	Veiled giving (£)					Total
	0	1	2	3	4	
Not revise	140	86	75	99	74	474
0	62	85	83	66	48	344
1	90	80	83	–	–	253
2	–	–	–	92	64	156
3	72	82	77	–	50	281
4	–	–	–	68	57	125
5	73	73	99	66	74	385
6	–	–	–	–	58	58
7	72	76	67	95	64	374
Total	509	482	484	486	489	2450

behaviour we used a partial variant of the strategy method, in which 19 different veiled–unveiled giving combinations were asked from each of our truster subjects.

When we report or discuss truster behaviour (the rest of this section and section 5.2) and we say that someone gave £X in the veiled DG, we mean that this is the sum the truster sees on the card, regardless of whether someone gave £X, or the noise we introduced makes it seem with 1/3 probability as if they gave £X, while in reality they gave £X ± 1.

H4a. More trusters will send money to naturally low generosity trustees who revise and increase their giving than to naturally low generosity trustees who do not revise.

We find evidence in support of this hypothesis. Using a logistic regression, we find that for every £1 of extra giving upon revision relative to a naturally miserly non-revisionist, the odds of being trusted for those who initially gave £0 increases

giving of £4. This is because a giving of £5 or more is so rare and would have complicated the analysis with no gain. Merging the responses that trusters give in response to veiled giving of £4 and £5 has no meaningful impact on our results.

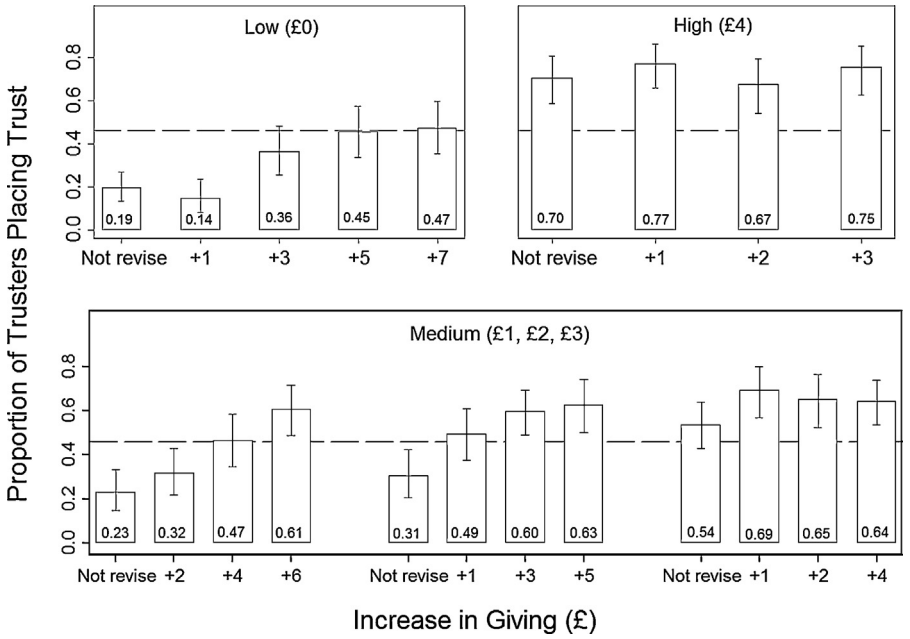

Fig. 6. The proportion of trusters placing trust according to veiled and unveiled giving combinations presented to them using a partial strategy method. The relative increase in generosity refers to the amount extra given by a revisionist in the *unveiled* dictator game relative to the amount that revisionist naturally gave in the *veiled* dictator game. The dashed reference line corresponds to 46%, the proportion trusting those about whom they have no information in Gambetta and Przepiorka (2014). Error bars indicate 95% confidence intervals.

by 27% ($p < 0.001$). In other words, trusters pay attention to a trustee's strategic generosity even when the trustees initially revealed themselves to be miserly (Fig. 6).

Statements made by low type revisionists (all of whom revised upwards) suggest that they recognise the effect that increasing generosity can have: 8 out of 10 wrote that they revised upwards to give the impression of trustworthiness to trusters (although only 3 out of the 10 were then actually trustworthy!).

H4b. Low types will have to invest more in signalling than medium types to achieve the same level of trust.

As we expected, low types have to invest more in signalling than intermediate types to achieve the same level of trust. For low types, no amount of giving makes those who gave £0 at first more trusted than a stranger—that is someone about whom there is no information (as we explained above, our measure for trust in strangers is taken from Gambetta and Przepiorka (2014)). In contrast, all medium types can surpass the trust a stranger receives (Fig. 6).

H5. More trusters will choose to trust medium generosity trustees who revise and increase their giving than medium types who do not revise.

For trustees who naturally gave £1, every one pound increase upon revising corresponded to a 31% increase in the odds of being trusted ($p < 0.001$) and for those who gave £2 naturally every £1 increase corresponded to a 29% in the odds of being trusted ($p < 0.001$); all relative to a non-revisionist. There was a non-significant gain of revised upward giving for those who gave £3 in the veiled DG ($p = 0.161$). Thus, we find support for this hypothesis for trustees whose veiled giving was £1 and £2, but not for those whose giving was £3. Most medium types could increase their probability of being trusted by revising (Fig. 6).

Almost half of the medium types who revised upwards explicitly stated that increasing their giving upon revision could make them more trusted (5/12, 42%).

H6a. Trusters will not treat high type trustees who revise and re-give the same amount differently to high type trustee who do not revise.

We find no difference in the probability of being trusted between naturally generous trustees who do not revise and naturally generous trustees who revise and give again the same amount they gave in the veiled DG ($p = 0.304$).

H6b. Trusters will also treat high type trustees who revise and increase their giving the same as high type trustees who do not revise. Increasing giving once unveiled has no effect on the probability of being trusted for naturally generous trustees ($p = 0.868$). Both results support H6a and H6b, showing that the initially generous have no incentive to revise (Fig. 6).

When asked why they chose not to revise when given the possibility, many stated that the reason was because they were 'happy' or 'confident' with their natural decision, and some added that their natural choice should be enough to make them trusted. Two trustees gave reasons of principle '. . .a choice is a choice is a choice. . .', while two others made a similar point but did so by denying that the situation had changed for them: 'I could not see any reason for doing so [revising]—nothing had changed since the first decision [veiled giving] was made (in my mind anyway).' This would suggest that generous trustees expected their decision not to cause them harm, and trusters' responses proved them to be correct.

5. Discussion

Generosity turns out to be a formidable indicator of trustworthiness as trustees who naturally give more are substantially more trustworthy: 75% of trustees who gave a high amount returned the money in the TG, while 89% of trustees who gave £0 did not return the money. Trusters recognise that (lack of) natural generosity is a reliable sign of (un)trustworthiness and respond appropriately by rewarding trustees who were more generous and shunning those who were not generous: those who gave nothing naturally were only trusted by 29% of trusters, while those who gave £4 were trusted by 60% of them (Fig. 3).[38,39]

As argued by Gambetta and Przepiorka (2014), whose results match ours, this finding could have far-reaching implications. Given the strong link between generosity and trustworthiness (and meanness and untrustworthiness), acts that reveal one's generosity should be an important source of information about people's pro-social dispositions, and a source with marked advantages: they are easy to observe, easy to produce in everyday life; and, by benefitting others, acts of sharing are, all else equal, more efficient as signals of trustworthiness than wasteful displays in which people 'burn' resources. Spontaneous gestures of sharing (or not) valuable resources could have been functioning as the first signs of (un)reliability in cooperation in the course of human evolution, and could have 'spread through positive assortment'. The information value of generosity might also explain the ubiquity, across a wide variety of cultures, of ritualised gift-exchange and sharing ceremonies.

5.1. Trustees' decisions

On discovering that they will play a TG and that the DG choice they made under the veil of ignorance will be revealed to trusters, despite the noisiness we injected, most trustees (72%) felt no need to revise the split they chose naturally. Many of them seem to believe that trusters would not be convinced by a belated signal of generosity; as one participant said, 'to increase the amount given would not be a guarantee that this would prompt the receiver in Part 3 [the TG] to give more back. A revision might make it obvious that the sender is only giving in order to receive—a cynical attitude which might not inspire generosity from the recipient.'

We found however marked differences in the proportion of trustees who choose to revise. As we expected, fewer of those at the extremes revised, appearing either content or resigned. Those who gave £4 seem confident that their natural action should suffice to persuade trusters to trust them, confirming our expectation that naturally generous trustees have no incentive to increase giving.[40] Those who gave nothing seem to believe either that it would be pointless to be more generous, for savvy trusters would doubt their motives, or expect, rightly, that the cost necessary to convince a truster is simply too high to be worth it.

As expected, it is medium types who revise the most. Once we take into account both the proportion of each type who revise *and* the proportion who take a communicative action, we find that medium types, while signalling in similar proportions to low types, signal more than high types who hardly bother to revise at all. The first choice of medium types does not conclusively show them to be generous and the possibility of looking even worse, given the uncertainty we introduce in the information, could prevent trusters from trusting them.

We had two unexpected results. First, just under a fifth of low types (18%) were not at all resigned and chose to revise. They all increased their giving of course,[41] 8 out of 10 mentioning that they did so to induce trusters to trust them in the TG (more on them below). Second, a considerable proportion of medium type revisionists (48%) *decreased* their investment—as Feltovich et al. find, 'too few' individuals of our medium types signal. It is as if discovering that they are in a strategic situation destroys their initial modest generosity, encouraging them to go for immediate gain, may be because they are pessimistic about the effect that a belated increase would have on trusters, and may just as well make the most of it while they can.

[38] The value of 60% refers to the proportion of the time that trusters choose to pass in encounters with trustees who display a veiled giving of £4 and any level of unveiled giving. Trusters pass more of the time to trustees who have a veiled giving of £4 but who did not revise, or who revised and increased (as can be seen in Fig. 6).

[39] It is worth noting that, as can be seen from Fig. 6, a staunch 30% of trusters never trusted, irrespective of trustees' veiled and unveiled giving, and that trusters who themselves were untrustworthy when they played as trustees were less likely to place trust—i.e. they were more sceptical—than trusters who had been trustworthy as trustees.

[40] Very few naturally generous trustees chose to revise (4/24), and almost all who did so chose to decrease their giving in the unveiled DG (3/4). Only a lone subject chose an action that suggests a motive of communication, by re-giving the same amount he or she had naturally given—a costless signal that removes the noise.

[41] Given the design the low type revisionists could not decrease. But none of them re-gave what they naturally chose either: removing noise around their miserliness would not inspire more trust, and besides, it would obviate the possibility that they are confused with medium types, which can happen when they rely simply on their signs.

Several of them say that this was the reason why they revised downward—one remarking: 'I was not sure A person [the truster] would give me any money even if I increase (as the stakes for A are very different) while this [giving in the revised DG] was a sure amount to keep.'[42]

Whether revising in order to lower their giving and go for short-term gain, or to restate or increase their giving to influence trusters' beliefs, medium types appear to be the group most 'uneasy' about their experimental past. On the one hand, they appear as relatively more fretful to have their qualities recognised by trusters—which may correspond to the lack of calm that, casual observation suggests, is displayed in real-life by 'intermediate' agents. On the other hand they seem a little conflicted normatively—first they show some generosity, though not going for the 'fair' even split, but then seem to repent and revert to a more greedy behaviour.

5.2. Trusters' decisions

As we move to discuss trusters' behaviour, there are two points to bear in mind:

- As mentioned before, the reference point which we use to evaluate the level of trust bestowed on trustees about whom trusters have no information is that of Gambetta and Przepiorka (2014), in which 46% of trusters trusted trustees about whom they knew nothing.
- We used the strategy method hence trusters' responses are hypothetical and the combinations to which they responded were not always chosen in reality by trustees. Using this method we collected at least 48 observations for each of our veiled–unveiled combinations.

Trusters overall show a remarkable appreciation of the information they have on trustees.

They take a dim view of trustees who give 0 in the veiled DG and do not revise, trusting only 19% of them. Trusters respond somewhat positively but very prudently to those 18% in the low type group who revise and increase their giving. While giving £1 is a waste as it gains trustees no more trust (14%), giving £3 or £5 in the second DG makes them a little more trusted at 36% and 45% respectively. Trusters therefore seem to interpret unveiled DG giving as a signal of relative generosity even when the natural DG giving was zero. Yet, the additional expenditure does not push trust up by much and seems hardly worth it.[43] Trust in them is 'capped' as at most they can expect to be trusted in 47% of the interactions with trusters, and this only if they pour 7 additional pounds into their signal; this level of trust is nearly identical to the reference level of 46% bestowed upon strangers, the most redemption low types can hope for as their past actions brands them. Miserly trustees who revised and increased their giving are, it seems, overly optimistic about their chances of being trusted and do not appreciate the 'stigma' that the sign they naturally emitted placed on them. It may be a trait of the self-interested to underestimate the negative impact their behaviour has on others, and to think that their image can be quickly brightened up again.

Trusters respond more trustingly to the signals of those who gave an intermediate amount in the veiled DG and chose to revise upward. For £1 and £2 natural givers each additional pound makes a substantial difference in their probability of being trusted; this is most apparent for the £2 givers, as, relative to not revising, a *single* pound increase pushes the probability that they will be trusted up from 31% to 49%—for them a £1 increase gains more in probability of being trusted than a £3 increase gains for those who naturally gave £0. The gain does not materialise, however, for those who gave £3 first and then revised upward later. One interpretation is that this group see themselves as falling in the intermediate group, and thus in need of sending a stronger message of generosity and hence trustworthiness. Yet, trusters already seem to perceive £3 givers as being on the borderline of the generous side—not as starkly as the £4 givers (who are trusted 60% of the times), but enough to be deserving of a modicum of trust (51%).

Although overall medium type trustees seemed to have gained significant additional trust by revising and increasing, most individuals in this group chose not to revise (61%). This is puzzling for not revising for these trustees would seem a dangerous strategy: if they rely on their sign as the sole source of information, 1/6th of the times they are perceived, due to the noise we introduced, as having given only £2 naturally. This corresponds to a large decrease in the probability of a truster bestowing trust: only 31% of trusters bestow trust on a trustee who gave £2 naturally and did not revise, while 54% bestow trust in a trustee who gave £3 naturally and did not revise. One could imagine that those at the top of the medium type group who naturally gave £3 anticipated (correctly as it turns out) the futility for them of revising and increasing. This does not seem to be the case however as the revision rate of this sub-group is very close to the rate of the rest of medium types.

[42] Others made no mention of the impact their action would have on trusters' decisions, suggesting some failed to realise that by sending a more generous signal they could induce trusters to trust them and increase their gain in the TG.

[43] Establishing precisely whether a choice paid off is difficult with our experiment because we did not formally model subjects' costs and benefits. Still we can propose some tentative conclusions about the rationality of trustees' behaviour using the changes in the probability of being trusted they receive according to deviations away from their natural level of generosity. While we cannot say which strategy was best for all trustees in an absolute sense using this logic, what we can say is that a certain strategy made more sense for one group of trustees than it did for others.

But there is a subtler reason for not revising: consider that 1/6th of the time these trustees who gave £3 naturally would be mistaken for trustees who gave £4. By mimicking the strategy of the truly generous and do nothing, they can be treated as one. At least one trustee, who gave £3 naturally, followed this strategic reasoning in an attempt to persuade trusters that he was a high type:

> In part 1 [veiled dictator game], I chose to give the Recipient £3. I assumed that if I didn't choose to revise in part 2 [unveiled dictator game], then the A person [truster] would never know how much exactly did I give to the Recipient. Even if he clicked on a card with £3 underneath, he would probably assume that he had chosen to reveal a card that was showing £1 less than the real value; the fact that I hadn't chosen to revise would have made an overall good impression, having person A to think that I have probably had made the fairest possible choice, handing down £4. Thus it would make an impression of giving out £1 more than I actually did, and would be able to save myself £1 what I might have lost, had I chosen to revise my decision in part 2.

Trusters reward all trustees who display a natural giving of £4 in the same way, regardless of whether they signal or not. Most of the naturally generous trustees chose not to revise (83%), as our sophisticated participant above expected, and they appear to have got it right. Using an additional precise source of information did not produce benefits for the naturally generous. Trustees here seem to clash against the ceiling of unconditionally distrustful trusters, which leaves little room for improvement. Decreasing though is certainly not appreciated by trusters: a £2 reduction results in trust sinking to a measly 23%, well below the level achieved with no communication.

To summarise, we find that:

(a) generosity functions as a sign of trustworthiness and trusters recognise this;
(b) trustees who naturally give an intermediate amount revise the most frequently;
(c) a majority of intermediate type revisionists increase their giving, and more intermediate types signal than high types;
(d) naturally miserly trustees can increase their probability of being trusted by using a costly signal but only to the level in which strangers are trusted;
(e) intermediately generous trustees can make substantial gains by signalling and increasing their giving; and
(f) naturally generous trustees cannot gain by utilising a signal of the same intensity as their sign, or even from increasing their signal.

How do our findings fit with countersignalling theory? We found that low type trustees decided to revise more than predicted, and medium types revised less than predicted. Still, more medium types revised than both low and high types, however medium types signalled at similar levels to low types. For both low and medium types, signalling by revising and increasing giving was effective, as countersignalling predicted, but for the former it was too expensive to be affordable, for the latter it is likely to have been worthwhile to signal. Regarding the behaviour of high types, they conformed to the predictions of the model and did not revise—had they revised and signalled they would not have gained further trust.

We set out to answer the question, when people realise that a past behaviour of theirs provides information about them to others, and that this may have consequences for them, good or bad, what do they do? The answer for a majority of subjects is simply 'nothing'. Others do decide to take action: to either remedy negative information receivers have about them, or to cut their losses, decrease their generosity and gain more in the short term. Yet what people do varies by their conditions. Individuals who naturally behave in a generous way, and consequently emit the 'right kind' of sign primarily rely on their sign to get their message across. They eschew the use of a strategic signal. Rightly so, as giving the same as their preferred level but through signalling is wasteful in terms of time and energy, while giving more than what they prefer is wasteful also in monetary terms. Others, who behave in a naturally miserly fashion also mainly avoid the use of a signal. For them, it is costly—they prefer to give nothing—and their gains to doing so are capped at the level of a stranger. For those who give an intermediate amount naturally, signalling seems to be a profitable strategy as they can be perceived as either trustworthy or untrustworthy, and so a proportion of these revise in an effort to be trusted.

Appendix A. Regression Tables

See Tables A1–A4.

Table A1
Logistic regression outputs for the odds of trustees choosing to return according to their veiled dictator game giving (H1a) and the odds of trusters bestowing trust according to trustees' veiled dictator game giving (H1b). The coefficients are odds ratios.

Dependent variable	Return exp(*b*)	Trust exp(*b*)
Veiled giving	2.22***	1.40***
	(0.34)	(0.05)
Constant	0.15**	0.20***
	(0.15)	(0.10)
Controls		
Age	0.98	1.02
	(0.03)	(0.02)
Female	0.85	1.05
	(0.34)	(0.21)
Session 1	3.03	1.87
	(2.21)	(0.71)
Session 2	1.38	1.20
	(1.02)	(0.41)
Session 3	1.86	1.30
	(1.38)	(0.44)
Session 4	0.48	0.74
	(0.42)	(0.28)
Session 5	1.29	0.84
	(0.97)	(0.33)
Session 6	2.76	1.09
	(1.99)	(0.38)
Session 7	1	1
	(Omitted)	(Omitted)
Observations	155	2450

Standard errors in parentheses. Standard errors adjusted for clustering at subject level in the regression in which Trust is the dependent variable.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Table A2
Logistic regression output for the odds of trustees choosing to revise according to their veiled dictator game giving (H2). The coefficients are odds ratios.

Dependent variable	Revise exp(*b*)	Revise exp(*b*)
Veiled giving	4.50***	
	(2.17)	
Veiled giving squared	0.68***	
	(0.09)	
High generosity		0.26**
		(0.17)
Medium generosity		1
		(Omitted)
Low generosity		0.29***
		(0.13)
Constant	0.12**	0.42
	(0.12)	(0.39)
Controls		
Age	1.02	1.01
	(0.03)	(0.03)
Female	0.62	0.63
	(0.24)	(0.24)
Session 1	1.23	1.26
	(0.93)	(0.95)
Session 2	2.43	2.40
	(1.76)	(1.74)
Session 3	1.87	1.94
	(1.42)	(1.47)
Session 4	1.68	1.77
	(1.27)	(1.32)

Table A2 (*Continued*)

Dependent variable	Revise exp(*b*)	Revise exp(*b*)
Session 5	1.41	1.41
	(1.11)	(1.11)
Session 6	0.57	0.61
	(0.47)	(0.50)
Session 7	1	1
	(Omitted)	(Omitted)
Observations	155	155

Standard errors in parentheses. *N.B.* Only one observation per trustee was collected. Therefore, we do not use cluster robust standard errors here.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Table A3
The logistic regressions show the effect that an increase in unveiled giving, relative to not revising, has on the trusting behaviour of trusters (H4a, H5, H6b). The unveiled giving variable shows the impact of increasing; the constant corresponds to a decision to not revise. The coefficients are odds ratios.

Dependent variable	Veiled £0 Trusting exp(*b*)	Veiled £1 Trusting exp(*b*)	Veiled £2 Trusting exp(*b*)	Veiled £3 Trusting exp(*b*)	Veiled £4 Trusting exp(*b*)
Unveiled giving	1.27***	1.31***	1.29***	1.12	1.02
	(0.06)	(0.08)	(0.08)	(0.09)	(0.15)
Not revise (constant)	0.04***	0.07***	0.43	2.42	3.04
	(0.03)	(0.07)	(0.30)	(2.14)	(2.85)
Controls					
Age	1.05**	1.06	1.00	0.98	1.00
	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)
Female	1.18	1.16	1.43	1.09	0.98
	(0.36)	(0.36)	(0.44)	(0.38)	(0.36)
Session 1	2.31	0.63	2.45	1.51	2.01
	(1.37)	(0.36)	(1.55)	(1.05)	(1.31)
Session 2	1.90	0.66	1.35	0.87	1.40
	(0.92)	(0.38)	(0.79)	(0.60)	(0.92)
Session 3	0.79	2.41*	1.27	2.69	0.48
	(0.50)	(1.26)	(0.70)	(1.92)	(0.32)
Session 4	0.85	0.56	0.62	0.36	0.53
	(0.47)	(0.37)	(0.37)	(0.23)	(0.32)
Session 5	0.81	1.78	0.92	0.40	0.35
	(0.63)	(0.97)	(0.54)	(0.25)	(0.23)
Session 6	1.57	1.04	1.03	0.62	1.26
	(0.81)	(0.56)	(0.62)	(0.39)	(0.85)
Session 7	1	1	1	1	1
	(Omitted)	(Omitted)	(Omitted)	(Omitted)	(Omitted)
Observations	447	317	318	328	270

Standard errors in parentheses, adjusted for clustering at subject level. *N.B.* Observations where unveiled giving was equal to or below its corresponding veiled giving value were dropped (e.g. observations which had a veiled giving of £3 and an unveiled giving of £3 or less were dropped). The decision to not revise was then coded to replace the situation where unveiled giving was equal to veiled giving and—veiled giving values were centred on the unveiled giving values.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Table A4
Output for the logistic regression showing the difference between the odds of trusters placing trust in trustees who appear generous (those appearing to give £4 when veiled) and do not revise and trustees who appear generous and re-give the same amount (H6a). The coefficients are odds ratios.

Dependent variable	Trusting exp(*b*)
Revise and re-give the same amount	0.68
	(0.26)
Not revise (constant)	9.13**
	(9.19)
Controls	
Age	0.94*
	(0.03)
Female	0.73
	(0.33)
Session 1	3.00
	(2.55)
Session 2	1.13
	(1.00)
Session 3	0.59
	(0.51)
Session 4	1.35
	(1.11)
Session 5	0.99
	(0.77)
Session 6	1.55
	(1.28)
Session 7	1
	(Omitted)
Observations	131

Standard errors in parentheses, adjusted for clustering at subject level.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2014.07.009.

References

Ashraf, N., Bohnet, I., Piankov, N., 2006. Decomposing trust and trustworthiness. Exp. Econ. 9, 193–208.
Belot, M., Duch, R., Miller, L., 2010. Who Should be Called to the Lab? A Comprehensive Comparison of Students and Non-Students in Classic Experimental Games (Discussion Paper No. 2010001). Nuffield College, University of Oxford.
Binmore, K., Shaked, A., Sutton, J., 1985. Testing noncooperative bargaining theory: a preliminary study. Am. Econ. Rev. 75, 1178–1180.
Bourdieu, P., 1984. Distinction: A Social Critique of the Judgement of Taste. Harvard University Press, Cambridge, MA.
Bracht, J., Feltovich, N., 2009. Whatever you say, your reputation precedes you: observation and cheap talk in the trust game. J. Public Econ. 93, 1036–1044.
Diggle, S.P., Gardner, A., West, S.A., Griffin, A.S., 2007. Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? Philos. Trans. R. Soc. B: Biol. Sci. 362, 1241–1249.
Ellingsen, T., Johannesson, M., Tjøtta, S., Torsvik, G., 2010. Testing guilt aversion. Games Econ. Behav. 68, 95–107.
Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14, 583–610.
Ermisch, J., Gambetta, D., Laurie, H., Siedler, T., Noah Uhrig, S.C., 2009. Measuring people's trust. J. R. Stat. Soc.: Ser. A (Stat. Soc.) 172, 749–769.
Faris, R., Felmlee, D., 2011. Status struggles: network centrality and gender segregation in same- and cross-gender aggression. Am. Sociol. Rev. 76, 48–73.
Fehrler, S., Przepiorka, W., 2013. Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. Evol. Hum. Behav. 34, 139–145.
Feltovich, N., Harbaugh, R., To, T., 2002. Too cool for school? Signalling and countersignalling. RAND J. Econ. 33, 630–649.
Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.
Gambetta, D., 2009. Signaling. In: Hedström, P., Bearman, P. (Eds.), The Oxford Handbook of Analytical Sociology. Oxford University Press, Oxford, pp. 168–194.
Gambetta, D., Przepiorka, W., 2014. Natural and strategic generosity as signals of trustworthiness. PLoS ONE 9 (e97533), http://dx.doi.org/10.1371/journal.pone.0097533.
James, E., 2003. A Life Inside: A Prisoner's Notebook. Atlantic Books, London.
Powell, B., Wilson, B.J., 2008. An experimental investigation of Hobbesian jungles. J. Econ. Behav. Organ. 66, 669–686.
Pradel, J., Euler, H.A., Fetchenhauer, D., 2009. Spotting altruistic dictator game players and mingling with them: the elective assortation of classmates. Evol. Hum. Behav. 30, 103–113, http://dx.doi.org/10.1016/j.evolhumbehav.2008.09.003.
Schniter, E., Sheremeta, R.M., Sznycer, D., 2013. Building and rebuilding trust with promises and apologies. J. Econ. Behav. Organ. 94, 242–256.
Schotter, A., Weiss, A., Zapater, I., 1996. Fairness and survival in ultimatum and dictatorship games. J. Econ. Behav. Organ. 31, 37–56.
Scott-Phillips, T.C., 2008. Defining biological communication. J. Evol. Biol. 21, 387–395.
Spence, M.A., 1973. Job market signaling. Quart. J. Econ. 87, 355–374.
Spence, M.A., 1974. Market Signaling: Informational Transfer in Hiring and Related Screening Processes. Harvard University Press, Cambridge, MA.
Sykes, G.M., 1958. The Society of Captives: A Study of a Maximum Security Prison. Princeton University Press, Princeton.
Thaler, R.H., 1988. Anomalies: the ultimatum game. J. Econ. Perspect. 2, 195–206.