Casting Doubt: Image Concerns and the Communication of Social Impact

Foerster, M.; van der Weele, J.J.

DOI
10.1093/ej/ueab014
Publication date
2021
Document Version
Final published version
Published in
Economic Journal

THE ECONOMIC JOURNAL

The Economic Journal, **131** (*October*), 2887–2919 https://doi.org/10.1093/ej/ueab014

Advance Access Publication Date: 13 February 2021

CASTING DOUBT: IMAGE CONCERNS AND THE COMMUNICATION OF SOCIAL IMPACT*

Manuel Foerster and Joël J. van der Weele

We investigate strategic communication about the social impact of costly prosocial actions. A 'sender' with noisy information about impact sends a cheap-talk message to a 'receiver', upon which both agents choose whether to act. In the presence of social preferences and image concerns, the sender trades off *persuasion*, exaggerating impact to induce receiver action, and *justification*, downplaying impact to cast doubt on the effectiveness of action and excuse her own passivity. In an experiment on charitable giving we find evidence for both motives. In line with our theory and a justification motive, increasing image concerns reduces communication of positive impact.

Our moral reputation depends on the social impact of our actions. When we sacrifice for other people, we are more likely to be viewed as altruistic and benevolent. However, our social impact is often unclear, allowing us to misrepresent it and cast our actions in a better light. For instance, we may downplay the importance of climate change in order to excuse our current lifestyle, gloss over the plight of disadvantaged minorities to evade the burden of helping, or trivialise the impact of politically inconvenient policies. These misrepresentations come at a cost however, because they encourage others to behave selfishly in turn, compounding the negative outcomes to society.

How people make such trade-offs in interpersonal communication and how they affect the subsequent decisions is unknown. We investigate this issue in the context of charitable giving. This is an ideal setting, as the impact of giving is an important determinant of donations (Meer, 2014). At the same time, impact is often opaque, affecting the perception of the giving decision and the donation itself. Some donors exploit uncertainty as a personal excuse not to give (Exley, 2016; 2020), or even avoid impact information altogether to mitigate the demands of conscience (Dana *et al.*, 2007; Grossman and van der Weele, 2017). By contrast, enthusiastic donors cultivate an inflated sense of efficacy, in order to better savour their own generosity (Niehaus, 2020).

To understand communication about social impact in this context, we develop a formal theory of strategic communication and conduct an experiment to test it. Our model features two agents who may take a prosocial action. The 'sender' receives a noisy but informative signal about the

* Corresponding author: Joël J. van der Weele, University of Amsterdam, Tinbergen Institute, Roeterstraat 11, 1018WB Amsterdam, The Netherlands. Email: vdweele@uva.nl

This paper was received on 9 January 2020 and accepted on 10 February 2021. The Editor was Rachel Kranton.

The data and codes for this paper are available on the Journal website. They were checked for their ability to reproduce the results presented in the paper.

An earlier version of this paper has circulated under the title 'Persuasion, justification and the communication of social impact'. We would like to thank the editor, three anonymous referees, Roland Bénabou, Jason Dana, Christine Exley, Anke Gerber, Gero Henseler, Mark Le Quement, Yves Le Yaouanq, Andreas Nicklisch, Timo Promann, Karl Schlag, Peter Schwardmann, Ivan Soraperra, Jeroen van de Ven, Achim Voss, Florian Zimmermann and a large number of seminar participants for useful comments. We thank Alejandro Miranda Salas, Davide Pace and Ivar Kolvoort for excellent research assistance. Joël van der Weele gratefully acknowledges financial support from the NWO through VIDI grant 452-17-004. The work presented in this paper was carried out while Manuel Foerster was affiliated to the University of Hamburg, Germany. We report all data gathered in the context of this research project. Materials can be found at the Open Science Framework https://osf.io/94juy/files/. Ethical Approval was granted by the University of Amsterdam with reference number EC 20180205020227.

social return associated with the action. She then submits a cheap-talk report about the return to the other agent, the 'receiver', upon which both agents decide whether to take the action. In line with empirical evidence, we assume that agents differ in their intrinsic motivation to donate and would like to be perceived as a prosocial actor. Our model generates a central trade-off in the sender's communication decision. Publicly communicating a high social impact may *persuade* the receiver to take the action and increase social welfare. At the same time, it raises moral pressure on the sender to take the prosocial action herself. To escape such pressure and *justify* inaction, the sender may instead downplay the social returns.

The strength of image concerns emerges as a crucial determinant of equilibrium communication. When image concerns are low, the persuasion motive dominates and induces opportunistic exaggeration of the social return on the action. Some of this exaggeration is 'hypocritical', as some senders report high impact but do not act themselves. Exaggeration and hypocrisy reduce the persuasiveness of communication in equilibrium. By contrast, when image concerns are high, the justification motive dominates. This deters (most) hypocritical reports of high impact and allows for equilibria with 'influential communication', in which the messages affect the receiver's action. Agents with low intrinsic motivation to give may even downplay social impact in equilibrium, thus justifying their own passivity, but reducing prosocial actions by the receiver. Taking these effects together, image concerns reduce the relative frequency of high signals about the impact of prosocial actions in equilibrium.

To test this prediction, we conduct a laboratory experiment on charitable giving. An informed 'sender' is matched with an uninformed 'receiver', both of whom may choose to make a donation to a charity, GiveDirectly. Before they do so, the sender receives a noisy signal about the impact of the donation. She can then communicate a message to the receiver, with the option to falsify the signal she observed. To test our predictions, the experiment varies the strength of reputation motives. While giving is anonymous in a *private* treatment, we make the donation of the sender visible and salient to the receiver and other participants in a *public* treatment, creating a justification motive.

We find evidence for a persuasion motive, as more than 40% of the subjects in the *private* treatment exaggerate impact at least in one round. This misrepresentation happens despite a lack of personal monetary gain from the receiver's donation. Moreover, in line with a motive to justify selfish inaction, senders in the *public* treatment are about 10 percentage points less likely to report high impact. As the theory predicts, senders in the *public* treatment are also more likely to donate after a high message to avoid looking hypocritical, making messages more costly than in the *private* treatment. The shift in communication affects behaviour of receivers, who are about 40 percentage points less likely to donate after receiving a message of low impact.

These results show that communication about impact is intimately tied up with image management, and adds to the theoretical and the empirical literature on communication in prosocial contexts. First, on the theoretical side, we characterise equilibrium communication about impact in the presence of reputation concerns. We show that such concerns allow for information transmission even if preferences regarding the third party are not aligned. They reduce the relative frequency of high signals but may also distort the transmitted information in the direction of low impact. Our model combines cheap-talk pre-play communication (Crawford and Sobel, 1982) with a signalling model of prosocial behaviour following Bénabou and Tirole (2006) and Ellingsen and Johannesson (2008). In related work, Morris (2001) shows how reputation concerns may lead unbiased advisors away from truth telling towards 'political correctness', in order to be perceived as a type with aligned preferences. Ottaviani and Sørensen (2006) show that if

the informativeness of the sender's signal is uncertain and the sender wants to be perceived as well informed, then she cannot reveal all her information in equilibrium.[1] By contrast, in our paper, we consider an environment where agents can signal both through their messages and their actions.

In our model, communication affects the signalling value of subsequent actions. A few other studies explore how agents may change the parameters of the game to influence subsequent signalling equilibria. Bénabou and Tirole (2011) and Ali and Bénabou (2020) investigated incentives to induce socially desirable behaviour by agents with prosocial and image concerns. Henry and Louis-Sidois (2020) showed that in a public good environment, agents may vote against sanctions for non-compliance to increase the signalling value of their contributions. Kuran (1997) discussed the concept of 'preference falsification', the public misrepresentation of private preferences or opinions to conform with the majority opinion. We enrich these settings by showing how communication itself may be strategically employed to affect subsequent signalling incentives, and the trade-off between persuasion and justification.

Most closely related is independent and contemporaneous theoretical work by Bénabou *et al.* (2018), which highlights a similar trade-off in communication. Agents with image concerns first search for and then disclose verifiable information about the size of an externality. Agents may withhold positive information to justify their inaction, similar to the justification motive in our framework. Furthermore, agents have an 'influence motive' to increase actions by others, mirroring our persuasion motive. There are multiple differences between the papers: whereas Bénabou *et al.* (2018) mostly abstract from modelling the persuasion technology, we use standard communication models in economics where the receiver is a Bayesian agent. In terms of results, our paper shows that image concerns reduce the relative frequency of high signals about the social returns in a one-shot interaction, whereas Bénabou *et al.* (2018) focus on the diffusion of ideas in linear networks, as well as the endogenous search for narratives and differences between various forms of communication ('narratives' versus 'imperatives').

Our second contribution is on the empirical side, where we are the first to study the trade-off between persuasion and justification. Our main finding is that image concerns induce subjects to report a positive impact of a donation less often; exaggeration largely disappears and some participants even 'downplay' impact. By contrast, the literature on audience effects mainly emphasises the positive impact of image concerns on prosocial behaviour (Bursztyn and Jensen, 2017).[2] More generally, the experiment demonstrates how communication about social impact is intimately tied to public image management. It helps explain why efficient information aggregation cannot be taken for granted when it comes to morally charged topics like charitable giving, or other prosocial actions and public good contributions.

In related empirical work, Hillenbrand and Verrina (2018) showed how 'narratives' about giving expressed by experimental participants affect the giving of their peers. Bursztyn *et al.* (2020) conducted a series of survey experiments, where they measured the willingness of US residents to publicly choose socially stigmatised, anti-immigrant action. They did not consider communication but varied the visibility of a non-racist 'excuse'. Participants choose the stigmatised action

[1] There is a considerable literature on cheap talk with many players (Hagenbach and Koessler, 2010; Galeotti *et al.*, 2013), on networks (Foerster, 2019) and in committees prior to voting (Coughlan, 2000; Austen-Smith and Feddersen, 2006; Deimen *et al.*, 2015). Foerster (2020) investigated cheap talk when the sender observes multiple noisy signals.

[2] In line with theories of costly signalling of altruism, increasing the visibility of donations raises prosocial behaviour in various contexts, as has been demonstrated in the lab (e.g., Andreoni and Petrie, 2004; Rege and Telle, 2004; Ariely *et al.*, 2009) and in the field (e.g., Harbaugh, 1998; Soetevent, 2005; Lacetera and Macis, 2010; Karlan and McConnell, 2014). Soraperra *et al.* (2019) highlighted a situation where image concerns are detrimental.

more often when the audience is informed that they had the excuse, as this reduces the social cost of the action. The focus on misrepresentation of information also relates our paper to a literature on lying and deception. Many studies show that people lie for money, although not everyone does so (e.g., Gneezy, 2005; Abeler *et al.*, 2019). Our experiment demonstrates that people are willing to lie for other reasons as well: about half of the participants are willing to misrepresent information without *any* pecuniary benefit to themselves. In particular, some people are willing to lie to convince others to donate or to protect their image as a prosocial actor.

Finally, we speak to a literature on how donors react to the impact of charitable donations. Previous literature has generally found that perceived effectiveness matters for giving. As noted above, Meer (2014) reviewed a literature showing that increasing the return to giving through matching grants, tax rebates or lower organisational overhead increases donations. Gneezy *et al.* (2014) found that people are especially reluctant to pay for overhead costs. In a recent, large scale study about charitable motives among Canadians, 61% said they would give more if they had more confidence in charities and where the money is going (Angus Reid Institute, 2017).

Research on communicating factual aspects of effectiveness however shows mixed results. Gordon *et al.* (2009) found that positive ratings from watchdog organisations were associated with higher givings to the rated charities, Yörük (2016) used a regression discontinuity design to show a causal effect, but only for smaller charities. In a field experiment on charitable fund raising Karlan and Wood (2017) found no effect of communicating effectiveness, although there is a positive effect on large donors. In another field study, Karlan and List (2020) showed that emphasising the name of big donors can help signal the quality of the charity and raise donations. Metzger and Günther (2019a) investigated the framing of effectiveness and found that the perceived effectiveness of a donation increases giving, but detailed knowledge about the projects decreases it. Metzger and Günther (2019b) found that information about the type of the recipient and administrative costs had a strong impact on giving, while information about impact did not. Our study differs from this literature by focusing on communication among donors instead of the communication from charity to donor.

Finally, Butera and Horn (2020) varied the public visibility of both donations and the information of charity effectiveness in the laboratory. Information about effectiveness increases giving in private conditions, but reduces giving when it is publicly received. They concluded that image-conscious donors strategically reduce the quantity of giving to signal that they give 'smart'. By contrast, our study focuses on communication decisions themselves, and how they are distorted by image concerns.

1. Theory

1.1. *Model*

In this section we outline our theoretical model.[3] There are two agents, a sender and a receiver, who are denoted by subscripts s and r, respectively. Both agents choose to take a prosocial action

[3] In a working paper we model more general assumptions on the payoff and information structure (Foerster and van der Weele, 2018). First, both agents have information and are both sender and receiver. Second, it allows for monetary spillovers between both agents, i.e., the action is a traditional public good. This last feature increases the benefit of persuading the other player to contribute compared to the model in this paper, but the models yield very similar insights. The specific restrictions on the benefits from prosocial actions and the conditional probabilities of the signals in this setting ease the exposition and match the setup of the experiment.

$\hat{a} = 1$, or not, $\hat{a} = 0$.[4] This action has a cost $c > 0$ and generates a benefit $W \in \{0, 1\}$ for the agent. In addition, the action has a positive spillover γW, with $\gamma > 0$, to a passive third party, e.g., a charity. As we explain below, this spillover may confer psychological benefits to the agent. A priori, the value of W is unclear, but there is a common prior that $W = 1$ with probability $1/2$.

The timing of the game is as follows. The sender receives an unbiased but noisy signal $\sigma \in \Sigma = \{0, 1\}$ about W, where $\sigma = W$ with probability $2/3$ and $\sigma = 1 - W$ with probability $1/3$. We consider noisy signals for the sake of realism, as the presence of uncertainty even among experts characterises almost all policy debates; qualitatively, our results would not change with precise signals, because the noise simply compresses the sender's posteriors. After the sender has received the noisy signal, she submits a report $\hat{m} \in M = \{0, 1\}$ about her signal σ via cheap talk, that is, her report is costless, unverifiable and non-binding with respect to the action.

The receiver observes the sender's report $\hat{m}$ and both agents decide whether to take the action. Finally, each agent observes the action of the other agent. Note that we employ binary state and choice variables to keep the analysis simple and to allow for a straightforward implementation of the model in the laboratory.

Turning to the preferences of the sender, we introduce several behavioural elements. First, we introduce a social preference parameter $\theta_s \in \Theta = \{\underline{\theta}, \overline{\theta}\}$ that determines the degree of 'altruism' or 'intrinsic motivation' toward the third party. Heterogeneity in such preference is one of the key findings of the literature on prosocial behaviour and public goods (e.g., Fischbacher *et al.*, 2001; Burlando and Guala, 2004; Kurzban and Houser, 2005). We assume that θ_s is private information and takes the value of $\overline{\theta} > 0$ with prior probability $\pi \in (0, 1)$ and the value of $\underline{\theta} \in (0, \overline{\theta})$ with prior probability $1 - \pi$. Thus, we follow Bénabou and Tirole (2006) in defining prosocial preferences of the sender towards a third party or abstract social good, rather than the payoff of the receiver. We refer to θ_s as the sender's *type*, and refer to $\theta_s = \underline{\theta}$ as a low type and $\theta_s = \overline{\theta}$ as a high type. Hence, the sender is a type-signal pair (θ_s, σ).

Second, we introduce *image concerns*, i.e., the sender cares about the receiver's expectation of her type. We assume that the receiver's inference about the sender can depend on the sender's report $\hat{m}$ and action $\hat{a}_s$. The parameter $\mu \geq 0$ measures the importance of image concerns to the sender.[5] Formally, the preferences of the sender are given by

$$u_s(\theta_s, \sigma, \hat{m}, \hat{a}_s, \hat{a}_r) = (W - c)\hat{a}_s + \theta_s(\hat{a}_s + \hat{a}_r)\gamma W + \mu E_r[\theta_s \mid \hat{m}, \hat{a}_s].$$

Here, the first term represents the direct benefits and costs from the action, the second term represents indirect benefits from the actions taken by both agents, and the last term captures image concerns. This model is closely related to other models of signalling social preferences, e.g., Bénabou and Tirole (2006; 2011), Ellingsen and Johannesson (2008), Andreoni and Bernheim (2009), Grossman and van der Weele (2017) and Ali and Bénabou (2020).[6]

The preferences of the receiver are the same as those of the sender. To match the setup of the experiment, we abstract away from the image concerns of the receiver, but our results would

[4] To distinguish actual decisions of the agents from strategies, we indicate them by a 'hat' symbol.

[5] There is some evidence that image concerns are negatively correlated with social preferences (Friedrichsen and Engelmann, 2018; Henry and Sonntag, 2019). Our results are qualitatively robust to the introduction of such a negative correlation.

[6] The main difference from traditional signalling models like Spence (1973) is that the sender cares directly about the beliefs of the observer instead of the observer's actions. Formally, this turns the model into a psychological game à la Geanakoplos *et al.* (1989) and Battigalli and Dufwenberg (2009); see Battigalli and Dufwenberg (2019) for a recent survey. One could see this as a proxy for the continuation value in a game in which agents with a good reputation will reap additional benefits from future interactions. We abstract from image concerns that depend on the identity of the observer, as in Levine (1998) or Ellingsen and Johannesson (2008).

not change qualitatively by adding these. Thus, preferences are given by $u_r(\theta_r, \hat{a}_s, \hat{a}_r) = (W - c)\hat{a}_r + \theta_r(\hat{a}_s + \hat{a}_r)\gamma W$, where $\theta_r = \overline{\theta}$ with prior probability π and $\theta_r = \underline{\theta}$ with prior probability $1 - \pi$.

The solution concept we employ is perfect Bayesian equilibrium. In the main analysis, we restrict our attention to pure strategies, while we consider mixed strategies in Online Appendix B. A (pure) strategy (m, a_s) for the sender is a pair of mappings

$$m\colon \Theta \times \Sigma \to M \quad \text{and} \quad a_s\colon \Theta \times \Sigma \times M \to \{0, 1\}$$

that assign a report to each type-signal pair (first stage) and an action to each type-signal pair and report submitted to the receiver (second stage), respectively. For the receiver, a (pure) strategy a_r is a mapping $a_r : \Theta \times M \to \{0, 1\}$ that assigns an action to each type and report received.

1.2. *Equilibrium Analysis*

Our main interest is the sender's equilibrium behaviour. We categorise strategies according to four communication patterns that play an important role in our analysis.

DEFINITION 1. *Consider any sender strategy* (m, a_s).

(*i*) Honesty. *Type* $\theta_s \in \Theta$ *is 'honest' or 'truthful' if she always submits a report that corresponds to her signal,* $m(\theta_s, \sigma) = \sigma$ *for all* $\sigma \in \Sigma$.
(*ii*) Exaggeration. *Type* $\theta_s \in \Theta$ *'exaggerates impact' if she submits a high report regardless of her signal,* $m(\theta_s, \sigma) = 1$ *for all* $\sigma \in \Sigma$.
(*iii*) Downplaying. *Type* $\theta_s \in \Theta$ *'downplays impact' if she submits a low report regardless of her signal,* $m(\theta_s, \sigma) = 0$ *for all* $\sigma \in \Sigma$.
(*iv*) Hypocrisy. *Type* $\theta_s \in \Theta$ *is a 'hypocrite' if she submits a high report and does not take the action for some signal* $\sigma \in \Sigma$, $(m(\theta_s, \sigma), a_s(\theta_s, \sigma, m(\theta_s, \sigma))) = (1, 0)$.

Note that honesty, exaggeration and downplaying are mutually exclusive. By contrast, honesty and hypocrisy as well as exaggeration and hypocrisy may occur together, but are conceptually distinct. In particular, while honesty, exaggeration and downplaying only relate to communication, hypocrisy also relates to actions.

We are interested in situations with 'influential communication': at least some information is transmitted and affects the action of the receiver.

DEFINITION 2. *Consider any strategy profile* $((m, a_s), a_r)$.

(*i*) Information transmission. *We say that there is 'information transmission' or 'truthful communication' if at least one type* $\theta_s \in \Theta$ *is truthful.*
(*ii*) Influential communication. *We say that there is 'influential communication' if receiving a high report increases the likelihood of taking the action,* $E(a_r(\cdot, \hat{m}) \mid \hat{m} = 1) > E(a_r(\cdot, \hat{m}) \mid \hat{m} = 0)$.

We focus our analysis on situations in which high-type senders and receivers may take the action, while low types do not. That is, we assume that the cost of the action exceeds the low type's benefit from it, which is bounded from above by $(1 + \underline{\theta}\gamma)E[W \mid \sigma = 1] + \mu(\overline{\theta} - \underline{\theta}) = 2(1 + \underline{\theta}\gamma)/3 + \mu(\overline{\theta} - \underline{\theta})$. Furthermore, to ease the exposition, we assume that $\overline{\theta} - 2\underline{\theta} \leq 1/\gamma$,

which implies that the action is weakly more beneficial to a low type with a high signal $\sigma = 1$ than to a high type with a low signal $\sigma = 0$.[7]

ASSUMPTION 1. *Suppose that*

(*i*) $c > 2(1 + \underline{\theta}\gamma)/3 + \mu(\overline{\theta} - \underline{\theta})$,
(*ii*) $\overline{\theta} - 2\underline{\theta} \leq 1/\gamma$.

Assumption 1 allows us to state the following lemma, which narrows down the potential equilibria with influential communication.

LEMMA 1. *Under Assumption 1, the high-type receiver acts conditional on a high report in any equilibrium with influential communication.*

This result establishes that influential communication implies a 'persuasion motive': the sender has incentives to turn to hypocrisy/exaggeration to persuade the receiver to take the action. We first analyse how this motive affects the equilibria of the game if image concerns are low. Suppose that both sender types are honest and that the high-type sender acts conditional on a high signal. Then a sender of type θ_s with a low signal $\sigma = 0$ has no incentives to exaggerate impact (which would yield a low image instead of the prior image $\pi\overline{\theta} + (1 - \pi)\underline{\theta}$) if and only if

$$\mu(\pi\overline{\theta} + (1 - \pi)\underline{\theta}) \geq \pi\theta_s\gamma E[W \mid \sigma = 0] + \mu\underline{\theta} \quad \Longleftrightarrow \quad \mu \geq \frac{\theta_s\gamma}{3(\overline{\theta} - \underline{\theta})}. \tag{1}$$

Equation (1) shows that senders have incentives to deviate from honesty if image concerns are low enough, precluding influential communication. The following result pins down the exact bound on μ below which influential communication cannot be an equilibrium.

PROPOSITION 1. *Under Assumption 1, there does not exist an equilibrium with influential communication if*

$$\mu < \frac{\max\{\overline{\theta}, 2\underline{\theta}\}\gamma(2 - \pi)}{3(\overline{\theta} - \underline{\theta})}.$$

The proof of all results is presented in Appendix A. The intuition behind Proposition 1 is that, given influential communication and low-image concerns, senders prefer to exaggerate impact and induce actions by the receiver. These actions yield psychological benefits that are determined by θ_s and γ. Another insight from Proposition 1 is that influential communication is ruled out if the difference in social preferences, $\overline{\theta} - \underline{\theta}$, and hence the potential loss in image from being perceived as a low type, is too small relative to image concerns. Naturally, the result in Proposition 1 might be weakened by the presence of lying costs, which we discuss below in Subsection 1.3.

Although incentives to exaggerate impact preclude influential communication, there are equilibria in which both types downplay impact, as hypocrisy does not induce receiver action in absence of influential communication. However, such equilibria are not plausible—at least when image concerns are very low. To see this, suppose that both sender types downplay impact and that the receiver mistakenly takes the sender's report at face value with a small probability $\varepsilon > 0$.

[7] We thereby rule out the case in which a high type would always take the action, while a low type would never do so. Including it would not change the set of equilibria with influential communication, as in this case influential communication is not possible in equilibrium.

Hypocrisy now induces (some) receiver action if costs are not too high, $c \leq 2(1+\overline{\theta}\gamma)/3$. Hence, it is beneficial to turn to hypocrisy/exaggeration if image concerns are not too high. As a result, both sender types exaggerate impact and the high-type sender acts conditional on a high signal in the unique equilibrium. The following remark pins down the exact threshold on image, below which either the high-type or the low-type sender with a high signal $\sigma = 1$ has incentives to turn to hypocrisy/exaggeration.

REMARK 1. *Suppose that the receiver takes the sender's report at face value with probability $\varepsilon > 0$. Under Assumption 1, the unique equilibrium is such that both sender types exaggerate impact and the high-type sender acts conditional on a high signal if and only if*

$$c \leq \frac{2(1+\overline{\theta}\gamma)}{3} \quad \text{and} \quad \mu < \frac{2\varepsilon \max\{\pi\overline{\theta}, (2-\pi)\underline{\theta}\}\gamma}{3(\overline{\theta}-\underline{\theta})}.$$

Next, we show that, with sufficient image concerns, there exists an equilibrium with influential communication. The increased importance of reputation introduces a 'justification motive': the need to explain or 'justify' inaction. The justification motive deters exaggeration and hypocrisy, as the latter is associated with a low image in equilibrium. This causes the high type to be honest to make her report match her action, while the low type, who never contributes, now has incentives to downplay impact in order to avoid the loss in image from hypocrisy. To see this last point, suppose that both sender types are honest and that the high-type sender acts conditional on a high signal. Then a low-type sender with a high signal $\sigma = 1$ has no incentives to downplay impact (which would yield the prior image $\pi\overline{\theta} + (1-\pi)\underline{\theta}$ instead of a low image) if and only if

$$\pi\underline{\theta}\gamma E[W \mid \sigma = 1] + \mu\underline{\theta} \geq \mu(\pi\overline{\theta} + (1-\pi)\underline{\theta}) \quad \Longleftrightarrow \quad \frac{2\underline{\theta}\gamma}{3(\overline{\theta}-\underline{\theta})} \geq \mu.$$

Thus, high-image concerns imply that the low-type sender prefers downplaying impact over honestly sharing a high signal, even if this is likely to depress prosocial behaviour by the receiver. The following result pins down the exact equilibrium conditions and shows that this is the unique equilibrium with influential communication.

PROPOSITION 2. *Under Assumption 1, there exists an equilibrium with influential communication, in which the high-type sender is honest and acts conditional on a high signal and the low-type sender downplays impact and does not act, if and only if*

$$\frac{(3-2\pi)(1+\overline{\theta}\gamma)}{3(2-\pi)} \leq c \leq \frac{2(1+\overline{\theta}\gamma)}{3}, \tag{2}$$

$$\mu \geq \frac{\max\{\overline{\theta}, 2\underline{\theta}\}\gamma(2-\pi)}{3(\overline{\theta}-\underline{\theta})}. \tag{3}$$

No other equilibrium with influential communication exists.

This result shows that unlike in the low-image case, influential communication is possible when image concerns are high, reducing the relative frequency of high signals. The bounds on the cost of the action in (2) ensure that acting conditional on a high report is incentive compatible for the high-type receiver. The lower bounds on image concerns in (3) deter hypocrisy by senders who do not act, in particular the high-type sender with a low signal and the low-type sender with

a high signal.[8] Since a higher prior increases the 'outside image' obtained if the sender submits a low report and does not take the action, hypocrisy is less beneficial if the prior is high, i.e., the lower bounds on image concerns decrease in the prior. Note that the only upper bound on image concerns is implicitly given by Assumption 1 and precludes actions by low-type senders and the high-type sender with a low signal.[9]

Proposition 2 also establishes that there is a unique equilibrium with influential communication. As high-type senders submit a weakly higher report than low-type senders (conditional on the signal) in equilibrium, we only need to show that honesty by both types is not an equilibrium and that exaggeration by the high type and honesty by the low type is ruled out. In the first case, either exaggeration by the high type or downplaying by the low type is beneficial under Assumption 1. In the second case, hypocrisy yields the prior image $\pi\overline{\theta} + (1-\pi)\underline{\theta}$, while submitting a low report and not acting yields a low image, which implies that the low type has incentives to exaggerate impact. The following example illustrates our results, using the parameters for the costs and the spillover to the third party that we employ in the experiment.

EXAMPLE 1. *Suppose that* $c = 2$, $\gamma = 3$, $\underline{\theta} = 1/5$ *and* $\overline{\theta} \in [2/3, 11/15]$. *Under Assumption 1(i), which now reads* $\mu < 14/(3(5\overline{\theta} - 1))$, *there exists an equilibrium with influential communication, in which the high-type sender is honest and acts conditional on a high signal and the low-type sender downplays impact and does not act, if and only if*

$$\mu \geq \frac{5\overline{\theta}(2-\pi)}{5\overline{\theta} - 1}.$$

Note that these conditions require that the prior probability π *is larger than* $3/5$.

In summary, our model shows that image concerns play a central role in equilibrium communication, as they determine the relative importance of persuasion and justification motives. In particular, (a) incentives for exaggeration preclude influential communication in absence of image concerns, and (b) image concerns allow influential communication by deterring hypocrisy. In doing so, image concerns also generate incentives for downplaying social impact, reducing the relative frequency of high signals in equilibrium. The suppression of high signals leads to an (ex ante) lower likelihood of receiver actions and lower receiver welfare compared to truthful communication.

In Online Appendix B we analyse mixed strategy equilibria. We show that there exists another equilibrium with influential communication, in which the low-type sender is honest, while the high-type sender randomises between honesty and exaggeration and contributes conditional on a high signal. This equilibrium requires lower image concerns than the equilibrium in Proposition 2 and features more high reports, even though these are less credible due to partial exaggeration. This partial exaggeration by high types comes in the form of hypocrisy, which therefore does not carry such a high stigma or loss of image. In turn, this makes downplaying less attractive for low types, who are honest. Overall, the conclusion that image concerns allow influential communication and reduce the relative frequency of high signals by deterring hypocrisy still holds,

[8] Note that hypocrisy is off equilibrium. The equilibrium is supported by beliefs that attribute this deviation to the low type, which yields the largest possible parameter range on which it exists. In particular, we show that there does not exist an equilibrium with influential communication unless the image associated with hypocrisy is sufficiently low.

[9] The introduction of image benefits for the receiver would relax the equilibrium conditions in Proposition 2. It would relax the upper bound on costs (weakly) more than it would tighten the lower bound, which already depends on image concerns through Assumption 1.

although some hypocrisy may persist despite high image concerns. Like downplaying, partial exaggeration/hypocrisy leads to lower receiver welfare compared to truthful communication.

1.3. *Lying Costs*

A large empirical literature has documented that people do not like to lie, even if the exact reasons for this are still under scrutiny (Gneezy, 2005; Abeler *et al.*, 2019). Our model can easily incorporate costs of lying, by assuming that there is a common understanding that message $\hat{m} \in M$ means '$\sigma = \hat{m}$'. Following the definition in Sobel (2020), message $\hat{m}$ then is a *lie* given σ if $\hat{m} \neq \sigma$. We can hence model lying costs for the sender as a disutility $\widetilde{c} \geq 0$ if $\hat{m} \neq \sigma$. Although simplifying, this approach is broadly in line with findings that many people are lying averse independent of social preferences (Erat and Gneezy, 2012).

There are two main takeaways from such an exercise. First, if lying costs are high enough, truthful and influential communication becomes easier to sustain, resulting in new equilibria with (partial) honesty as well as exaggeration. To see this, suppose that the high-type sender exaggerates impact and the low-type sender is honest, so that there is influential communication. Under Assumption 1, the high-type receiver acts conditional on a high report (Lemma 1), which implies that the high-type sender acts conditional on a high signal. Both sender types have no incentives to deviate if and only if

$$\frac{\underline{\theta}\gamma\pi}{3} + \mu\pi(\overline{\theta} - \underline{\theta}) \leq \widetilde{c} \leq \frac{\overline{\theta}\gamma\pi}{3} + \mu\pi(\overline{\theta} - \underline{\theta}). \tag{4}$$

The upper bound in (4) is the expected gain in receiver action and image from hypocrisy for the high type with a low signal as compared to honesty, which would yield a low image as only the low type is expected to submit a low report. Hence, the high type with a low signal has no incentives to deviate to honesty if lying costs do not exceed this bound; this implies that both types with a high signal also submit a high report. Similarly, the lower bound in (4) is the expected gain in receiver action and image from hypocrisy for the low type with a low signal as compared to honesty. It hence deters the low type with a low signal from turning hypocrite. Submitting a low report is optimal for her already at a rather low level of lying costs, because she would benefit the least from hypocrisy. Furthermore, an equilibrium with influential communication in which both types are honest may exist, as long as types are not too different ($\overline{\theta} \leq 4\underline{\theta}$).

Second, as long as lying costs are not so high as to completely dominate image concerns, the tension between persuasion and justification remains. Thus, a qualitatively similar version of Proposition 2 will continue to hold, such that equilibrium communication features downplaying impact by the low type. If $\underline{\theta}$ is small, so that the low type has a strong incentive to lie, the range of the equilibrium in Proposition 2 would actually get larger. The reason is that lying costs reduce the high type's incentive to exaggerate, thus loosening constraint (3). Hence, if lying costs are such that they deter the low type but not the high type from exaggeration ($\underline{\theta}\gamma\pi/3 < \widetilde{c} < \overline{\theta}\gamma\pi/3$), then the equilibrium with partial exaggeration exists for low-image concerns (lower bound of (4)), while the equilibrium with partial downplaying of impact does so for high-image concerns (a weaker version of (3)).

These results show that if high types can exploit the honesty of others induced by lying costs, exaggeration and hypocrisy may be part of an equilibrium with influential communication. Moreover, they show that our result that an increase in image concerns reduces the relative frequency of high signals holds also in this version of the model.

2. Experimental Design and Hypotheses

We now turn to the experimental test of our theory in the laboratory. Like the model, our experiment centres around an individual prosocial act associated with image concerns. We choose a charitable giving decision, since image concerns are known to play an important role in this context. For instance, subjects in the laboratory exert more effort to generate charitable donations when these are public (Ariely *et al.*, 2009), and give more generously to the public good when they are identifiable by other subjects (Andreoni and Petrie, 2004). In the field, it has been shown that people are more likely to give blood when they receive public recognition in the local newspaper (Lacetera and Macis, 2010), give more to their church when others can see their donation (Soetevent, 2005) and, if reporting categories are discrete, give just enough to get into a higher category (Harbaugh, 1998).

The charity in our experiment is GiveDirectly. GiveDirectly makes direct cash transfers to poor recipients in East Africa, and 91% of each donation ends up with the recipient. This charity is suitable as its activities are easy to explain and have a concrete and deserving recipient.[10] While the results are of course specific to this charity, the donations represent a trade-off between self and other that is at the core of any social decision. In the instructions, we provided subjects with some information about recipients, activities and efficiency of the charity by citing an excerpt from their website. Subjects were given assurances that each donation would actually be transferred on their behalf by the experimenter, referring to the no-deception policy of the CREED laboratory. All instructions are available in Online Appendix D.[11]

We conducted the experiment at the CREED laboratory at the University of Amsterdam. Subjects were recruited using the online CREED recruitment system, and consisted of university students, with a majority having a background in economics, business or related fields. More detail about the participant's gender and age is given in Online Appendix Table C.1. In total, 228 subjects participated in 14 sessions: 7 sessions with a total of 116 subjects in the *public* treatment and 7 sessions with a total of 112 subjects in the *private* treatment. Each session had between twelve and eighteen participants, always in even numbers, depending on the show-up rate and lasted about one hour each. Fourteen sessions of the current experiment were run in February 2018. We had to discard the data of one session in which a technical problem occurred, so we ran one additional session in July 2018. All results are robust to the inclusion of the discarded data, or the inclusion of a dummy for the additional session.

2.1. *Interaction Between Sender and Receiver*

In the main part of the experiment, subjects were randomly matched in pairs and allocated the role of 'sender' and 'receiver'. Each subject had to make a binary choice between *option 1* and *option 2*. Figure 1, which is taken from the experimental instructions, depicts the associated payoffs. The payoff of *option 1* depends on the 'type of the interaction', which corresponds to W in the model and is uncertain. If the type was 'red' (low impact), neither the subject nor the charity would earn anything. If the type was 'green' (high impact), the charity would earn €15 and the participant €5. The payoff from *option 2* was independent of the type of interaction and yielded €10 for the participant and €0 for the charity. Below, we sometimes refer to *option 1* as a

[10] See www.givedirectly.org for more details of the charity's activities.

[11] We pre-registered the experiment. The pre-registration and accompanying notes are available in Online Appendix E.

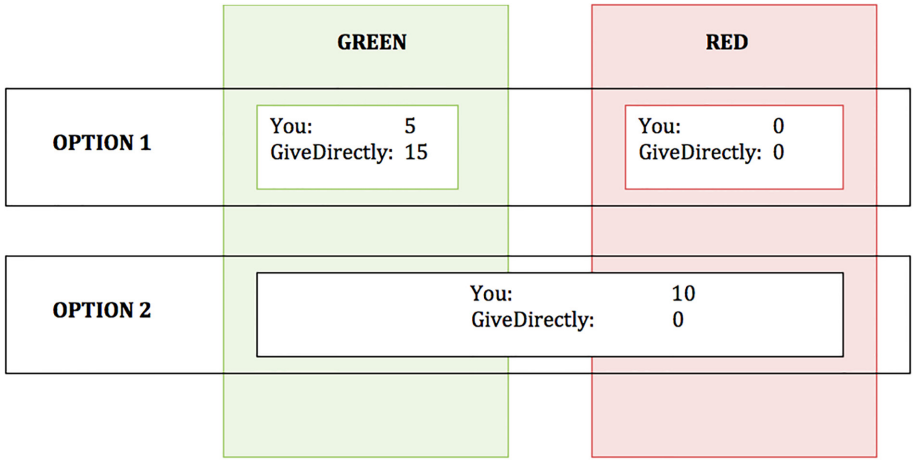


Fig. 1. *Payoffs in the Interaction Phase of the Experiment.*

'donation', since the participant gives up at least €5 to potentially donate €15 to the charity.[12] This interaction directly implements the model, including the binary nature of information, signals and choices. To see this, note that *option 1* corresponds to action $\hat{a} = 1$ and this payoff structure to the cost of the action $c = 2$ and relative spillovers $\gamma = 3$ in the model, as we normalise the potential direct benefit W from the action to 1.

Before making their choices, the sender received a noisy signal about the type of interaction, and communicated with the receiver. To determine the signal, the experimenter first privately rolled a die to determine the type of interaction. Both red or green were equally likely to be selected and this was known to both sender and receiver. Participants did not learn the result of the die roll, but senders received a noisy signal by drawing a card from a deck. The deck consisted of two red cards and one green card if the true type was red, and two green cards and one red card if the true type was green. Thus, the signal was correct with a probability of two-thirds. Upon receiving the signal, the sender communicated with the receiver by showing either a red or a green card. Finally, both sender and receiver chose between the two options described in detail above. The interaction was repeated several times, each time with a new interaction partner.

2.2. *Treatments*

Our aim is to test the prediction that an increase in image concerns (parameter μ in the model) raises the justification motive relative to the persuasion motive, and hence depresses the communication that the state is green, i.e., the communication of high impact. To do so, our main experimental conditions vary the visibility of the senders' actions between a *private* and a *public* treatment.

Image concerns are generally not easy to implement in the lab, where people interact with strangers, most of whom they will never see again. However, previous literature cited at the

[12] Note that, conditional on a green signal, the expected gain is $2/3 \cdot 15 = 10$ euro for the charity, while the expected cost is $2/3 \cdot 5 + 1/3 \cdot 10 = 6.67$. Thus, in expectation, a donation is multiplied by 1.5, which cannot be achieved by donating after the experiment.

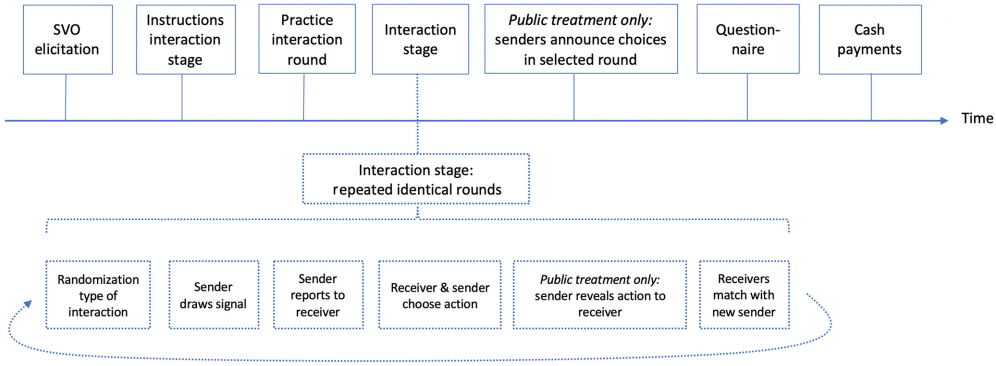


Fig. 2. *Timeline of the Experiment and the Interaction Rounds.*

beginning of this section shows that people are generally concerned about an audience's impression, even if it is composed of strangers. To make this audience as salient as possible, we chose to implement face-to-face interactions (details of the implementation follow below). Because implementing such interactions affects many aspects of the procedures, all treatments are conducted face to face. Instead, the treatments vary the information that is relayed to the audience, and hence the inferences that the audience can make.

In the *private* treatment, participants' choices between *option 1* and *option 2* were not visible to any other participant. By contrast, the *public* treatment featured two procedures that were designed to make image concerns as strong and salient as possible for the sender. First, at the end of each round, senders would communicate their choice to the receiver in front of them. Second, at the end of the experiment, senders would stand up to announce to all participants in the session their choices in the round that was randomly drawn for payment. In particular, they would publicly announce (a) their communication choice (green/red) and (b) their choice between *option 1* and *option 2*. Note that senders did not announce their observed card, which remains private information, as in the model. This procedure was announced in the instructions and applied to the practice round. The treatments are administered between subjects, as within subject variation might lead to spillovers between the treatments due to concerns for consistency or a difficulty to switch reputation concerns on and off.[13]

2.3. *Timing and Procedures*

The timeline of the experiment is illustrated in Figure 2. Upon entering the lab, participants were randomly allocated a seat at a computer terminal. The experimenter read aloud the instructions for the first part of the experiment, which contained information about the activities of GiveDirectly. Subjects then engaged in a social value orientation (SVO) task with GiveDirectly as recipient. The SVO is a standard way to measure social preferences in experimental economics and social psychology (see, e.g., Offerman *et al.*, 1996; Balliet *et al.*, 2009). We use this as a separate measure of the 'type' θ and its distribution π in our theoretical framework, which models the 'altruism' or

[13] Note that, due to the repetition of the game, the sender observes a new signal in every round, which generates random within-subject variation in the information of the sender. However, we do not think of this variation as a 'treatment', because it does not change any parameters of the game but rather follows naturally from the resolution of uncertainty within the game.

'prosocial motivation' toward the third party. In our experiment, this party is represented by the charity GiveDirectly, which is therefore also the recipient in the SVO task. This allows us to test the model's predictions about the relation between prosocial preferences and communication. To make sure the treatment does not affect the SVO task, we conducted it before the main part of the experiment was introduced.

To elicit the SVO, we followed the slider-based SVO design by Murphy *et al.* (2011) and used the program by Crosetto *et al.* (2012), programmed in the software z-Tree (Fischbacher, 2007). The task consists of six consecutive allocation tasks between the participant and the charity, where up to €1 is at stake for both recipient and charity and the possible allocations change in each decision. The instructions in Online Appendix D show an example of one such allocation task. The SVO score is measured as an angle from -16° to 62°, where, roughly speaking, a higher score translates into a higher weight on the charity's payoff (see Murphy *et al.*, 2011 for details of the computation).

After the SVO task, participants played several rounds of the interaction stage, explained above, each time with a different partner. After receiving and reading the instructions about this stage, participants answered a few control questions to test their understanding of the payment scheme. They then learned their role as sender or receiver, and moved to the interaction tables in the lab. Senders and receivers were seated opposite each other. To further raise social pressure to give, each table featured a sheet with a testimonial and photo of a potential recipient, taken from the website of GiveDirectly (see Online Appendix D.3 for an example of the reminder sheet and Online Appendix D.4 for all testimonials). At the beginning of each round, participants were invited to read the testimonial. While senders remained seated, receivers changed table and read a new testimonial each time. Between each sender-receiver pair there was a divider, so participants had no contact with the adjacent pair. Subjects were told that communication in other ways than described in the instructions was not allowed, and would result in exclusion from payment in the experiment. No participant was caught in a violation of these instructions.

Before the first interaction round, participants completed a single practice round to familiarise themselves with the procedures. The interaction stage was not computerised. Both participants in the interaction recorded all choices on a private decision sheet, where a screen on the table made sure their sheet was not visible to the interaction partner. During the practice round, subjects gained experience with filling in the decision sheet. To ensure truthful recording of the communicated card on the decision sheet, subjects were told verbally they would be paid only if both members of the pair recorded the same colour in the appropriate column on the decision sheet. After each interaction round, receivers left their seat and moved one place to the left. After the last interaction round, participants then returned to their cubicle and answered a short questionnaire (see Online Appendix D.2), while the experimenter collected the decision sheets.

The experiment concluded with private payment of the participants in cash. Payment consisted of a show-up payment of €6, one randomly drawn choice (from a total of six choices) in the SVO task, paying between €0.50 and €1, and the earnings from one randomly drawn round in the interaction stage, paying between €0 and €10, with payment based on the decision sheets. The average subject earned €14.40 (minimum €6.50, maximum €17). The money generated for the charity in the selected round was summed up after the experiment and transferred by the experimenter.

Table 1. *Interaction Parameters and Decision Variables in the Model and Their Experimental Implementation.*

Parameter	Definition	Experiment value
$c > 0$	Cost of action/donation	2
$\gamma > 0$	Relative spillover to third party/charity	3
$\overline{\theta} > \underline{\theta} > 0$	Social preferences of high/low type	Measured by SVO
$\pi \in (0, 1)$	Probability of high type	Measured by SVO
$\mu \geq 0$	Importance of image concerns	Varied by treatment
Decision	Definition	Implementation
$\hat{m} \in \{0, 1\}$	Sender's report	Card shown to receiver
$\hat{a}_s, \hat{a}_r \in \{0, 1\}$	Sender/receiver action	Donation decision

2.4. *Hypotheses*

Our experiment directly implements the model in Section 1. In line with our model, signals, messages and decisions are binary. Table 1 provides a full overview of the interaction parameters and decision variables in the model and their experimental implementation.

The close correspondence between experiment and model allows us to derive hypotheses. The model makes predictions about both the communication strategies for each individual across all rounds of the game, as defined in Definition 1, as well as the communication decisions in a given round. There are thus two ways to look at the observed behaviour. To reflect this, we give an overview of the communication strategies and provide (non-parametric) tests where we average individuals' behaviour over rounds. In addition, we analyse subjects' decisions in each individual round, as the theory generates crisp predictions about the decisions that should follow a given signal or report that are straightforward to analyse. This focus also brings additional statistical power, even when controlling for within-subject dependence. We use the following terminology to describe the communication decisions of the sender in a given round. First, the sender is 'honest' in a particular round if she communicates the signal accurately. Second, she 'underreports' if she reports a red card after seeing a green card. Third, she 'overreports' if she reports a green card after seeing a red card. These two ways of looking at the data give a robust view of behaviour in our experiment.[14]

As a point of departure, we are interested in the absolute levels of honesty, under- and overreporting. The theory does not generate detailed hypotheses on these absolute levels, but we can make some informed remarks. First, following our discussion of lying costs above, we know from previous literature that many people are honest even if this is not in their monetary interest. In this experiment, there is no monetary incentive for misreporting, so we would expect substantial amounts of honesty as well. Second, the model predicts that persuasion will occur if image concerns are low (i.e., in the *private* treatment) and agents benefit psychologically from others' donations. Third, in the *private* treatment there is no justification motive, so we do not expect (much) underreporting in this condition.

[14] These communication decisions thus refer to a single decision, rather than a full reporting strategy. This means that we can only distinguish downplaying (exaggerating) impact from honesty in case the sender has seen a green (red) card and reported a red (green) card. We therefore treat instances in which the report matches the card seen as honesty.

Our main interest is how changes in image concerns affect communication. All variations of our model yield unambiguous predictions for the comparative statics of the main treatment manipulation. Raising image concerns in the *public* treatment will increase the justification motive, making the report of a green card more costly.[15]

HYPOTHESIS 1 (THE JUSTIFICATION MOTIVE). *Senders will show fewer green cards in the public treatment than in the private treatment.*

The theory predicts two further behavioural patterns stemming from the trade-off between justification and persuasion. First, justification arises because there is a 'price' on the communication of a green card: the sender is now supposed to act prosocially. Not doing so is considered 'hypocritical', and leads to a low image. Since justification is more important in the *public* treatment, we expect hypocrisy to be lower. Note that this logic holds in all variations of our model, including the case of lying costs and mixed equilibria (Online Appendix B, Proposition 3).

HYPOTHESIS 2 (THE PRICE OF PERSUASION). *Senders are more likely to choose option 1 after showing a green card in the public treatment than in the private treatment.*

Second, the theory predicts that the communication differs by the type θ: the persuasion motive is higher for more enthusiastic givers. Since we measure θ by SVO type, high SVO types are predicted to show more green cards than low SVO types. In particular, Proposition 2 shows that they are less likely to underreport when image concerns are high, while our analysis of lying costs shows that they are also more likely to overreport when image concerns are low. Note that while we do not have experimental variation in the character type θ, these predictions are correlational in nature.

HYPOTHESIS 3 (SORTING). *Senders with a high SVO type are more likely than senders with a low SVO type to show green cards. In particular, senders with a high SVO type are both less likely to underreport in the public treatment and more likely to overreport in the private treatment.*

Our last hypothesis relates to the impact of communication on the receivers. Receivers react to messages only if at least some types tell the truth in equilibrium. This may arise for several reasons. First, for low-image concerns, partial truth telling may arise from the presence of lying costs, as we discussed in Subsection 1.3. However, equilibria arising from such lying costs also feature partial hypocrisy, so high signals are diluted. For high-image concerns, hypocrisy is punished and high signals become more informative, as in the equilibrium described in Proposition 2 or the mixed equilibria discussed in Online Appendix B. Thus, the model predicts that subjects should react to a high signal by taking the prosocial action more often, in particular in the *public* treatment.

HYPOTHESIS 4 (COMMUNICATION IMPACT). *Receivers are more likely to choose option 1 after seeing a green card than after seeing a red card, and more so in the public treatment than in the private treatment.*

[15] To illustrate this, take the parameter values $c = 2$ and $\gamma = 3$. If, e.g., $\underline{\theta} = 1/5$ and $\overline{\theta} \in [2/3, 11/15]$, then our treatment manipulation needs to raise image concerns μ above $5\overline{\theta}(2-\pi)/(5\overline{\theta}-1)$ for the equilibrium with influential communication (Proposition 2) to exist; see Example 1 for details. The bar for image concerns μ is $5\overline{\theta}/(5\overline{\theta}-1)$ and hence lower for a mixed equilibrium with influential communication in which the high-type receiver acts conditional on a high report (Online Appendix B, Proposition 3) to exist. Hence, our model predicts that sufficient image concerns allow influential communication by deterring (most) hypocrisy; only some hypocrisy by high types may persist.

Table 2. *Descriptive Statistics of Choice Variables, Reported as 'Mean (Number of Observations) [SD]' for the Pooled Observations and the Treatments.*

Variable	Definition	Pooled	Private	Public
SVO $\in [-16.3, 61.4]$	Preference for GD	24.1 (227) 16.3	25.9 (111) 15.1	22.4 (116) 17.3
$\hat{m} \in \{0, 1\}$	Sender's report	0.52 (931)	0.57 (445)	0.49 (486)
$\hat{a}_s \in \{0, 1\}$	Sender's donation	0.29 (931)	0.27 (445)	0.31 (486)
$\hat{a}_r \in \{0, 1\}$	Receiver's donation	0.25 (931)	0.26 (445)	0.24 (486)

Table 3. *Fraction of Green Cards Shown by Senders. SDs in Parentheses, Followed by the Number of Total Observations.*

	Private	Public	Total
Green card drawn	0.90	0.79	0.84
	(0.29)	(0.41)	(0.36)
	230	263	493
Red card drawn	0.22	0.13	0.17
	(0.41)	(0.33)	(0.38)
	215	223	438
Total	0.57	0.49	0.52
	(0.50)	(0.50)	(0.50)
	445	486	931

3. Experimental Results

We start our analysis by providing a descriptive overview of the choice variables used in the analysis. Table 2 presents the definitions and summary statistics of these variables for the pooled observations, the *private* and the *public* treatments. Online Appendix Table C.2 provides an overview of the SVO types based on the classification of Murphy *et al.* (2011), as well as a comparison with their findings.

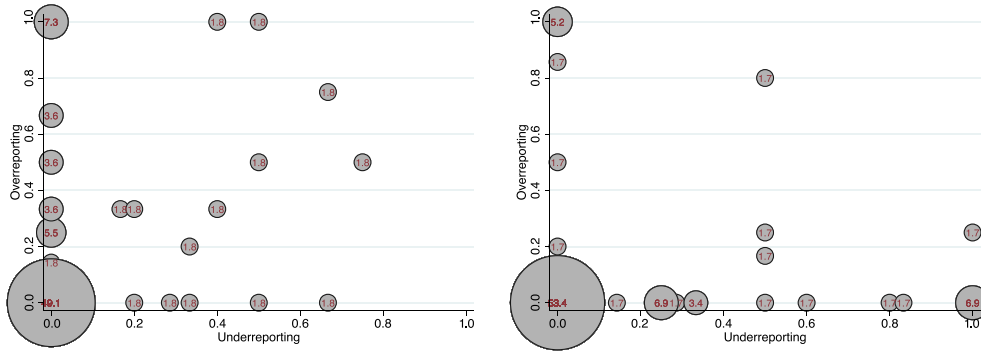
To test our hypotheses below, we provide two kinds of tests. First, we conduct non-parametric tests based on averaging each participant's behaviour across rounds, reflecting the overall strategies of individuals throughout the experiment. Second, we show linear ordinary least square (OLS) regressions with decisions in each round as a unit of observation, and include random effects for individuals to take into account the possible correlation of a subject's behaviour. Our main results are robust to the inclusion of dummies for the individual testimonials, as well as probit model specifications.[16] Unless otherwise stated, all our statistical tests are two sided. Online Appendix C provides additional figures and analysis.[17]

3.1. *Communication Across Treatments*

We start by analysing the composition of reported cards across treatments. Table 3 provides details on the senders' communication. It shows descriptive statistics of the fraction of green

[16] This statement refers to the sign and significance of the coefficients in the probit model. We have not computed marginal effects, as some of our hypotheses concern interaction terms that are notoriously difficult to compute or even define for non-linear models (Greene, 2010).

[17] Throughout the analysis, we excluded observations from one participant in the role of sender who did not follow the instructions. The participant could not answer the questions to check understanding. In the role of sender, the subject drew multiple cards from the deck more than once, invalidating the signal. In addition, we excluded data from two interactions where the sender's and receiver's report of the communicated message did not match, so we do not know which message was communicated.



(a) *Private* treatment (b) *Public* treatment

Fig. 3. *Overview of Communication Patterns.*

Notes: The *x* axis shows the level of underreporting: the fraction of green cards that are misreported as a red card. The *y* axis shows the level of overreporting: the fraction of red cards that are misreported as a green card. A person who is honest in all rounds is located at the origin. The size of the observation marker reflects the percentage of total observations in the treatment, which is stated inside the marker. The number of observations is 58 in the *public* treatment and 56 in the *private* treatment.

cards shown split by treatment and by the colour of the card drawn by the sender. Senders communicate more green cards after drawing a green card, showing high levels of honesty and modest levels of over- and underreporting of about 13% and 10% of reports, respectively. More green cards are shown in the *private* treatment, regardless of which card was drawn. Figure C.1 in the Online Appendix provides a further overview of the number of sender observations across treatments and choices.

Table 3 provides an overview of the individual decisions only. Instead, Figure 3 provides a more general overview of the communication patterns in the two treatments, where we represent each participant as an individual data point. On the horizontal axis, we plot the level of underreporting—the fraction of green cards that are misrepresented as red cards (the underreporting or 'justification' axis), and on the vertical axis we plot the level of overreporting—the fraction of red cards misrepresented as green cards (the overreporting or 'persuasion' axis). Thus, an honest individual is located at the south-west corner, someone who always overreports and never underreports is in the north-west corner, whereas someone who always underreports and never overreports is in the south-east corner.[18] The size of the observation marker reflects the percentage of total observations in the treatment, which is stated inside the marker.

Figure 3 shows that about half of the participants in both treatments is always honest. Given the clear correspondence between the signal and report space, these senders likely perceived inconsistency as a form of lying, which they wanted to avoid. The questionnaire provides evidence of this: fourteen senders explicitly motivate their reporting strategies using normative words like 'truthful' or 'honest' (see Subsection 3.5 for more detail). Many others write simply that

[18] If we assume that participants did not experiment with different strategies, we can interpret the level of overreporting (underreporting) of a participant located on the vertical (horizontal) axis as the probability of exaggeration (downplaying) in a mixed strategy (conditional on type) between exaggeration (downplaying) and honesty. In particular, the behaviour of individuals located in the South-West, North-West and South-East corner is consistent with the pure strategy of honesty, exaggeration and downplaying, respectively (see Definition 1).

Table 4. *Regressions of Green Card Shown on Treatment and Card Drawn.*

	(1) Green reported	(2) Green reported
Public treatment (PT)	−0.103***	−0.114**
	(0.0399)	(0.0455)
Green card drawn (GCD)	0.680***	0.670***
	(0.0218)	(0.0313)
PT × GCD		0.0200
		(0.0436)
Constant	0.217***	0.223***
	(0.0307)	(0.0325)
Observations	931	931
Overall R^2	0.463	0.463

Notes: Results are from a linear probability model with individual random effects. SEs in parentheses. $^{**}p < 0.05$, $^{***}p < 0.01$.

they reported the card they saw, which is consistent with striving for either consistency or honesty.

Of the remaining half of subjects, there is a shift from overreporting to underreporting between the *private* and *public* treatments. In the *private* treatment 42% overreport at least once, and 11% always do so, whereas in the *public* treatment these numbers are 18% and 5%, respectively. When it comes to underreporting, 25% do so at least once and no subject always does so in the *private* treatment, while in the *public* treatment these numbers are 36% and 9%, respectively.

There are several ways to statistically compare the two-dimensional distributions in Figure 3. The most straightforward way is to test Hypothesis 1 and ask how many green cards senders reported in each treatment; see Table 3. For an adequate test, we should control for the amount of green cards drawn in each treatment, as there turned out to be a slightly higher (but not statistically significant) proportion especially in the *public* treatment.[19] Regression analysis allows us to estimate the effect of the treatment while controlling for the colour of the card drawn.

Table 4 shows the result of such a multivariate analysis, using a linear probability model with random effects for the sender's ID to take into account the dependence between an individual sender's choices. The results show that drawing a green signal makes a sender about 68 percentage points more likely to show a green card, which is compatible with the observation that senders are honest most of the time. In line with Hypothesis 1, the *public* treatment reduces the likelihood of showing a green card by about 10 percentage points, a result that is significant at the 1% level.

Column (2) adds an interaction term for the treatment and the colour of the card drawn by the sender. This shows that the decline in green cards shown is smaller after the sender drew a green card, although the coefficient is not statistically significant. To directly evaluate the effect of the treatment after drawing a green card, we perform a Wald test for the significance of the sum of the coefficients of the treatment dummy and interaction term in column (2), which is significant at the 5% level ($p = 0.035$). In Online Appendix C.4, we provide a more detailed breakdown of the statistical results separated by the colour of the card drawn by the sender.

A more conservative statistical approach is to compare both reporting distributions non-parametrically. To do so, we summarise the communication into a single dimension in a way that

[19] These differences are not significant at the 10% level on a Fisher exact test ($p = 0.47$). As subjects had typically picked the upper card from the deck, the asymmetry may have been caused by imperfect shuffling of the card deck.

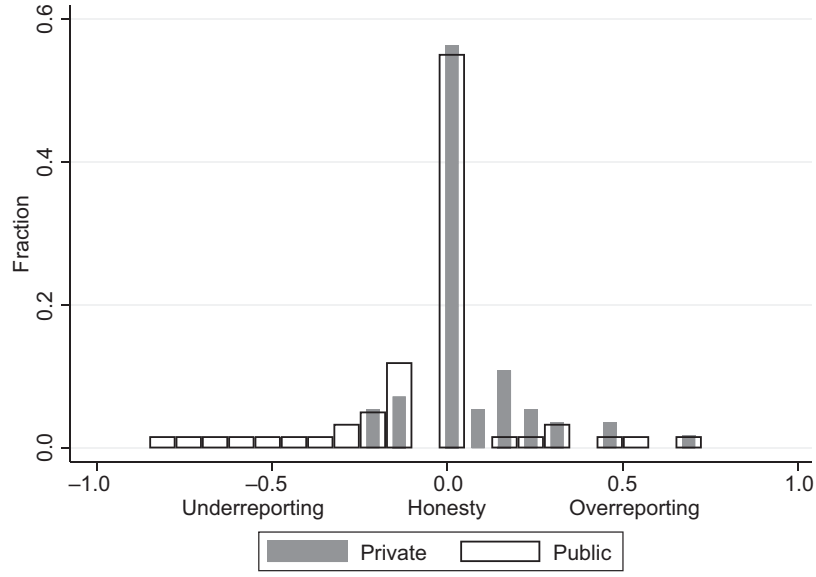


Fig. 4. *One-dimensional Communication Patterns.*
Notes: An individual observation is the average communication of an individual over the interaction rounds, where honesty is coded as 0, overreporting as 1 and underreporting as −1.

controls for the dependence in observations for a given sender and the amount of green cards drawn. We first code an honest representation as 0, overreporting as 1 and underreporting as −1. We then average these numbers over rounds to generate an individual-specific communication score. The higher the score, the more overreporting by the individual.

Figure 4 shows the resulting distribution of scores. We see a spike at 0, which includes honest subjects as well as a few subjects whose over- and underreporting exactly offset each other. The remaining subjects show a shift towards more negative scores (underreporting) in the *public* treatment. The two distributions differ significantly on a Wilcoxon rank-sum test ($p = 0.0025$, two sided). Thus, senders in the *public* treatment do indeed act as if they are in need of justification.

SUMMARY 1. *We find subjects are honest most of the time, with under- and overreporting making up a modest 10% and 13% of reports, respectively. On the individual level, about half of the subjects over- or underreport at least once. In line with Hypothesis 1, senders are about 10 percentage points less likely to show a green card in the public treatment. This is due to both a drop in overreporting and a rise in underreporting, with the former being more pronounced.*

3.2. *The Price of Persuasion*

We now look deeper into the reasons behind the treatment difference in communication patterns. As explained in Subsection 2.4, the logic of the justification motive is that reporting a green card has a 'price'. The visibility of her actions in the *public* treatment forces the sender to either incur the cost of a donation (choosing *option 1*), or to reveal that she is not motivated to donate despite a high return of the donation (choosing *option 2*), i.e., appearing 'hypocritical' in the terminology

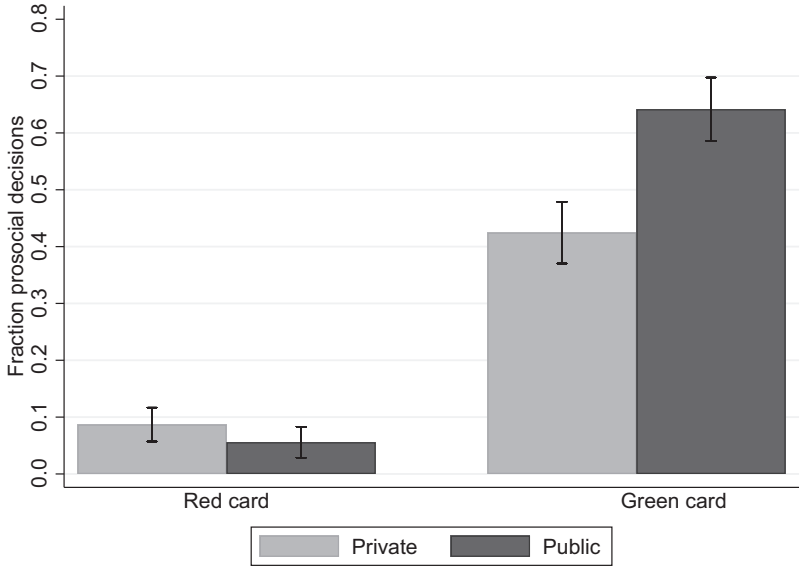


Fig. 5. *Relation Between Reports and Prosocial Actions.*
Notes: The fraction of senders' prosocial decisions by card shown and treatment. One observation is the fraction of prosocial decisions for each individual. Bars show SEs.

of Section 1. Hypothesis 2 thus specifies that in the *public* treatment, more senders will follow the report of a green card with a donation.

Figure 5 shows the rates of prosocial behaviour (choosing *option 1*) among senders in both treatments, after reporting either a red or a green card. The fraction of prosocial decisions for each individual constitutes one observation. Figure 5 shows that in either treatment, few senders donate after showing a red signal, but much more so after they show a green card. Moreover, in line with Hypothesis 2, they do so more often in the *public* treatment (64% on average) than in the *private* treatment (42%), a difference which is significant on a Wilcoxon rank-sum test ($p = 0.0050$, two sided). Thus, senders in the *public* treatment do indeed act as if they face a higher cost of showing 'frivolous' green cards.

The corresponding regression results are reported in Table 5, which shows OLS panel regressions of sender prosocial behaviour on a dummy for the treatment and the card shown by the sender. The models used are the same as those in Table 4. The positive and highly significant interaction term in column (2) confirms that prosocial behaviour after showing a green card is higher in the *public* treatment.

SUMMARY 2. *In line with Hypothesis 2, senders in the public treatment are more likely to follow up a green card with a donation.*

3.3. *Sender Type and Misreporting*

The model also makes predictions about the relation between the sender's prosocial preferences and her communication choices. While the analysis above already provides a correlation between prosocial behaviour and the report of green cards, it is based on prosocial behaviour that is

Table 5. *Regressions of Sender Prosocial Behaviour on Treatment and Card Shown.*

	(1) Prosocial choice	(2) Prosocial choice
Public treatment	0.0854*	−0.0506
	(0.0511)	(0.0564)
Green card reported	0.481***	0.349***
	(0.0226)	(0.0318)
PT × GCR		0.259***
		(0.0443)
Constant	−0.00333	0.0717*
	(0.0388)	(0.0410)
Observations	931	931
Overall R^2	0.233	0.246

Notes: Results are from a linear probability model with individual random effects. SEs in parentheses. $^{*}p < 0.10$, $^{***}p < 0.01$.

measured post-treatment. To provide a cleaner test of the theoretical predictions formulated in Hypothesis 3, we correlate communication behaviour with our pre-treatment measure of SVO. Because the raw SVO angle does not have intuitive units, we create a 'low' and a 'high' SVO group of roughly equal size based on the classification in Murphy *et al.* (2011). Murphy *et al.* (2011) proposed thresholds for SVO values to classify individuals as 'competitive', 'individualistic', 'prosocial' and 'altruistic'. To create a binary distinction, we merge the first two and the last two categories. We refer to those categories of subjects as 'low SVO type' and 'high SVO type', respectively. The results also hold when we use correlations with the raw SVO scores instead.

Figure 6 shows the average fraction of green cards shown for individual senders in both treatments, split by SVO type and card drawn. The overall difference between the high SVO and low SVO types, shown in the two leftmost bars, goes in the hypothesised direction. However, it is small and not statistically significant (53% versus 48%, Wilcoxon rank-sum test, $p = 0.49$). When we focus on the senders who drew a green card, high-SVO-type senders on average underreport a lower fraction of green cards than low-SVO-type senders (91% versus 75%, $p = 0.043$). In addition, the rightmost bars show that high-SVO-type senders are also less likely to overreport after seeing a red card (20% versus 14%, $p = 0.56$). Although this last difference is small and not statistically significant, it contradicts Hypothesis 3 and explains why the reporting of green cards overall does not differ much between types.

Turning to the parametric results, Table 6 shows OLS panel regressions of green cards shown on a dummy for the SVO classification of the sender, controlling for the colour of the card drawn. Columns (1) and (2) confirm the visual and non-parametric results. In particular, high SVO types are about 8 percentage points less likely to overreport, while the interaction between green card drawn and high SVO type is large and positive, indicating a drop in underreporting of about 25 percentage points.

Columns (3)–(4) and (5)–(6) show the results in isolation for the *private* and *public* treatments, respectively. This shows that the effect of SVO on overreporting and underreporting is substantial and statistically significant only in the *public* treatment. This provides mixed evidence for Hypothesis 3, which states that high SVO types would overreport more in the *private* treatment, and underreport less in the *public* treatment. The regressions confirm the latter but not the former

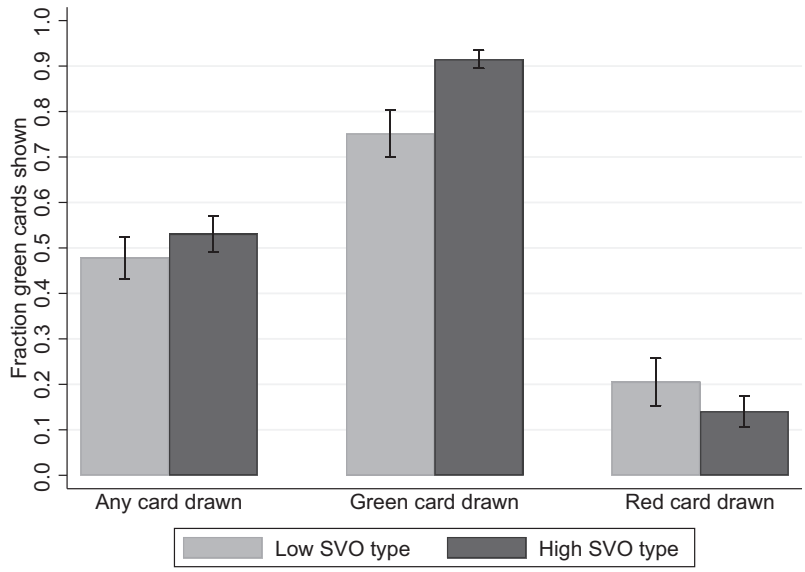


Fig. 6. *Relation Between Reporting and SVO Type.*
Notes: The fraction of senders' green cards shown by SVO type and card drawn. One observation is the fraction of green cards shown for each individual. Bars show SEs.

Table 6. *Regressions of Green Cards Reported on SVO Type.*

	(1) All data	(2) All data	(3) Private	(4) Private	(5) Public	(6) Public
Green card drawn (GCD)	0.679***	0.536***	0.675***	0.629***	0.693***	0.489***
	(0.0218)	(0.0327)	(0.0326)	(0.0577)	(0.0292)	(0.0373)
High SVO type (HST)	0.0500	−0.0813*	0.00380	−0.0308	0.0472	−0.194***
	(0.0411)	(0.0455)	(0.0487)	(0.0592)	(0.0637)	(0.0682)
GCD × HST		0.252***		0.0688		0.442***
		(0.0434)		(0.0700)		(0.0549)
Constant	0.136***	0.208***	0.219***	0.242***	0.0852*	0.188***
	(0.0330)	(0.0341)	(0.0433)	(0.0485)	(0.0459)	(0.0459)
Observations	931	931	445	445	486	486
Overall R^2	0.454	0.475	0.480	0.482	0.442	0.499

Notes: Results are from a linear probability model with individual random effects. SEs in parentheses. $^{*}p < 0.10$, $^{***}p < 0.01$.

hypothesis. In Subsection 3.5, we discuss potential reasons for the violations of part of Hypothesis 3, in particular the reasons why high SVO types seem to overreport less than low SVO types.

SUMMARY 3. *We find mixed evidence for Hypothesis 3. High-SVO-type senders are only slightly more likely to show green cards and the result is not statistically significant. This result arises because, contrary to the hypothesis, high-SVO-type senders are less likely to overreport. However, in line with the hypothesis, we find strong evidence that high SVO types are less likely to underreport in the public treatment.*

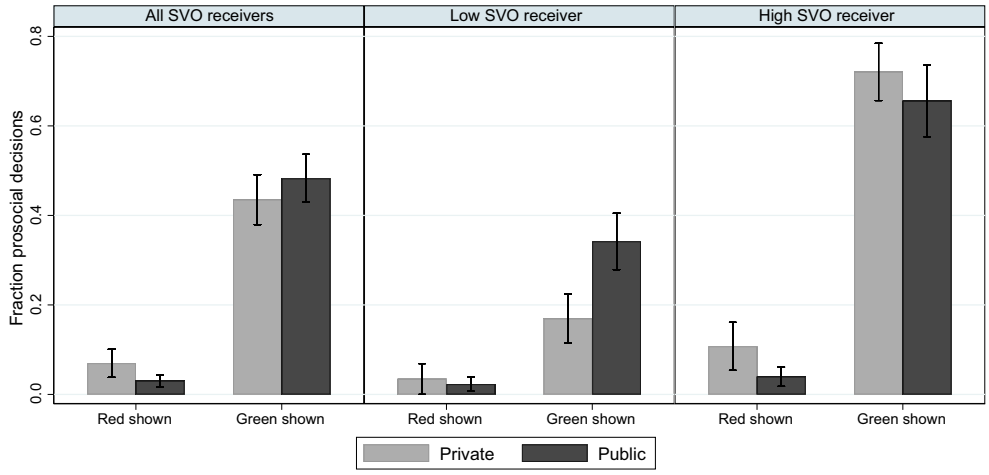


Fig. 7. *Receivers' Prosocial Decisions, Split by Card Seen, Treatment and SVO Type.*
Notes: One observation is the fraction of prosocial decisions for each individual; bars show SEs. The left panel shows all data, the middle panel shows the patterns for 'individualistic' and 'competitive' individuals and the right panel for 'prosocial' and 'altruistic' individuals, using the SVO typology in Murphy *et al.* (2011).

In Subsection 3.5, we discuss in more detail why high-SVO-type senders are less likely to overreport, including potential confounds like lying costs and other measurement problems related to the SVO task.

3.4. *Impact of Communication on Receivers*

We now turn to our final hypothesis, and ask whether communication matters for the behaviour of receivers. To answer this question, we can simply compare the behaviour of receivers who have seen a green card with those who have seen a red card. The left panel of Figure 7 does just that, by showing the proportion of prosocial decisions for receivers, split both by treatment and by the card seen. In line with Hypothesis 4, receivers are about 40 percentage points more likely to donate after being shown a green card, as measured within subject (Wilcoxon signed-rank test, $p < 0.001$).

The left panel also shows that this tendency is slightly more pronounced in the *public* treatment, in line with the second part of Hypothesis 4. This makes sense given that observing a green card is slightly more informative in this condition. However, the effect is small, and the difference-in-difference comparison is not statistically significant (Wilcoxon rank-sum test, $p = 0.28$). To better understand this result, we split it by the SVO type of the receiver. The middle panel of Figure 7 shows the behaviour of low SVO types, who are in fact more likely to respond to communication of high impact in the *public* treatment, a result that is statistically significant (difference-in-difference, Wilcoxon rank-sum test, $p = 0.033$). By contrast, the high SVO types (right panel) react similarly in both treatments.

To check the robustness of these results, Table 7 shows the result of a linear regression analysis of receiver prosocial behaviour on a dummy for the treatment and green card reported. The first two columns show all data, whereas columns (3)–(4) and (5)–(6) show data for the low-SVO-

Table 7. *Regressions of Receiver Prosocial Behaviour on Treatment and Green Card Reported.*

	(1) All data	(2) All data	(3) Low SVO	(4) Low SVO	(5) High SVO	(6) High SVO
Public treatment	0.00861	−0.0302	0.0940*	−0.0187	−0.0650	−0.0462
	(0.0459)	(0.0516)	(0.0547)	(0.0612)	(0.0592)	(0.0689)
Green card reported	0.400***	0.361***	0.224***	0.108***	0.602***	0.620***
	(0.0225)	(0.0326)	(0.0275)	(0.0404)	(0.0336)	(0.0474)
PT × GCR		0.0738		0.210***		−0.0367
		(0.0450)		(0.0546)		(0.0671)
Constant	0.0397	0.0617	−0.0153	0.0502	0.0916**	0.0812
	(0.0352)	(0.0376)	(0.0427)	(0.0453)	(0.0456)	(0.0497)
Observations	931	931	501	501	430	430
Overall R^2	0.196	0.198	0.105	0.130	0.387	0.387

Notes: Results are from a linear probability model with individual random effects. SEs in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

and high-SVO-type receivers, respectively. Column (1) shows that there is a strong effect of communication of high impact, while column (2) shows that there is no statistically significant interaction effect, confirming the non-parametric results. Furthermore, columns (4) and (6) show that there is a highly significant interaction effect for low SVO but not for high SVO types.

One explanation for the interaction effect among low SVO types is that because lower SVO types are less motivated to give, they can be persuaded to do so only by the more informative report in the *public* treatment. By contrast, high SVO types are motivated to contribute in either case, despite a lower quality report in the *private* treatment. This may be because GiveDirectly is seen as a high-quality and uncontroversial charity, or because the high SVO types may underestimate the tendency for subjects to overreport. This latter explanation is plausible, since high SVO subjects in the role of sender do not engage in such overreporting. Clearly, these are post-hoc rationalisations in need of further confirmation in future research.

SUMMARY 4. *In line with Hypothesis 4, the colour of the card shown by the sender has a big impact on giving rates of receivers. This impact of a green card is higher in the public treatment overall, but the effect is not statistically significant and appears to be pronounced only among less motivated givers.*

In the discussion section, we further comment on how the treatment affects overall donations and charity receipts, showing that its positive impact on senders is offset by the negative impact on receivers, through the change in communication.

3.5. *Discussion*

In line with our hypotheses, raising image concerns induces subjects to report a positive impact of a donation less often; exaggeration largely disappears and some participants even downplay impact. Moreover, those subjects who report a positive impact follow up on their report with a donation more often. However, there are also discrepancies between our results and hypotheses that we discuss below.

3.5.1. *Underreporting*

We first discuss the incidences of underreporting in the *private* treatment. The motives for this behaviour are unclear. It cannot be explained by image concerns, as these are ruled out by design. We also do not find clear patterns between a sender's underreporting, his/her SVO score and the

rating of the 'deservingness' of the charity (rated during the questionnaire by each subject on a 10 point scale). Rather, a part of the explanation may be that some subjects reported somewhat carelessly in the *private* condition. Three subjects write in the questionnaire that they reported 'randomly', while one writes that the only round where (s)he underreported was a 'mistake'. Since no person engaged in this communication consistently, it thus seems unwise to seek too much behind it.

3.5.2. *Overreporting*

Second, we turn to the interpretation of overreporting. The questionnaire provides anecdotal support for the idea that overreporting is related to the persuasion motive: five participants who always chose to show the green card state explicitly that their aim was to persuade the receiver to benefit the charity, even if they did not do so themselves. Nevertheless, we do find some deviations from the theoretical prediction on overreporting. In particular, we do not confirm part of Hypothesis 3, that overreporting occurs among senders who have a high utility weight on the income of the charity. In Subsection 3.3 we even find some evidence for the opposite relation. To further investigate the relation between overreporting and preferences towards the charity, we look at the fraction of senders who follow up on their overreport with a donation. Across treatments, only 42% follow up with a donation at least once, and no sender always follows up. This casts further doubt on the idea that overreporting is driven by strong concerns for the charity.

The results on SVO are correlational, as it is hard to change subject's social preferences experimentally. It is thus possible that the SVO is confounded with other traits that matter for behaviour in the experiment. One possibility is that lying costs are higher among high SVO types, explaining why they are more likely to tell the truth, even after a green signal. Lying costs could also be driven by the weight on the receiver's payoffs, which we did not measure in this experiment. Because we do not have an independent measure of lying costs, we cannot conduct a sharp test of this idea. To provide some anecdotal evidence, we look at whether high SVO types are more likely to invoke normative explanations in the final questionnaire when asked to describe their reporting strategy.[20] We find fourteen such senders, interestingly enough, twelve of them in the *private* treatment. We count nine incidences among high-SVO-type senders, and five among low-SVO-type senders, providing some evidence for the idea that high SVO types are more lying averse.

An alternative motive behind overreporting is that making the receiver donate may generate some form of 'warm glow', perhaps by reducing guilt for a sender who does not want to donate herself. A closely related idea is that of 'moral licensing', where the sender feels 'licensed' not to contribute herself because she induces contributions by others, essentially considering contributions by the self and others as substitutes, possibly even across rounds. These explanations differ from the hypothesised one in that warm glow derives not from the effect of the donation, which is likely to be low, but from the act of inducing others to donate. Such warm glow may not be fully captured by the SVO task, which measures the sender's willingness to donate with his or her *own* money. A better measure may be whether the sender thinks the charity is generally doing good work and is deserving of donations by others.

To investigate this explanation, we look at the correlation between the sender's average overreporting and his or her rating of the 'deservingness' of the charity. We find a modest but significant

[20] We limit our search among senders who are always honest, and mark them when they use one of the following words: 'honest', 'honesty', 'correct', 'truth', 'truthful', 'deceive', 'false' or 'wrong', where the last three are used as a contrast with the actual choice.

correlation (Pearson $\rho = 0.20$, $p = 0.033$). By contrast, correlations of deservingness with individual measures of honesty and underreporting are negative and not statistically significant. In line with the idea that persuasion helps assuage guilt about a lack of personal contributions, the correlation is especially high among those who never follow up their overreport with a donation (Pearson $\rho = 0.52$, $p = 0.021$).

In summary, we find anecdotal evidence to explain why overreporting occurs more among low SVO subjects: subjects who are less prosocial towards the charity have lower lying costs, and subjects who think the charity is deserving of donations but do not want to donate themselves may derive utility from making others contribute. We cannot investigate these explanations further here, but encourage further research to understand the drivers behind the exaggeration of impact.

3.5.3. *Total giving by treatment*

From a policy perspective, one may wonder whether raising image concerns is a good or a bad thing for the charity. To evaluate this question, we look at the effect of raising image concerns on giving. In line with our other results and the theory, we find that the answer depends on the subjects' role: senders' donations increase by 17% in the *public* treatment, as their generosity is now advertised to others. By contrast, receivers' donations decrease by 10%, because they now see more red cards. In aggregation, there remains only a small and not statistically significant positive treatment effect of about 4%, with overall donations rising from 26.5% to 27.6% of decisions. Online Appendix C.5 provides a more detailed breakdown of these results by subjects' roles, states of the world and SVO values.

Making senders' donations public thus has very little effect on overall giving in our experiment, because its positive impact on senders is offset by the negative impact on receivers, through the change in communication. By contrast, the literature on audience effects in prosocial behaviour has generally emphasised a robust increase in prosocial behaviour (Bursztyn and Jensen, 2017). Our results thus provide a qualification to this finding in cases where communication is important.

3.5.4. *Noise*

One concern is that responses in the experiment may be noisy. Note that subjects completed a series of control questions and engaged in a practice round, which should reduce noise. We also asked subjects in the exit questionnaire if they found the instructions were clear. Just over 90% of subjects confirm this explicitly, although some say the practice round was necessary to fully understand the interaction phase. The results are robust to excluding the 10% of subjects who express some kind of doubt. Thus, we do not believe that noise is an important driver of the results.

4. Conclusion

We investigate communication about the impact of personally costly actions on third parties. Our model predicts that concern for the third party introduces a persuasion motive to exaggerate impact. The wish to maintain a good reputation introduces a motive to justify one's actions. This justification motive facilitates truthful communication by putting a price on exaggerated reports that are not followed by costly donations. However, it also introduces an incentive to downplay impact in order to cast doubt on the effectiveness of giving and excuse inaction. This depresses prosocial behaviour by receivers of the information.

In an experiment in a charitable giving context, we find that although half of the subjects are always honest, communication among the remaining half is systematically distorted: in

situations where actions are not observable and talk is cheap, participants overreport impact to persuade others to take the action. When talk is not cheap because donations are made observable, exaggeration is reduced and some participants underreport impact, reducing giving among receivers in the process. These effects of increased observability occur despite the fact that senders derive no direct material gain from miscommunication.

While our experiment took place within the laboratory, there are several reasons to think that persuasion and justification motives will affect communication behaviour in other contexts as well. First, the signal in our experiment is relatively unambiguous. To the extent that signals outside the lab are multi-interpretable, people can more easily convince *themselves* of the truth of their misrepresentations. This is in line with a well-documented tendency for people to form self-serving beliefs that rationalise self-interested decisions (Kunda, 1990; Exley, 2016; Gino *et al.*, 2016). Thus, our results may even apply to an intra-personal communication game, such as proposed in the self-signalling literature, where the sender and receiver reside in the same person (Bénabou and Tirole, 2011).

Second, the communication in our experiment was highly stylised, consisting only of binary signals. Natural languages offer much richer shades of persuasion and deception. Both these features may decrease lying costs and increase strategic communication. Third, image concerns in the laboratory are likely to be limited, as participants interact with strangers. In less artificial environments, like online social networks, the motive for sharing exculpatory content may be stronger. Politics is another area where reputation is paramount; politicians may downplay the impact of social actions to excuse both their own inaction and that of their voters. Partisan sorting in political and online environments may also play a role, putting pressure on individuals to align with prevailing ingroup ideas, which introduce additional strategic motives for misrepresentation that may reinforce or counter those identified here.

Future research should determine the applicability of our results outside of the laboratory, as well as in other applications than charitable giving. Generally, Bolderdijk *et al.* (2017) argued that 'effectiveness skepticism' arises for policies that consumers consider personally unattractive. One promising application is climate change. The oil industry has been involved in a well-documented effort to manufacture doubt and obfuscate the impact of their products through funding of contrarian 'research' on climate change (e.g., Conway and Oreskes, 2011). There is some evidence of similar tendencies among consumers, who engage in denial in order to avoid changes to their lifestyle (Stoll-Kleemann *et al.*, 2001; Norgaard, 2006). Thus, the justification motive may help explain why substantial minorities in many countries do not believe in the scientific consensus on climate change. Future research could also explore how different contextual factors affect strategic communication. For instance, unlike charitable giving, the climate change context involves direct externalities among citizens, which may increase the persuasion motive (see Foerster and van der Weele, 2018).

Appendix A. Proofs of Proposition 1 and Proposition 2

We first determine possible equilibrium candidates with influential communication. Assumption 1(*i*) implies that low-type senders and receivers do not take action $\hat{a}_s = 1$. Furthermore, by Assumption 1(*ii*), $\overline{\theta} - 2\underline{\theta} \leq 1/\gamma \Leftrightarrow (1 + \underline{\theta}\gamma)E[W \mid \sigma = 1] \geq (1 + \overline{\theta}\gamma)E[W \mid \sigma = 0]$, i.e., also high-type senders with $\sigma = 0$ do not take action $\hat{a}_s = 1$. By Lemma 1, the high-type receiver acts conditional on a high report, which implies that the high-type sender acts conditional on a high signal, as the assigned image increases in the action. Moreover, influential communication

requires that at least one sender type is honest. Note that, conditional on the signal, the high-type sender will submit a weakly higher report than the low-type sender in equilibrium. This implies that there are three equilibrium candidates.

(*i*) Suppose that the high type is honest and that the low type downplays impact. Then contributing, conditional on a high report, $a_r^*(\overline{\theta}, \hat{m}) = \hat{m}$, is incentive compatible for the high-type receiver if and only if

$$\begin{aligned}
& (1+\overline{\theta}\gamma)E[W \mid \hat{m}=1] \geq c \geq (1+\overline{\theta}\gamma)E[W \mid \hat{m}=0] \\
\iff \quad & (1+\overline{\theta}\gamma)E[W \mid \sigma=1] \\
& \quad \geq c \\
& \quad \geq (1+\overline{\theta}\gamma)\frac{\Pr(\sigma=0)E[W \mid \sigma=0]+\Pr(\sigma=1)(1-\pi)E[W \mid \sigma=1]}{\Pr(\sigma=0)+\Pr(\sigma=1)(1-\pi)} \\
\iff \quad & \frac{2(1+\overline{\theta}\gamma)}{3} \geq c \geq (1+\overline{\theta}\gamma)\frac{3-2\pi}{3(2-\pi)}, \qquad \text{(A1)}
\end{aligned}$$

which are the desired bounds on c. Second, consider the second stage and let $U_s^*(m, a_s \mid \theta_s, \sigma) \equiv E_s[u_s(\theta_s, \sigma, m, a_s, a_r^*) \mid \theta_s, \sigma]$ denote the sender's expected utility from strategy (m, a_s) conditional on her type θ_s and signal σ and the receiver's strategy a_r^*. Note that (not) contributing after submitting a low (high) report is off equilibrium. We assume that this induces a high (low) belief of the receiver about the sender's type, $E_r[\theta_s \mid \hat{m}=0, \hat{a}_s=1] = \overline{\theta}$ ($E_r[\theta_s \mid \hat{m}=1, \hat{a}_s=0] = \underline{\theta}$). Condition (A1) implies that a high-type sender with $\sigma = 1$ does not have incentives to take action $\hat{a}_s = 0$ regardless of the submitted report $\hat{m} \in \{0, 1\}$, as the assigned image increases in the action. Furthermore, recall that low-type senders and a high-type sender with $\sigma = 0$ have no incentives to take action $\hat{a}_s = 1$.

Next, consider the first stage and take a low type. As deviations that involve an action are ruled out, the only possible deviation is submitting a high report and not taking the action, $(m', a_i') = (1, 0)$. Recall that this strategy yields a low image, while submitting a low report and not taking the action yields an image of

$$\begin{aligned}
E_r[\theta_s \mid \hat{m}=0, \hat{a}_s=0] &= \frac{\Pr(\sigma=0)(\pi\overline{\theta}+(1-\pi)\underline{\theta})+\Pr(\sigma=1)(1-\pi)\underline{\theta}}{\Pr(\sigma=0)+\Pr(\sigma=1)(1-\pi)} \\
&= \frac{\pi\overline{\theta}+2(1-\pi)\underline{\theta}}{2-\pi}.
\end{aligned}$$

The low-type sender with signal $\sigma \in \{0, 1\}$ does not have incentives to do this deviation if and only if

$$\begin{aligned}
& U_s^*(m^*, a_i^* \mid \theta_s, \sigma) \geq U_s^*(m', a_i' \mid \theta_s, \sigma) \\
\iff \quad & \mu E_r[\theta_s \mid \hat{m}=0, \hat{a}_s=0] \geq \pi\underline{\theta}\gamma E[W \mid \sigma] + \mu\underline{\theta} \\
\iff \quad & \mu\frac{\pi\overline{\theta}+2(1-\pi)\underline{\theta}}{2-\pi} \geq \frac{\pi\underline{\theta}\gamma(1+\sigma)}{3} + \mu\underline{\theta} \\
\iff \quad & \mu \geq \frac{\underline{\theta}\gamma(1+\sigma)(2-\pi)}{3(\overline{\theta}-\underline{\theta})}, \qquad \text{(A2)}
\end{aligned}$$

which is the first desired lower bound on μ. Since, for the high-type sender with $\sigma = 0$, taking the action is ruled out, the only possible deviation is to $(m', a_i') = (1, 0)$. She has no incentives to do this deviation if and only if

$$\begin{aligned} & U_s^*(m^*, a_i^* \mid \theta_s, \sigma) \geq U_s^*(m', a_i' \mid \theta_s, \sigma) \\ \iff \quad & \mu \frac{\pi\overline{\theta} + 2(1-\pi)\underline{\theta}}{2-\pi} \geq \pi\overline{\theta}\gamma E[W \mid \sigma] + \mu\underline{\theta} \\ \iff \quad & \mu \geq \frac{\overline{\theta}\gamma(2-\pi)}{3(\overline{\theta}-\underline{\theta})}, \end{aligned} \tag{A3}$$

which is the second desired lower bound on μ. Finally, if $\sigma = 1$ then the high-type sender will take the action in any case. As downplaying is ruled out because it only lowers actions by the receiver compared to truthful reporting, the sender does not have incentives to deviate. Hence, we have established that neither the receiver nor the sender (for any type-signal pair) has incentives to deviate if Assumption 1, (A1), (A2) and (A3) hold. Moreover, these conditions are also necessary, as (A2) and (A3) tighten if we take other assumptions on off-equilibrium beliefs (hypocrisy becomes more profitable if it yields a higher image).

(*ii*) Suppose that both types are honest. Then the incentive compatibility condition for the receiver is given by

$$2(1 + \overline{\theta}\gamma)/3 \geq c. \tag{A4}$$

Next, consider the first stage and take a high type with $\sigma = 0$. Consider a deviation to exaggeration and not taking the action, $(m', a_i') = (1, 0)$. Note that this strategy yields a low image, while submitting a low report and not taking the action yields the prior image $\pi\overline{\theta} + (1-\pi)\underline{\theta}$. The sender does not have incentives to do this deviation if and only if

$$\begin{aligned} & U_s^*(m^*, a_i^* \mid \theta_s, \sigma) \geq U_s^*(m', a_i' \mid \theta_s, \sigma) \\ \iff \quad & \mu(\pi\overline{\theta} + (1-\pi)\underline{\theta}) \geq \pi\overline{\theta}\gamma E[W \mid \sigma] + \mu\underline{\theta} \\ \iff \quad & \mu \geq \frac{\overline{\theta}\gamma}{3(\overline{\theta}-\underline{\theta})}. \end{aligned} \tag{A5}$$

Finally, take a low type with $\sigma = 1$ and consider a deviation to downplaying and not taking the action, $(m', a_i') = (0, 0)$. The sender does not have incentives to do this deviation if and only if

$$\begin{aligned} & U_s^*(m^*, a_i^* \mid \theta_s, \sigma) \geq U_s^*(m', a_i' \mid \theta_s, \sigma) \\ \iff \quad & \pi\underline{\theta}\gamma E[W \mid \sigma] + \mu\underline{\theta} \geq \mu(\pi\overline{\theta} + (1-\pi)\underline{\theta}) \\ \iff \quad & \frac{2\underline{\theta}\gamma}{3(\overline{\theta}-\underline{\theta})} \geq \mu. \end{aligned} \tag{A6}$$

Note that (A5) and (A6) imply that $2\underline{\theta} \geq \overline{\theta}$. Moreover, Assumption 1(i) and (A4) imply that

$$\mu < \frac{3c - 2(1 + \underline{\theta}\gamma)}{3(\overline{\theta} - \underline{\theta})} \leq \frac{2\gamma(\overline{\theta} - \underline{\theta})}{3(\overline{\theta} - \underline{\theta})} = \frac{2\gamma}{3},$$

which, together with (A5), implies that $\overline{\theta} > 2\underline{\theta}$, i.e., this strategy profile cannot be an equilibrium under Assumption 1.

(iii) Suppose that the high type exaggerates impact and that the low type is honest. Consider the first stage and take a low type with $\sigma = 0$. Consider a deviation to exaggeration and not taking the action, $(m', a_i') = (1, 0)$. Note that this strategy yields the prior image $\pi\overline{\theta} + (1 - \pi)\underline{\theta}$, while submitting a low report and not taking the action yields a low image. The sender does not have incentives to do this deviation if and only if

$$\begin{aligned} & U_s^*(m^*, a_i^* \mid \theta_s, \sigma) \geq U_s^*(m', a_i' \mid \theta_s, \sigma) \\ \iff \quad & \mu\underline{\theta} \geq \pi\underline{\theta}\gamma E[W \mid \sigma] + \mu(\pi\overline{\theta} + (1 - \pi)\underline{\theta}) \\ \iff \quad & \mu \leq \frac{-\underline{\theta}\gamma}{3(\overline{\theta} - \underline{\theta})}, \end{aligned}$$

i.e., this strategy profile cannot be an equilibrium.

Hence, there exists at most one equilibrium with influential communication, which is such that the high type is honest and the low type downplays impact. This equilibrium exists if and only if conditions (A1), (A2) and (A3) hold, which establishes Proposition 2. Moreover, there does not exist an equilibrium with influential communication if either condition (A2) or (A3) does not hold, which establishes Proposition 1.

Bielefeld University, Germany
University of Amsterdam, The Netherlands

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

References

Abeler, J., Nosenzo, D. and Raymond, C. (2019). 'Preferences for truth-telling', *Econometrica*, vol. 87(4), pp. 1115–53.
Ali, S.N. and Bénabou, R. (2020). 'Image versus information: changing societal norms and optimal privacy', *American Economic Journal: Microeconomics*, vol. 12(3), pp. 1–49.
Andreoni, J. and Bernheim, D.B. (2009). 'Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects', *Econometrica*, vol. 77(5), pp. 1607–36.
Andreoni, J. and Petrie, R. (2004). 'Public goods experiments without confidentiality: a glimpse into fund-raising', *Journal of Public Economics*, vol. 88(7–8), pp. 1605–23.
Angus Reid Institute. (2017). 'What stops Canadians from donating more to charitable organizations', See https://angusreid.org/giving-tuesday/ (accessed: April 10, 2020).
Ariely, D., Bracha, A. and Meier, S. (2009). 'Doing good or doing well? Image motivation and monetary incentives in behaving prosocially', *American Economic Review*, vol. 99(1), pp. 544–55.
Austen-Smith, D. and Feddersen, T.J. (2006). 'Deliberation, preference uncertainty, and voting rules', *American Political Science Review*, vol. 100(2), pp. 209–17.
Balliet, D., Parks, C. and Joireman, J. (2009). 'Social value orientation and cooperation in social dilemmas: a meta-analysis', *Group Processes & Intergroup Relations*, vol. 12(4), pp. 533–47.
Battigalli, P. and Dufwenberg, M. (2009). 'Dynamic psychological games', *Journal of Economic Theory*, vol. 144(1), pp. 1–35.

Battigalli, P. and Dufwenberg, M. (2019). 'Psychological game theory', Working Paper 646, Innocenzo Gasparini Institute for Economic Research.
Bénabou, R., Falk, A. and Tirole, J. (2018). 'Narratives, imperatives and moral reasoning', Working Paper 24798, National Bureau of Economic Research.
Bénabou, R. and Tirole, J. (2006). 'Incentives and prosocial behavior', *American Economic Review*, vol. 96(5), pp. 1652–78.
Bénabou, R. and Tirole, J. (2011). 'Laws and norms', Working Paper 17579, National Bureau of Economic Research.
Bolderdijk, J.W., Steg, L., Woerdman, E., Frieswijk, R. and De Groot, J.I. (2017). 'Understanding effectiveness skepticism', *Journal of Public Policy & Marketing*, vol. 36(2), pp. 348–61.
Burlando, R.M. and Guala, F. (2004). 'Heterogeneous agents in public goods experiments', *Experimental Economics*, vol. 8, pp. 35–54.
Bursztyn, L., Haaland, I.K., Rao, A. and Roth, C.P. (2020). 'I have nothing against them, but...', Working Paper 27288, National Bureau of Economic Research.
Bursztyn, L. and Jensen, R. (2017). 'Social image and economic behavior in the field: identifying, understanding, and shaping social pressure', *Annual Review of Economics*, vol. 9, pp. 131–53.
Butera, L. and Horn, J. (2020). ''Give less but give smart' experimental evidence on the effects of public information about quality on giving', *Journal of Economic Behavior and Organization*, vol. 171, pp. 59–76.
Conway, E.M. and Oreskes, N. (2011). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, London: Bloomsbury Press.
Coughlan, P.J. (2000). 'In defense of unanimous jury verdicts: mistrials, communication, and strategic voting', *American Political Science Review*, vol. 94(2), pp. 375–93.
Crawford, V.P. and Sobel, J. (1982). 'Strategic information transmission', *Econometrica*, vol. 50(6), pp. 1431–51.
Crosetto, P., Weisel, O. and Winter, F. (2012). 'A flexible z-Tree implementation of the social value orientation slider measure', *Jena Economic Research Papers*, vol. 62, pp. 1–8.
Dana, J., Weber, R.A. and Kuang, J.X. (2007). 'Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness', *Economic Theory*, vol. 33(1), pp. 67–80.
Deimen, I., Ketelaar, F. and Le Quement, M.T. (2015). 'Consistency and communication in committees', *Journal of Economic Theory*, vol. 160, pp. 24–35.
Ellingsen, T. and Johannesson, M. (2008). 'Pride and prejudice: the human side of incentive theory', *American Economic Review*, vol. 98(3), pp. 1–40.
Erat, S. and Gneezy, U. (2012). 'White lies', *Management Science*, vol. 58(4), pp. 723–33.
Exley, C.L. (2016). 'Excusing selfishness in charitable giving: the role of risk', *Review of Economic Studies*, vol. 83(2), pp. 587–628.
Exley, C.L. (2020). 'Using charity performance metrics as an excuse not to give', *Management Science*, vol. 66(2), pp. 553–63.
Fischbacher, U. (2007). 'z-Tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics*, vol. 10(2), pp. 171–8.
Fischbacher, U., Gächter, S. and Fehr, E. (2001). 'Are people conditionally cooperative', *Economics Letters*, vol. 71(3), pp. 397–404.
Foerster, M. (2019). 'Dynamics of strategic information transmission in social networks', *Theoretical Economics*, vol. 14(1), pp. 253–95.
Foerster, M. (2020). 'Strategic transmission of imperfect information—why revealing evidence (without proof) is difficult', Working Paper, SSRN.
Foerster, M. and van der Weele, J.J. (2018). 'Denial and alarmism in collective action problems', Discussion Paper, TI 2018–019/I, Tinbergen Institute.
Friedrichsen, J. and Engelmann, D. (2018). 'Who cares about social image?', *European Economic Review*, vol. 110, pp. 61–77.
Galeotti, A., Ghiglino, C. and Squintani, F. (2013). 'Strategic information transmission networks', *Journal of Economic Theory*, vol. 148(5), pp. 1751–69.
Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989). 'Psychological games and sequential rationality', *Games and Economic Behavior*, vol. 1(1), pp. 60–79.
Gino, F., Norton, M.I. and Weber, R.A. (2016). 'Motivated Bayesians: feeling moral while acting egoistically', *Journal of Economic Perspectives*, vol. 30(3), pp. 189–212.
Gneezy, U. (2005). 'Deception: the role of consequences', *American Economic Review*, vol. 95(1), pp. 384–94.
Gneezy, U., Keenan, E.A. and Gneezy, A. (2014). 'Avoiding overhead aversion in charity', *Science*, vol. 346(6209), pp. 632–5.
Gordon, T.P., Knock, C.L. and Neely, D.G. (2009). 'The role of rating agencies in the market for charitable contributions: an empirical test', *Journal of Accounting and Public Policy*, vol. 28(6), pp. 469–84.
Greene, W. (2010). 'Testing hypotheses about interaction terms in nonlinear models', *Economics Letters*, vol. 107(2), pp. 291–6.
Grossman, Z. and van der Weele, J.J. (2017). 'Self-image and willful ignorance in social decisions', *Journal of the European Economic Association*, vol. 15(1), pp. 173–217.

Hagenbach, J. and Koessler, F. (2010). 'Strategic communication networks', *Review of Economic Studies*, vol. 77(3), pp. 1072–99.
Harbaugh, W.T. (1998). 'What do donations buy? A model of philanthropy based on prestige and warm glow', *Journal of Public Economics*, vol. 67(2), pp. 269–84.
Henry, E. and Louis-Sidois, C. (2020). 'Voting and contributing when the group is watching', *American Economic Journal: Microeconomics*, vol. 12(3), pp. 246–76.
Henry, E. and Sonntag, J. (2019). 'Measuring image concern', *Journal of Economic Behavior and Organization*, vol. 160, pp. 19–39.
Hillenbrand, A. and Verrina, E. (2018). 'The differential effect of narratives on prosocial behavior', *Discussion Papers of the Max Planck Institute for Research on Collective Goods No. 2018/16*.
Karlan, D. and List, J.A. (2020). 'How can Bill and Melinda Gates increase other people's donations to fund public goods?', *Journal of Public Economics*, vol. 191.
Karlan, D. and McConnell, M.A. (2014). 'Hey look at me: the effect of giving circles on giving', *Journal of Economic Behavior and Organization*, vol. 106, pp. 402–12.
Karlan, D. and Wood, D.H. (2017). 'The effect of effectiveness: donor response to aid effectiveness in a direct mail fundraising experiment', *Journal of Behavioral and Experimental Economics*, vol. 66, pp. 1–8.
Kunda, Z. (1990). 'The case for motivated reasoning', *Psychological Bulletin*, vol. 108(3), pp. 480–98.
Kuran, T. (1997). *Private Truths, Public Lies: the Social Consequences of Preference Falsification*, Cambridge, MA: Harvard University Press.
Kurzban, R. and Houser, D. (2005). 'Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations', *Proceedings of the National Academy of Sciences*, vol. 102(5), pp. 1803–7.
Lacetera, N. and Macis, M. (2010). 'Social image concerns and prosocial behavior: field evidence from a nonlinear incentive scheme', *Journal of Economic Behavior and Organization*, vol. 76(2), pp. 225–37.
Levine, D.K. (1998). 'Modeling altruism and spitefulness in experiments', *Review of Economic Dynamics*, vol. 1(3), pp. 593–622.
Meer, J. (2014). 'Effects of the price of charitable giving: evidence from an online crowdfunding platform', *Journal of Economic Behavior and Organization*, vol. 103, pp. 113–24.
Metzger, L. and Günther, I. (2019a). 'Is it what you say or how you say it? The impact of aid effectiveness information and its framing on donation behavior', *Journal of Behavioral and Experimental Economics*, vol. 83, article ID 101461.
Metzger, L. and Günther, I. (2019b). 'Making an impact? The relevance of information on aid effectiveness for charitable giving. A laboratory experiment', *Journal of Development Economics*, vol. 136, pp. 18–33.
Morris, S. (2001). 'Political correctness', *Journal of Political Economy*, vol. 109(2), pp. 231–65.
Murphy, R.O., Ackermann, K.A. and Handgraaf, M.J. (2011). 'Measuring social value orientation', *Judgement and Decision Making*, vol. 6(8), pp. 771–81.
Niehaus, P. (2020). 'A theory of good intentions', Discussion paper, UC San Diego.
Norgaard, K.M. (2006). ''People want to protect themselves a little bit': emotions, denial, and social movement nonparticipation', *Sociological Inquiry*, vol. 76(3), pp. 372–96.
Offerman, T., Sonnemans, J. and Schram, A. (1996). 'Value orientations, expectations and voluntary contributions in public goods', Economic Journal, vol. 106(437), pp. 817–45.
Ottaviani, M. and Sørensen, P.N. (2006). 'Reputational cheap talk', *The Rand Journal of Economics*, vol. 37(1), pp. 155–75.
Rege, M. and Telle, K. (2004). 'The impact of social approval and framing on cooperation in public good situations', *Journal of Public Economics*, vol. 88(7–8), pp. 1625–44.
Sobel, J. (2020). 'Lying and deception in games', *Journal of Political Economy*, vol. 128(3), pp. 907–47.
Soetevent, A.R. (2005). 'Anonymity in giving in a natural context—a field experiment in 30 churches', *Journal of Public Economics*, vol. 89(11–12), pp. 2301–23.
Soraperra, I., Suvorov, A., Van de Ven, J. and Villeval, M.C. (2019). 'Doing bad to look good: negative consequences of image concerns on pro-social behavior', *Revue Economique*, vol. 70(6), pp. 945–66.
Spence, M. (1973). 'Job market signaling', *The Quarterly Journal of Economics*, vol. 87(3), p. 355.
Stoll-Kleemann, S., O'Riordan, T. and Jaeger, C.C. (2001). 'The psychology of denial concerning climate mitigation measures: evidence from Swiss focus groups', *Global Environmental Change*, vol. 11(2), pp. 107–17.
Yörük, B.K. (2016). 'Charity ratings', *Journal of Economics and Management Strategy*, vol. 25(1), pp. 195–219.