TI 2023-062/I
Tinbergen Institute Discussion Paper

Spoiling the party. Experimental evidence on the willingness to transmit inconvenient ethical information

Revision: April 2024

Jantsje M. Mol[1]
Ivan Soraperra[2]
Joël J. van der Weele[3]

1 University of Amsterdam and Tinbergen Institute

2 Max Planck Institute for Human Development, Berlin

3 University of Amsterdam and Tinbergen Institute

Spoiling the party. Experimental evidence on the willingness to transmit inconvenient ethical information

Jantsje M. Mol[1,2], Ivan Soraperra[3], and Joël J. van der Weele[1,2]

April 2024

Abstract

Information about the consequences of our consumption choices can be unwelcome, and people sometimes avoid it. We investigate a situation where one person possesses information that is inconvenient for another, and study why and when they decide to transmit that information. We introduce a simple and portable experimental game to analyze transmission of inconvenient information. In this game, a Sender can inform a Receiver at a small cost about a negative externality associated with a tempting and profitable action for the Receiver. The results from our online experiment (N = 1,512) show that Senders transmit more information for large negative externalities. Sender's decisions are largely driven by their own preferences for information. However, some Senders take the Receiver's feelings into account, by transmitting evidence of positive externalities or by suppressing negative information upon the Receiver's request. Understanding the decision to share inconvenient information matters, as it will affect the personal and political will to address important externalities and can inform strategies to encourage the transmission of inconvenient information within organizations.

Keywords: willful ignorance, information avoidance, unethical behavior, lab experiment
JEL Codes: B41, C91, C93

[1]Center for Research in Experimental Economics and Political Decision Making (CREED), University of Amsterdam, The Netherlands.
[2]Tinbergen Institute, Amsterdam
[3]Max Planck Institute for Human Development, Berlin

1 Introduction

In many contexts, people have preferences over information, and may go to some length to avoid it (Golman et al., 2017). Information avoidance often serves to protect cherished beliefs, for instance the protection of one's ego from bad feedback (Castagnetti and Schmacker, 2022), or the avoidance of bad financial news to reduce disappointment or stress (Sicherman et al., 2016). In particular, previous research has shown that some people try to escape responsibility for ethical decisions and maintain a good self-image by remaining uninformed about the consequences of their decisions (Dana et al., 2007; Grossman and van der Weele, 2017; Vu et al., 2023). Such willful ignorance has important consequences for everyday consumption behavior, for instance the decision to buy products with adverse impacts on the environment or manufactured in exploitative conditions.

Information avoidance also has an interpersonal side that has received much less attention. People often have information that is potentially inconvenient for *others*, and must decide whether to share it. For instance, a vegetarian may ponder whether to give her friends detailed information about the environmental costs associated with meat eating. In doing so, she may weigh several considerations. First, a concern for environmental consequences might motivate her to influence her friends' diets in the "right" direction. A second, more procedural reason to share would be to make sure her friends take an informed decision, whatever it is. Finally, she may hold back information out of consideration for the Receiver of the information. She may assess that transmitting information may make her friends feel judged, and even lead to confrontations that she may wish to avoid. Indeed, there is evidence that vegetarians and vegans sometimes experience backlash for sharing information about their diets, which causes some to keep a low profile (De Groeve and Rosenfeld, 2022; MacInnis and Hodson, 2017).

Other applications may also occur within organizations or markets. For instance, when-

ever employees have knowledge of organizational practices with negative external consequences, such inconvenient news might not be transmitted up the decision-making chain. In buyer-seller interactions, buyers may be less informed than sellers, who could disclose ethical information about their products. While our experiment does not feature a transaction or an incentive for the seller to convince the buyer, it measures the intrinsic motives a seller might have to share such information.

Understanding the decision to share inconvenient information matters, as a reluctance to pass on inconvenient information will affect the dispersion of knowledge and the personal and political will to address important externalities. Understanding the reasons behind such reluctance may inform strategies to encourage the spread of inconvenient information. To study the trade-offs facing the sender, we design an experiment that we call the "Button game". The game involves two participants in the role of a Sender and a Receiver. The Receiver can press a large red button on the screen, which yields a bonus payoff of £1 for the Receiver, but may or may not degrade a fund destined for donation to a worthy cause. If the Receiver does not press the button, there are no additional payoffs for the Receiver or for the charity. The button is designed to be tempting; indeed, in the absence of specific information about the externality, all Receivers in our experiment press the button and pocket the £1.

Our primary interest is the decision of the Sender: Before the Receiver presses the button, the Sender can send information about the size of the externality at a small cost. To study how Senders trade off different interests, we experimentally vary the decision environment, and also measure Sender's preferences for the charity and for information in a separate game based on Dana et al. (2007). In the Baseline treatment, Senders make multiple decisions for different sizes of the externality, where one of their decisions is randomly implemented. We find most Senders are willing to pay to send information. Relatively large

externalities (the cost to the charity is 2.5 times the benefit to the Receiver) are shared by 71.8% of subjects, while smaller externalities (e.g. $-£1$, where cost equals benefit) are shared by 58.3%. This indicates that some Senders trade off the payoff for the charity with the (small) cost of sending. We also find evidence that procedural motives to transmit information are pervasive: Almost 40% of Senders share when the externality is positive, and information does not alter Receivers' decisions. Moreover, personal preferences for information, measured on a separate task, explain sharing, even controlling for a concern for the charity. Finally, almost 30% of Senders say explicitly that information sharing is the right thing to do.

To further investigate the Sender's willingness to accommodate the Receiver, we designed a treatment in which we vary the Senders information about the latter. Before the Sender makes a decision, the Receiver can request either information or ignorance. We find that 71.2% of the Receivers request information, while the remainder requests ignorance. A request for information does not have much effect in our setting – perhaps because most Senders already share information. By contrast, requests for ignorance are effective as they reduce information sharing. This shows that at least some Senders are willing to protect the Receiver from inconvenient information at the potential expense of the charity. We find that the only Senders who are sensitive to ignorance requests are those who themselves avoid information in a similar setting, indicating an interaction between the information preferences of Senders and Receivers. To reinforce the power of the request and mimic the possibility of conflict, we add a treatment with an option for the Receiver to punish the Sender at the end of the game by denying part of the Sender's participation payment. Contrary to our expectations, the threat of punishment does not make either type of requests more effective, even though we observe some punishment by Receivers.

The key take-away that emerges from our dataset is that sharing of inconvenient infor-

mation is a complex decision driven by multiple motives. Utilitarian motives to affect the outcome of the decision are prevalent, but personal attitudes towards information play an important role, and a small fraction of participants is willing to withhold information to protect the Receiver. To the extent the findings from our stylized setting capture behavior outside the lab, the prevalence of sharing provide reasons for optimism about the spread of inconvenient information in society or in organizations. Nevertheless, the results also indicate the limits of sharing. The central role of Sender's own preferences for information in the decision to share and respond to requests suggest a role for assortative matching. If social networks are characterized by homophily, i.e. people interact with others who have similar preferences, this might lead to information bubbles. For instance, a group of carnivores might not share any information about the negative impact of meat. We discuss these implications in more detail in the conclusion section.

Our paper contributes to a fast-growing literature on information avoidance on ethical dilemmas in experimental economics (Dana et al., 2007; Grossman, 2014; Vu et al., 2023), and a smaller literature on how people share inconvenient information. Closest to our paper is Soraperra et al. (2023), who examine the demand and supply of willful ignorance in a market setup. Over multiple rounds, Senders choose to release information or not, and decision makers can choose to match with the Sender they prefer. In this setting, Senders suppress about 25% of inconvenient information on average, which correlates with their own preferences. However, the market setting is noisy and there is not much control over the strategic incentives of the Senders or their beliefs about the decision-makers preferences, making it hard to disentangle various explanations for information transmission and suppression.

Lind et al. (2019) allow Senders to force information on decision makers after they made their own decision to avoid or obtain information. They find that the option to be

"overruled" by the Sender results in more information seeking by decision makers. Lane (2022) investigates a setting in which subjects can inform others about the externalities of their actions *after* they have taken a decision, so the information has no instrumental value but may reduce the happiness of the decision maker. Most Senders reveal information, despite the potentially negative impact on the Receiver. The setup of the paper differs from ours in various ways; in particular, the information is transmitted after the decision and hence has no instrumental value, and there is no information about the Receiver's preference. Finally, Foerster and van der Weele (2021) also look at the communications of uncertainty about the impact of a charitable donation. An informed Sender communicates with an uninformed Receiver, after which both players can make a donation. The Sender can distort the signal about impact upward or downwards. They find that when the Sender's donation is revealed, Senders are more likely to "downplay" the impact to excuse their own selfishness.

The main contribution of our paper to this literature is to introduce a simple and portable setting to analyze transmission of inconvenient information. Our finding that information is shared most of the time is broadly in line with previous literature, and we offer new evidence on the determinants of sharing decisions and the willingness of Senders to accommodate Receiver's information avoidance.

2 Method and Experimental Design

The experiment consists of two tasks and a final survey. The first task measures participants' preferences for information in the binary dictator game proposed by (Dana et al., 2007, DWK hereafter). The second and main task, a novel two-person game we call the "Button game", disentangles different motives to share information.

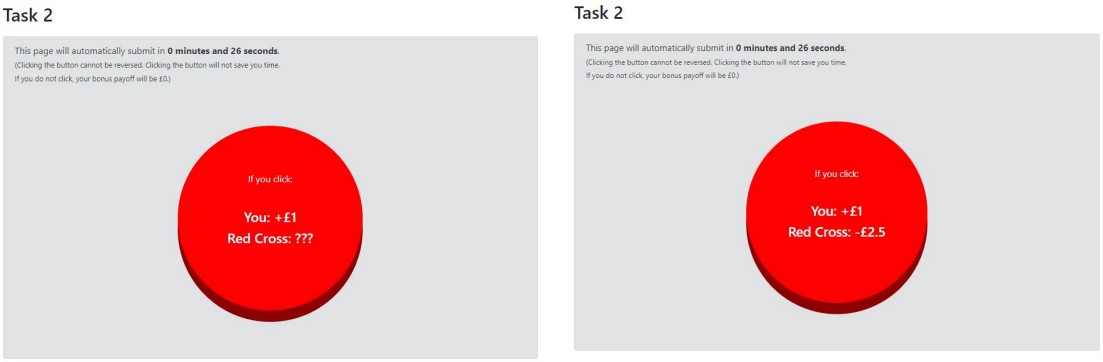
2.1 The DWK binary dictator game

The binary dictator game closely follows the Hidden Information treatment proposed by Dana et al. (2007). In this task, the participant has to choose between two options, i.e., Option A and Option B, that have consequences for their payoff and for the donation to a charity, the Red Cross. The payoffs of Option A and Option B for the participant are £0.60 and £0.50, respectively. The payoffs for the Red Cross, instead, depend on the scenario: in the conflicting scenario A and B pay £0.10 and £0.50 to the charity; in the aligned scenario the payoffs for the charity are flipped, with A and B paying £0.50 and £0.10, respectively. Participants are informed that each scenario is randomly selected with equal probability and they can find out the realized scenario by clicking a *Reveal* button. Alternatively, they can select their preferred option directly, without knowledge of whether the payoffs for the charity follow the aligned or the conflicting scenario.

2.2 The Button Game

As the main task, we designed a two-person game in which a Receiver with limited knowledge about the consequences of their actions for a third party interacts with a Sender possessing superior information. The Sender can inform the Receiver before the latter makes their choice. We consider three variants of the game that manipulate how the two parties interact and define our treatments — i.e., the *Baseline* treatment, the *Request* treatment, and the *Request + Punishment* treatment.

In all versions of the game, the Receiver has to decide whether to press a button, see Figure 1 for an example. The button is displayed for a total time of 30 seconds during which the Receiver can press it. When the button is pressed, the timer continues on the next page, to avoid button pressing for time efficiency concerns. Pressing the button pays a bonus of £1 to the Receiver but it has also consequences for a third party, the Red Cross.



(a) Uninformed (b) Informed

Figure 1: Decision screen for the Receiver in the Button Game.

The consequences for the Red Cross can range from +0.5 to -2.5 pounds, but the Receiver has no information about them; neither about the actual value nor about the range. The Receiver only knows that the consequences for the Red Cross can be either positive or negative. Not pressing the button bears no consequences for the Receiver and for the charity. The Sender, who earns a bonus of £0.25, is informed of the consequences for the charity, and this is common knowledge among the players. The Sender's task is to decide whether to share this piece of information with the Receiver before the latter makes their choice. The decision to pass information comes at a small cost of £0.10 for the Sender. In the experiment, we implemented the Sender's decision using the strategy method. Each Sender had to choose whether to share information for three negative impact levels (-2.5, -1.0 and -0.5 pounds) and one positive impact level (+0.5).

Figure 2 shows the timeline of the Button Game and highlights the differences between the treatments. In the *Baseline* treatment, the Sender moves first and decides whether or not to inform the Receiver. After the decision of the Sender, the Receiver decides whether to press the button. The Receiver chooses with or without information about the consequences for the Red Cross, depending on the decision of the Sender. The *Request*

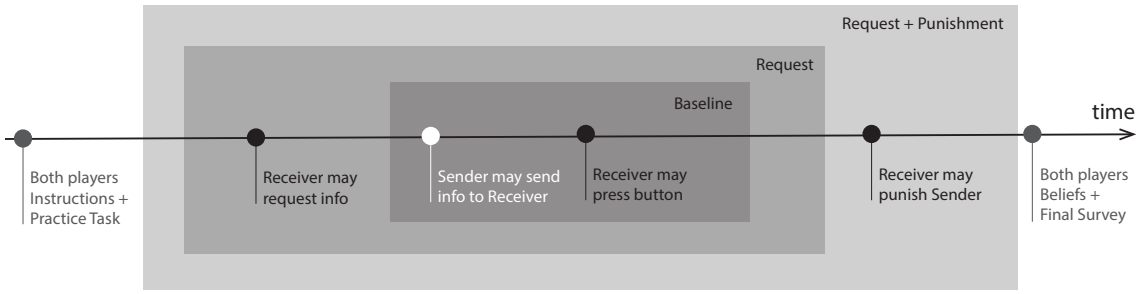


Figure 2: Timeline of the different variants of the Button game.

treatment extends the Button game by adding a stage at the beginning where Receivers can either request information or ignorance about the payoffs for the charity. The Receiver selects from two pre-specified messages options: there is no option not to send a request message. Finally, the *Request + Punishment* treatment extends the Request treatment by adding a stage at the end. In this final stage, the Receiver chooses whether to confirm or to cancel the bonus payment of the Sender. In the experiment, this decision was neutrally framed as 'A final choice' to avoid normative connotations related to the word 'punishment'.

2.3 Hypotheses

Here we explain how the treatment differences are interpreted and discuss our hypotheses about the Sender's motives. All hypotheses were preregistered prior to data collection (https://aspredicted.org/X8Y_Q7T and is also included in Appendix D for ease of reference).[1]

Before diving into the hypotheses about the Sender's behavior, we briefly discuss what we expect for the Receiver. For the time being, we assume that these expectations reflect

[1]Our hypotheses are not based on a formal model. In order to keep the experiment simple and intuitive, we did not tell the Receiver the possible outcomes for the charity nor the probabilities associated with these outcomes. This makes it hard to model this as a Bayesian (disclosure) game.

the Sender's beliefs about Receivers. As for the Receivers, we expect that virtually all of them press the button when uninformed. When informed, instead, we expect that some of the Receivers will decide not to press the button to avoid generating harm to the charity. Moreover, we expect the likelihood of pressing the button to be weakly decreasing with the size of the consequences. Intuitively, if someone is willing to give up £1 to avoid a given level of the consequences the same person should also be willing to give up £1 to avoid more serious consequences.

Since our main interest is in the Sender's decision to share inconvenient information, we will focus only on Senders' choices for the negative consequences for the Red Cross, i.e., we ignore Sender's decisions for the +0.5 pounds.[2] For each sender, we define a "sender-index" that measures the point at which consequences for the Red Cross become too serious not to share information. The index ranges from 0, when the Senders do not share information for any negative consequence, to 3, when the Sender shares information for all negative consequences. An index of 1 identifies those Senders that share information only for the most extreme (-2.5 pounds) consequence and an index of 2 identifies those Senders that share information for the -1.0 and -2.5 pounds consequences, but not for the least extreme (-0.5 pounds) consequence.

The *Baseline* treatment measures whether Senders have preferences for sharing inconvenient information of the Senders that are strong enough to overcome the small cost of sharing. As mentioned in the introduction, such preferences could depend on various motives, e.g. (1) outcome-based reasons, e.g. a concern for the charity, increasing in the size of the externality, (2) procedural reasons, e.g. the belief that one ought to make an informed choice, or (3) on the desire to help the Receiver and they believe the Receiver would like to be informed, e.g. to prevent guilt. Accordingly, our first hypothesis is that a

[2]The main reason to include a positive value is dictated by the need to truthfully tell the Receiver that consequences could be either positive or negative.

non negligible fraction of Senders decides to share inconvenient information.

Hypothesis 1 *Senders send inconvenient information about the charity to their partners.*

The comparison of the *Request* and the *Baseline* treatment allows us to examine the strength of the desire to help the Receiver (motive 3). Indeed, if this motive is strong enough, one would observe a shift in the decision to share information in the direction of the request. Therefore, we expect Senders to follow these requests:

Hypothesis 2 *The likelihood to share inconvenient information increases with a request for information and decreases with a request for ignorance.*

Finally, the comparison of the *Request + Punishment* and the *Request* treatment may be used to assess the strength of the outcome-based or moral preferences of the Senders. In the *Request + Punishment* treatment the Receivers can actually harm the Senders when they are unhappy about the provided or hidden information. Since Receivers can affect the Senders' payoffs, we expect Senders with weak outcome-based motives to follow the request of the Receiver more closely. Furthermore, the *Request + Punishment* treatment provides a measure of the elasticity of the supply of inconvenient information.

Hypothesis 3 *The possibility of Receiver punishment amplifies the impact of the requests on the likelihood to share information.*

Along with hypotheses about the Sender's behavior, we derive secondary hypotheses regarding the impact on the overall welfare of the charity. Based on the previous hypotheses about Receivers' and Senders' behavior, we expect that Receivers requesting information are motivated by a willingness to avoid harming the charity. Therefore, the possibility to make a request, when followed by the Senders, has an impact on the final outcome for the charity. Specifically, we expect the following:

Hypothesis 4 *A request for information is associated with higher earnings for the charity, and request for ignorance with lower earnings for the charity. These effects are amplified in the punishment treatment.*

2.4 Procedure

The main study was run on Prolific in November 2022. In total, N = 1,512 participants completed the study ($n_{Baseline}$ = 302, $n_{Request}$ = 610, $n_{Request+Punishment}$ = 600). All participants gave informed consent before participation. Participants were rewarded a £1.30 show-up fee plus the bonus earned in one of the two tasks, which was randomly selected at the end of the experiment. The experiment was programmed and data was collected via oTree software (Chen et al., 2016). The analysis code can be found at https://www.jantsje.nl/files/analysis_button.html.

The experiment started with the binary dictator game, followed by the Button Game and the final survey. In the Button Game participants were matched in pairs by the software, which meant that they had to wait for another player to join. If no other player appeared within 5 minutes, the software moved on to the survey and the bonus payment was based on the results of the binary dictator game. When a match was possible, players were randomly assigned to the role of Sender and Receiver to start the Button Game.

After reading the instructions, both Senders and Receivers faced a practice round to experience the decision of the Receiver button page. In the practice round, no information about the consequences for the charity was communicated (see panel (a) of Figure 1). After the practice round, Senders had to state their beliefs about the number of people pressing the button and answered a short set of comprehension questions. No comprehension questions were asked to the Receiver due to the simplicity of the button task.

At the end of the Button Game, Senders completed a belief elicitation page where they

were asked to guess the likelihood that the Receiver pressed the button for each possible consequence. In the *Request + Punishment* treatment, Senders were further asked to guess the likelihood of punishment. At the end of the experiment, all participants completed a demographics questionnaire, which included some open questions about their motivations in the Button Game and a 10-point slider to indicate how much they identified with the Red Cross (inspired by Ariely et al., 2009). Finally, each player was shown an overview of payoffs and was informed about the task that was randomly selected for the bonus payment.

Several days after the end of the main study, all participants who completed at least the Dictator game (in one of the pilots or the main study) were messaged[3] via Prolific with a proof of the donation to the charity.

3 Results

3.1 Preliminaries

3.1.1 Randomization check

Table 1 provides summary statistics about the participants' demographics and other variables. The table allows us to assess the quality of the randomization across treatments. Overall, the sample is balanced regarding age, income, identification with the charity, button pressing in the practice round, and most importantly own preferences for information (measured by the decision to reveal in the binary dictator game). Gender distribution (more women in the *Request + Punishment* treatment) and the device used (more mobile

[3]Dear participant, In [month] 2022, you participated in our decision-making experiment on Prolific. As part of this experiment, we scheduled a donation to the British Red Cross. Based on the decisions in the experiment, positive and negative payoffs could be collected for the Red Cross. We would like to inform you that the donation to the Red Cross has been made. You can find the donation receipt and more details here: https://figshare.com/s/d684b47812a2585174f4 Thanks again for your participation. You do not need to respond to this message. The researchers.

devices in the *Request + Punishment* treatment) are slightly unbalanced across treatments. Due to the live matching procedure of the Button Game, treatments were run sequentially on the Prolific platform. It may be the case that different user groups log into Prolific on different times and days of the week. To control for such differences, we will thus add gender and device type as covariates in all further analyses.

Table 1: Descriptive statistics by treatment

	Baseline	Request	Req. + Pun.	Overall	p-value
Gender (%)					0.054
Female	136 (45.0)	316 (51.8)	323 (53.8)	775 (51.3)	
Male	161 (53.3)	291 (47.7)	273 (45.5)	725 (47.9)	
Non-binary/not say	5 (1.7)	3 (0.5)	4 (0.7)	12 (0.8)	
Age in years (SD)	39.5 (12.9)	39.1 (13.0)	39.0 (13.0)	39.1 (13.0)	0.825
Monthly income (%)					0.209
< £999	29 (9.6)	55 (9.0)	47 (7.8)	131 (8.7)	
£1000-£1999	68 (22.5)	156 (25.6)	130 (21.7)	354 (23.4)	
£2000-£2999	80 (26.5)	175 (28.7)	162 (27.0)	417 (27.6)	
£3000-£3999	61 (20.2)	103 (16.9)	123 (20.5)	287 (19.0)	
> £4000	51 (16.9)	88 (14.4)	88 (14.7)	227 (15.0)	
Rather not say	13 (4.3)	33 (5.4)	50 (8.3)	96 (6.3)	
Identify with charity[a] (SD)	0.2 (2.9)	0.2 (2.7)	0.2 (2.8)	0.2 (2.8)	0.947
Browser type Desktop (%)	281 (93.0)	566 (92.8)	517 (86.2)	1364 (90.2)	<0.001
Reveal in DWK (%)	119 (39.4)	256 (42.0)	223 (37.2)	598 (39.6)	0.232
Press btn in the test round (%)	292 (96.7)	579 (94.9)	572 (95.3)	1443 (95.4)	0.477
Observations	302	610	600	1512	

Notes: The table reports the means for the continuous and the counts for the categorical variables with, respectively, SD and percentages in parentheses. [a] Response to the question *How much do you identify with the charity Red Cross?* ranging from -5 = not at all to 5 = very much. The column "p-value" reports the results of a test comparing the different treatments. A Chi-squared test is used for categorical variables and an Anova for the continuous variables.

3.1.2 Receiver's behavior

As a first step, we look at the Receiver's behavior and check if it broadly aligns with our assumptions. Indeed, almost all Receivers press the button when uninformed (96.4%; n = 364) across all treatments. Moreover, all 61 Receivers that saw good news — i.e., saw that the button increased the donation by an additional £0.5 — pressed the button. Finally,

we observed that the likelihood of pressing the button decreases with the severity of the negative consequence for the charity: 73.8% (n = 107) of the informed Receivers clicked the button when the consequences were -0.5 pounds, 66.1% (n = 112) when they were -1.0 pounds, and 51.8% (n = 112) when they were -2.5 pounds. This shows that, overall, the behavior of Receivers is in line with our predictions, suggesting that they trade off the consequences for the charity with the cost of sharing.[4]

3.1.3 Sender's beliefs

In order to interpret the decision to share bad news as an attempt to help the charity, it must be true that Senders believe that sharing bad news leads indeed to a lower likelihood to press the button. On average, Senders believe that 79.0% (SD = 18.1) of the Receivers press the button when not informed about the consequences. Moreover, Senders (correctly) expect a lower percentage of Receivers to press the button when consequences become more serious. Statistical support for this conjecture is provided in Table 2.

3.2 Information Sharing in the Baseline Treatment

On the aggregate, Senders' decisions to share information increase with the size of the consequence. A positive consequence of £0.5 is shared by 38.6% of Senders. Negative consequences of -0.5, -1.0 and -2.5 pounds are shared by 44.6%, 58.3% and 71.8% of Senders, respectively. To test our hypotheses, we will use the sender-index, which measures for which consequences Senders share information. Note that the index is not well-defined if a Sender decides to share information for less serious consequences and not to share information for more serious ones. Empirically, this is the case for only 40 out of 756 Senders which,

[4]The pattern is confirmed also when looking at the Receiver's behavior separated by treatment.

Table 2: Senders' beliefs about Receivers' button pressing, by consequence and treatment.

	Dependent variable: Beliefs about Receivers' button pressing			
	All tmts	Baseline	Request	Req. + Pun.
	(1)	(2)	(3)	(4)
consequence+0.5	5.344^{***}	4.715°	4.593^{**}	6.423^{***}
	(0.967)	(2.496)	(1.426)	(1.511)
consequence-0.5	-23.134^{***}	-20.808^{***}	-24.636^{***}	-22.777^{***}
	(0.981)	(2.132)	(1.517)	(1.608)
consequence-1.0	-31.354^{***}	-26.596^{***}	-33.272^{***}	-31.800^{***}
	(1.075)	(2.360)	(1.631)	(1.776)
consequence-2.5	-40.458^{***}	-36.172^{***}	-42.466^{***}	-40.573^{***}
	(1.210)	(2.671)	(1.868)	(1.968)
Observations	3,780	755	1,525	1,500
R^2	0.465	0.409	0.490	0.469
Adjusted R^2	0.331	0.258	0.361	0.334

Notes: Dependent variable: Response to the statement *I believe ... in 100 players will press the button.* Reference category: uninformed. Linear model with individual level fixed effects and heteroscedasticity robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

following the preregistration protocol, are excluded from the main analysis.[5]

Figure 3 shows the distribution of the sender-index in the *Baseline* treatment. This shows that we have substantial heterogeneity in the preferences for sharing information among our participants. On the one hand, one forth of the Senders never share information with the Receiver (sender-index = 0), another 40% of Senders always share information about the consequences (sender-index = 3). The remaining Senders have intermediate preferences and share only when consequences are serious (sender-index 1 and 2). Overall, the distribution suggests that the majority of the Senders are willing to pay a small cost to inform the Sender when consequences are serious and there is a high likelihood that

[5]Table B3 examines the pre-registered robustness check of a binary sender-index (including these 40 participants with non-monotonic sharing behavior). The appendix shows that results are robust to the inclusion of these participants. A detailed analysis of the sender-index can be found in Appendix A.

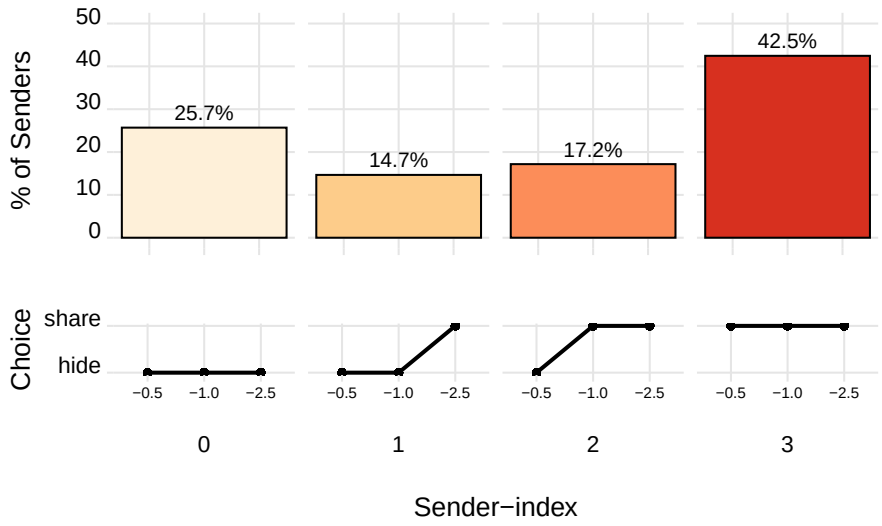


Figure 3: Percentage of Senders by sender-index in the *Baseline* treatment.

sharing information can make a difference. This provides support for Hypothesis 1.

3.3 The Effect of Requests

We now turn to the *Request* treatment, which allows us to investigate whether Senders take into account the preferences of the Receivers when sharing information. Figure 4 (the three middle panels) show the distribution of the sender-index across the various treatments. The Request and Punishment treatments are split by the nature of the request. According to our Hypothesis 2, if Senders are mainly motivated to help the Receivers, we should observe an increase in the sender-index when the request is for information and a decrease when the request is for ignorance. The figure, however, does not show clear evidnece in favor of this hypothesis. We do not find evidence of a shift in the frequency of Senders from lower to higher levels of the sender-index when moving from the left to the right panel in Figure 4. Indeed, a non-parametric Jonckheere-Terpstra trend test fails to reject the hypothesis of no

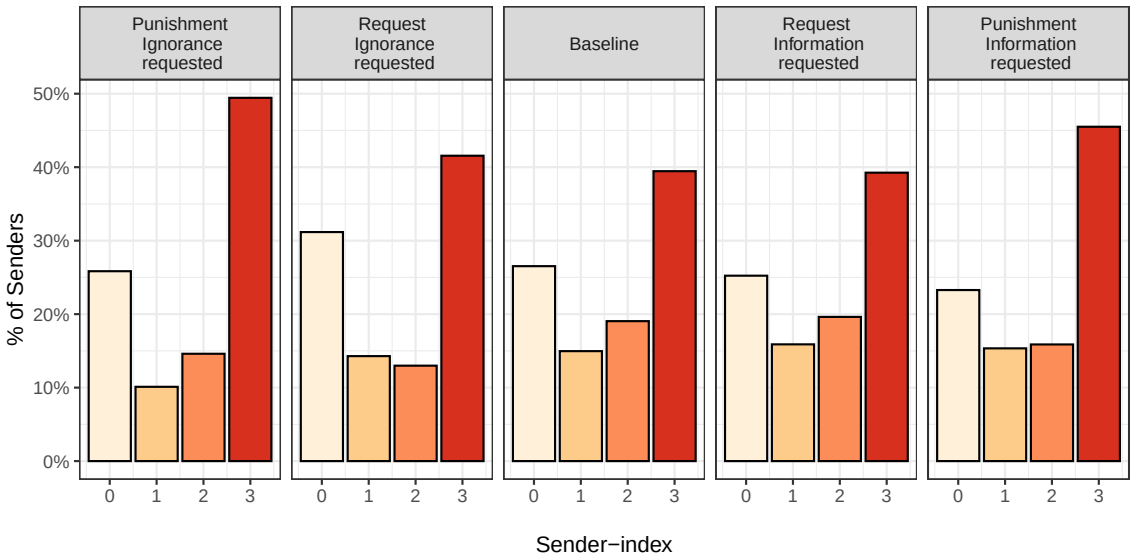


Figure 4: Percentage of Senders by sender-index and treatment.

difference in the sender-index across different requests ($z = 0.31$, $p = 0.377$).[6] The average sender-index gives a similar picture, with an average of 1.73 when information is requested, of 1.65 when ignorance is requested, and of 1.71 when the request was not present.

Regression evidence. To examine the sharing decision more closely and with additional statistical power, we regress the sender-index on the treatments, as well as variables that measure the Senders' preferences for information and their identification with the charity. We also include control variables such as gender, age, income, device, and the number of attempts to get the comprehension questions correctly. Since the sender-index is ordinal by nature, we employ an ordinal probit model to explore the correlation between such variables and the decision to share information. Model (1) in Table 3 presents the results of the regression using the *Baseline* data.

[6] Also testing the more general assumption of a difference across distributions does not support the idea that the request has an effect on the decision to share information. A χ^2 test cannot reject the null hypothesis of no differences in the distributions of the sender-index ($\chi^2(6) = 2.38$, $p = 0.882$)

Table 3: Ordered probit regressions of Sender-index

	Dependent variable: Sender-index					
	(1)	(2)	(3)	(4)	(5)	(6)
Information preference						
Request info	−0.011	−0.111	0.085	−0.028	−0.072	0.023
	(0.092)	(0.121)	(0.148)	(0.082)	(0.107)	(0.132)
Request ignorance	−0.130*	−0.255***	0.023	−0.150***	−0.232***	−0.063
	(0.053)	(0.065)	(0.101)	(0.035)	(0.041)	(0.072)
Request info under punishment threat				0.096	−0.048	0.356*
				(0.087)	(0.107)	(0.153)
Request ignorance under punishment threat				0.083	0.097	−0.008
				(0.053)	(0.072)	(0.083)
Control variables						
Identify with charity	0.099***	0.123***	0.076*	0.073***	0.088***	0.055*
	(0.021)	(0.027)	(0.036)	(0.017)	(0.021)	(0.028)
Revealed in DWK	0.257*			0.323***		
	(0.119)			(0.093)		
Log likelihood	-522	-312.5	-203.6	-842.8	-523.1	-311.9
Pseudo R^2 (McFadden)	0.038	0.048	0.028	0.029	0.028	0.026
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Revealed in DWK		No	Yes		No	Yes
Observations	412	244	168	667	402	265

Notes: Ordinal probit model of the Sender-index. Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Model 1, 2 and 3 include all participants in the Request treatment. Model 4, 5 and 6 include all participants in the Request and Request + Punishment treatment. Robust standard errors in parentheses ($^\circ p < 0.10$; $^* p < 0.05$; $^{**} p < 0.01$; *** $p < 0.001$).

Column (1) includes data from the *Baseline* and *Request* treatments. A majority of Senders (225; 73.8%) received a Request for information, while the rest (80; 26.2%) received a request for ignorance. It shows that requests for information have little effect, but requests for ignorance have a statistically significant and negative impact on the sender-index (compared to the Baseline without requests). Furthermore, sharing is positively related to how close the Senders feel to the charity, and whether they themselves revealed in the DWK game. This latter result shows that preferences about the information one would like to have for oneself play a large role in sharing information with others.

To better understand the channels through which the request for ignorance affects the

sender, we consider two possibilities. First, it could be that senders are likely to oblige the request if it aligns with their own preferences for information. To test this, we run the same model for those who remained ignorant in the DWK game (column 2) and those who informed themselves (column 3). Comparing these two models shows that the Sender's preferences for information interact with the request: only those Senders who remained ignorant themselves accommodate a request for ignorance, whereas those who informed themselves are, if anything, a bit less likely to send ignorance. Thus, the graphical evidence in Figure 4 hides an important interaction in the response to request. We see a similar asymmetric reaction to requests for information, but here the results are not statistically significant.

A second channel by which requests may affect the sender's decision is by changing sender's beliefs about what the receiver will do upon receiving the information. For instance, a request for ignorance may signal that the receiver will not use the information, so the sender may be less willing to bear the cost of sending it. To understand if this channel is at work, Appendix Table D2 examines the posterior beliefs of the Sender about their Receiver pressing a button. However, we find no statistically significant evidence that requests change sender beliefs, so we conclude the belief channel is not operative.

3.4 The Role of Punishment

The *Request + Punishment* treatment allows us to test whether Senders stick to their preferences for sharing even when they risk being punished when they do not follow the request (Hypothesis 3). Note that punishment rates were low, but not negligible. Information requests were more likely to be followed, and deviations were more likely to be punished compared to requests for ignorance.[7] In line with the pre-registration, we test

[7]The distribution of requests observed in the punishment treatment is similar to the one observed in the treatment without punishment. In *Request + Punishment* treatment, 206 Senders (68.7%) received a

the null hypothesis that punishment does not change the pressure to follow the request of the Receiver against the alternative hypothesis that it increases the pressure to follow the request of the Receiver. Specifically, we test whether, under the threat of punishment, the sender-index increases when information is requested and decreases with the threat of punishment when ignorance is requested. As for the Senders' decision in the *Request + Punishment* treatment, the left- and rightmost panels of Figure 4 show the distribution of the sender-index when a request for ignorance and for information are received, respectively. Visually, these distributions do not differ substantially from the ones observed in the *Request* treatment, which are reported in the second and fourth panel, respectively. Indeed, a one-sided Wilcoxon rank-sum test fails to reject the null hypothesis that punishment has no effect on the sender-index both when ignorance is requested ($p = 0.865$) and when information is requested ($p = 0.164$).[8]

Regression evidence. In column (4) of Table 3, we investigate whether requests combined with the threat of punishment induce different patterns from the baseline. In line with the graphical and non-parametric evidence, we do not see significant evidence for an effect of requests when punishment is present. As column (5) and (6) demonstrate, we also do not see the effect of the Sender's preferences of information that is present in the *Request* treatment. All coefficients are insignificant at the 5% threshold, although Senders' who reveal in the DWK game appear slightly more responsive to a request for information, with marginal statistical significance.

request for information and 94 (31.3%) received a request for ignorance. 52 (17.3%) of the 300 Receivers punished the Sender for (not) responding to their request. Information requests were followed by 113 Senders and ignored by 93 Senders, of which 30 were punished (32.2%). Conversely 48 Senders sent information when ignorance was requested, but only 7 of those were punished (14.6%). In a few cases, Receivers were punished when following the request for information (6 cases) or ignorance (9 cases).

[8] A Jonckheere-Terpstra trend test using all 5 combinations of treatments and requests does not reject the null hypothesis that the sender-index is not increasing in the pressure to follow the request ($z = 0.38$, $p = 0.352$). Similarly, a χ^2 test cannot reject the null hypothesis of no differences in the distributions of the sender-index across all 5 combinations of requests and treatments ($\chi^2(12) = 7.58$, $p = 0.817$).

3.5 Motives for sharing information

To further investigate the motives for sharing information, we study the open-ended questionnaire responses of Senders. At the very end of the experiment, we asked Senders across treatments who shared at least one consequence with Receivers to explain in words why they did so. We hired two research assistants, blind to the hypotheses, to manually classify these answers into categories.

First, a concern for the charity appears to be a dominant motive. This is evident from the fact that many Senders condition on the size of the externality when deciding to send or not, and the strong correlation between sending information and self-reported identification with the charity. Indeed, the questionnaire reveals that the main motivations to send are a wish to help the charity either unconditionally (30.3% "*I wanted Red Cross to get as much money as possible*"), or conditional on the impact not being too negative (30.1% "*when the consequence was too large*").

However, concern for the outcome does not explain all sending decisions. For instance, 38.6% of Senders share even when the externality is positive, and information does not change the Receiver's decision. This is consistent with a more procedural concern, whereby the Sender thinks that the Receiver should make an informed decision. Indeed, we find that a large number of Senders mention such procedural concerns (28.8% "*it is the right thing to do*").

However, sharing information about positive externalities is also consistent with an attempt to protect the Receivers' feelings, as this information would help the Receiver to feel good about pressing the button. Indeed, a few Senders mention helping the other player explicitly in the questionnaire (1.3% "*probably they would be happy to be informed, instead of feeling bad*"). Another piece of evidence that Senders are concerned about the Receiver's feelings, is the willingness to suppress information on request. At the same

time, some Senders appear to be doing the opposite, and use information as a way to make Receivers feel bad about pressing the button (0.4% "*to make them feel guilty*").

The fact that participants do not try to protect the Receiver's feelings is in line with Lane (2022). However, we cannot rule out that Senders care to some degree about the Receiver's preferences. The willingness to suppress information on request is in line with a wish to help the Receiver avoid information, not by a threat of punishment, which has no impact on Sender's responses to requests.

Figure 5 summarizes the responses to the questionnaire by sender-index. Senders with a sender-index of 1 or 2 report mainly conditional charity help (i.e. helping to prevent the largest loss) while those with the maximum sender-index report mainly unconditional charity help and procedural motivations (i.e. one should be informed). Generally, the results show that the decision to share inconvenient information is a complex one, in which several different motives come together.

Furthermore, we asked the 156 respondents with a sender-index of 0 why they did not send at all (*Can you explain why you did not inform player A about the consequences for the Red Cross?* - open answer box with forced response). Their main motivations are a request for ignorance (30.1%, '*they specifically asked for me not to*') and the cost of sending (29.5%, '*I wanted to maximise my return*'), followed by procedural motivations (14.7%, '*to see what they would do*') and indifference (14.1%, '*the other person would be just as likely to press the button*'). Nine participants (5.8%) indicated that they wanted to help the other person to be selfish, and another 5.8% indicated various other reasons.

3.6 Consequences for the charity

After studying the Senders' sharing decisions, we move to analyzing the consequences of sharing information for the Red Cross. In Hypothesis 4 we predicted that a request for

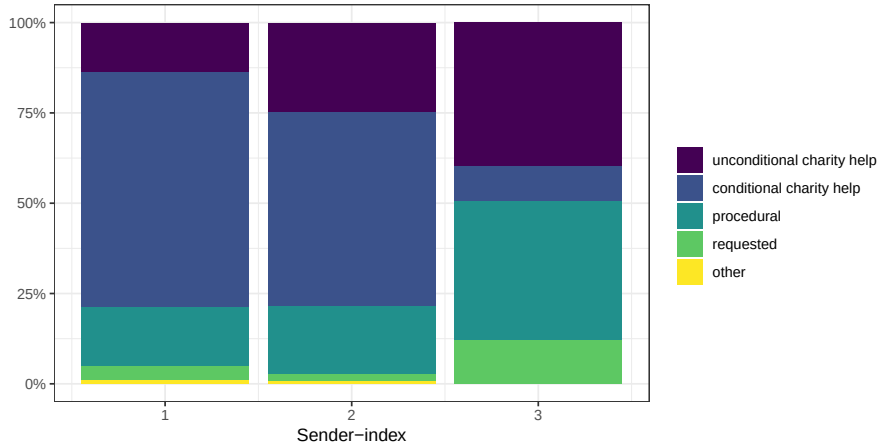


Figure 5: Self-reported motivations to send, by sender-index

information is associated with higher earnings for the charity, and request for ignorance with lower earnings for the charity. We also predicted amplified effects under the threat of punishment. Figure 6 shows the average transfer to the charity from the Receiver in the different treatments. The left panel examines the aggregate effect of the treatment. The results show no difference in the average payoff of the Red Cross, and therefore we cannot reject the null hypothesis that the aggregated outcome is the same across the three treatments ($F(2, 753) = 0.87$, $p = 0.419$).[9] So overall, the average outcome for the charity does not change across treatments.

Dis-aggregating the *Request* and *Request + Punishment* treatments by the Receiver's request allows us to see whether asking for information is correlated with the payoff of the charity. The right panel of Figure 6 and the OLS regressions in Table C1 suggest this. Senders who ask for ignorance seem to be causing more harm to the charity. However, a Wald test comparing the means across the five groups gives a non-significant result ($F(4, 751) = 1.70$, $p = 0.148$).[10] Overall, we cannot confirm Hypothesis 4.

[9] Comparing the distribution of payoffs leads to the same conclusion ($\chi^2(8) = 4.55$, $p = 0.804$).

[10] Comparing the distributions, instead, shows a significant difference in the outcomes for the charity ($\chi^2(16) = 34.13$, $p = 0.005$), but this test is not the most suitable as some cells have less than 5 observations.

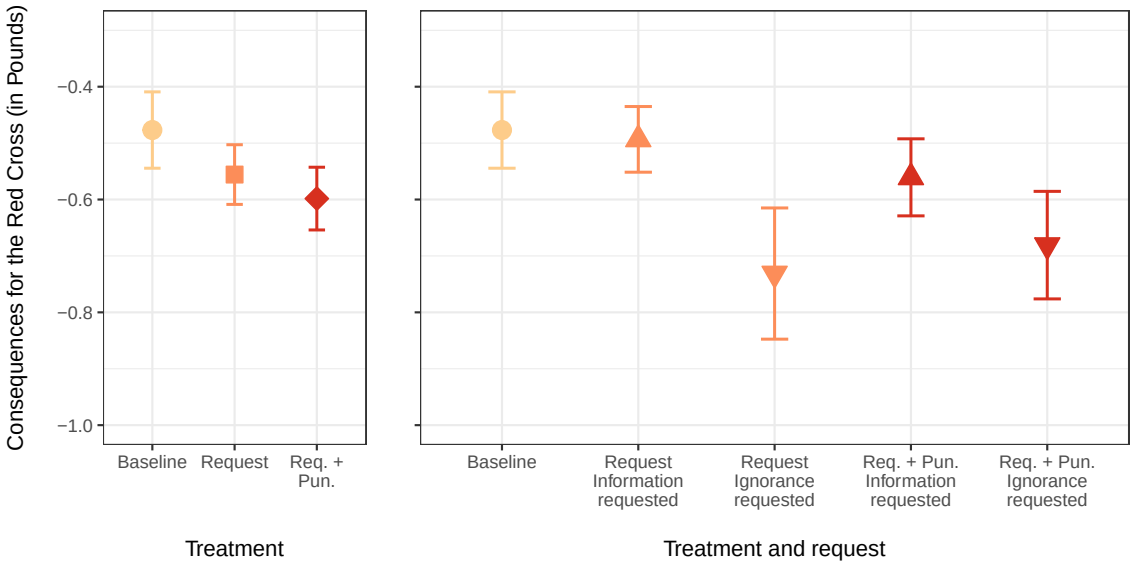


Figure 6: Consequence for the Charity. Average transfer of the Receiver (means and SE).

4 Conclusion

We investigated the willingness to share inconvenient information in the laboratory. The key take-away that emerges from our dataset is that Senders share inconvenient news out of concern for the otherwise negative consequences. The Sender's own preferences for information play a large role in both the Sending decision and in the response to requests, thus showing that people share information that they consider in their own decision making. We find some evidence that a minority of Senders consider the feelings of the receiver, either by transmitting positive information about impact or by suppressing negative information. If these results replicate in the field, it implies that there is scope for inconvenient information to spread in society or organizations. However, the results also point to the limits of information sharing. The fact that people share what they think is valuable for themselves and are willing to reduce their sharing upon request, implies that people will share information if they know the receiver does not like it.

There are a number of avenues for further research to address the limitations of the current study. One interesting extension would be to consider less anonymous interactions between Senders and Receivers. Here, there may be more scope for conflict and hence more self-censorship by Senders. Second, one could look at different audiences: perhaps people would be more motivated to share information to multiple Receivers, as the potential impact is bigger. One could also look at more extensive sharing networks to understand how inconvenient information spreads, perhaps testing predictions in Bénabou et al. (2020). Finally, one might look at various formats for the information sharing, perhaps also including advice on the decision, which is the focus of Coffman and Gotthard-Real (2019).

Acknowledgments

We would like to thank Yves Le Yaouanq, Lenka Fiala, as well as conference participants at ESA Online Global 2021, TIBER 2021, ICSD 2022 and ESA World 2023 for valuable comments. Sam Walsh, Britt van der Elsken, Ruben Bijl and Paulina Schulte-Vels provided excellent research assistance. This research has received financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 865931) and the Dutch Research Council (NWO; Vi.Vidi.195.137 and 452.17.004).

References

Ariely, D., Bracha, A., and Meier, S. (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99(1):544–555.

Bénabou, R., Falk, A., and Tirole, J. (2020). Narratives, Imperatives, and Moral Persuasion. *Working Paper*, (249429).

Castagnetti, A. and Schmacker, R. (2022). Protecting the ego: Motivated information selection and updating. *European Economic Review*, 142:104007.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Coffman, L. C. and Gotthard-Real, A. (2019). Moral Perceptions of Advised Actions. *Management Science*, 65(8):3904–3927.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

De Groeve, B. and Rosenfeld, D. L. (2022). Morally admirable or moralistically deplorable? a theoretical framework for understanding character judgments of vegan advocates. *Appetite*, 168:105693.

Foerster, M. and van der Weele, J. J. (2021). Casting Doubt: Image Concerns and the Communication of Social Impact. *The Economic Journal*, (February):1–33.

Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information avoidance. *Journal of economic literature*, 55(1):96–135.

Grossman, Z. (2014). Strategic Ignorance and the Robustness of Social Preferences. *Management Science*, 60(11):2659–2665.

Grossman, Z. and van der Weele, J. J. (2017). Self-Image and Willful Ignorance in Social Decisions. *Journal of the European Economic Association*, 15(1):173–217.

Lane, T. (2022). Intrinsic preferences for unhappy news. *Journal of Economic Behavior & Organization*, 202:119–130.

Lind, J. T., Nyborg, K., and Pauls, A. (2019). Save the planet or close your eyes? Testing strategic ignorance in a charity context. *Ecological Economics*, 161(March):9–19.

MacInnis, C. C. and Hodson, G. (2017). It ain't easy eating greens: Evidence of bias toward vegetarians and vegans from both source and target. *Group Processes & Intergroup Relations*, 20(6):721–744.

Sicherman, N., Loewenstein, G., Seppi, D. J., and Utkus, S. P. (2016). Financial attention. *The Review of Financial Studies*, 29(4):863–897.

Soraperra, I., van der Weele, J., Villeval, M. C., and Shalvi, S. (2023). The social construction of ignorance: Experimental evidence. *Games and Economic Behavior*.

Vu, L., Soraperra, I., Leib, M., der Weele, J. J. V., and Shalvi, S. (2023). Ignorance by choice: A meta-analytic review of the underlying motives of willful ignorance and its consequences. *Psychological Bulletin*, In press.

Appendix A Sender-index figure

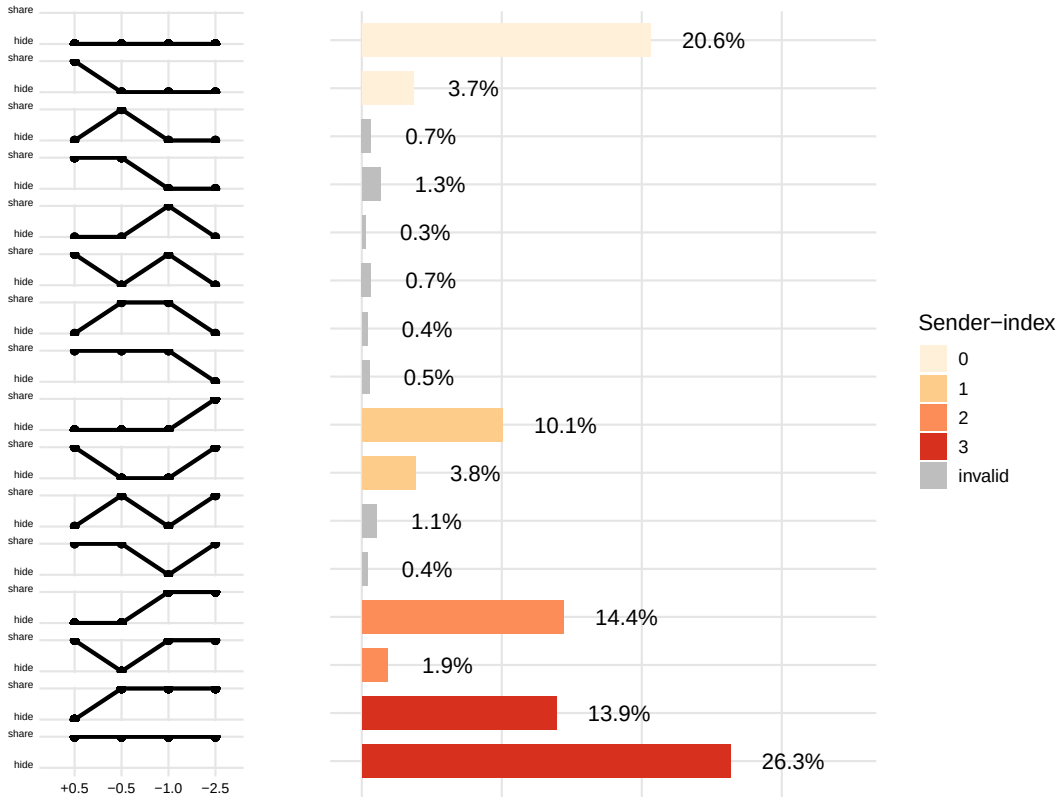


Figure A1: Percentage of Senders by switching structure.

Appendix B Sender-index regressions - Robustness checks

Table B1: OLS regressions of Sender-index

	Dependent variable: Sender-index				
	(1)	(2)	(3)	(4)	(5)
Constant	1.252**	1.714***	1.225***	1.714***	1.164***
	(0.395)	(0.102)	(0.248)	(0.103)	(0.209)
Information preference					
Request Info		0.015	−0.010	0.015	−0.030
		(0.133)	(0.133)	(0.133)	(0.133)
Request Ignorance		−0.065	−0.166	−0.065	−0.190
		(0.175)	(0.176)	(0.175)	(0.176)
Request Info under punishment threat				0.122	0.086
				(0.137)	(0.140)
Request Ignorance under punishment threat				0.162	0.076
				(0.167)	(0.169)
Control variables					
Identify with charity	0.104**		0.105***		0.079***
	(0.036)		(0.021)		(0.017)
Revealed in DWK	0.162		0.251*		0.330***
	(0.215)		(0.124)		(0.097)
Observations	140	438	412	716	667
R^2	0.129	0.001	0.099	0.003	0.074

Notes: OLS model of the Sender-index. Model 1 includes observations in the Baseline treatment. Models 2 and 3 include observations in the Baseline and Request treatments. Models 4 and 5 include all observations. Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

Table B2: Tobit regressions of Sender-index

	Dependent variable: Sender-index				
	(1)	(2)	(3)	(4)	(5)
Constant	0.846	1.480***	0.842*	1.483***	0.771**
	(0.518)	(0.139)	(0.334)	(0.137)	(0.277)
Information preference					
Request Info		0.028	−0.017	0.027	−0.042
		(0.179)	(0.177)	(0.177)	(0.176)
Request Ignorance		−0.114	−0.247	−0.113	−0.277
		(0.236)	(0.235)	(0.234)	(0.233)
Request Info under punishment threat				0.157	0.117
				(0.182)	(0.184)
Request Ignorance under punishment threat				0.177	0.071
				(0.222)	(0.222)
Control variables					
Identify with charity	0.130**		0.142***		0.104***
	(0.047)		(0.029)		(0.023)
Revealed in DWK	0.260		0.322*		0.410**
	(0.276)		(0.164)		(0.128)
Observations	140	438	412	716	667

Notes: Tobit model of the Sender-index (left = 0). Model 1 includes observations in the Baseline treatment. Models 2 and 3 include observations in the Baseline and Request treatments. Models 4 and 5 include all observations. Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

Table B3: Probit regressions of binary Sender-index

	Dependent variable: Sender-index binary				
	(1)	(2)	(3)	(4)	(5)
Constant	0.216	0.569***	0.307	0.569***	0.102
	(0.479)	(0.109)	(0.291)	(0.109)	(0.233)
Information preference					
Request Info		0.027	−0.030	0.027	−0.056
		(0.141)	(0.154)	(0.141)	(0.148)
Request Ignorance		−0.115	−0.253	−0.115	−0.253
		(0.182)	(0.200)	(0.182)	(0.191)
Request Info under punishment threat				0.039	0.021
				(0.143)	(0.156)
Request Ignorance under punishment threat				0.024	−0.066
				(0.176)	(0.192)
Control variables					
Identify with charity	0.082°		0.096***		0.059**
	(0.043)		(0.025)		(0.019)
Revealed in DWK	0.190		0.188		0.242*
	(0.260)		(0.144)		(0.109)
Log likelihood	-78.8	-273.1	-226.6	-449.1	-449.1
Pseudo R^2 (McFadden)	0.06	0.001	0.056	0.001	0.033
Covariates	Yes	Yes	Yes	Yes	Yes
Observations	143	456	412	756	703

Notes: Probit model of binary Sender-index (1 if sender sends worst outcome, 0 otherwise) - note that this includes the 40 responses which were considered invalid for the regular sender-index. Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Model 1 includes observations in the Baseline treatment. Models 2 and 3 include observations in the Baseline and Request treatments. Models 4 and 5 include all observations. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

Appendix C OLS regression of charity outcomes

Table C1: OLS regression of the outcome for the charity

	Dependent variable: Charity outcome in £	
	(1)	(2)
Constant	-0.477^{***}	-0.661^{***}
	(0.068)	(0.148)
Information preference		
Request Info	-0.254°	-0.207
	(0.134)	(0.137)
Request Ignorance	-0.084	-0.119
	(0.096)	(0.098)
Request Info under punishment threat	-0.017	-0.055
	(0.089)	(0.090)
Request Ignorance under punishment threat	-0.204°	-0.175
	(0.117)	(0.127)
Control variables (Receiver)		
Identify with charity		0.031^{*}
		(0.012)
Revealed in DWK		0.186^{*}
		(0.073)
Observations	756	713
R^2	0.009	0.036
Adjusted R^2	0.004	0.022

Notes: OLS regression of the outcome for the charity. Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

Appendix D Exploratory analysis of beliefs

In this appendix, we present an exploratory analysis of beliefs, focusing on two aspects. Firstly, we investigate whether Senders believe that Receivers who make different requests also exhibit different likelihoods of pressing the button. Secondly, we explore whether Senders' apparent disregard for the risk of punishment can be attributed to their belief that Receivers do not actually follow through with punishment when requests are not complied with.

The two leftmost panels of Figure D1 display the difference in the belief regarding pressing the button when sharing versus not sharing information, conditional on the treatment and the request of the Receiver. A value lower than 0 indicates that the Sender believes that the button is more likely to be pressed when uninformed compared to when informed. In other words, it means that sharing information is believed to reduce the likelihood of pressing the button.

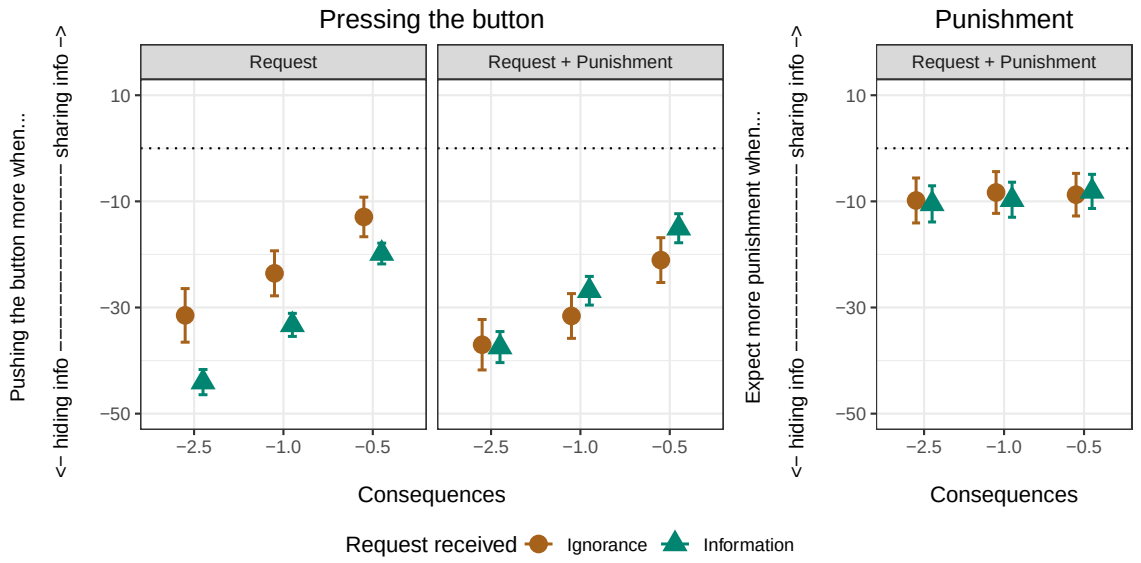


Figure D1: Difference in beliefs of pressing the button and difference in beliefs of punishment by treatment and consequences. Figure shows means and standard errors.

The results reveal that Senders believe that having information about more serious consequences reduces the likelihood of pressing the button. Furthermore, in the *Request* treatment, Senders believe that this effect is more pronounced for Receivers who ask for information compared to those who ask for ignorance. In the *Request + Punishment* treatment, instead, Senders believe that the likelihood of pressing the button does not change with the request made by the Receiver, which is somewhat surprising. These observations are supported by the results of linear regression models (1) to (4) reported in Table D1. All models employ the difference between the belief that the Receiver will press the button when sharing and the belief that the Receiver presses the button when not sharing as a dependent variable. Explanatory variables include the request of the receiver, two dummies capturing the consequences for the red cross, and the interactions. Model (2) and (4) include control variables as well. All models control for the correlation of the beliefs coming from the same Sender by including a random effects at individual level. As it is apparent from the results, the request of the Receiver has an impact on the beliefs in the *Request* treatment, but not in the *Request + Punishment* treatment.

The rightmost panel of Figure D1 displays the difference in the belief regarding being punished when sharing versus not sharing information conditional on the request of the Receiver. Here a value below 0 indicates that the Sender believes that punishment is more likely to occur when information is not shared, compared to when it is shared.

The figure shows that, in general, Senders believe that it is more likely to get punished for not sharing information. Moreover, the request seem to play a minor role in shaping beliefs. In other words, Senders believe that the decision to punish is not correlated with the request for information. These conclusions are supported by the results of linear regression models (5) and (6) reported in Table D1. Both models employ the difference between the beliefs to be punished for sharing and for not sharing information as a dependent variable. Explanatory variables include the request of the receiver, two dummies capturing the consequences for the red cross, and the interactions. Model (6) includes control variables as well. Both models control for the correlation of the beliefs coming from the same Sender by including a random effects at individual level. As it is apparent from the results, the request of the Receiver does not contribute explaining the Senders beliefs about punishment.

Table D1: OLS regression of the Risk Difference of the Senders' belief about pressing the button and punishment

	Dependent variable: Risk Difference of the beliefs					
	Push the button				Punishment	
	Req.	Req.	Req + Pun.	Req. + Pun.	Req + Pun.	Req. + Pun.
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-30.050^{***}	-36.730^{***}	-37.468^{***}	-30.431^{**}	-12.404^{**}	-18.766
	(5.135)	(9.723)	(4.637)	(11.304)	(4.232)	(12.047)
Request Information	-12.794^{*}	-11.167°	-0.347	-2.514	2.433	2.609
	(5.643)	(5.854)	(5.418)	(5.850)	(5.327)	(5.674)
Consequence -1.0	8.062^{***}	8.867^{***}	5.606^{**}	5.730^{**}	1.543	1.517
	(2.387)	(2.410)	(1.747)	(1.844)	(2.195)	(2.324)
Consequence -0.5	18.937^{***}	20.267^{***}	17.255^{***}	17.640^{***}	1.883	1.820
	(3.860)	(3.900)	(3.227)	(3.405)	(2.868)	(3.036)
Request Information × Consequence -1.0	2.649	2.196	5.195^{*}	5.649^{*}	-0.455	-0.143
	(2.685)	(2.728)	(2.350)	(2.488)	(2.826)	(3.027)
Request Information × Consequence -0.5	4.800	4.446	5.347	5.841	-0.708	-0.537
	(4.290)	(4.351)	(4.018)	(4.233)	(3.693)	(3.908)
Control variables (Receiver)						
Female		3.481		0.618		-2.844
		(3.823)		(4.452)		(4.971)
Age		0.109		-0.012		0.287
		(0.155)		(0.182)		(0.181)
Income[a]		-0.001		-0.002		-0.002
		(0.001)		(0.002)		(0.002)
Mobile/Tablet (ref = Desktop)		4.211		5.565		6.992
		(10.276)		(7.379)		(7.376)
Observations	915	852	900	828	900	828
R^2	0.241	0.264	0.169	0.178	0.001	0.006
Adjusted R^2	0.237	0.256	0.164	0.169	-0.004	-0.005

Notes: OLS regressions of the risk difference (RD) of the Senders' beliefs that the receiver presses the button (model 1 to 4) and that the receiver punishes. The dependent variable is the difference of the belief when sharing and when not sharing information. [a] Respondents could indicate their after-tax income category, starting at £0 to £999, increasing in steps of £1,000. Continuous values of income variables were constructed by setting the income value of each respondent to the midpoint of the interval. £4,500 was used for the highest income category (>£4,000) Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

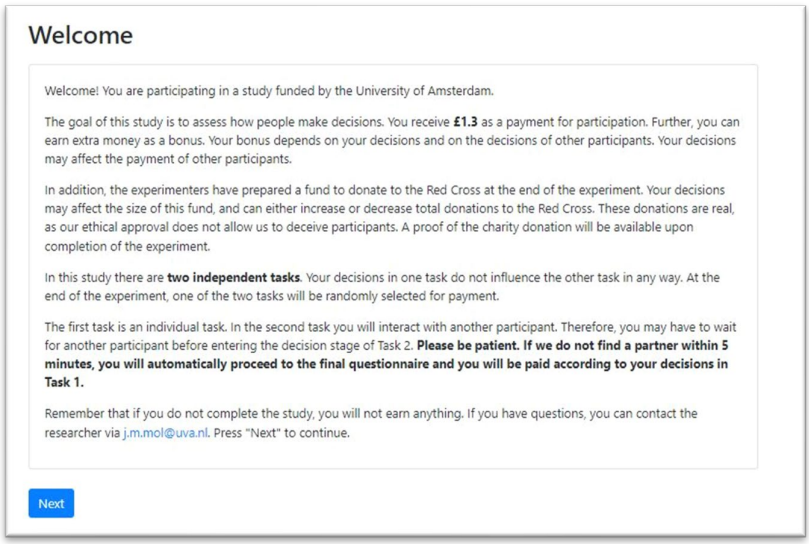
Table D2 examines the posterior beliefs of the Sender about their Receiver pressing a button. The first column shows that on average, Senders believe that Receivers will press the button in 86.1% of the cases when they are uninformed. This belief decreases with the size of the consequence, from 73.3% for the least serious consequence (-0.5 pounds) to 62.9% for -1.0 pounds and almost half (51.5%) for -2.5 pounds. These beliefs are not affected by received requests, not even under punishment threat.

Table D2: Posterior belief of button press, by consequence

	Dependent variable: Posterior belief of button press			
	uninformed	consequence -0.5	consequence -1.0	consequence -2.5
Constant	86.127**	73.290**	62.905**	51.497**
	(3.090)	(4.246)	(4.565)	(5.104)
Information preference				
Request Info	3.182	−0.404	−2.976	−3.110
	(1.999)	(2.746)	(2.952)	(3.301)
Request Ignorance	2.784	1.278	−1.412	−3.928
	(2.636)	(3.622)	(3.893)	(4.353)
Request Info under punishment threat	−0.124	0.215	−3.482	−3.014
	(2.070)	(2.844)	(3.057)	(3.419)
Request Ignorance under punishment threat	0.327	−3.473	−3.733	−8.249*
	(2.508)	(3.446)	(3.704)	(4.142)
Control variables				
Identify with charity	−0.151	−0.514	−0.383	−0.207
	(0.253)	(0.348)	(0.374)	(0.418)
Revealed in DWK	2.407°	−2.348	−2.401	−2.537
	(1.444)	(1.985)	(2.133)	(2.386)
Observations	703	703	703	703
R^2	0.046	0.051	0.029	0.018
Adjusted R^2	0.031	0.036	0.013	0.002

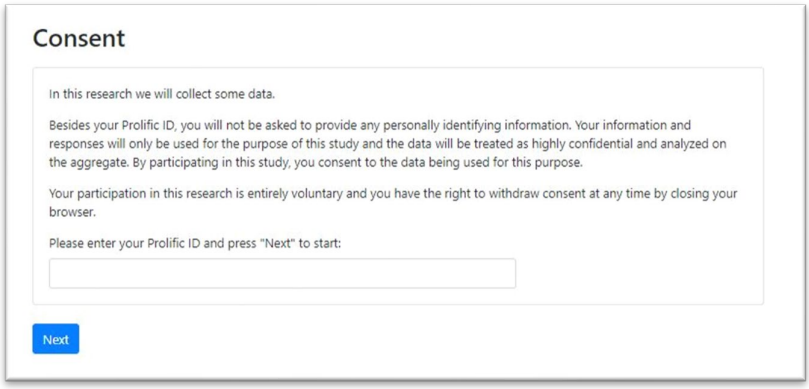
Notes: Linear probability models of posterior beliefs of button pressing with panel specification (plm package). Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

Appendix E Experimental instructions



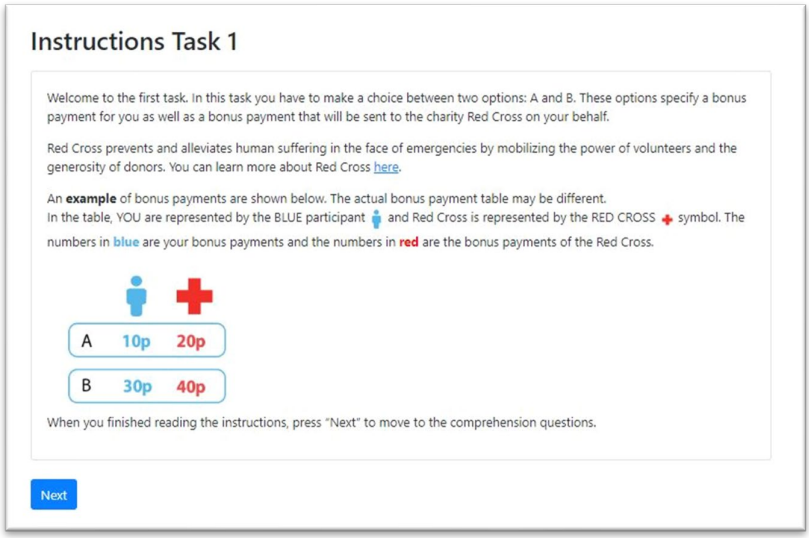
Welcome

Welcome page, including information about payoffs and the waiting app between the two games.



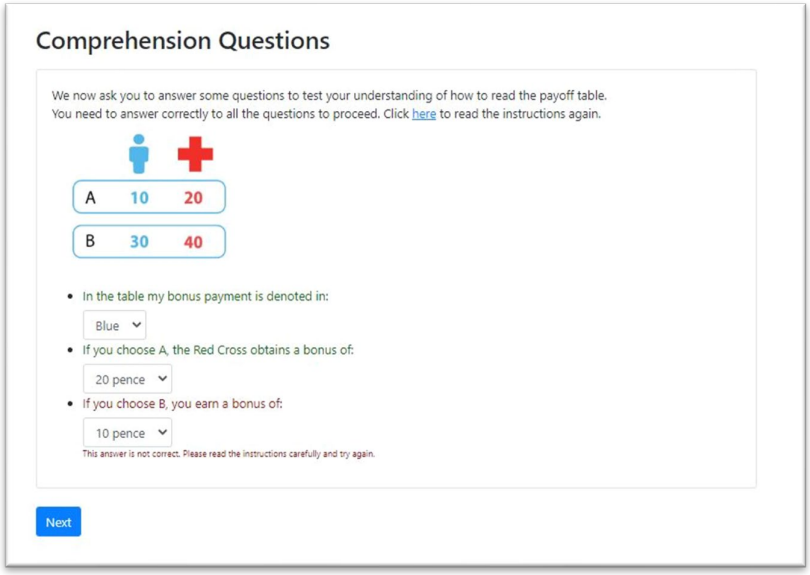
Consent

Informed consent page.
Also asking for Prolific ID for payment purposes.



Instructions DWK (1)

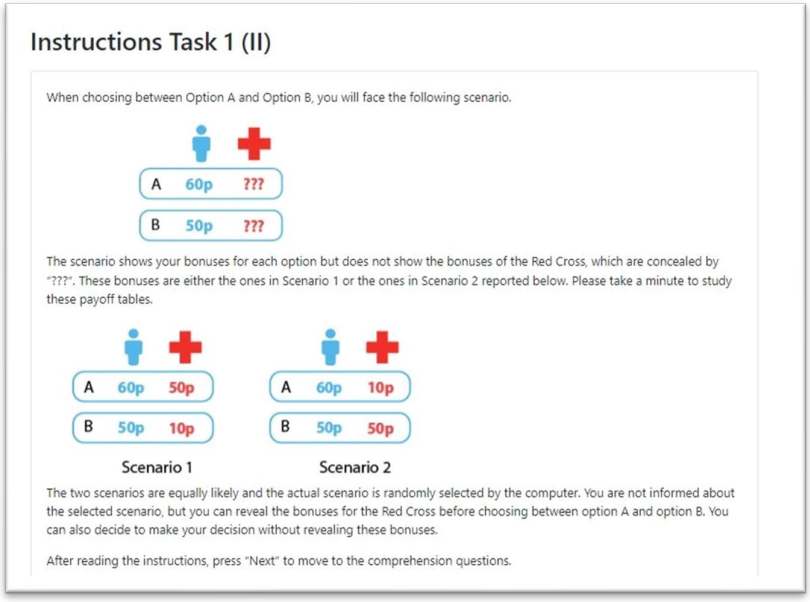
First instructions page for Dana Weber Kuang dictator game.



Comprehension Questions

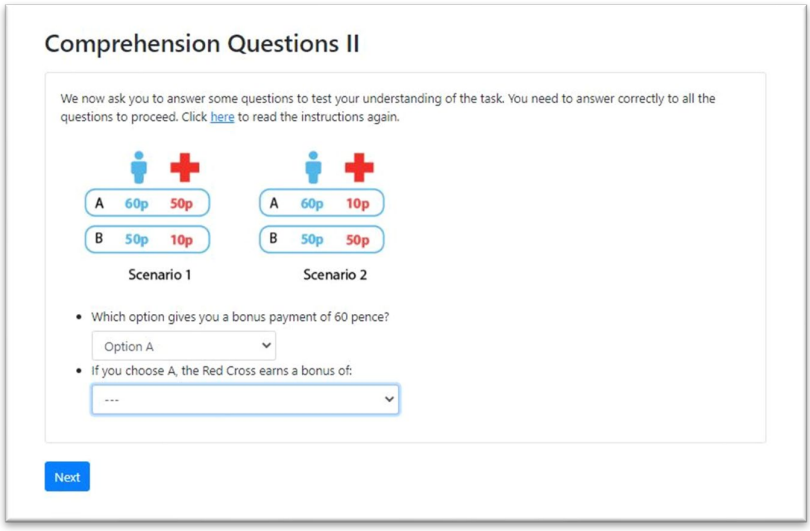
Comprehension questions for DWK dictator game. Correct answers:

- Blue
- 20 pence
- 30 pence



Instructions DWK (2)

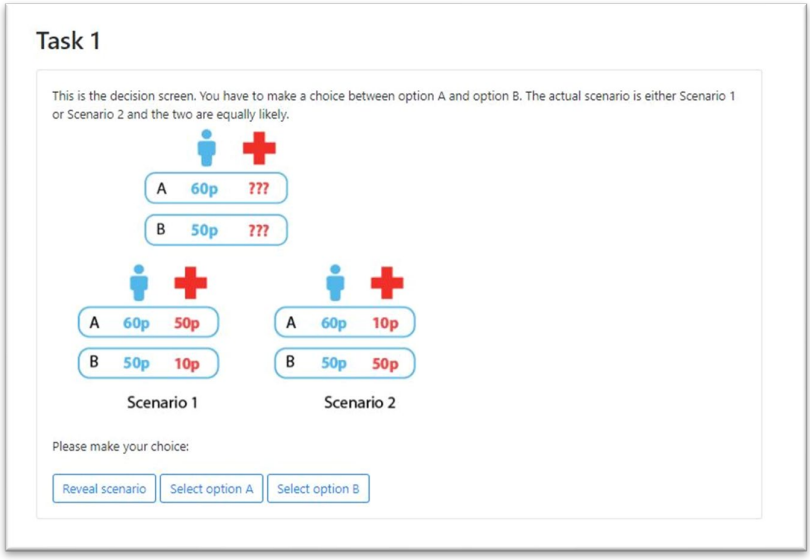
Second instructions page for Dana Weber Kuang dictator game.



Comprehension Questions (2)

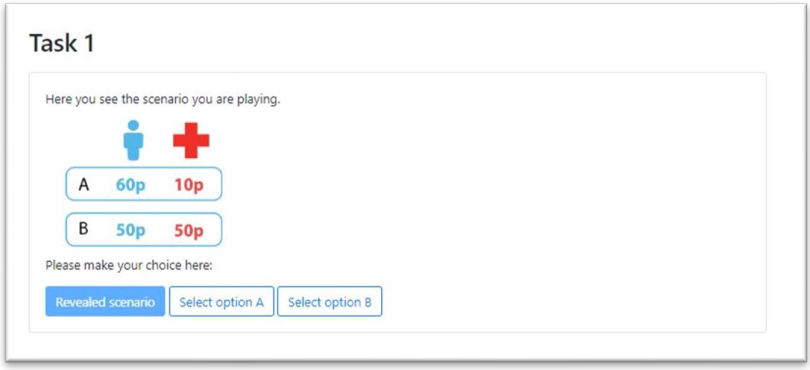
More comprehension questions after the second set of instructions. Correct answers:

- Option A
- Either 50 pence or 10 pence, depending on the scenario



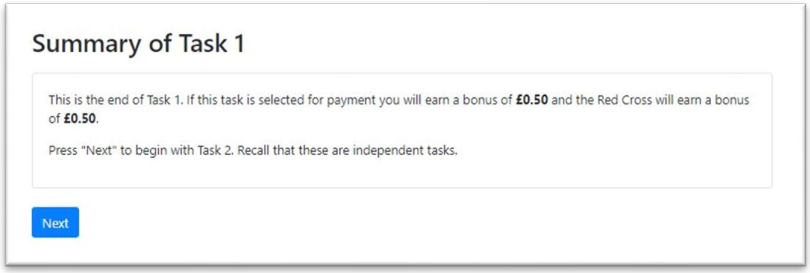
DWK dictator game

Main page for DWK dictator game.



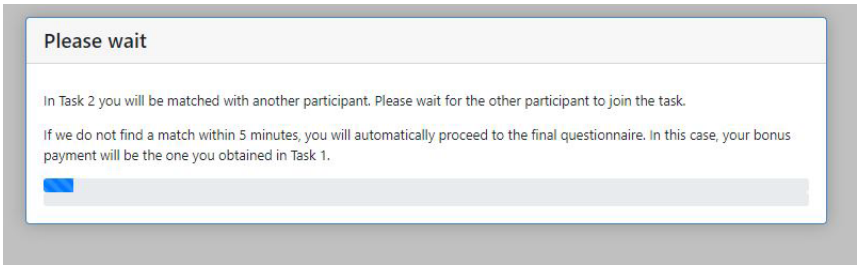
Revealed page

Only shown if a participant decided to reveal the selected scenario. Participant still has to select the preferred scenario under the revealed payoffs for both parties.



DWK summary

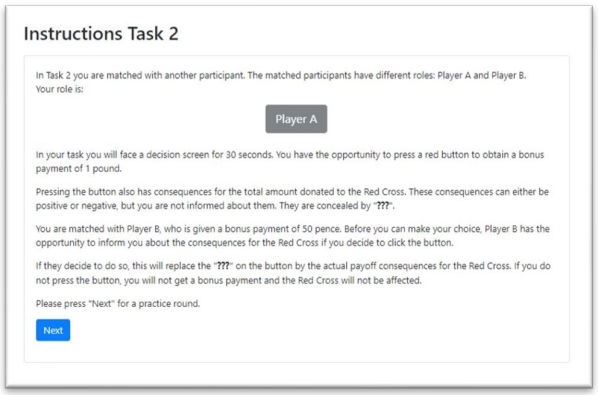
Results of the DWK dictator game.

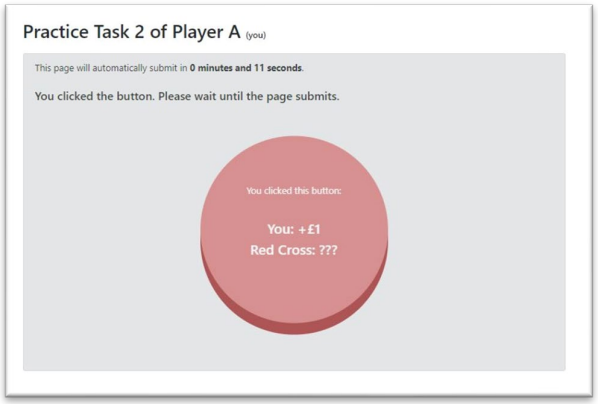


Waitpage for next game

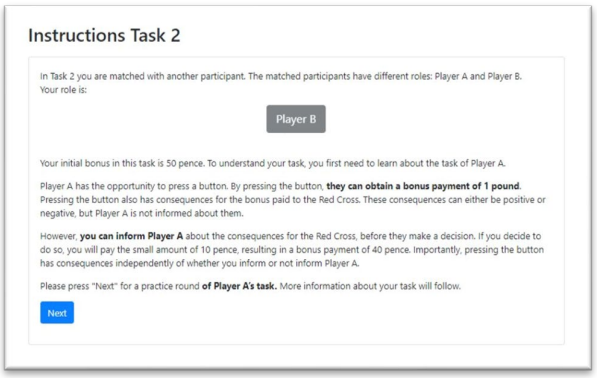
Page with progress bar of 5 minutes to allow matching in pairs for the Button Game.

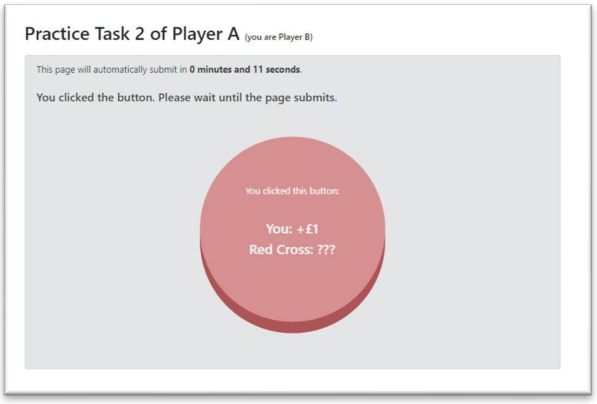
Button Game for the Receiver (A)

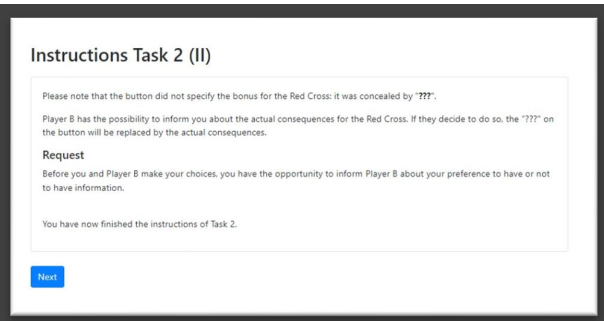




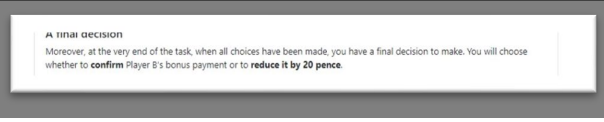
Button game for the Sender (B)







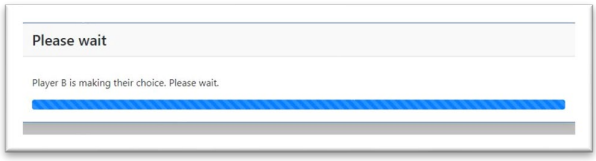
(only in Request and Request + Punishment treatments)



(only in Request + Punishment treatment)

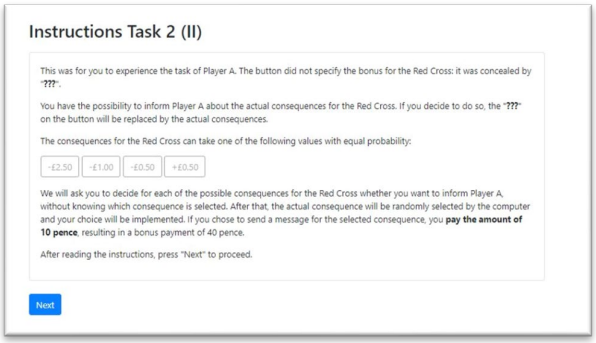
Button Game for the Receiver (A)

Button game for the Sender (B)



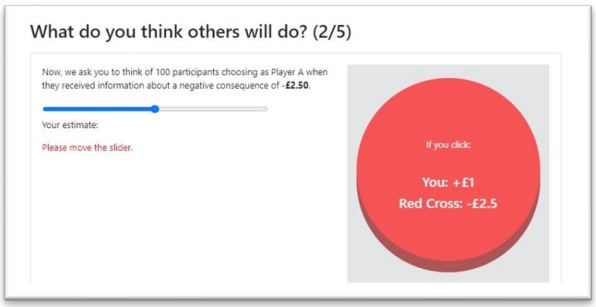
While the Receiver waits, the Sender completes the instructions, comprehension questions, beliefs and sending choice. Screen capture on the right shows: belief elicitation Sender for an uninformed Receiver.

Extra instructions for the Sender



Belief elicitation of the Sender

Slider for each of the 4 possible consequences (-£2.5. -£1.0, -£0.5, +£0.5) Slider must be moved.

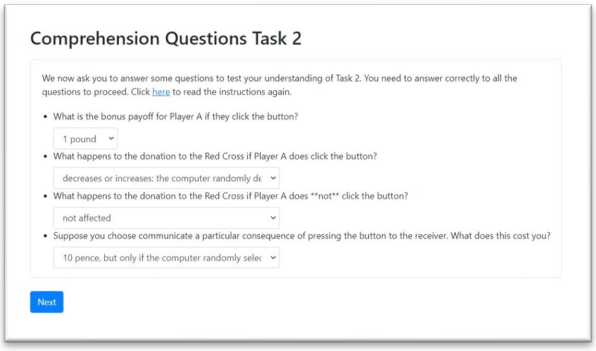


Button Game for the Receiver (A)

Button game for the Sender (B)

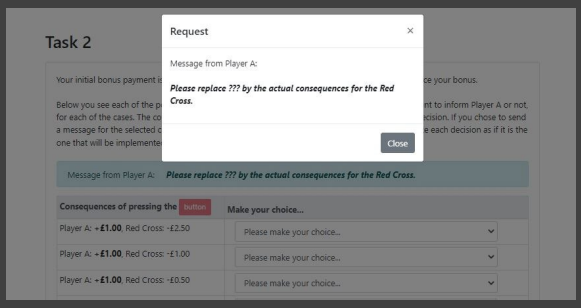
Comprehension Questions of the Sender

Correct answers currently selected.



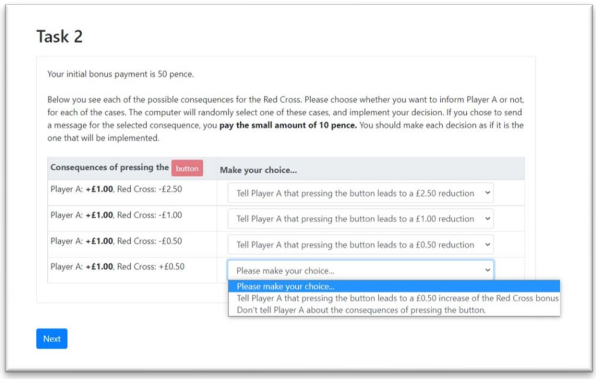
(only in Request and Request + Punishment treatments)

Request for the Sender

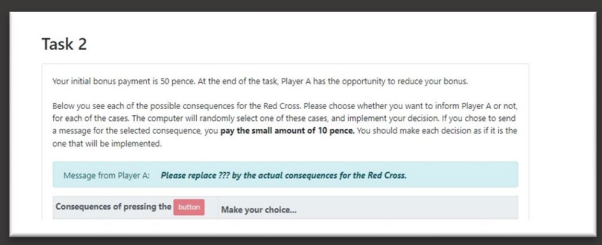


(Baseline treatment interface)

Interface of the Sending decision



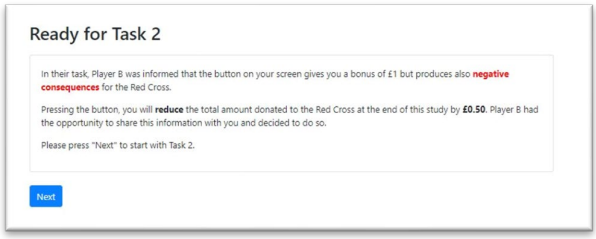
(additional blue request bar in Request and Request + Punishment treatments)



Button Game for the Receiver (A)

Button game for the Sender (B)

Information sharing from Sender to Receiver



The Button Page

The Button page (if clicked)

Button Game for the Receiver (A)

only in Baseline and Request treatments

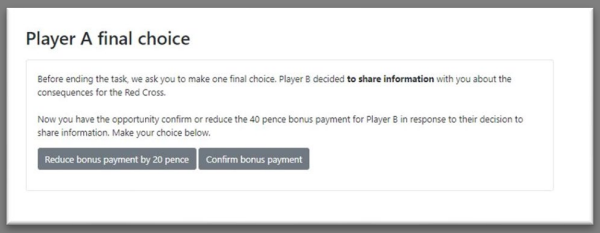
Button game for the Sender (B)

Belief elicitation button pressing

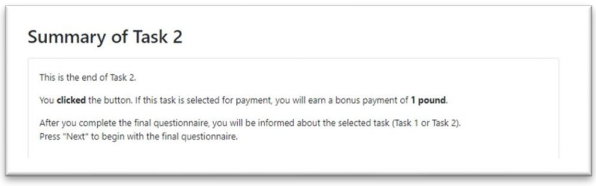
only in Request + Punishment treatment

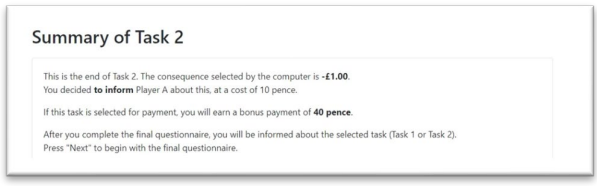
Belief elicitation button pressing and punishment

Punishment page



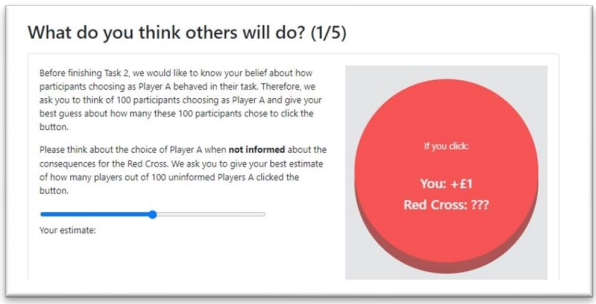
only in Request + Punishment treatment



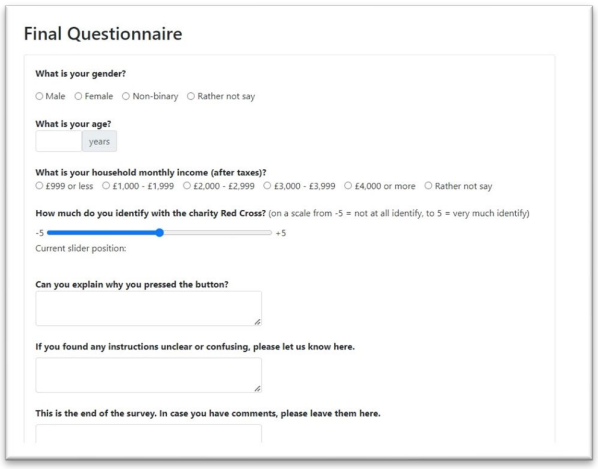


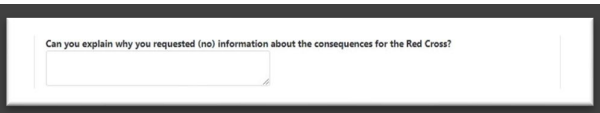
Belief elicitation of the Receiver

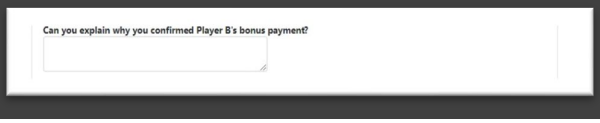
Slider for each of the 4 possible consequences (-£2.5. -£1.0, -£0.5, +£0.5) Slider must be moved.

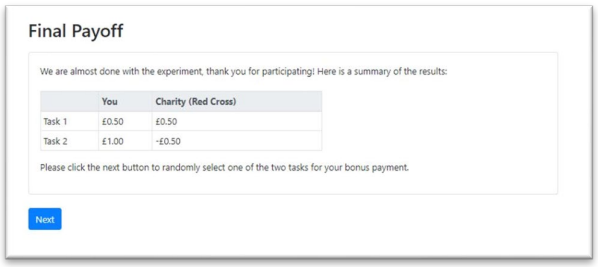


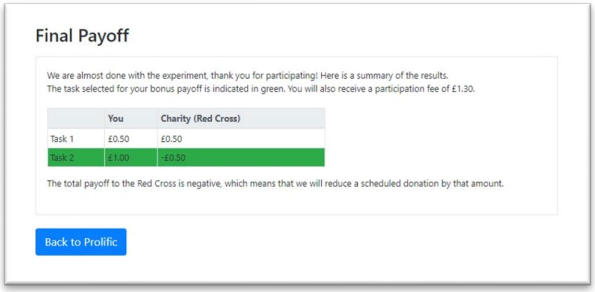
Button Game for the Receiver (A)



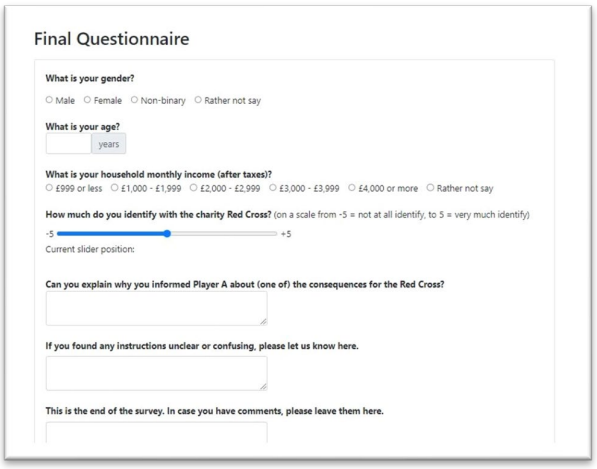






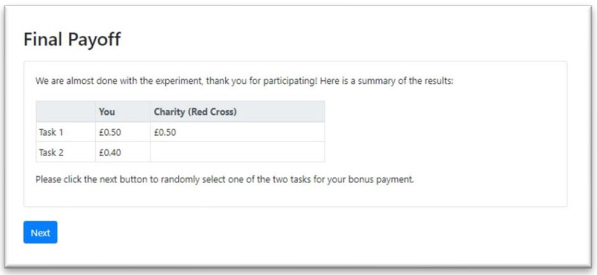


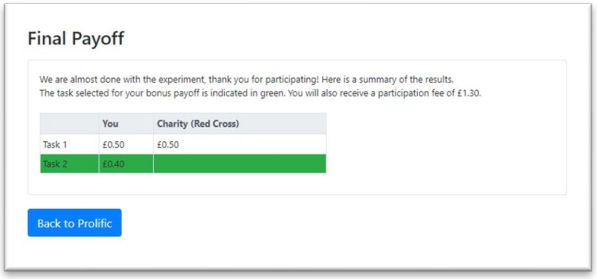
Button game for the Sender (B)



(only in Request and Request + Punishment treatments)

only in Request + Punishment treatment





Appendix F Preregistration

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Spoiling the party. On the willingness to transmit inconvenient ethical info (#111393)

Created: 11/01/2022 07:37 AM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review.
A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

---

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Are people willing to send others 'inconvenient' information about the consequences of their actions. Do they withhold such information under social pressure?

3) Describe the key dependent variable(s) specifying how they will be measured.
- We consider an interaction between a sender and a receiver. The receiver can take an action that is personally profitable, but leads to an unknown monetary impact for a charity (the Red Cross). The sender can send precise information about these consequences to the receiver at a small cost. Using the strategy method, each sender makes decisions to send (yes/no) for three negative impact levels (-2.5, -1.0 and -0.5 pounds), and one positive impact level (+0.5).
- Our first key interest is in the sender's decision to send information about negative consequences (i.e. we ignore sender's decisions for positive consequences). For each sender, we define a "sender-index" that measures how much information they send for different impact levels (0=no information sent for any impact level, 1=information sent only for the worst impact level, 2 =information sent for the worst two impact levels, 3=information sent for all three impact levels).
- Our second outcome variable is the payoff for the Red Cross resulting from the task.

4) How many and which conditions will participants be assigned to?
We have 3 between-subject treatments:
1) Baseline: after sending, senders face no further interaction with the recipient
2) Request: before the sender's decision, receivers make a request to senders for information or non-information. These requests thus create two groups of senders, one with a request for information and one with a request for ignorance.
3) Request + Punishment: in addition to making the request, receivers can punish senders by reducing their payoffs by a small amount, after they learn whether information had been sent.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
H1: People send inconvenient information about the charity to their partners.
We evaluate the sender-index in the Baseline treatment. We regress the sender-index using ordered probit, on a number of control variables that measure their preferences for information and for the charity, as well as demographics (gender, age, income, identification with the charity, time spent in experiment, number of attempts to get comprehension questions correct, reveal behavior in DWK task).
H2: The sender-index increases with request for information and decreases with request for ignorance.
We use a non-parametric Jonckheere trend test to test the alternative hypothesis that the sender-index follows: Request Info > Baseline > Request Ignorance. In addition, we use parametric ordered probit regressions to examine robustness of this effect, under several control variables (see above).
H3: The possibility of receiver punishment increases the impact of requests on the sender-index.
We use a non-parametric Wilcoxon rank-sum test to test the alternative hypothesis that the sender-index is: Request Info Punishment > Request info and Request Ignorance Punishment < Request Ignorance. In addition, we use parametric ordered logit regressions examine robustness of this effect under several control variables (see above).
H4: Request for information is associated with higher earnings for the charity, and request for ignorance with lower earnings for the charity. These effects are amplified in the punishment treatment.
We will use a chi-square test to compare the distribution of payoffs across all treatments, as well as split by receiver request (three and five groups, respectively). We will also do an OLS regression with the payoffs for the charity as a dependent variable and as explanatory variables a dummy equal to one when a request for information is made, a dummy equal to one in the punishment treatment, and their interaction.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude subjects who are not able to answer basic understanding questions. We will exclude interactions in which the sender-index cannot be computed because senders switch twice (e.g. no/yes/no or yes/no/yes for the three negative consequences).

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will recruit 750 senders (150 in Baseline, 300 in Request, and 300 in Request + Punishment) and 750 receivers. Sampling will stop when we have 750 pairs that completed the main task. We collect twice the amount of observation in the Request and Request + Punishment treatment to be able to analyze the behavior of senders conditional on the request for information or ignorance. The sample size allows us to detect an effect on the sender-index of size d = 0.30 based on one-sided Mann-Whitney Rank Sum Test (under the assumption of normality of the distributions, power = 0.8, and alpha = 0.05).

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

- We will do robustness checks of the results of the sender-index using parametric regressions with different specifications: OLS and Tobit.
- We will do robustness checks with an alternative sender-index, which is 1 if the sender sends information for the worst outcome, and zero otherwise.
- We will investigate the role of beliefs about receiver behavior in sender decisions. Specifically, we will elicit sender's beliefs about receivers behavior towards the Red Cross as well as their expectations of receivers to implement punishment in the Punishment treatments.
- For the receivers, we will correlate their request behavior with their willingness to reveal information in an individual decision making task with hidden information, as well as the identification with the charity, age, income and gender.
- We conducted a pilot session of the Request + Punishment treatment to test the software, the comprehension of the task, the size of the incentives, and to assess the proportion of people requesting information to determine the sample size.