TI 2023-062/I
Tinbergen Institute Discussion Paper

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

Revision: November 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]

November 2024

Abstract

Information about the consequences of our consumption choices can be unwelcome, and people sometimes avoid it. Thus, when people possess information that is inconvenient for *another* person, they may face a dilemma about whether to inform them. We introduce a simple and portable experimental game to analyze the transmission of inconvenient information. In this game, a Sender can, at a small cost, inform a Receiver 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 when negative externalities are larger and that Sender's decisions are largely driven by their own preferences towards the charity and their own use of information. We do not find evidence that Senders take the Receiver's preferences into account, as they largely ignore explicit requests for information or ignorance, even if Receivers have the option to punish the Sender.

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 sometimes try 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 may have important consequences for everyday consumption behavior, such as the decision to buy products that have adverse impacts on the environment or are manufactured in exploitative conditions (Ehrich and Irwin, 2005; Amasino et al., 2024).

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 carnivorous 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 know the truth, whatever they end up doing. Finally, she may hold back information out of consideration for her friends' feelings. 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 occur within politics, organizations or markets. For instance, politicians may have to decide whether to inform their voters about difficult trade-offs. Employees who have knowledge of organizational practices with negative external consequences, must decided whether to pass it up the decision-making chain. In buyer-seller interactions, buyers may voluntarily disclose ethical information about their products. Understanding the decision to share inconvenient information matters as it affects the spread of ethically relevant information.

To study the trade-offs facing a sender of information, we designed 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, virtually 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. 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, but only when externalities are relatively large. This indicates that some Senders trade off the payoff for the charity with the cost of sending. We also find evidence that personal preferences for inconvenient information, measured on a separate task, explain sharing. Finally, procedural concerns matter, as some senders share even if it does not change the Receiver's decision, and almost 30% of the Senders that share information say explicitly that this 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 Sender's information about the Receiver's preferences. Before the Sender makes a decision, the Receiver can request either information or ignorance. We find no evidence that Senders respond to the Receiver's preferences, as neither the request for ignorance nor the request for information significantly affects information sharing. 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 by denying part of the Sender's participation payment. The threat of punishment does not make either type of request more effective, even though we observe some punishment by Receivers.

The key takeaway that emerges from our dataset is that sharing of inconvenient information is driven by the Sender's personal attitudes towards information and the externality. To the extent the findings from our stylized setting capture behavior outside the lab, the prevalence of sharing shows most people are motivated to share unethical information when it can have a significant impact. Nevertheless, the results also indicate the limits of sharing. The central role of Sender's own preferences for information in the decision to share suggests that sharing will be less prevalent for topics in which people are widely averse to information, like in the meat-eating example above (Epperson and Gerster, 2024; Onwezen and van der Weele, 2016).

Our paper contributes to a fast-growing experimental literature on information avoidance in ethical dilemmas (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. Another closely related study is Vellani et al. (2024). In an online experiment, they examine the motives of sharing potentially unpleasant information about monetary losses for the receiver. The results, which are in line with ours, show that participants use their own information-seeking preferences when deciding to share such information with others. Our study instead focuses on information sharing in the ethical domain.

A number of further studies look at information transmission. Lind et al. (2019) allow Senders to force ethical 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. In our paper, information hasno instrumental value, but our finding that Senders do not cater to the preferences of the Receiver is in line with Lane (2022).

Our paper also has a link to research on paternalism. In particular, Ambuehl et al. (2021) find that people engage in an "ideals-projective" paternalism, where they assume their preferences are relevant for others, and restrict others' options accordingly. While Senders in our study do not restrict any options, we do find that the Sender's own preferences for information and the charity are the main predictors of what they share with others.

The main contribution of our paper to this literature is to introduce a simple and

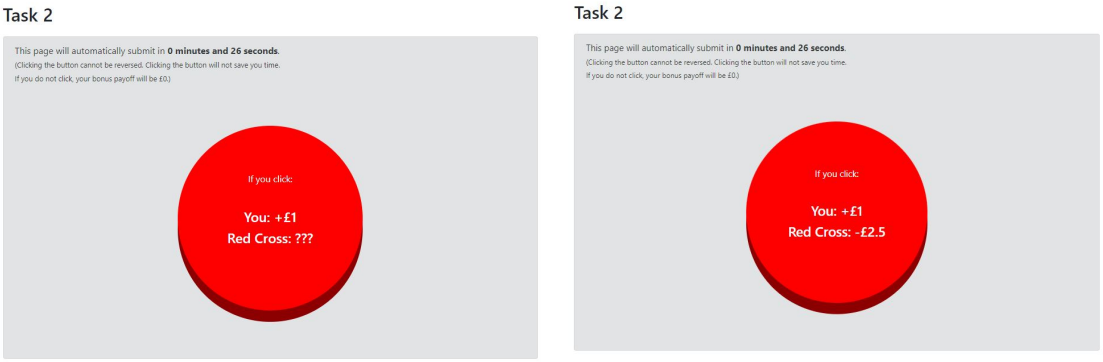
portable setting to analyze the transmission of inconvenient information. We offer new evidence on the determinants of sharing decisions and the willingness of Senders to accommodate the Receiver's information avoidance.

2 Method and Experimental Design

The experiment consists of two tasks and a final survey. The first task measures participants' preferences for information in an adaptation of the binary dictator game in 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

Every participant played the binary dictator game, regardless of their role (Sender or Receiver) in the Button Game. The binary dictator game is inspired by the Hidden Information treatment proposed by Dana et al. (2007), with all participants acting as dictators and a charity as recipient as in Lind et al. (2019). 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.



(a) Uninformed (b) Informed

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

2.2 The Button Game

As the main task, we designed a two-person game in which a Receiver interacts with a Sender. The Sender possesses superior information about the consequences of the Receiver's action for a third party. The Sender can inform the Receiver before the latter chooses an action. 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. It was designed to be attractive to press: large and red. Pressing the button pays a bonus of £1 to the Receiver. In addition, it has consequences for a third party, the Red Cross, which range from +0.5 to -2.5 pounds. Crucially, the Receiver has no information about the charity payoffs; neither about the actual value nor about the possible values.[1] Not pressing the button means that the Receiver will not get

[1] From the instructions: "*Pressing the button also has consequences for the total amount donated to the Red Cross. These consequences can be either positive or negative, but you are not informed about them. They are concealed by ???*". A complete set of screenshots of all instructions can be found in Appendix E

a bonus payment, but ensures that the Red Cross will not be affected. To avoid Receivers pressing the button in order to speed up the experiment, subjects still have to spend the remaining time on the page. During this time, the button press cannot be reversed. On top of the button, we displayed the bonus of £1, alongside the payoff consequences of the charity – depending on the decision of the Sender.

The Sender 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 (see Appendix E for screenshots). 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). If the Sender decided to send information for a certain impact level and that level was randomly selected for implementation, the Sender's payoff for participation was reduced from £0.50 to £0.40. We tested understanding of these consequences with a comprehension question. A complete set of screenshots of all instructions and comprehension questions can be found in Appendix E.[2]

The button game was designed to keep the strategic aspects of the Sender's decision relatively simple. In particular, inspired by the sender-receiver game in Gneezy (2005), we kept the information to the Receiver about the payoff consequences for the charity down to a minimum. This feature encourages the Receiver to press the button in the absence of information. Moreover, since the Receiver does not know what kind of information

[2]The decision environment features several aspects that are likely to affect the size of sending. This includes the size of the externality and the size of the cost of sending. Moreover, the binary nature of the sending decision and the use of the strategy method may induce some experimenter demand effect by suggesting that sending is important, at least for some externality levels. We cannot test whether these elements affect senders' decisions, but since they are kept constant in the experiment, they should not affect our comparisons between treatments.

the Sender can communicate, it limits the degree to which the Receiver can form beliefs about the externality in the absence of information, or form higher-order beliefs about the Sender's intentions. This simplifies the analysis, where we will (mostly) abstract from such higher-order beliefs. It also simplifies the Sender's decision problem, as she can assume that Receivers will press the button without information. To make sure that the Sender understands the decision environment of the Receiver, both Senders and Receivers start the game with an (unincentived) practice round, where they can choose to press the button as an uninformed Receiver.

2.3 Timeline and treatments

Figure 2 shows the timeline of the Button Game and highlights the differences between the treatments. The software randomly allocates participants between the roles of Senders and Receivers. At the start of the game, all participants play a test round as an uninformed Receiver. In the *Baseline* treatment, the Sender moves first and decides whether or not to inform the Receiver. After the decision of the Sender, the software randomly selects one of the four possible consequences for the charity. The information about these consequences is transmitted to the Receiver (or not), depending on the Sender's decision. If it is transmitted, it is displayed at the top of the red button.

The Receiver thus decides whether to press the button with or without information about the consequences for the Red Cross, depending on the decision of the Sender for the selected consequence. The *Request* 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 message 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

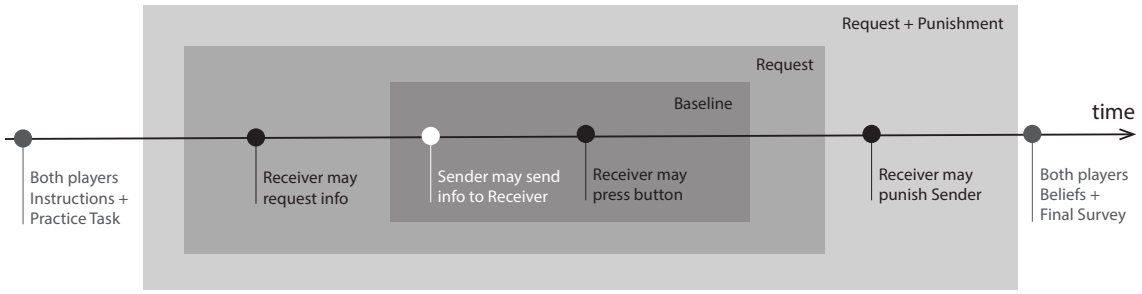


Figure 2: Timeline of the different variants of the Button game.

chooses whether to confirm or 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'. Finally, we administer a closing questionnaire.

2.4 Hypotheses

Here, we explain how we interpret the treatment differences and discuss our hypotheses. We preregistered the hypotheses prior to data collection.[3]

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 the Sender's beliefs about Receivers. As for the Receivers, we expect that virtually all of them press the button when uninformed, given that they earn a sure £1 bonus and the consequence to the Red Cross is ambiguous and possibly positive. 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

[3]For the preregistration, see Appendix F orhttps://aspredicted.org/X8Y_Q7T. Our hypotheses are not based on a formal model. 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 feature makes it unfeasible to model this as a Bayesian (disclosure) game.

to give up £1 to avoid a given level of 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 (mostly) ignore Sender's decisions for the +0.5 pounds.[4] Moreover, we expect Senders to understand that in the absence of information, Receivers will press the button. This means that sending information about positive consequences is unlikely to make a difference to the outcome for the charity, although it may help the Receiver feel better about her choice.

For each sender, we define a "sender-index" that measures the point at which consequences for the Red Cross become too large 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) a concern for the charity, (2) procedural reasons like the belief that Receivers ought to make an informed choice, or (3) the desire to help the Receiver, combined with a belief the Receiver would like to be informed. Accordingly, our first hypothesis is that a non-negligible fraction of Senders decides to share inconvenient information.

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

[4]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.

and are more likely to do so as the externality becomes more negative.[5]

To hypothesize the impact of the *Request* treatment, we consider both requests for information and for ignorance. The former request is straightforward to interpret: the main reason Receivers would like information is in order to decide whether to push the button. A request for information is, therefore, a signal that the information is likely to be used by the Receiver. Since sending helps both the Receiver and the charity, we thus expect the Sender to increase the likelihood of sending information compared to the *Baseline*.

The effect of a request for ignorance is more complex. First, it may change the Sender's beliefs about the impact of information on the charity. The request may be a signal that the Receiver will not use the information, which may make the Sender less willing to send it. There is a caveat to this reasoning however: the literature on moral wiggle room shows that a sizable fraction of subjects who choose to avoid information would nevertheless use it when they are confronted by it (Dana et al., 2007; Grossman and van der Weele, 2017). To the extent the Sender anticipates this, she may still perceive the potential impact on the charity to outweigh the cost of sending. To better understand how requests change beliefs, we therefore measure the Sender's beliefs about the Receiver's action in each condition. In addition, if the Sender cares about helping the Receiver, who expressed a wish for ignorance, one would expect the Sender to be more likely to suppress information.

Taken together, these considerations lead us to expect that Senders' decisions follow the direction of the request:

Hypothesis 2 *Relative to the Baseline treatment, the likelihood to share inconvenient information increases with a request for information and decreases with a request for ignorance.*

[5]The second part of this hypothesis was not preregistered but is implied by our use of the "sender-index", as explained above.

Finally, the comparison of the *Request + Punishment* is meant to amplify the strength of the *Request* treatment . 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 to follow the request of the Receiver more often. Furthermore, to the degree that the request affects the perceived cost of sharing information, the *Request + Punishment* treatment provides a measure of the cost sensitivity 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 nature of the request, when obliged by Senders, will be correlated with 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 a request for ignorance with lower earnings for the charity. These effects are amplified in the punishment treatment.*

2.5 Procedure

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 end of the experiment, 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 by moving a slider from 0-100 (*"'we ask you to think of 100 participants choosing as player A and give your best guess about how many of these chose to press the button").* To keep the game simple and payment quick, beliefs were not incentivized. Next, Senders 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 their Receiver pressed the button for each possible consequence, again unincentivized and on a scale from 0-100. In the *Request + Punishment* treatment, Senders were further asked to guess the likelihood of punishment. Receivers were also asked about their belief of other players A pressing the button. This page was identical to the Sender's first belief elicitation page, but it was placed after the Receiver's own choice to avoid spillover effects. At the very 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.

The main study was run on Prolific in November 2022, where 1,796 participants started

the study. 71 of them dropped out before starting the DWK task, 5 could not be matched with another player, 171 dropped out during the DWK task, and 37 participants finished the study without a partner, leaving N = 1,512 responses (84.2% completion rate) for the analysis ($n_{Baseline}$ = 302, $n_{Request}$ = 610, $n_{Request+Punishment}$ = 600). Due to practical constraints of the live matching into Sender-Receiver pairs, the treatments were run sequentially. To account for time effects, we started each treatment session approximately on the same time of day. 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. On the first page of the study, participants were informed about the payoffs to the Red Cross.[6] We did not inform participants about the size of the original fund (which was £100). 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.

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[7] via Prolific with a proof of the donation to the charity.

[6]The experimenters have prepared a fund to donate to the Red Cross at the end of the experiment. Your decisions may affect the size of this fund, and can either increase or decrease the total donations to the Red Cross. These donations are real, as our ethical approval does not allow us to deceive participants. A proof of the charity donation will be available upon completion of the experiment.

[7]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.

3 Results

3.1 Preliminaries

3.1.1 Randomization check

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 at different times and days of the week. 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 devices in the *Request + Punishment* treatment) are slightly unbalanced across treatments. To control for such differences, we added gender and device type as covariates in all further analyses.

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. As mentioned in the hypothesis section, this high rate is unsurprising, given the monetary payoff of pressing the button and the absence of any information about the charity.[8]

Moreover, all 61 Receivers that saw good news — i.e., saw that the button increased the

[8]Furthermore, uninformed Receivers may have inferred that the lack of information from the Sender constituted good news about the externality. However, given that Receivers had no information about the possible consequences that the Sender might send, the design discouraged such inferences.

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.

donation by an additional £0.5 — pressed the button. Finally, 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.[9] We will investigate the behavior of Receivers in more detail in Section 3.7.

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

3.1.3 Sender's beliefs

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 of pressing the button. Before making any decisions, Senders believe on average that 80.0% (SD = 18.1) of the Receivers press the button when not informed about the consequences. Table 2 regresses Senders' beliefs that Receivers will press the button, conditionally on being informed, for the different possible consequences. It reveals that Senders' beliefs about Receivers pressing the button decline as the severity of negative consequences increases, compared to uninformed Receivers across all treatments. By contrast, Senders expect Receivers to press the button about 5 percentage points more often when they are informed about positive consequences. This shows that, on average, Senders (correctly) believed that sharing information would be effective, and more so when the externality was more negative. In Section 3.5, we provide more details on the role of Sender beliefs in decision-making.

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 32.5% of Senders in the Baseline treatment. Sending information about a positive consequence may reassure Receivers that the charity benefits from their decision to press the button. However, there is a cost of sending information. Since most Senders (correctly) expect uninformed Receivers to press the button regardless, this can explain why most senders did not share this information. We discuss Sender's motives to send positive information further in Section 3.6.

Negative consequences of -0.5, -1.0, and -2.5 pounds are shared by 40.4%, 57.6%, and 71.5% of Senders, respectively. The differences between these proportions are statistically

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 ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

significant (Pairwise McNemar tests: consequence -0.5 vs consequence -1.0: $\chi^2(1) = 19.5$, $p < 0.001$; consequence -1.0 vs consequence -2.5: $\chi^2(1) = 17.4$, $p < 0.001$). Moreover, almost all Senders act consistently with a strategy where sharing small (negative) externalities implies sharing larger negative externalities. Only 40 out of 756 Senders decide to share information for less serious consequences and but not for more serious ones.

This provides a rationale for our (preregistered) use of a "sender-index", which reflects the smallest consequence for which the Senders decide to share information.[10] Figure 3 shows the distribution of the sender-index in the *Baseline* treatment. This shows that

[10]Following the preregistration protocol, we exclude the 40 non-monotonic participants from the analysis of the sender-index. 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.

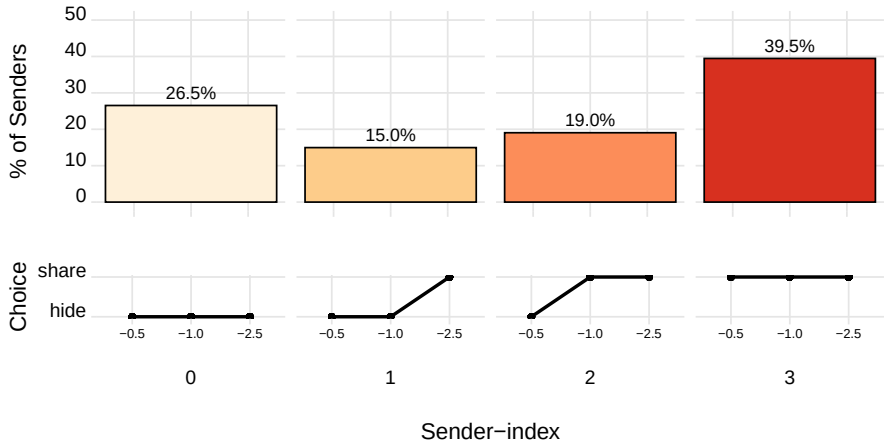


Figure 3: Distribution of the Sender-index in the *Baseline* treatment.

we have substantial heterogeneity in the preferences for sharing information among our participants. On the one hand, 26.5% of the Senders never share information with the Receiver (sender-index = 0), and another 39.5% of Senders always share information about the consequences (sender-index = 3). The remaining Senders have intermediate preferences and share only when consequences are sufficiently negative (sender-index 1 and 2).

Overall, these numbers show that the majority of the Senders trade off the cost of sharing with the potential consequenes for the charity, and 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. In this treatment, the majority of Senders (225; 73.8%) received a Request for information, while the rest (80; 26.2%) received a request for ignorance. Figure 4 (the three middle panels) shows the distribution of the sender-index across the various treatments. The Request and Punish-

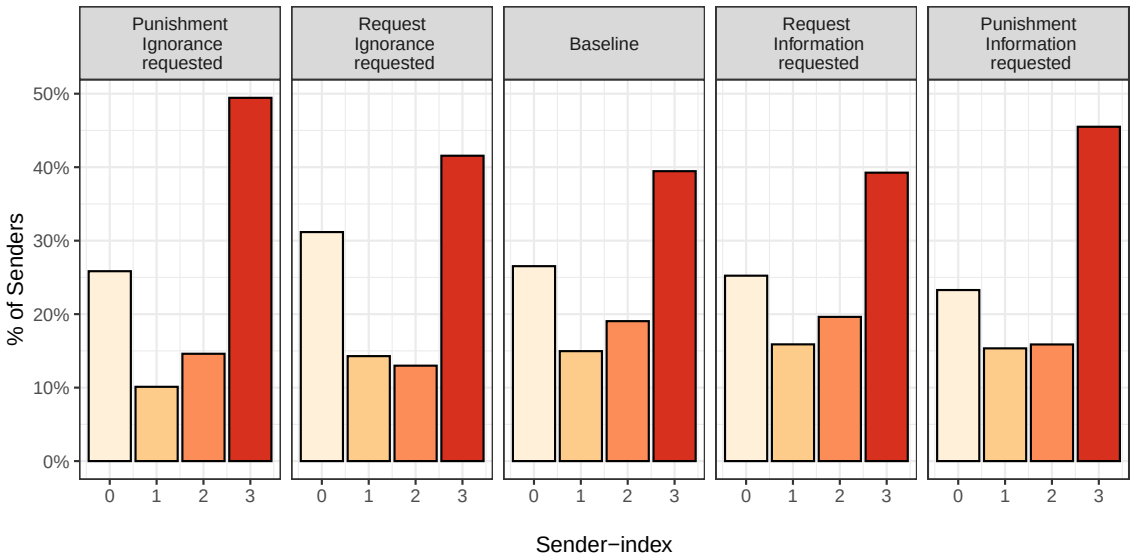


Figure 4: Distribution of the sender-index by treatment.

ment treatments are split by the nature of the request. According to our Hypothesis 2, we should observe an increase in the sender-index when the request is for information and a decrease when the request is for ignorance. However, the data do not show an increase in the frequency of Senders with higher sender-index values when transitioning from the left to the right panel in Figure 4. 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 is not present. Indeed, a non-parametric Jonckheere-Terpstra trend test fails to reject the hypothesis of no difference in the sender-index across different requests ($z = 0.31$, $p = 0.377$).[11]

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

[11] Testing the more general assumption of a difference across distributions also 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 ($\chi^2(6) = 2.38$, $p = 0.882$)

measure the Senders' preferences for information and their identification with the charity. We also include control variables such as gender, age, income, type of device used, 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* and *Request* data. It shows that requests for information have little effect, but requests for ignorance have a negative impact on the sender-index (compared to the Baseline without requests), although this is not statistically significant. Furthermore, sharing is positively related to how close the Senders feel to the charity, and whether they themselves revealed in the DWK game. This result is highly statistically significant, and shows that preferences about the information one would like to have for oneself play an important role in sharing information with others.

Table 3: Ordered probit regressions of Sender-index

	Dependent variable: Sender-index					
	(1)	(2)	(3)	(4)	(5)	(6)
Information preference (ref = Baseline)						
Request info	−0.011	−0.111	0.085	−0.028	−0.072	0.023
	(0.124)	(0.162)	(0.202)	(0.122)	(0.157)	(0.197)
Request ignorance	−0.130	−0.255	0.023	−0.150	−0.232	−0.063
	(0.172)	(0.227)	(0.272)	(0.169)	(0.219)	(0.273)
Request info under punishment threat				0.096	−0.048	0.356
				(0.130)	(0.164)	(0.224)
Request ignorance under punishment threat				0.083	0.097	−0.008
				(0.167)	(0.228)	(0.245)
Control variables						
Identify with charity	0.990***	1.234***	0.764*	0.731***	0.884***	0.555
	(0.214)	(0.269)	(0.365)	(0.167)	(0.212)	(0.287)
Revealed in DWK	0.257*			0.323***		
	(0.119)			(0.094)		
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 Baseline and Request treatments. Model 4, 5 and 6 include all participants across all treatments. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

To further understand the channels through which the request for ignorance affects the sender, we investigate whether senders are more likely to oblige when the request aligns with their own preferences for information. To test this, we run the same model restricting the data to those who remained ignorant in the DWK game (column 2) and those who informed themselves (column 3). The results show no significant interaction between the Sender's preferences and the request: while Senders who chose to remain ignorant are more likely to accommodate a request for ignorance, this effect is not statistically significant. Section 3.5 further examines the role of Sender beliefs in responding to such requests.

3.4 The effect of adding punishment

The *Request + Punishment* treatment allows us to test whether Senders stick to their preferences for sharing even when they risk punishment for not following the request (Hypothesis 3). Note that punishment rates were low, but not negligible. Requests were followed about half of the time, and deviations in responses to information requests were more likely to be punished (32%) compared to deviations after requests for ignorance (14.6%).[12] In line with the pre-registration, we test 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 the threat of punishment increases the sender-index when information is requested and decreases the sender-index when ignorance is requested compared to the *Request* treatment.

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 ($p = 0.164$).[13]

[12] 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 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).

[13] 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 ($\chi^2(12) = 7.58$, $p = 0.817$).

We also use regressions to investigate whether requests combined with the threat of punishment induce different patterns from the baseline. Column (4) of Table 3 provides results that are in line with the graphical and non-parametric evidence: we do not observe any significant effect of requests when punishment is present. Columns (5) and (6) further reveal that the Sender's preferences for information do not interact with the request when the threat of punishment is introduced. All treatment coefficients in these regressions are statistically insignificant at the 5% threshold.

3.5 The role of Sender beliefs

We previously showed that, prior to making a decision or learning about the requests, most Senders believed that Receivers will press the button when uninformed and that they are more likely to refrain from doing so when informed about the impact of externalities (see Table 2). In this section, we aim to better understand the role of beliefs in the decision to send information and in its lack of responsiveness to the requests. Appendix D provides the visual and statistical evidence corresponding to the claims in this section.

Did the nature of the request affect Sender's beliefs in the *Request* treatment? We find that the nature of the request does affect Sender's beliefs about the impact of sharing information on Receiver's behavior. Model (1) and (2) of Appendix Table D1 report regressions studying how the belief about the effect of sharing information — measured by the difference between the belief of pressing the button when informed and when uninformed — changes with the request and consequence levels. Focusing on the *Request* treatment and the largest externality level (-2.5 euros), Senders believe that sharing information when Receivers request it reduces the likelihood of Receivers pressing the button by almost 48 percentage points. This difference becomes smaller after receiving a request for ignorance: the estimated drop in the coefficient ranges from 9 to 11 percentage points,

depending on the model, which is statistically significant at the 10% level.

While requests for ignorance decrease Sender's expectations that the Receiver will use the information, Senders still expect information sharing to reduce button pressing by more than 37 percentage points. Thus Senders expect that Receivers who prefer to remain ignorant may nevertheless refrain from pushing the button when they are informed of the consequences. This expectation is partially correct (see Section 3.7) and helps explain the lack of response to requests.

Did the presence of punishment affect Sender beliefs? In Model (5) and (6) of Table D1, we look at the difference in the Senders' beliefs about getting punished when sharing and when not sharing information. Regression results show that Senders do not perceive much difference in the likelihood of being punished when sharing or when withholding information, as the constants in the model are not significantly different from zero (although expected punishment is slightly higher after withholding information). Moreover, the nature of the request barely moves this expectation, suggesting that Senders were not sure how to interpret the request of the Receiver in this treatment. This may explain why punishment was not effective in enforcing responses to the request.[14]

Do beliefs explain Sender decisions? To further understand the impact of beliefs on decisions, we regress the Senders' decisions on their beliefs that information makes a difference to the charity. We measure this as the decrease in the subjective Sender belief that the Receiver will press the button when informed, compared to being uninformed. We control for the treatments, the nature of the request, and the externality size.

The results of this exercise are presented in Appendix Table D2. There is a clear

[14]Moreover, the presence of punishment changes the Sender's beliefs about button pressing. As we discussed above, in the *Request* treatment, Senders correctly expected that information was less likely to have an impact on Receivers who requested ignorance. Model (3) and (4) in Table D1 and the left panel of Figure D1 show that this is no longer the case in the *Request + Punishment* treatment.

effect of Sender beliefs: for the case where the externality is -2.5 euros, i.e., Model (1), an increase of 1 percentage point in the Sender's beliefs that information will sway the Receiver's behavior is associated with a statistically significant increase in the likelihood to send information of about 0.138 percentage point. Similar effects are observed in Models (2) and (3) for the other negative externalities.

Interestingly, the effect of the Sender's identification with the charity remains statistically significant, even when controlling for beliefs about the impact on the charity. One interpretation of this result is that Senders who care about the charity wish to signal to themselves that, regardless of whether Receivers act on the information, they have fulfilled their "duty" by providing information about the consequences. The effect of the Senders' own preferences for information, as measured by revelation in the DWK task, is also robust, underscoring the conclusion that Senders want others to have information that they value for themselves.

3.6 Stated 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.

The results confirm the conclusions from the behavioral and belief data. 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 charity 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 simply know the truth. Indeed, we find that a large number of Senders mention such procedural concerns (28.8% "*it is the right thing to do*").

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. 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*"). 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*"), but in either case, there is only a small number of responses that contain such references.

To get a more complete overview of motives, 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*

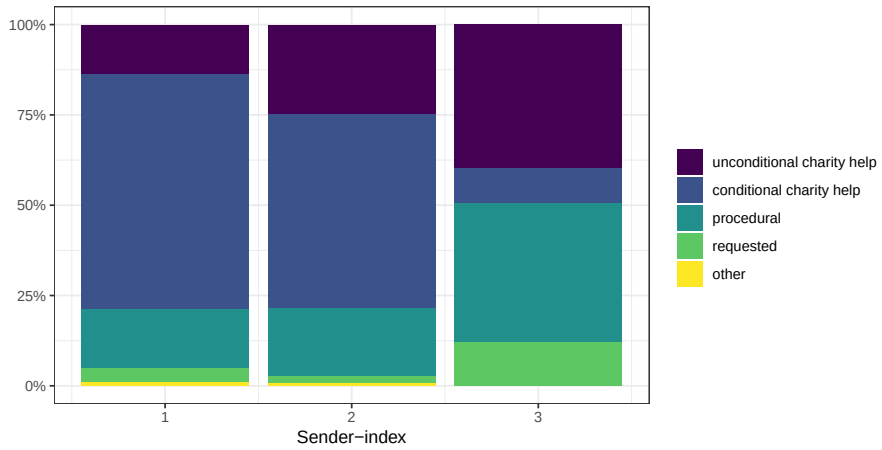


Figure 5: Self-reported motivations to send, by sender-index

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.

Overall, these responses reveal that some Senders were concerned with the Receiver's preferences. However, it does not appear to be a dominant motive, as witnessed by the muted responses to the requests. This is in line with evidence in Lane (2022), which shows that most Senders will send information about negative externalities, after Receivers have already made a decision, and the information is likely to have negative hedonic value. It is also consistent with Arrieta and Bolte (2023), who show that a majority of people think that having false beliefs is detrimental to a person's welfare.

3.7 Receiver behavior and consequences for the charity

We now analyze the consequences of sharing information for the charity. Figure 6 shows the average Receiver impact on charitable donations in the different treatments. The left panel examines the aggregate effect of the treatment. The results show relatively small differences in the average payoff of the Red Cross. Statistically, we cannot reject the null hypothesis that the aggregated outcome is the same across the three treatments ($F(2, 753) = 0.87$, $p = 0.419$).[15]

In Hypothesis 4, we predicted that a request for information is associated with higher earnings for the charity and a request for ignorance with lower earnings for the charity. To visually evaluate the hypothesis, we split the results for the charity at the request of the Receiver. The right panel of Figure 6 suggests that Receivers who ask for ignorance cause more harm to the charity. A test that the distribution of charity outcomes is the same across all five groups shows a significant difference in the outcomes for the charity ($\chi^2(16) = 34.13$, $p = 0.005$). The OLS regressions in Table C1 show a negative effect of requesting ignorance (relative to Baseline) in both *Request* and *Request + Punishment* treatments, although the coefficients are only significant at $p < 0.1$. Moreover, we do not find evidence that punishment amplifies the effect of requests.

Receiver selection and response to information. Above, we have shown that there are no statistical differences in Sender behavior in response to the request. Yet Figure 6 reveals that the impact on the charity varies with the request, suggesting that the Receivers' behavior correlates with the request they make. To understand this, we investigate the impact of both making a *request* for information and *actually* becoming informed on the behavior of Receivers. In Table 4, we regress the decision to press the button on dummies

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

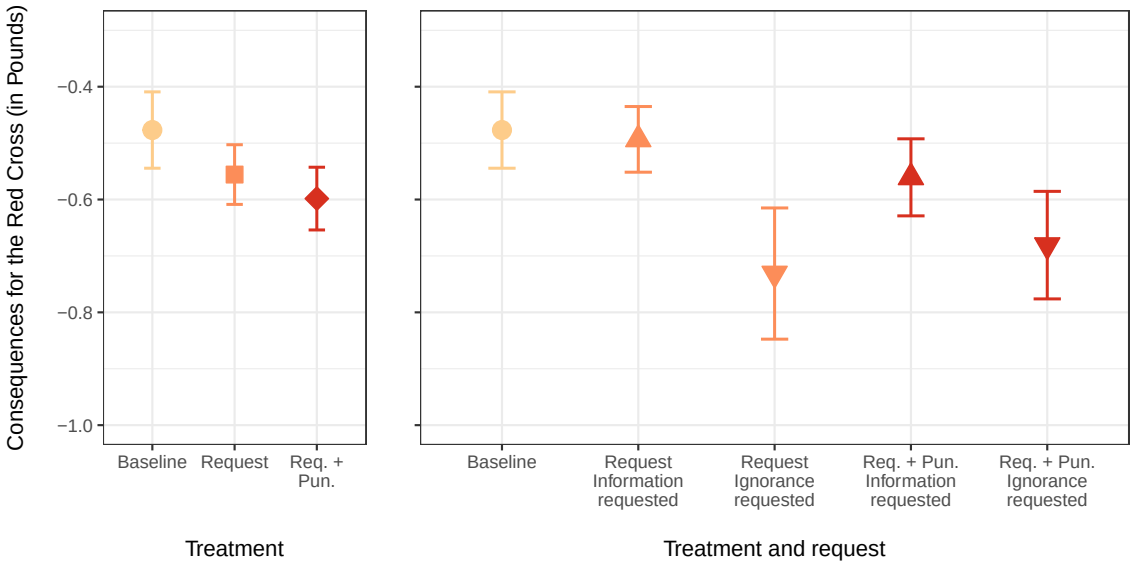


Figure 6: Consequences for the Charity. Average transfer of the Receiver to the charity fund (means and SE).

capturing information requested and received. To fully control for Senders' behavior, which varies with the size of the externality, we run separate regressions for each level of the externality. Moreover, we pool the data of the *Request* and *Request + Punishment* treatments to increase statistical power. Appendix Figure B1 shows the proportion of button clicks across treatment, type of request, and received message (again pooling the *Request* and *Request + Punishment* treatments).

The results show that uninformed Receivers have a similar behavior independently of their requests, as they universally press the button. Moreover, we observe that receiving information has a strong impact on pressing the button in the *Baseline* and that this impact decreases with the size of the externality, as the estimated parameter of "Information received" gets smaller moving from model (1) to Model (3). These results are in line with the raw data reported in Section 3.1.2.

Table 4: Receivers' button pressing by request and information obtained.

	Dependent variable: Button pressing		
	Cons. -2.5	Cons. -1.0	Cons. -0.5
	(1)	(2)	(3)
Constant	1.019^{***}	1.012^{***}	1.020^{***}
	(0.061)	(0.056)	(0.057)
Information preference (ref = Baseline)			
Request info	0.011	0.021	0.013
	(0.032)	(0.032)	(0.032)
Request Ignorance	0.016	0.019	0.026
	(0.030)	(0.030)	(0.030)
Information (ref = Ignorance)			
Information received	-0.546^{***}	-0.393^{***}	-0.192°
	(0.105)	(0.109)	(0.100)
Interactions			
Request Info × Information received	0.027	0.028	−0.126
	(0.123)	(0.127)	(0.117)
Request Ignorance × Information received	0.362^{**}	0.202	0.101
	(0.134)	(0.130)	(0.114)
Observations	447	447	442
R^2	0.374	0.223	0.181
Adjusted R^2	0.358	0.203	0.160

Notes: Dependent variable: Button pressed. Covariates suppressed for brevity: revealed in DWK, age, income, browser type, comprehension questions. Linear model with heteroscedasticity robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

When looking at the interaction of information preference and information received, we conclude that supplying information to those Receivers who request information has the same effect as supplying information in the Baseline. By contrast, supplying information to people who requested ignorance has a smaller impact than supplying information in the Baseline, principally because Receivers who request ignorance do not use the information as much. Nevertheless, Receivers who request ignorance are not wholly unresponsive to

information, as they do not press the button as often as uninformed Receivers. This pattern is in line with previous research (Dana et al., 2007), and and Senders seem to anticipate it (see Section 3.5).

In summary, our data show that a) information has a clear impact, as on average, Receivers want to avoid imposing negative externalities, and b) Receivers who request ignorance are less responsive to information and act more selfishly. Nevertheless, even among this last group, information does reduce button clicking.

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 are willing to pay to share inconvenient news out of concern for the otherwise negative consequences. The Senders' own preferences for information also play a significant role, thus showing that people share information that they consider in their own decision-making, in line with (Vellani et al., 2024). We find little evidence that Senders try to cater to the preferences of Receivers. In particular, we do not find that they respond to explicit requests for ignorance or information, even when there is a threat of punishment. Indeed, Senders correctly anticipate that sharing information will still lead to better results for the charity, even if Receivers asked to remain ignorant.

If these results replicate in the field, it implies that there is scope for inconvenient information to spread in society or organizations, as long as there are enough people who care about the affected third party. However, the results also point to the limits of information sharing. The fact that people share what they think is valuable for themselves suggests that people may end up in information silos. There is evidence that people dislike obtaining information that casts their behavior in a negative light (Golman et al., 2017; Vu

et al., 2023). If social networks are characterized by homophily, i.e. people interact with others who have similar preferences or behavior, this might lead to information bubbles, in line with results in Soraperra et al. (2023). For instance, returning to the example in the introduction, there is evidence that meat eaters do not like to receive information about the negative consequences of meat (Epperson and Gerster, 2024; Onwezen and van der Weele, 2016). To the degree that carnivores seek each other out, they are unlikely to hear about the negative impacts of meat production.

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, as in Foerster and van der Weele (2021). More generally, stronger forms of Receiver punishment or opportunities for conflict may induce more self-censorship by Senders. Second, one could look at different audiences: perhaps people would be more motivated to share information with 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 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 their 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).

Conflict of interest statement

The authors declare no conflict of interest.

References

Amasino, D. R., Oosterwijk, S., Sullivan, N. J., and van der Weele, J. J. (2024). Seeking or ignoring ethical certifications in consumer choice. Technical report, Tinbergen Institute Discussion Paper.

Ambuehl, S., Bernheim, B. D., and Ockenfels, A. (2021). What motivates paternalism? an experimental study. *American economic review*, 111(3):787–830.

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.

Arrieta, G. and Bolte, L. (2023). What you don't know may hurt you: A revealed preferences approach. *Available at SSRN*.

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.

Ehrich, K. R. and Irwin, J. R. (2005). Willful ignorance in the request for product attribute information. *Journal of Marketing Research*, 42(3):266–277.

Epperson, R. and Gerster, A. (2024). Willful ignorance and moral behavior. *Available at SSRN: 3938994*.

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.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95(1):384–394.

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.

Onwezen, M. C. and van der Weele, C. N. (2016). When indifference is ambivalence: Strategic ignorance about meat consumption. *Food Quality and Preference*, 52:96–105.

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*.

Vellani, V., Glickman, M., and Sharot, T. (2024). Three diverse motives for information sharing. *Communications Psychology*, 2(1):107.

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*,

149(9-10):611.

Appendix A Sender-index figure

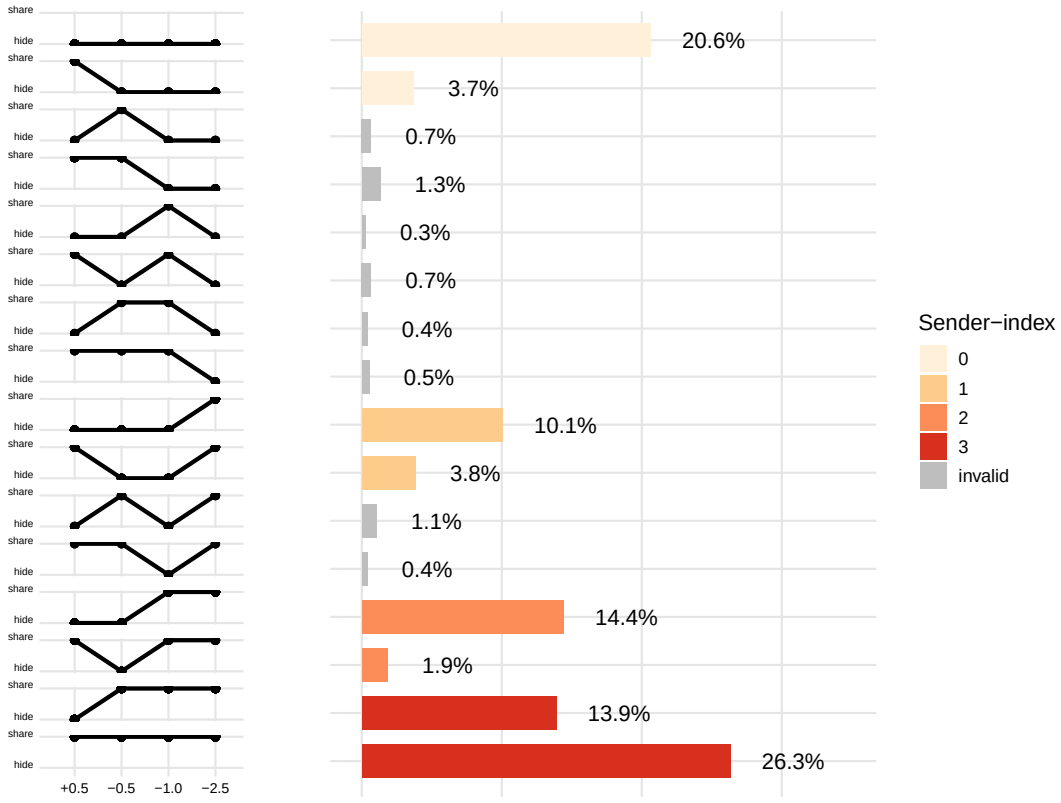


Figure A1: Percentage of Senders by switching structure.

Appendix B Additional results and 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 (ref = Baseline)					
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 (ref = Baseline)					
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 (ref = Baseline)					
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 ($^\circ p < 0.10$; $^* p < 0.05$; $^{**} p < 0.01$; $^{***}\ p < 0.001$).

Table B4: Ordered probit regressions of Sender-index, subsample of desktop users

	Dependent variable: Sender-index					
	(1)	(2)	(3)	(4)	(5)	(6)
Information preference (ref = Baseline)						
Request info	−0.005	−0.114	0.077	−0.027	−0.101	0.026
	(0.129)	(0.169)	(0.206)	(0.127)	(0.166)	(0.202)
Request ignorance	−0.142	−0.290	0.045	−0.169	−0.281	−0.027
	(0.175)	(0.231)	(0.282)	(0.174)	(0.225)	(0.281)
Request info under punishment threat				0.131	−0.034	0.312
				(0.137)	(0.176)	(0.230)
Request ignorance under punishment threat				0.072	0.063	−0.003
				(0.174)	(0.238)	(0.256)
Control variables						
Identify with charity	0.924***	1.133***	0.730*	0.721***	0.885***	0.538
	(0.216)	(0.272)	(0.369)	(0.173)	(0.223)	(0.291)
Revealed in DWK	0.241*			0.267**		
	(0.122)			(0.097)		
Log likelihood	-492.9	-292.4	-195	-769.4	-464.8	-299.1
Pseudo R^2 (McFadden)	0.034	0.045	0.022	0.028	0.033	0.016
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Revealed in DWK		No	Yes		No	Yes
Observations	386	228	158	606	359	247

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

Table B5: Ordered probit regressions of Sender-index, split by gender

	Dependent variable: Sender-index					
	(1)	(2)	(3)	(4)	(5)	(6)
Information preference (ref = Baseline)						
Request info	−0.011	0.151	−0.211	−0.028	0.145	−0.215
	(0.124)	(0.187)	(0.184)	(0.122)	(0.180)	(0.179)
Request ignorance	−0.130	−0.179	−0.077	−0.150	−0.185	−0.158
	(0.172)	(0.246)	(0.255)	(0.169)	(0.238)	(0.252)
Request info under punishment threat				0.096	−0.032	0.138
				(0.130)	(0.188)	(0.191)
Request ignorance under punishment threat				0.083	0.224	−0.081
				(0.167)	(0.242)	(0.242)
Control variables						
Identify with charity	0.990***	0.845**	1.223***	0.731***	0.788***	0.756**
	(0.214)	(0.301)	(0.308)	(0.167)	(0.232)	(0.238)
Female	0.089			0.143		
	(0.113)			(0.089)		
Log likelihood	-522	-255.5	-261.7	-842.8	-401.3	-435.8
Pseudo R^2 (McFadden)	0.038	0.045	0.045	0.029	0.037	0.03
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Gender	Both	Male	Female	Both	Male	Female
Observations	412	203	209	667	318	349

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

Table B6: Ordered probit regressions of Sender-index, probability weighted by gender and browser type

	Dependent variable: Sender-index	
	(1)	(2)
Information preference (ref = Baseline)		
Request info	−0.011	−0.025
	(0.122)	(0.119)
Request ignorance	−0.117	−0.124
	(0.166)	(0.166)
Request info under punishment threat		0.104
		(0.127)
Request ignorance under punishment threat		0.080
		(0.164)
Control variables		
Identify with charity	0.926***	0.667***
	(0.230)	(0.188)
Revealed in DWK	0.230	0.285**
	(0.130)	(0.103)
Observations	412	667

Ordered probit model with inverse probability weighting of gender and browser type. The propensity scores are obtained using a multinomial logit predicting the probability to be in each treatment as a function of Gender and browser type used (the variables that were unbalanced in our samples). In each treatment, the sample is rebalanced in order to match the overall frequency of females and mobile browser users observed in the study. Covariates suppressed for brevity: revealed in DWK, age, income, browser type, and comprehension questions. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

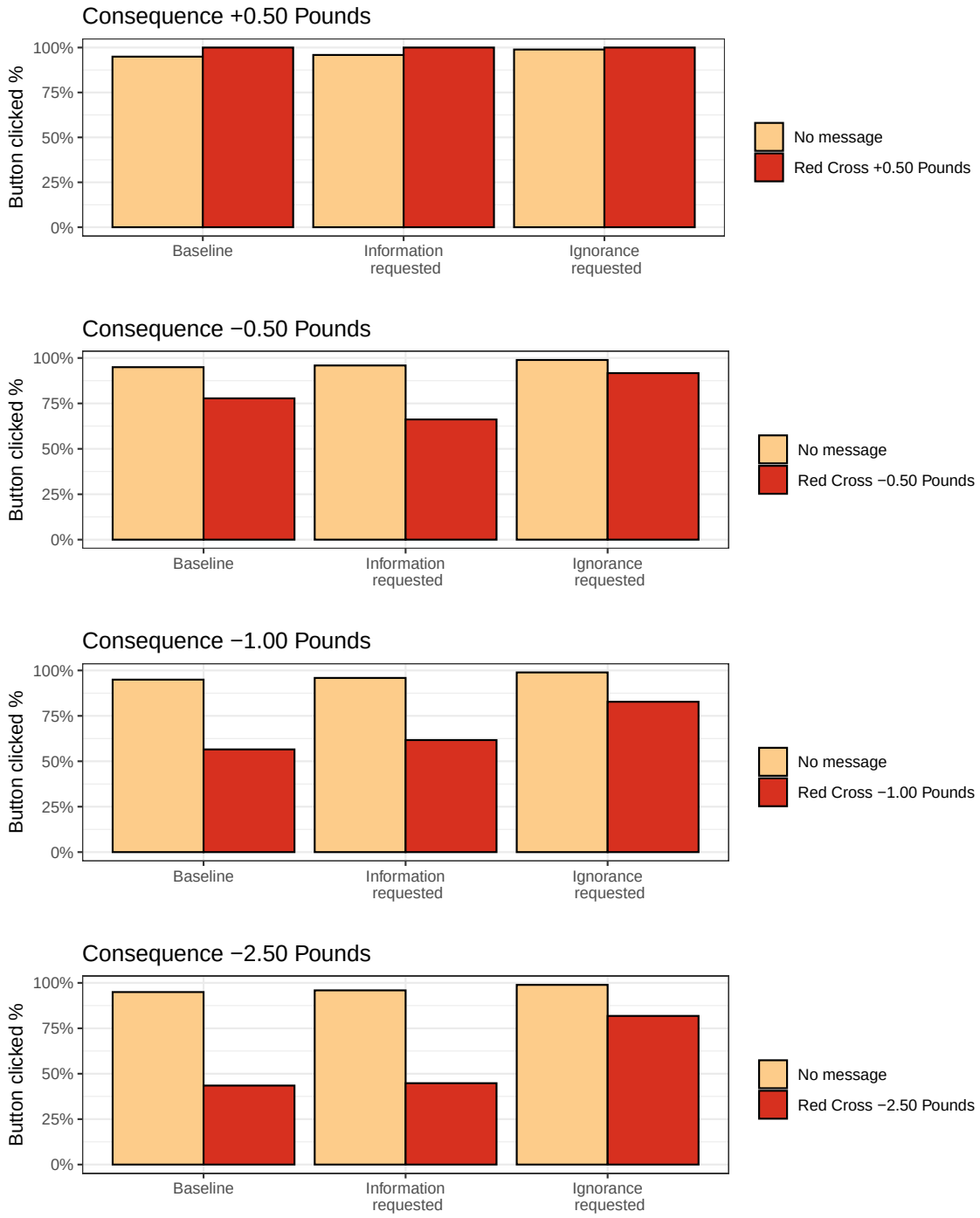


Figure B1: Proportion of button clicks by received treatment, request and message (pooling the *Request* and *Request + Punishment* treatments).

Appendix C OLS regression of charity outcomes

Table C1: OLS regression of the outcome for the charity

	Dependent variable: Charity outcome in pound	
	(1)	(2)
Constant	-0.477^{***}	-0.661^{***}
	(0.068)	(0.148)
Information preference (ref = Baseline)		
Request info	-0.017	-0.055
	(0.089)	(0.090)
Request ignorance	-0.254°	-0.207
	(0.134)	(0.137)
Request info under punishment threat	-0.084	-0.119
	(0.096)	(0.098)
Request ignorance under punishment threat	-0.204°	-0.175
	(0.117)	(0.127)
Control variables		
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. Model 1 includes all observations. Model 2 excludes participants with incomplete answers for 'identify with 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.

D.1 Sender beliefs about button pressing

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.

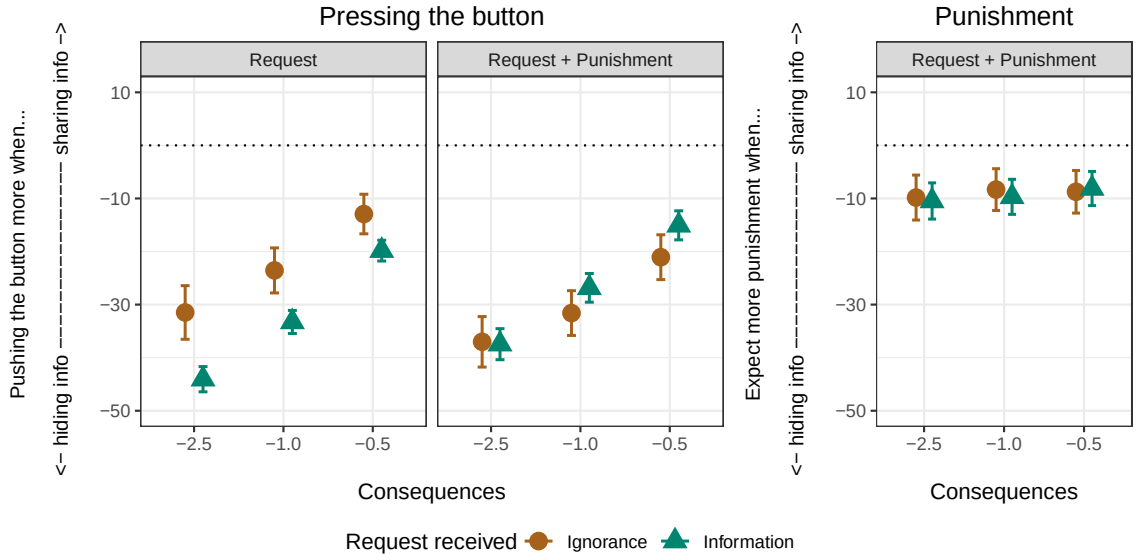


Figure D1: Difference in beliefs of pressing the button and difference in beliefs of punishment by treatment and consequences. Figure shows means and standard errors.

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.

Table D1, column 1-4 reports the results of regression models to investigate these patterns statistically. The dependent variable in these models is the difference between the Sender's belief that the Receiver will press the button after the Sender shared information, and the belief that the Receiver presses the button after the Sender did not share information. Explanatory variables include the request of the Receiver, two dummies capturing the consequences for the Red Cross. Model (2) and (4) include the interactions. All

models include control variables (revealed in DWK, age, income, browser type and comprehension questions) and control for the correlation of the beliefs coming from the same Sender by including a random effects at individual level. The results indicate that in the *Request* treatment, Senders' expect Receivers to push the button about 9 percentage points less often when they request information, although the result is significant only at the 10% level. In the *Request + Punishment* treatment, the effect disappears, in line with the visual evidence.

D.2 Sender beliefs about punishment

We then turn Sender beliefs about punishment. The rightmost panel of Figure D1 displays the difference in the belief regarding being punished after sharing information 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 suggests 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 partially supported by the results of linear regression models (5) and (6) reported in Table D1. The dependent variable in both models is the difference between the Sender's beliefs to be punished after sharing and after not sharing information . Explanatory variables include the request of the Receiver, two dummies capturing the consequences for the red cross, and the covariates. Model (6) includes interactions as well. Both models control for the correlation of the beliefs coming from the same Sender by including a random effects at individual level. The results show that the nature of the information request has no statistically significant impact on Sender's beliefs. Moreover, as we find non-significant estimated parameters for the Constant, we cannot reject the hypothesis that Senders believe there is a similar level of punishment when withholding and sharing information.

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	-47.312^{***}	-47.897^{***}	-31.711^{***}	-32.946^{***}	-16.230	-16.157
	(7.605)	(7.672)	(9.552)	(9.569)	(11.049)	(11.051)
Request Ignorance	8.953°	11.167°	-1.316	2.514	-2.383	-2.609
	(4.739)	(5.854)	(5.071)	(5.850)	(5.040)	(5.674)
Consequence -1.0	10.482^{***}	11.062^{***}	9.558^{***}	11.380^{***}	1.420	1.374
	(1.136)	(1.279)	(1.287)	(1.671)	(1.511)	(1.940)
Consequence -0.5	23.539^{***}	24.713^{***}	21.598^{***}	23.481^{***}	1.457	1.283
	(1.755)	(1.928)	(2.031)	(2.515)	(1.931)	(2.461)
Request Igno × -1.0		-2.196		-5.649^{*}		0.143
		(2.728)		(2.488)		(3.027)
Request Igno × -0.5		-4.446		-5.841		0.537
		(4.351)		(4.233)		(3.908)
Observations	852	852	828	828	828	828
R^2	0.262	0.264	0.175	0.178	0.006	0.006
Adjusted R^2	0.256	0.256	0.168	0.169	-0.003	-0.005

Notes: OLS regressions with random effects at individual level of the risk difference (RD) of the Senders' beliefs that the receiver presses the button (model 1 to 4) and the belief that the receiver would punish (model 5 and 6). The dependent variable is the difference of the belief when sharing and when not sharing information. Covariates suppressed for brevity: revealed in DWK, age, income, browser type, comprehension questions. Robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$).

D.3 Correlation between beliefs and decision to share

A key question is whether Senders' beliefs about Receivers' pressing behavior are correlated with their decision to share information. To investigate this, we calculate the difference between the belief that a Receiver will press the button when informed and the belief that they will press when uninformed about the consequences. For negative consequences, this difference reflects the belief in how much sharing information reduces the likelihood of pressing the button. For positive consequences, instead, it indicates how much sharing information increases the likelihood of pressing, as the belief in pressing when informed about positive consequences is observed to be 100%.

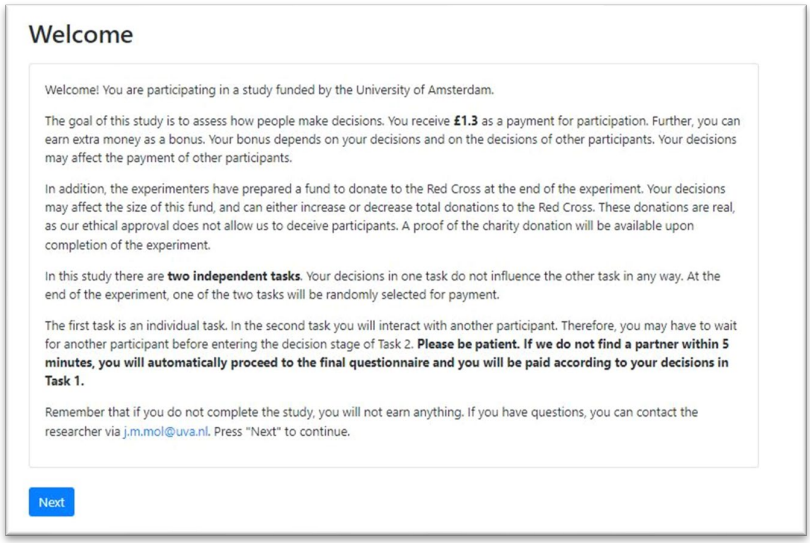
Therefore, in case of the negative consequences (positive consequences), a negative (positive) correlation between the gap in beliefs and the decision to share means that Senders are less willing to pay the cost of sending information when they believe that it is less effective. Table D2 shows that this is the case for all levels of the consequences.

Table D2: Senders' information sharing by request and beliefs.

	Dependent variable: Information sharing			
	Cons. -2.5	Cons. -1.0	Cons. -0.5	Cons. +0.5
	(1)	(2)	(3)	(4)
Constant	0.493***	0.278***	0.226**	0.295***
	(0.079)	(0.080)	(0.079)	(0.080)
Information preference (ref = Baseline)				
Request info	−0.021	−0.016	0.003	0.034
	(0.047)	(0.051)	(0.052)	(0.052)
Request ignorance	−0.072	−0.083	0.016	−0.018
	(0.065)	(0.067)	(0.069)	(0.067)
Request info under punishment threat	0.010	0.002	0.074	0.123*
	(0.049)	(0.053)	(0.055)	(0.055)
Request ignorance under punishment threat	−0.015	0.010	0.080	0.054
	(0.061)	(0.065)	(0.068)	(0.066)
Beliefs				
Belief when informed - Belief when not informed	−0.138**	−0.212***	−0.125*	0.138*
	(0.043)	(0.052)	(0.055)	(0.057)
Control variables				
Identify with charity	0.020**	0.028***	0.027***	0.020**
	(0.006)	(0.007)	(0.007)	(0.006)
Revealed in DWK	0.067*	0.108**	0.122**	0.018
	(0.034)	(0.037)	(0.039)	(0.038)
Observations	703	703	703	703
R^2	0.053	0.086	0.068	0.040
Adjusted R^2	0.036	0.070	0.051	0.023

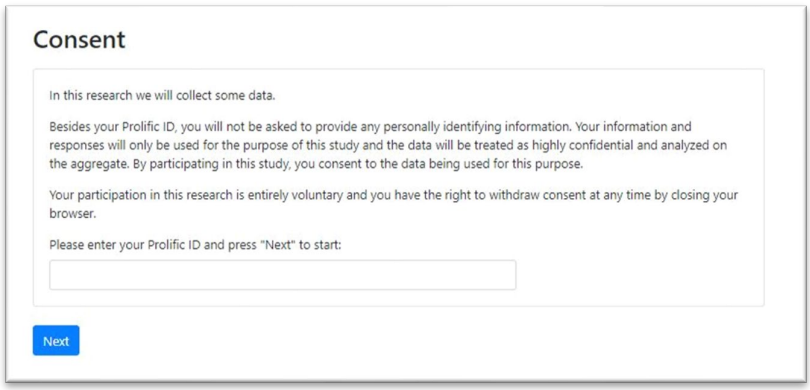
Notes: Dependent variable: Shared information. Reference category: Baseline. Covariates suppressed for brevity: gender, age, income, browser type, comprehension questions. Linear model with heteroscedasticity robust standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

Appendix E Experimental instructions



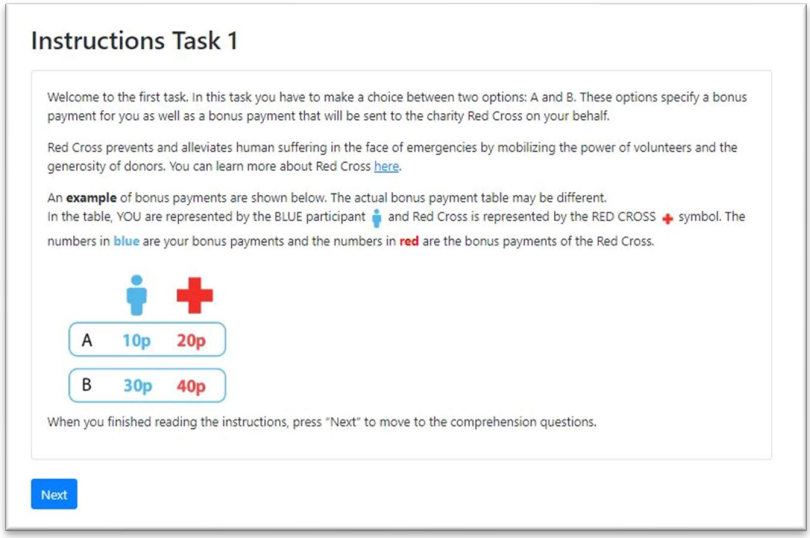
Welcome

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



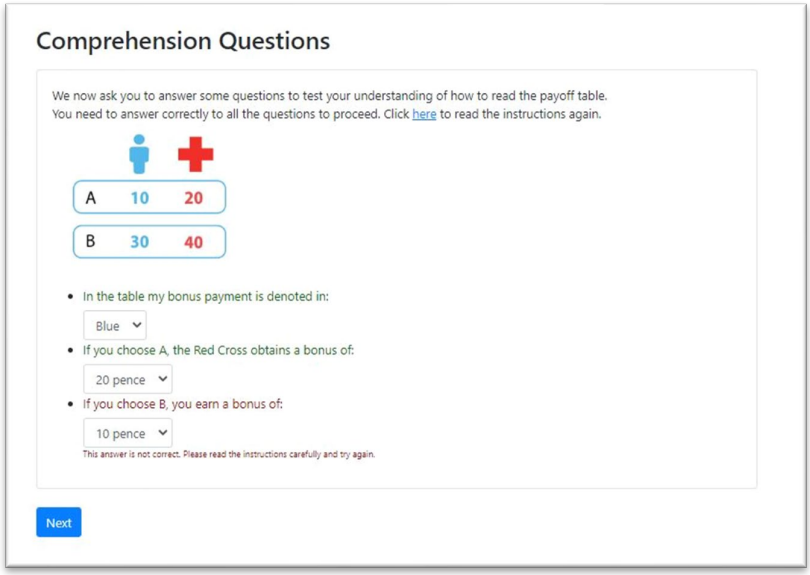
Consent

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



Instructions DWK (1)

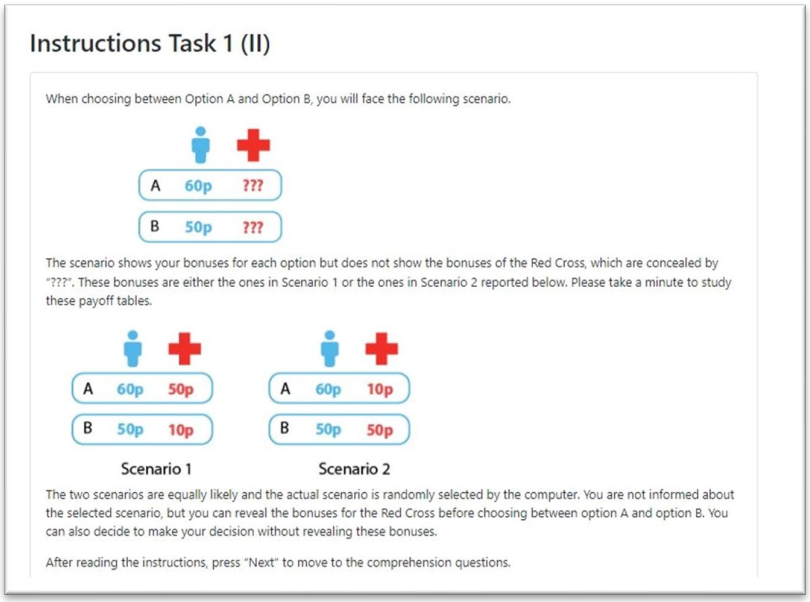
First instructions page for Dana Weber Kuang dictator game.



Comprehension Questions

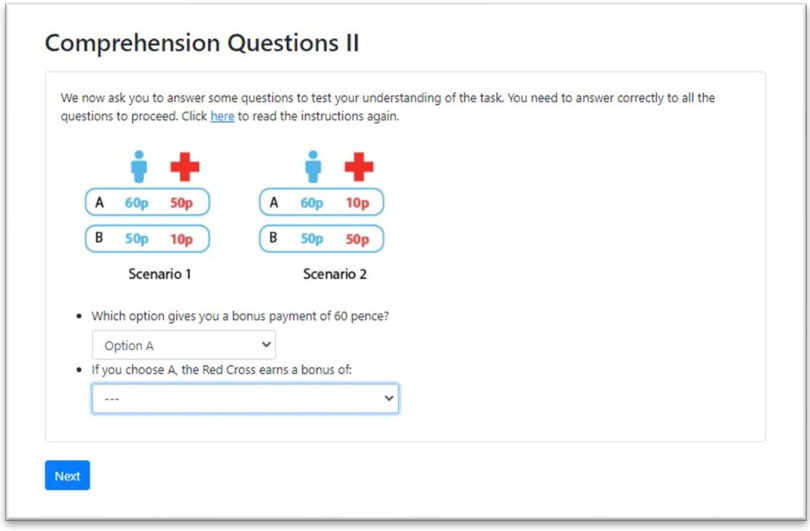
Comprehension questions for DWK dictator game. Correct answers:

- Blue
- 20 pence
- 30 pence



Instructions DWK (2)

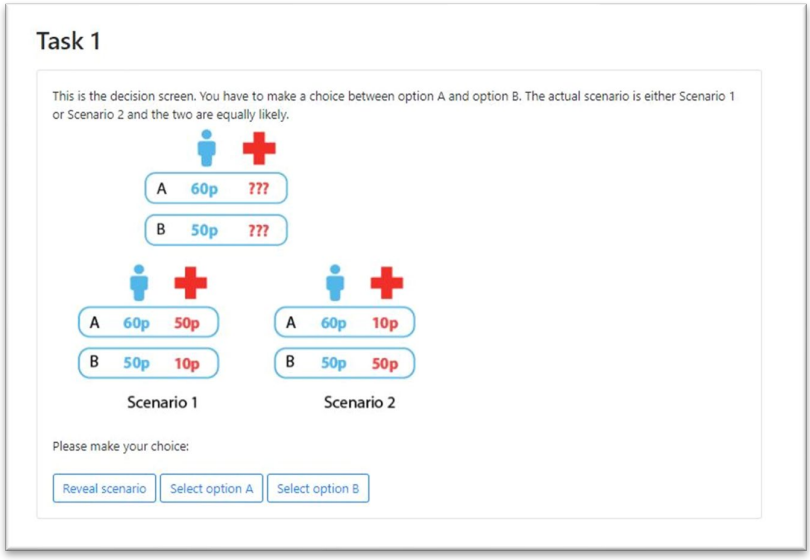
Second instructions page for Dana Weber Kuang dictator game.



Comprehension Questions (2)

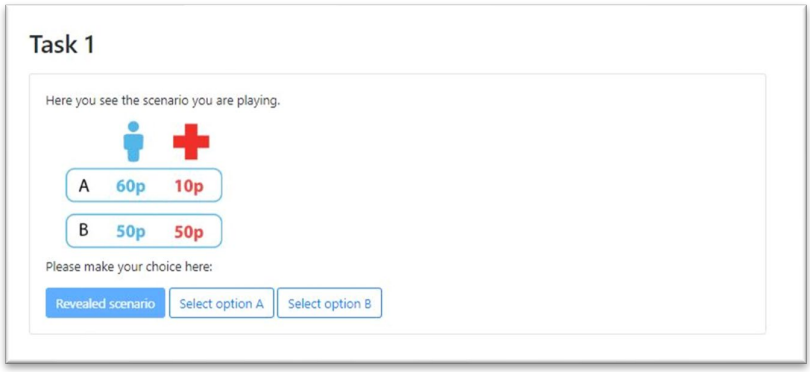
More comprehension questions after the second set of instructions. Correct answers:

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



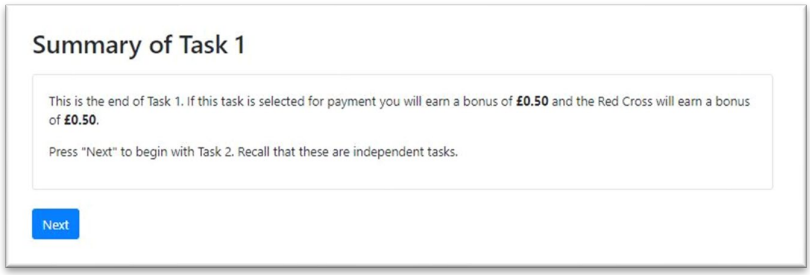
DWK dictator game

Main page for DWK dictator game.



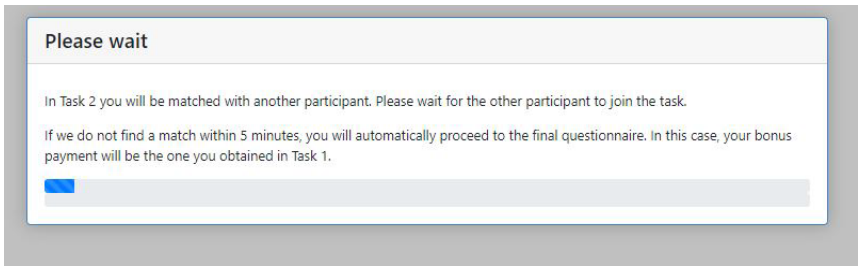
Revealed page

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

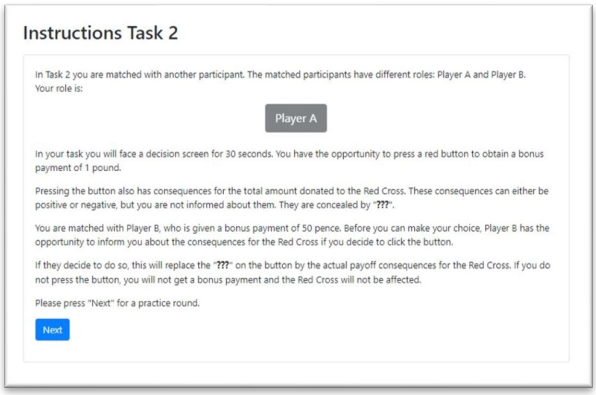
Results of the DWK dictator game.

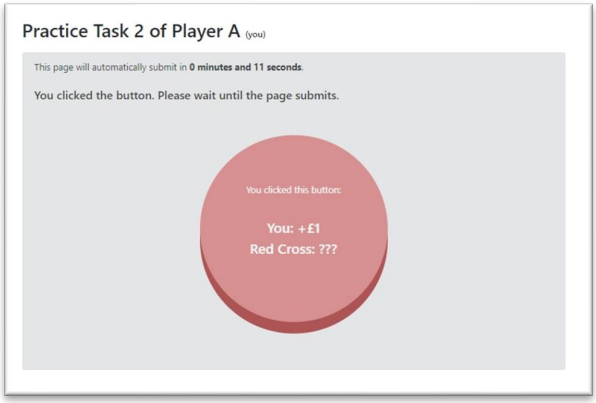


Waitpage for next game

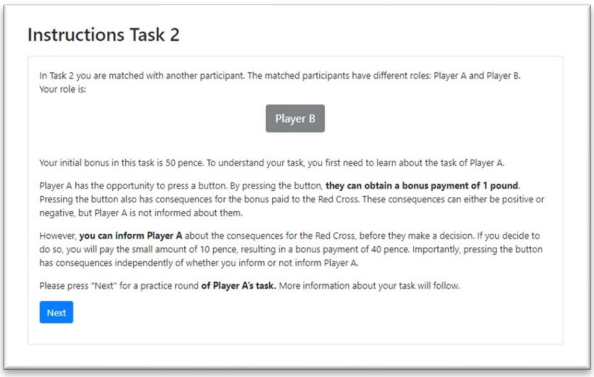
Page with progress bar of 5 minutes to allow matching in pairs for the Button Game.

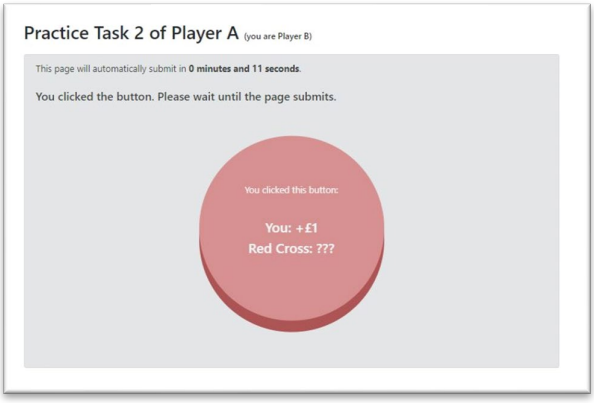
Button Game for the Receiver (A)

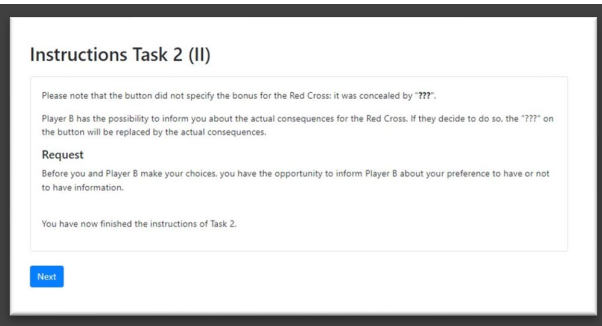




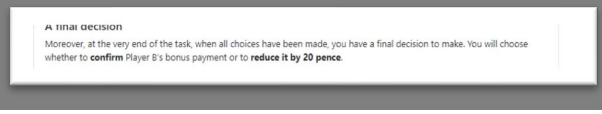
Button game for the Sender (B)





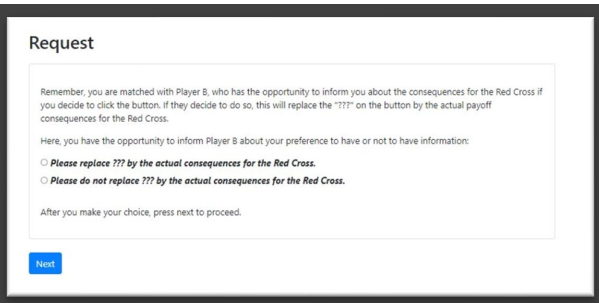


(only in Request and Request + Punishment treatments)



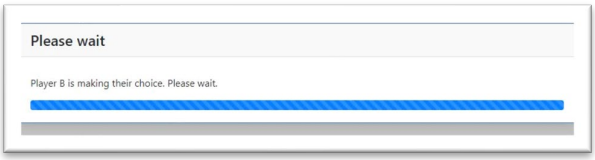
(only in Request + Punishment treatment)

Button Game for the Receiver (A)



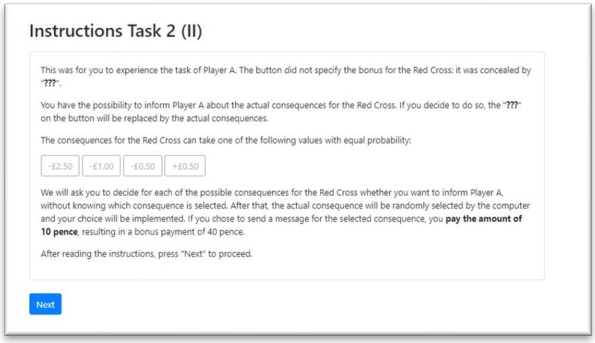
Button game for the Sender (B)

(only in Request and
Request + Punishment treatments)



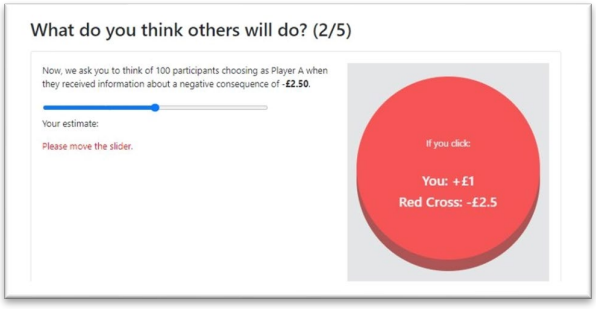
While the Receiver waits, the Sender completes the instructions, comprehension questions, beliefs and sending choice. Screen capture on the right shows: belief elicitation Sender for an uninformed Receiver.

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.

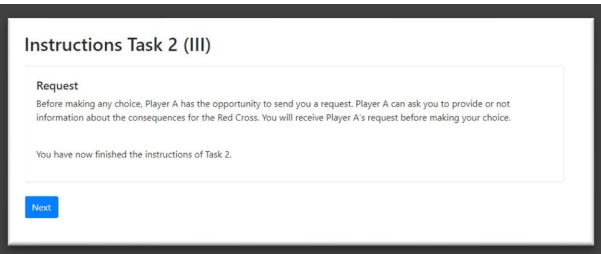


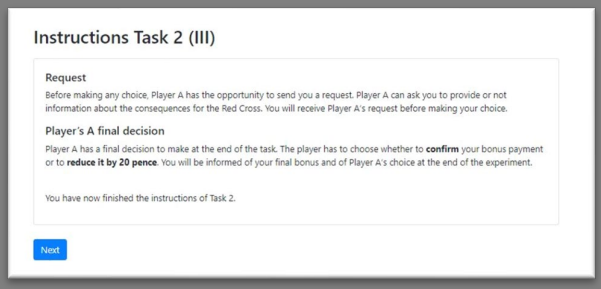
Button Game for the Receiver (A)

(only in Request treatment)

(only in
Request + Punishment treatment)

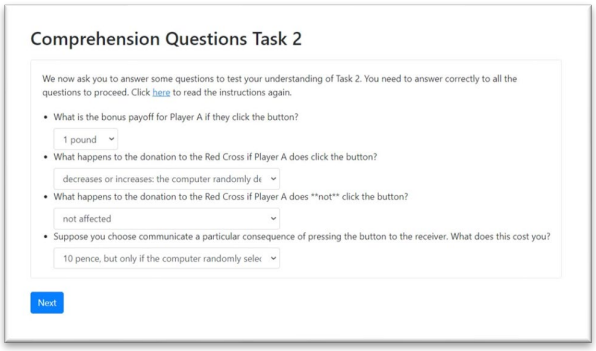
Button game for the Sender (B)





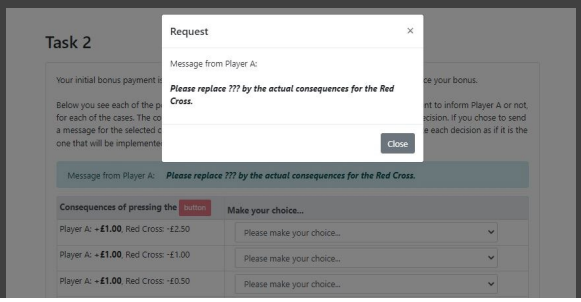
Comprehension Questions of the Sender

Correct answers currently selected.



(only in Request and
Request + Punishment treatments)

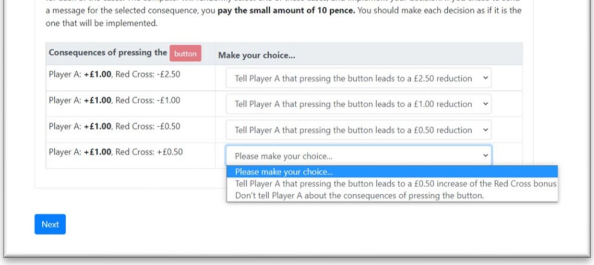
Request for the Sender

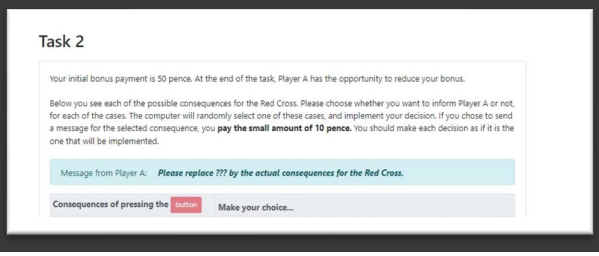


Button Game for the Receiver (A)

(Baseline treatment interface)

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

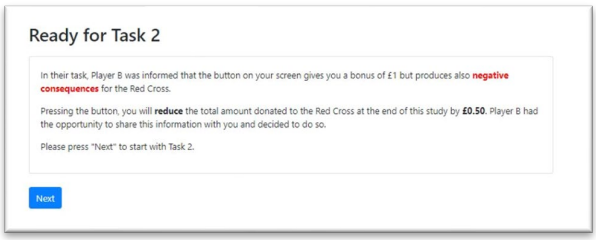




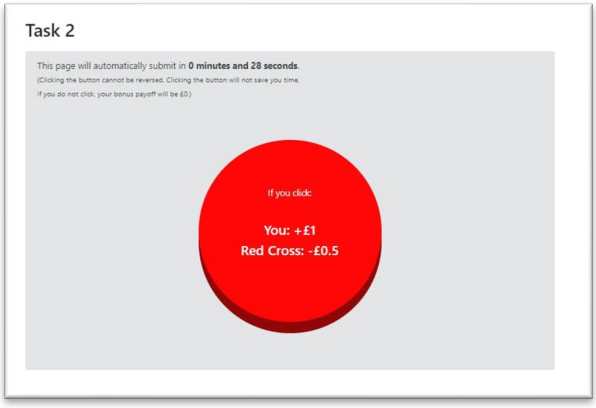
Button game for the Sender (B)

Interface of the Sending decision

Information sharing from Sender to Receiver



The Button Page



Button Game for the Receiver (A)

The Button page (if clicked)

only in Baseline and Request treatments

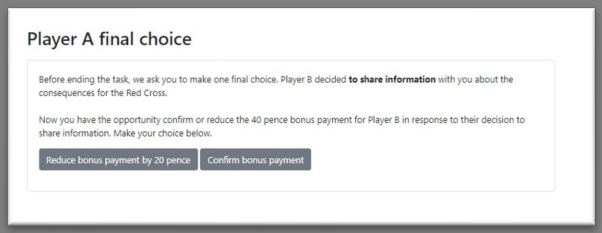
Button game for the Sender (B)

Belief elicitation button pressing

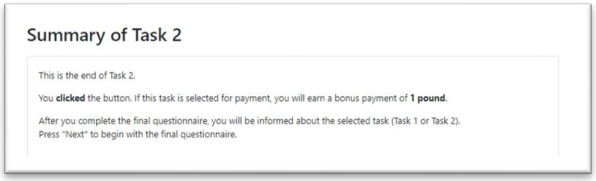
only in Request + Punishment treatment

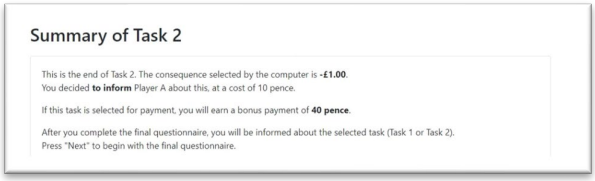
Belief elicitation button pressing and punishment

Punishment page



only in Request + Punishment treatment



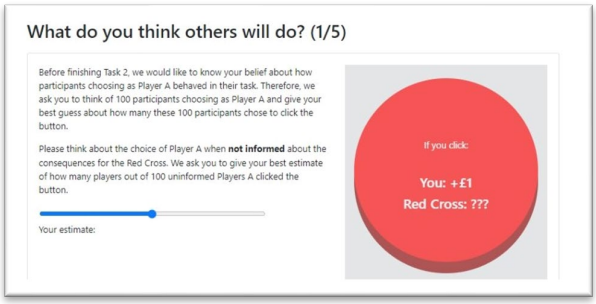


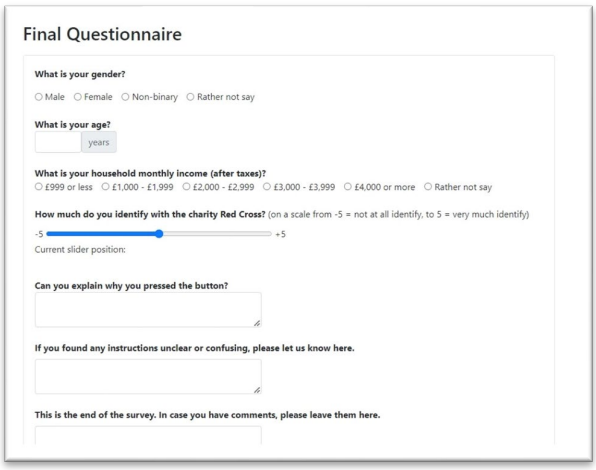
Button Game for the Receiver (A)

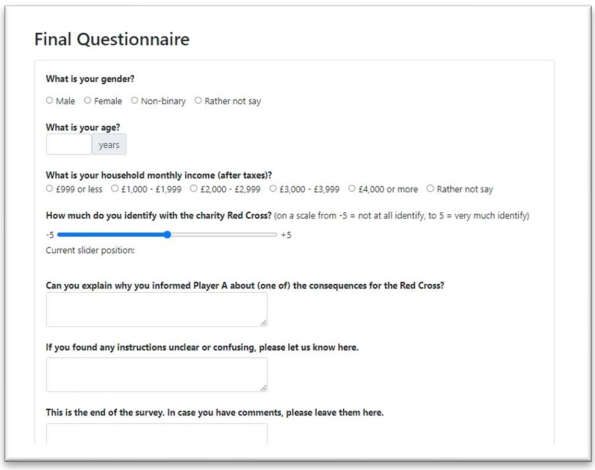
Button game for the Sender (B)

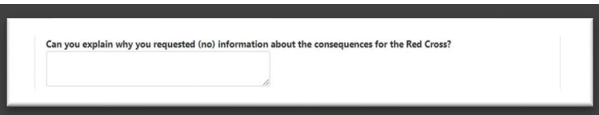
Belief elicitation of the Receiver

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

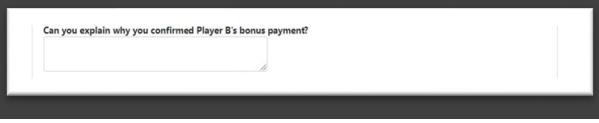




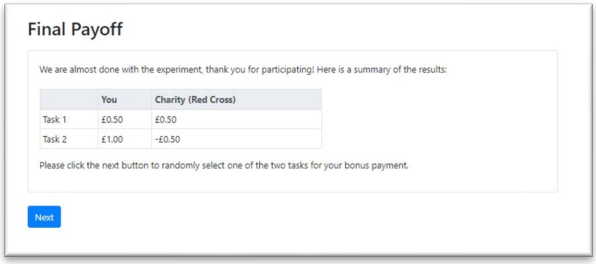


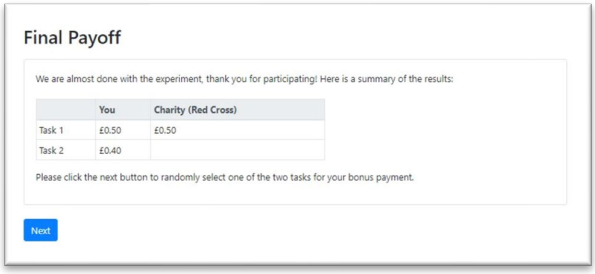


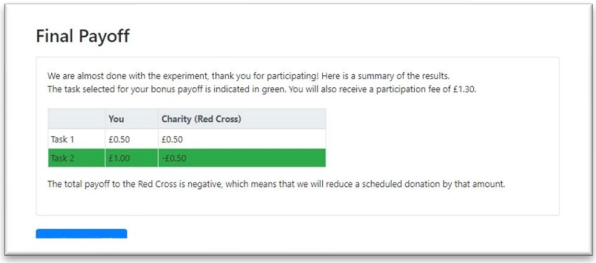
(only in Request and Request + Punishment treatments)

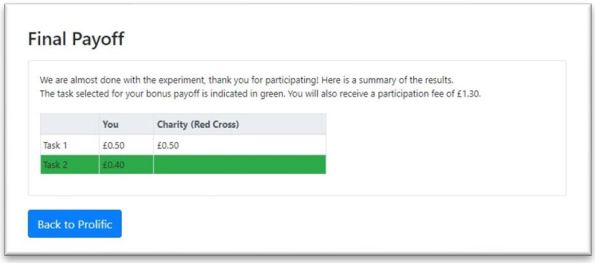


only in Request + Punishment treatment









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.