Does room for reflection reduce ignorance and increase pro-social behavior? An experimental study

Fabian Bopp* and Wendelin Schnedler+

Abstract

A lot of harm comes about because people ignore the consequences of their behavior on others. Experimental evidence suggests that people might even willfully ignore consequences so that they can act selfishly without a 'bad conscience'. In essence, such people 'kid themselves'. If I care about the consequences of my acts on others, I should not ignore them. Upon reflection, people may discover this inconsistency. De-biasing people may thus be an effective tool to prevent harm. We examine this idea experimentally. We find that inviting subjects to describe their aims and means makes them more likely to inform themselves and ultimately act more pro-socially.

JEL classification: C91

Keywords: willful ignorance, strategic ignorance, social preferences, de-biasing

*fabian.bopp@uni-paderborn.de, Paderborn University, Department of Management, Warburger Straße 100, D-33098 Paderborn, Germany.

+wendelin.schnedler@upb.de, Paderborn University, Department of Management, Warburger Straße 100, D-33098 Paderborn, Germany.

The authors wish to thank Nora Szech, Christine Exley, Linh Vu, Johannes Jarke-Neuert, Fiona tho Pesch, Katharina Momsen, Joël van der Weele and the participants of the IAREP 2022 and the ICSD 2022 conferences for helpful comments and Simon Müller for programming. All errors remain our own.

You can close your eyes to the things
you don't want to see,
but you can't close your heart to the
things you don't want to feel.

attributed to Johnny Depp

1 Introduction

Consumer power has been regarded as a potential force for a more equitable and environmentally healthy world.[1] While many consumers claim that they prefer fairly traded and environmentally friendly products, this is not reflected in their actual choices (Carrington et al., 2014; Hassan et al., 2016; Farjam et al., 2019).

One reason for this gap might be that people choose not to learn of the effects of their behavior. I may, for example, not inform myself about whether the beans for my favorite chocolate are harvested by child slaves. Then, I can continue purchasing this chocolate, while pretending that I would, of course, not buy it if I *knew* that it involved child labor. Dana et al. (2007), Ehrich and Irwin (2005) and many others[2] find experimental evidence consistent with this idea.

At the heart of this reasoning, however, is a logical inconsistency. Suppose I care about working conditions so much that I am willing to forgo my favorite chocolate once I know that conditions are bad. Then, I also have an incentive to obtain easily available information about these conditions. (We later provide the formal backing for this inconsistency in Proposition 1.)

Everyday small stakes consumption decisions, however, are often made quickly and intuitively and the inconsistency may not become apparent. It may only emerge after careful consideration, or, in the terminology of Stanovich and West (2000) popularized by

[1]For an example, see the contribution by Cristianne Close to the World Economic Forum in 2021, accessed on 29th of August 2022

[2]See, e.g., Feiler (2014); Grossman (2014); Larson and Capra (2009); Grossman and Van Der Weele (2017) for lab evidence and Andreoni et al. (2017); DellaVigna et al. (2012); Ganguly and Tasoff (2017); Knutsson et al. (2013); Li et al. (2021); Thunström et al. (2021) for field evidence.

Kahneman (2011) if decision makers employ System 2 rather than 1. Reflection may thus lead decision makers to inform themselves and use this information to act less harmfully (see Corollary 1, later).

Finding out whether reflection has this beneficial effect is not simple. First, telling subjects to 'think hard' or 'reflect carefully' may suggest to them that the experimenter expects a certain behavior. More pro-social behavior would then result from the specifics of the experiment and not from free deliberation. Second, any such intervention may not only improve how individuals understand available facts (i.e., move a decision from System 1 to 2) but introduce new facts. Being asked to reflect may be seen as a signal that one should obtain information. We, however, are interested in the effect of reflection in itself and not that of learning, for example, about a norm as in Cappelen et al. (2011) or Serra-Garcia and Szech (2022).

In this paper, we propose a simple design to address these issues and test whether individuals are more likely to inform themselves and act more pro-socially if they are given a neutral opportunity to reflect. The design builds directly on the canonical experiment by Dana et al. (2007), where subjects in their hidden information treatment can click on a button to learn whether their favorite choice has a negative impact for another subject. We use this treatment as a CONTROL. Since we want choices to reflect the character of a normal 'automated' small stakes consumption decision, like buying chocolate, we follow Exley and Kessler (2021) and run the experiment on a crowd working platform, where stakes are small and subjects typically decide fast and less carefully than in the standard laboratory setting. With control questions, we ensure that subjects have all the necessary facts to decide optimally, while we hope that they do not take the necessary time to come to such a decision.

Our DESCRIBE treatment aims at de-biasing subjects by helping them to clarify their goals and identifying appropriate behavior to reach these goals. We do so by adding an opportunity to describe their aims and how they want to achieve them. We also ask them to relate aims and intended behavior, so that potential inconsistencies become apparent to them. Since the respective screen follows the control questions, answering these additional questions seems natural and is arguably free from pushing subjects to act in a certain way.

Suppose subjects indeed want to prevent harm but do not inform themselves. If our intervention works, they reflect on their aims and means and discover the fundamental inconsistency in their behavior. Then, we expect that (i) more people will inform themselves in the DESCRIBE than the CONTROL treatment and (ii) informed subjects will behave more pro-socially in the DESCRIBE than in the CONTROL treatment.

None of the extant explanations for why subjects remain ignorant makes both these predictions. Staying ignorant may, for example, avoid moral constraints (Feiler, 2014) or obligations (Spiekermann and Weiss, 2016). These explanations as well as any outcome-based preference, such as money-maximizing or other-regarding preferences, require an alteration of the economic environment for a change in behavior to occur. Since choices, information about and presentation of these choices as well as payoffs are the same in both treatments, one would not expect any difference in behavior. Under certain assumptions, which we will discuss later, the self-signaling model by Grossman and Van Der Weele (2017) in the spirit of Bodner and Prelec (2003) can explain why more subjects inform themselves in the DESCRIBE than in the CONTROL treatment. It can, however, not explain our second prediction, namely, more pro-social behavior among informed subjects.

We replicate well-established patterns from the literature. In our CONTROL treatment, 54% of subjects remain ignorant. Since uninformed subjects tend to choose their favorite option in more than 80% of the cases a passive player to whom the favorite option is harmful is very likely to be harmed. If the same passive player encounters an informed subject, she only gets harmed with a probability of 36%. Getting more subjects to inform themselves is thus possible and might lead to more pro-social behavior.

Indeed, having to describe means and ends increases the share of informed subjects from 46% to 59%. The increase by more than 25% is economically sizable and statistically significant. It is consistent with the idea that caring subjects realize that they need to inform themselves to act pro-socially. Moreover, the drop in ignorance is accompanied by a (statistically significant) reduction of harmful behavior among informed subjects from 36% in the CONTROL to only 18% in the DESCRIBE treatment. The treatment thus leads to an economically meaningful halving of the share of harmful behavior among informed subjects. Our findings are in line with the notion that reflection leads subjects

to inform themselves in order to act more pro-socially.

Whatever the precise mechanism, giving room for reflection seems to be effective in preventing harm. For an individual who is harmed by the decision maker's favorite choice, the probability of coming out unharmed increases from roughly 40% in the CONTROL to over 60% in the DESCRIBE treatment. The sizable increase in harm preventing behavior by 50% is statistically significant.

Several articles use experimental variation to support or refute specific preference-based explanations for ignorance (Matthey and Regner, 2011; Feiler, 2014; Kajackaite, 2015; Spiekermann and Weiss, 2016; Grossman and Van Der Weele, 2017). While they take behavior as revealing subjects' preferences, we start from the idea that subjects engage in a systematic error or inconsistency. Similar to Cason and Plott (2014) and Bartling et al. (2020), we study how subjects' deliberation may affect their behavior. In contrast to them, we do not use group discussions for these deliberations. Group discussions would have introduced confounds like social image concerns. Instead, subjects here are given the opportunity to deliberate individually.

Grossman and Van Der Weele (2017) find that committing to a specific contingent behavior when being informed leads to less ignorance. Their intervention offers a clean check of their theory, in particular their Proposition 4. It is intended as such and not as a policy intervention. From a policy perspective, getting individuals to reflect on their actions seems easier than implementing contingent commitments. Although their intervention differs from ours, their finding may well be driven by subjects discovering possible inconsistency if they have to commit to a behavior once informed.

Probably the closest to our study is that by Grossman (2014). Grossman (2014) finds that subjects are much more likely to ignore information if clicking reveals this information rather than keeps them ignorant. Having to click in order to remain ignorant may imply that subjects think more carefully whether clicking is consistent with their aims. Another explanation put forward by Grossman is that subjects find it easier to defend their behavior if it coincides with the default. Our experiment isolates these two explanations from each other. We have the same default in both treatments, namely that subjects are uninformed. Still, we find a considerable effect. This suggests that reflecting

about the consistency of one's behavior may also play a role in Grossman's data.

2 Theoretical Consideration

In this section, we formalize the idea that ignoring easily available information is not consistent with acting pro-socially when being given the information (Proposition 1). We also show to which strategy one should switch upon realizing one's irrational behavior (Corollary 1).

We illustrate our argument using a simple model with other regarding preferences. The notation follows that of Grossman and Van Der Weele (2017), where we removed image concerns because our argument does not require them.

An economic agent or active player A can engage in an action $a = 1$ at costs c or not, $a = 0$. In terms of our example, $a = 1$ means that she forgoes her favorite chocolate. This action may or may not generate welfare w for a passive player P, where $w > c$. Denote A's prior belief that the action generates welfare, $W = w$, by $p := Prob(W = w)$, where there is some uncertainty, $p < 1$. Player A can inform herself about the consequences of her action ($I = 1$) or remain ignorant ($I = 0$). She learns whether the action generates welfare if and only if she informs herself. For examining whether A's choices are consistent, we represent the decision problem as a sequential game against nature. Nature chooses the state of the world $W = 0$ or $W = w$, the agent decides whether or not to inform herself, $I = 1$ or $I = 0$, and then either takes action $a = 1$ or $a = 0$ —see Figure 1.

Agents dislike the costs from forgoing their favorite choice $a = 1$ and value the welfare w to the passive player to different degrees $\theta \in [0, 1]$. Overall, the agent's utility amounts to:

$$u(a, I, W) = a(\theta W - c) - I \cdot k, \tag{1}$$

where costs k of informing oneself are negligible: $0 < k < \epsilon$ for some small $\epsilon > 0$. Consider the state of nature, where $a = 1$ generates w. If A chooses $a = 1$ and forgoes c to generate w in this state, we say that A's choice is *pro-social*. Such pro-social behavior naturally only makes sense if A cares about the passive player.

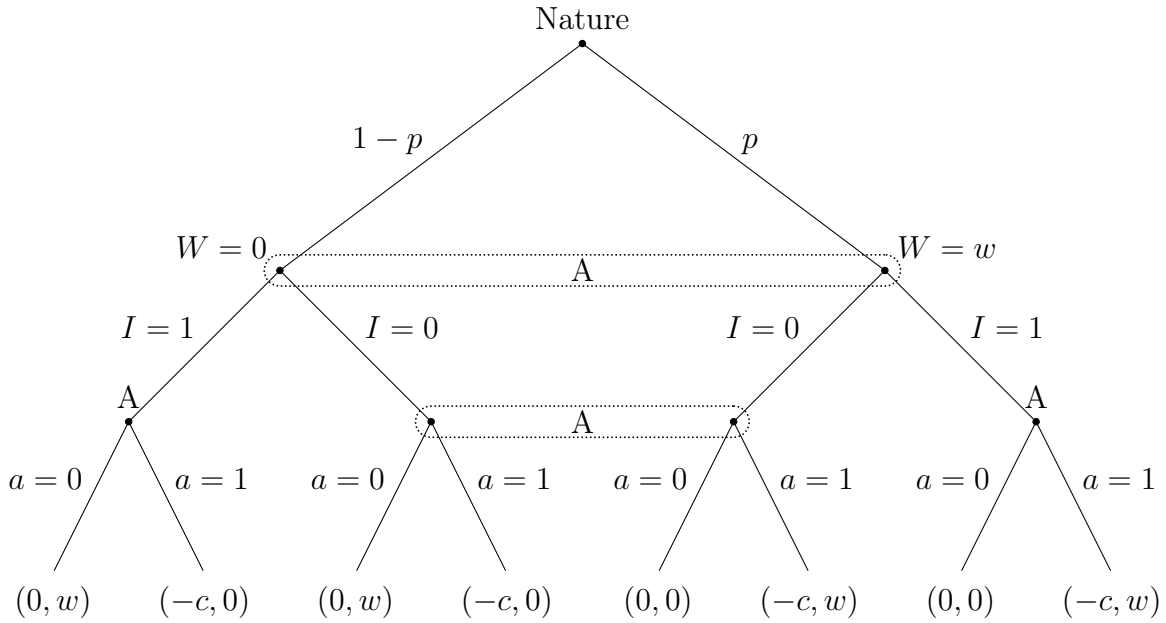


Figure 1: A's problem as a game against nature; the first number represents A's material payoff, the second that of the passive player P.

Lemma 1. *An agent has a strict preference for the pro-social choice whenever she cares sufficiently about the passive player:*

$$a = 1 \text{ given } W = w \Leftrightarrow \theta > \frac{c}{w} =: \underline{\theta}. \tag{2}$$

Proof. Knowing that $W = w$, the agent gains from $a = 1$ if and only if $\theta w > c \Longleftrightarrow \theta > \frac{c}{w}$. □

We are now in the position to formalize why ignoring information is irrational when you care about avoiding harmful behavior. The intuition is the following. Engaging in the pro-social action implies that the agent is willing to forgo some of her material value in order to preserve welfare w. Then, however, she should also be willing to incur very small costs for the option value to help.

Proposition 1 (Ignorance is irrational when you care enough to act)**.** *Suppose information is sufficiently easy to obtain (k small). Then, an agent with a strict preference for the pro-social choice also has a strict preference to inform herself $I = 1$.*

Proof. Consider an agent with a strict preference for the pro-social choice. By definition, this agent selects $a = 1$ when $W = w$, leading to gains of $\theta w - c$ and $a = 0$ whenever

$W = 0$ resulting in gains of zero. The agent can thus expect gains of $p(\theta w - c)$ and costs of k when deciding to inform herself ($I = 1$). If the agent remains ignorant and chooses $a = 1$, she gains $p\theta w - c$. If she remains ignorant while selecting $a = 0$, she obtains a payoff of zero. An agent with a strict preference for the pro-social choice thus has a strict preference to inform herself, $I = 1$, if and only if

$$\begin{aligned} &\max\{0, p\theta w - c\} < p(\theta w - c) - k \\ \Leftrightarrow &0 < p(\theta w - c) - k \text{ and } p\theta w - c < p(\theta w - c) - k. \end{aligned} \tag{3}$$

Under the assumption that information costs are negligible, $k < \epsilon$, we find that the agent wants to inform herself if

$$\epsilon < p(\theta w - c) \text{ and } p\theta w - c + \epsilon < p(\theta w - c). \tag{4}$$

For an agent to have a strict preference for $a = 1$ given $W = w$, she must care enough about the passive player $\theta > \underline{\theta}$—see Lemma 1. Since $\theta > \underline{\theta} = \frac{w}{c}$, we know that $(\theta w - c) > 0$ and because $p \in (0, 1)$ also $p(w\theta - c) > 0$ and there is an $\epsilon_1 > 0$ such that the left-hand side inequality in (4) is fulfilled. The right-hand side inequality in (4) simplifies to $\epsilon < (1 - p)c$. Since $p \in (0, 1)$ and $c > 0$, there is an $\epsilon_2 > 0$ that fulfills this inequality. Then, taking the smaller value $\epsilon := \min\{\epsilon_1, \epsilon_2\}$ fulfills both inequalities and an agent with a strict preference for $a = 1$ given $W = w$ also has a strict preference to inform herself. □

The proposition tells us that caring about welfare and remaining ignorant is not a consistent strategy when information is readily available. What would be a more consistent choice to which the agent may switch? If the agent cares enough about welfare, she will continue to behave pro-socially but also inform herself. If she doesn't, she remains ignorant and chooses the selfish option.

Corollary 1 (Resolving Irrationality)**.** *The agent's optimal strategy depends on her type* θ*.*

- *If she sufficiently cares* $\theta > \underline{\theta}$ *: she will inform herself,* $I = 1$*, and act pro-socially:* $a = 1$ *whenever* $W = w$*.*

- *If she does not sufficiently care $\theta < \underline{\theta}$: she will not inform herself and always choose $a = 0$.*

Proof. For $\theta > \underline{\theta}$, it follows directly from Lemma 1 that the agent has a strict preference for $a = 1$ whenever $W = w$ and from Proposition 1 that she informs herself. For $\theta < \underline{\theta}$, it follows again from Lemma 1 that $w\theta - c < 0$. The agent thus chooses $a = 0$ even if she knows that $W = w$. Informing herself thus has no value for such an agent but only results in costs of k. If the agent does not inform herself, we get from $\theta w - c < 0$ that the expected utility from $a = 1$ is $p\theta w - c < 0$ and hence smaller than the utility from $a = 0$, which is zero. The uninformed agent hence decides for $a = 0$ and gets utility zero. Since $-k < 0$, $I = 0$ generates the larger utility and the agent does not inform herself. □

3 Experiment

Here, we describe our experimental design, use the results from the previous section to derive predictions, and report on the implementation.

3.1 Design

In order to see whether giving people the room to reflect renders them more pro-social, we build on the seminal hidden information treatment by Dana et al. (2007) using as our CONTROL treatment the respective online version by Exley and Kessler (2021). Running the experiment online allows us to have the low stake environment in which most everyday shopping decisions are taken. In addition, subjects arguably feel less 'observed' and more anonymous in an online experiment because they are in their own home and money is not handed over personally but transferred automatically. We then add the DESCRIBE treatment in which subjects are given the space to reflect on their choice—see Figure 2.

Decisions and Payoffs. Subjects are randomly and anonymously paired. Each pair consists of an active and a passive player. The active player decides whether to choose 'high' ($a = 0$) and gain 60 cents or 'low' ($a = 1$) and get 50 cents. Accordingly, 'high' is A's favorite choice in terms of monetary payoff.

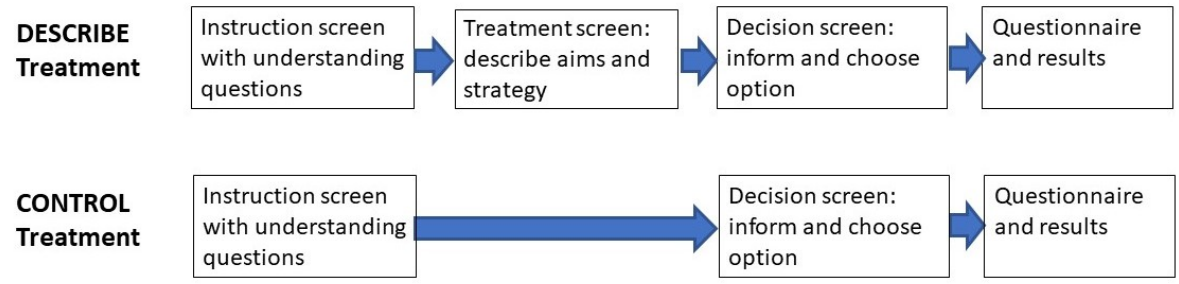


Figure 2: Experimental Sequence

The choice of 'high' and 'low' also determines the passive player P's payoff. There are two situations. Either, P gets 50 cents when A chooses 'high' and 10 cents when A chooses 'low' ($W = w$). Or, payoffs are flipped and P receives 10 cents for 'high' and 50 cents for 'low' ($W = 0$). Both cases are equally likely. The active player can inform herself about which situation is relevant by clicking on a 'Reveal' button ($I = 1$). A then learns the consequences of her choices—the screenshot is depicted in Figure 9 in the Appendix 6. If she does not click, all she knows is that both cases are equally likely.

Payments are equivalent to those of Exley and Kessler (2021) and in the range expected by subjects on the online platform that we use (MTurk); higher stakes might have felt suspicious to subjects.[3]

For more observations, we employ the strategy method. We ask both players to state their choices and only reveal at the end of the experiment who is the active and who the passive player. Sometimes when using the strategy method, subjects are asked for choices conditional on the behavior of the other player. Being confronted with possible behavior of other players can create a demand effect (Zizzo, 2010). Here, this caveat of the strategy method arguably does not apply. Apart from the instructions, none of the screens is altered by using the strategy method and accordingly subjects are unlikely to be affected by it.

Not knowing whether they are on the giving or receiving end may render subjects more empathic and lead to more pro-social behavior. This effect, if it exists, concerns both, CONTROL and DESCRIBE treatment equally, and hence not the treatment effect.

[3]Incidentally, Exley and Kessler (2021) find that raising stakes does not affect revelation decisions in their experiment.

Moreover, we later find no difference between the pro-social behavior observed in the literature and our control treatment.

Sequence and Intervention. Our intervention consists of an additional screen in the DESCRIBE treatment in relation to the CONTROL treatment. The screen appears after subjects have read the instructions and answered control questions and before they are shown the main decision screen—see Figure 2

The questions on this screen are framed in the spirit of a strategic planning tool—see Figure 3. We use the term 'plan' and highlight that any information given on this screen is non-binding. We recall the essence of the situation and then ask subjects about their aim. We deliberately chose this as an open question not to direct subjects or give a hint of what might be expected from them. Next, they can state their strategy. No default has been marked. Again, this was done to not guide subjects in a certain direction. If subjects want to specify a conditional strategy, they can click on the last answer involving 'if', as done in Figure 3 for the option 'I plan to choose High if...'. Then, a free format field opens in which subjects can write the respective condition. Finally, we give them the opportunity to link the strategy back to their aim. None of the fields on the screen have to be filled out and subjects decide how much time they want to invest in this questionnaire. Subjects can adapt their answers as often as they want before clicking on the 'OK' button. This allows them to correct logical inconsistencies between the formulated goal and the planned actions and to work toward a consistent strategy. While imposing as little in terms of aims and strategies as possible, we hope that this screen gives them a structure to come up with more consistent strategies.

The screen contains no new information about the decision situation. All information is being presented before and rigorously tested with control questions. Moreover, all information available on the describe screen is later repeated on the decision screen. Subjects in the DESCRIBE treatment who see the screen thus do not know more about the situation than those in the CONTROL treatment. Still, the screen might affect what inferences subjects draw from the information. In other words, their 'understanding' of the situation that goes beyond the mere factual information may well change through the

screen.

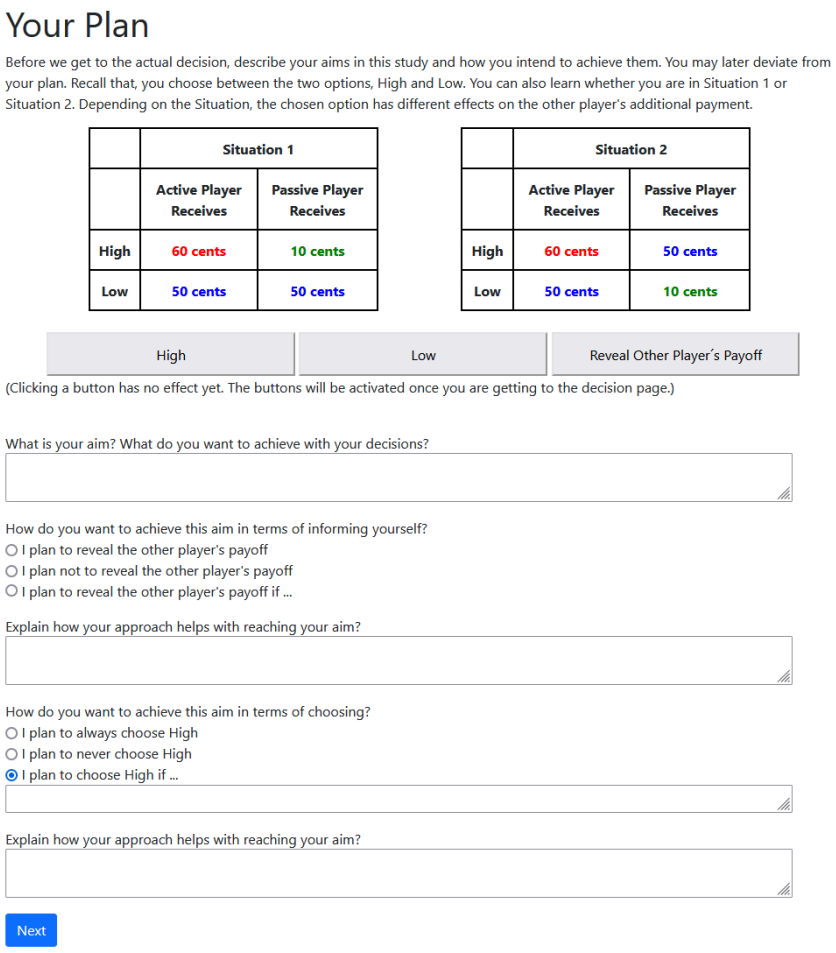


Figure 3: Screen used to encourage subjects to think more carefully about their choice (only in DESCRIBE treatment)

3.2 Predictions

Our key premise is that the additional opportunity to reflect might change behavior even though available choices, the information about these choices, the presentation of these choices and underlying preferences stay the same. More specifically, we believe that some subjects are happy to forgo 10 cents in order to increase the passive player's payoff by 40 cents but act irrationally and do not inform themselves, where the irrationality follows from Proposition 1. If such subjects indeed exist and our intervention reduces irrational

choices, they should start to inform themselves by Corollary 1.[4] The same outcome would be expected using a specific interpretation of our treatment in light of the model by Grossman and Van Der Weele (2017)—see Appendix 6.

Hypothesis 1 (Less ignorance in DESCRIBE treatment)**.** *In the DESCRIBE treatment, a larger share informs themselves about the consequences of their action than in the CONTROL treatment.*

Alternatively to the hypothesis, one might expect no change in behavior because neither preferences, nor choices, outcomes, or even the framing of choices changes. Other reasons why the proportion may remain the same is that all subjects who sufficiently care already inform themselves in the CONTROL treatment or that our description screen is incapable of inducing reflection that changes the understanding of the situation.

Suppose that in addition to subjects who care but do not inform themselves, other subjects inform themselves but do not act pro-socially, for example, because they are selfish but curious. If our argument holds, the pro-social subjects start to inform themselves in line with Hypothesis 1. Since these subjects are willing to forgo money for increasing the passive player's payoff, they will also act pro-socially by Corollary 1 and decide for 'low' ($a = 1$) whenever this protects the welfare of the passive player ($W = w$). In other words, the share of pro-social behavior among informed subjects increases.

Hypothesis 2 (More pro-sociality among informed in DESCRIBE treatment)**.** *In the DESCRIBE treatment, the share of informed subjects who engage in pro-social behavior, i.e., choose* $a = 1$ *given* $W = w$*, is higher than in the CONTROL treatment.*

Alternatively, the share of subjects who engage in pro-social behavior may stay the same. Once more, this might well be expected given that economic fundamentals of the game and its presentation are not different between the CONTROL and DESCRIBE treatment. Hypothesis 2 cannot be generated from the model by Grossman and Van Der Weele (2017)—see Appendix 6.

[4]In order to map the problem faced by subjects maps to that used for our theoretical considerations, take $p = .5$, $c = 10$ cents, $w = 40$ cents, and (without altering relative payoffs) add 60 cents to all A's and 10 cents to all P's payoffs in Figure 1.

The arguments leading to our hypotheses assume some observable pre-conditions. First, some subjects must not inform themselves in the CONTROL treatment. Otherwise, their share cannot increase in the DESCRIBE treatment as stipulated by Hypothesis 1. Second, some of those subjects who do inform themselves must not act pro-socially in the CONTROL treatment. Otherwise, the share of informed subjects who act pro-socially cannot increase as stipulated by Hypothesis 2). We summarize these observable prerequisites.

Prerequisite 1. *In the CONTROL treatment, some subjects do not inform themselves ($I = 0$). Among those subjects who do inform themselves, some do not act pro-socially, i.e. $a = 0$ although $W = w$ and $I = 1$.*

Recall that the CONTROL treatment closely follows existing treatments in the literature. Judging from the observations from these experiments, our prerequisites are likely to be met. Dana et al. (2007), for example, find in their hidden information treatment that 44% choose to reveal the consequences and that 25% of them do not act pro-socially. Exley and Kessler (2021) observe that 35% obtain the information and that 33% of them engage in harmful behavior.

3.3 Implementation

The experiment was conducted as an online experiment using the experimental software oTree (Chen et al., 2016). The sessions took place in April and May 2022. While we hoped to get fast and unreflected decisions in our CONTROL treatment from running the experiment online, this also bears the risk that subjects do not take the experiment seriously. They may, for example not read the instructions, so that the data becomes very noisy. Trying to reduce this risk, subjects were recruited from the pool of European rather than US American MTurk workers, which is reputed to have more attentive subjects. In addition, subjects were requested to have an approval rating above 95%. We verified their understanding of the instructions by control questions—see Figure 7 and 8. In order to prevent subjects from answering the multiple choice questions by 'trial-and-error', we only allowed them to make two mistakes per page and four in total. Of those subjects who were not excluded 23% made one mistake and 6% two or three mistakes before getting

the right answer. Our main results are robust to restricting the sample to those subjects who never committed errors in the control questions. Individuals had at most 15 minutes to finish all tasks but the actual duration is more likely to have been around 5 minutes.[5] After the experiment, we elicit demographics using a questionnaire. Median earnings were 1.50 USD,[6] leading to an hourly wage of about 18 USD. The experiment was approved by the ethical board of Paderborn University and preregistered using the American Economic Association's registry for randomized controlled trials (AEARCTR-0008172).

We conducted 13 sessions with a total of 359 Subjects. One subject did not answer the demographic questions, which reduces observation numbers in tables and regressions.[7] Subjects have been randomly assigned to treatments. All treatments took place at the same time and each subject only participated in one treatment.

4 Results

Before we turn to the effect on ignorance and pro-social behavior, we report on the success of randomization and check whether our pre-requisite condition is met.

4.1 Randomization and Prerequisites

Randomization seems to have worked well. Table 1 shows similar values for the observed demographics in the CONTROL and DESCRIBE treatment; none of the differences are significant at the 5% level (using Fisher's exact test). A potential worry could be that the share of women is higher in the CONTROL treatment; a difference that would be significant at a 10% level. Later, we will control for gender to see whether this overturns our results.

Recall our prerequisite, namely, that we need variation in behavior in the CONTROL treatment to test our predictions. We find that indeed a considerable share of subjects

[5]Unfortunately, we did not track the actual time. In a later study and a treatment very similar to our CONTROL treatment, subjects took on average 4:51 minutes for the whole experiment.

[6]Amazon MTurk uses USD as its default currency even for European workers.

[7]In the rest of the experiment the subject acted normally and i.e. answered open questions. We decided to not exclude the observation.

Table 1: Randomization Check

	CONTROL	DESCRIBE
Age	33.9	34.1
Students	36%	31%
Economics Student	18%	17%
Graduated	32%	30%
Female	35%	26%
Italian	44%	45%
British	29%	33%
Other European	27%	22%
# Observations	176	182

does not inform themselves in this treatment (53.98%)—see Table 2. This share is within the range typically found. Moreover, some of those who do inform themselves in this treatment do not act pro-socially and choose 'high' when this reduces the passive players' payoff—see Table 3. Again, this is what we would expect from the literature. All in all, the results in the CONTROL treatment match the existing ones. Crucially, they allow us to test our two hypotheses.

Table 2: Information Avoidance by Treatment

	CONTROL	DESCRIBE
Uninformed	95 (53.98%)	75 (40.98%)
Informed	81 (46.02%)	108 (59.02%)
# Observations	176	183

Note. In parentheses are the shares in the respective treatments.

4.2 Effect of DESCRIBE treatment on Information Acquisition

The share of informed subjects is almost 13 percentage points higher in the DESCRIBE treatment—see Table 2 and Figure 4; this amounts to an increase by more than 25%. Traditionally, Fisher's exact test is employed when dealing with such binary response data. The exactness of Fisher's test, however, crucially relies on the number of subjects who inform themselves to be known in advance. Since we did not know this marginal distribution beforehand, we use Boschloo's test (1970) as a uniformly more powerful

alternative to Fisher's test. We find that the reduction in ignorance is statistically significant (p-Value: .0073).[8]

Recall that we have more women in the CONTROL than in the DESCRIBE treatment. Women have been observed to act more pro-socially than men in various contexts (Rooney et al., 2005; Eckel and Grossman, 2008). Given that the rational choice for sufficiently pro-social individuals is to inform themselves, this might introduce a downward bias in our observed effect. Indeed, controlling for gender in a regression yields a slightly larger effect—see Model 2 in Table 4 in Appendix 6. If anything, the already sizable increase by 13 percentage points underestimates the effect.

Finally, we restrict the sample to all those 252 subjects who answered all control questions correctly at the first go; a group that may be more diligent. In this group, 51.2% inform themselves in the CONTROL as opposed to 66.4% in the DESCRIBE treatment. The difference amounts to 15.2 percentage points and is even larger than in the whole sample.

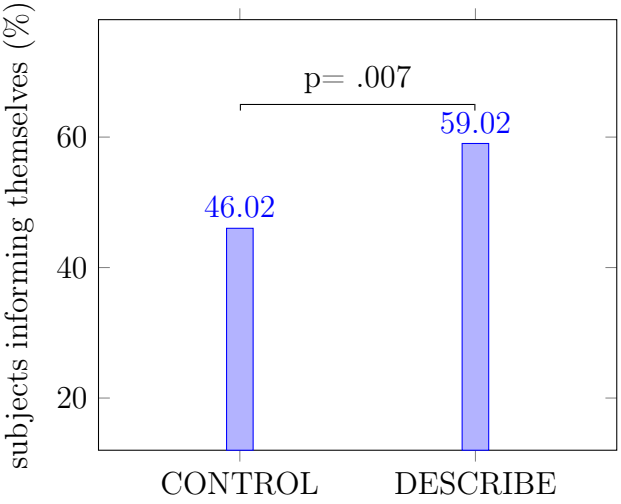


Figure 4: More subjects inform themselves when they are invited to describe their aims and means (DESCRIBE treatment) rather than in the standard benchmark (CONTROL treatment).

All in all, we find our first hypothesis confirmed.

Result 1. *More subjects obtain information in the DESCRIBE than in the CONTROL*

[8]Incidentally, significance does not rely on our choice of test. The conditional one-sided test by Fisher has a p-value of 0.009.

treatment.

4.3 Effect of DESCRIBE treatment on harm imposed

The relevant case in which harmful behavior can occur is when the active player has to give up her favorite choice 'high' for 'low' to prevent the passive player from losing out ($W = w$). In this case, the share of uninformed players who choose 'low' and prevent harm for the passive player is about 19% in the CONTROL treatment—see Table 3. The same probability in the same treatment increases to around 64% when the player is informed—an increase which is statistically significant (p-Value for Boschloo's test is lower than any conventional level).[9] All else being equal, reducing ignorance thus also decreases harm inflicted on the passive player because the active player chooses 'high'. This argument, however, assumes that informed individuals are equally likely in both treatments to choose the pro-social option 'low' when this prevents harm.

Table 3: Option 'low' selected by knowledge and treatment

	Uninformed	Informed	
		A prevents harm by 'low':	
		no ($W = 0$)	yes ($W = w$)
CONTROL	18/95 (18.9%)	2/39 (5.1%)	27/42 (64.3%)
DESCRIBE	9/75 (12.0%)	0/53 (0.0%)	45/55 (81.8%)
#Observations	170	92	97

Note. In parentheses are the shares by the knowledge and treatment. The fraction of uninformed subjects choosing low is comparable to Grossman and Van Der Weele (2017), who in a laboratory setting report a fraction of 15 % choosing low.

Hypothesis 2, however, even predicts that the share or pro-social behavior among informed subjects increases. Indeed, if subjects know that 'low' is preventing harm for the passive player, 64.3% of them choose 'low' in the CONTROL treatment as opposed to 81.8% in the DESCRIBE treatment—see Figure 5. Put differently, the share of harmful behavior in this group almost halves from 36% to 18%, which is relevant from an economic

[9]The same holds for Fisher's exact test.

perspective. This decrease is also statistically significant (p-Value for Boschloo's test is .0276).[10]

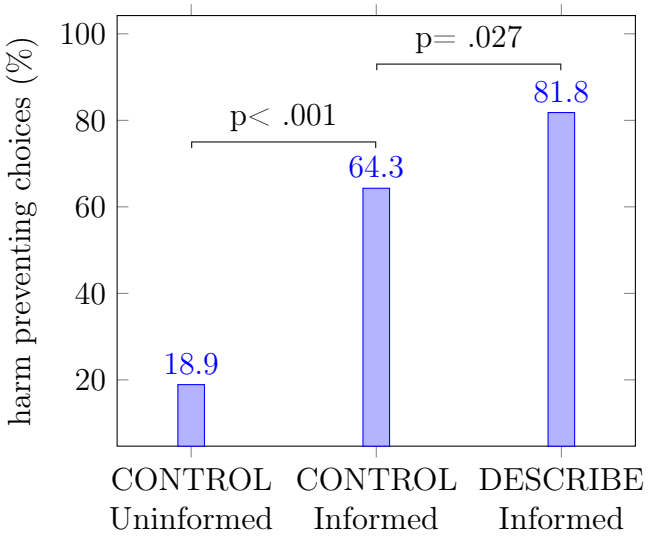


Figure 5: A passive subject is less likely to be harmed by a subject choosing 'high' in the CONTROL treatment when the subject is informed and even less likely by an informed subject in the DESCRIBE treatment.

Once more, we run a regression to see whether the gender composition of the treatments introduces a bias. We find that the observed treatment effect remains robust —see Table 5. To summarize, we also find our second hypothesis confirmed.

Result 2. *The share of informed subjects* $(I = 1)$ *who behave pro-socially* $(a = 1$ *in situation* $W = w)$ *is larger in the DESCRIBE than in the CONTROL treatment.*

Based on the data from the experiment, we can estimate how much harm is avoided in the DESCRIBE treatment, either because subjects inform themselves more often or because informed subjects act less harmfully. Note the probability that a passive subject who is harmed by the favorite 'high' choice $(W = w)$ remains unharmed $(a = 1)$ as $P(a = 1|W = w)$. We separate out the two cases in which harm can occur: if the active player is uninformed $(I = 0)$ or informed $(I = 1)$. This leads to the conditional probability:

$$P(a = 1|W = w) = P(a = 1|I = 0, W = w)\cdot P(I = 0)+P(a = 1|I = 1, W = w)\cdot P(I = 1).$$

[10]One -sided Fisher's exact test has a p-value of 0.043.

The respective terms on the right-hand side can be estimated for the CONTROL and DESCRIBE treatment using the data from Table 3 for the conditional probabilities and from Table 2 for the probability of encountering an informed subject. This yields:

$\widehat{P(a=1|W=w)}_{\text{CONTROL}} = .189 \cdot .540 + .643 \cdot .460 \approx .398$. for the CONTROL and
$\widehat{P(a=1|W=w)}_{\text{DESCRIBE}} = .120 \cdot .410 + .818 \cdot .590 \approx .532$. for the DESCRIBE treatment.

The probability for a subject exposed to harm to be unharmed is thus 13 percentage points or 33% higher in the CONTROL than in the DESCRIBE treatment—see Figure 6.

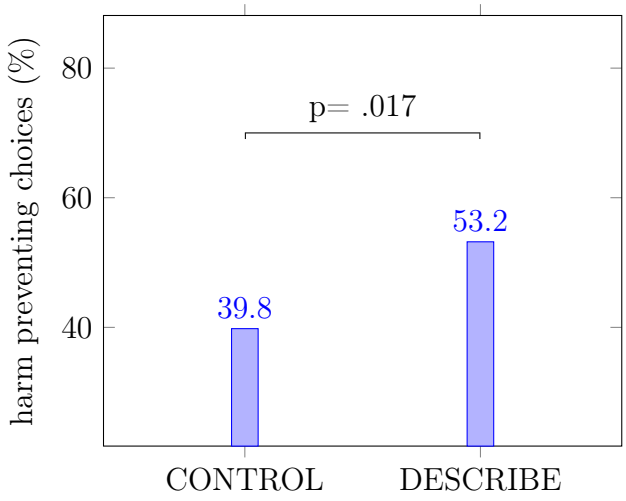


Figure 6: A passive subject who loses from 'high' is more likely to be harmed in the CONTROL than in the DESCRIBE treatment.

In order to examine whether this change is not only economically but also statistically significant, we construct a 2x2 table using the probability of remaining unharmed and the number of cases in which harm can occur. In the CONTROL treatment, we have a total of 137 cases in which the passive player may be harmed, 95 cases in which the active player is uninformed and 42 cases in which she is informed. Of these 137 cases, we expect 39.8% to be unharmed, i.e., 54.52 cases. In the DESCRIBE treatement, we have 75 uninformed and 55 informed cases, leading to a total of 130 cases in which the passive play may be harmed. Of these cases, we expect that the passive player remains unharmed in $130 \cdot .532 = 69.16$ cases. Since we cannot use fractions, we need to round these numbers. For a conservative test, we round up (and thereby increase) the cases in

which the passive player remains unharmed in the CONTROL treatment, while we round down in the DESCRIBE treatment—see Table 6 in Appendix 6. We find that the increase in harm preventing choices is statistically significant (p-Value for Boschloo's test .017).[11]

4.4 Other evidence for reduction of irrational behavior

The findings are consistent with our idea that subjects reflect on their behavior and act more rationally, i.e., they understand better what they want and which actions help with these goals. While the observed effects can be clearly attributed to the DESCRIBE screen and while the screen invites subjects to reflect on the decision problem, we do not know whether this actually happened or whether behavior changed for some other reason such as generally increased pro-sociality.[12]

The aim of this section is to bring some additional evidence that the screen helps subjects to act more rationally, in the sense of acting in line with their preferences, even if these preferences are not pro-social. Observe, for example, that an uninformed subject has no reason to choose 'low' irrespective of their preferences. If the subject is selfish, they clearly gain more from 'high'. If they are pro-social, they also gain from 'high' because 'high' and 'low' are equally likely to harm the passive player. Choosing 'low' is irrational. We observe that about 19% of the uninformed subjects in the CONTROL treatment nevertheless chose 'low'. In the DESCRIBE treatment, the share drops to 12%. The describe screen thus reduces this form of irrational behavior.

Next, consider informed subjects who choose 'low' when it not only reduces their benefit but also that of the passive player. Such behavior could be irrational or driven by anti-social preferences. If the behavior is driven by anti-social preferences, the share of subjects engaging in this behavior should remain constant across treatments. We find that the share of 'low' choices in the CONTROL treatment, where it harms active and passive subject ($W = 0$), is slightly above 5%. In the DESCRIBE treatment, none of the 53 subjects chooses 'low'. This is not consistent with anti-social behavior but suggests

[11]The p-value for Fisher's exact test is 0.023.

[12]Schnedler and Stephan (2020), for example, show that writing a letter can increase pro-social behavior. In contrast, subjects here write short descriptive sentences that are not addressed to another person.

once more that the DESCRIBE treatment increases rationality.

5 Discussion

Ignorance may result in harming others. In this study, we present evidence that harm can be reduced by giving people the opportunity to describe their aims and relate their choices to these aims. Our evidence is consistent with subjects acting more 'rationally' in the sense of better aligning their preferences and choices.

Our observation is made with small stakes in an online setting where decisions are usually made without deep contemplation like in many everyday situations. This suggests that interventions that encourage people to carefully consider the motives and options in a specific day-to-day decision problem, like buying chocolate, may help them to act more sustainably. Ultimately, one would want such a change to last. Of course, our study only concerns the short term and does not reveal whether the effect is still present half a year later. Nevertheless, our study provides useful input. Without a short term change in behavior, the intervention would have no chance to trigger a long term effect.

References

Andreoni, James, Justin M Rao, and Hannah Trachtman, "Avoiding the ask: A field experiment on altruism, empathy, and charitable giving," *Journal of Political Economy*, 2017, *125* (3), 625–653.

Bartling, Björn, Vanessa Valero, Roberto A Weber, and Lan Yao, "Public discourse and socially responsible market behavior," Technical Report, Working Paper 2020.

Bodner, Ronit and Drazen Prelec, "Self-signaling and diagnostic utility in everyday decision making," *The Psychology of Economic Cecisions*, 2003, *1*, 105–26.

Boschloo, RD, "Raised conditional level of significance for the 2× 2-table when testing the equality of two probabilities," *Statistica Neerlandica*, 1970, *24* (1), 1–9.

Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden, "The importance of moral reflection and self-reported data in a dictator game with production," *Social Choice and Welfare*, 2011, *36* (1), 105–120.

Carrington, Michal J, Benjamin A Neville, and Gregory J Whitwell, "Lost in translation: Exploring the ethical consumer intention–behavior gap," *Journal of Business Research*, 2014, *67* (1), 2759–2767.

Cason, Timothy and Charles Plott, "Misconceptions and game form recognition: Challenges to theories of revealed preference and framing," *Journal of Political Economy*, 2014, *122* (6), 1235 – 1270.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "'oTree' An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

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

DellaVigna, Stefano, John A List, and Ulrike Malmendier, "Testing for altruism and social pressure in charitable giving," *The Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

Eckel, Catherine C and Philip J Grossman, "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices," *Journal of Economic Behavior & Organization*, 2008, *68* (1), 1–17.

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

Exley, Christine L and Judd B Kessler, "Information avoidance and image concerns," Technical Report 28376, National Bureau of Economic Research 2021. Working Paper.

Farjam, Mike, Olexandr Nikolaychuk, and Giangiacomo Bravo, "Experimental evidence of an environmental attitude-behavior gap in high-cost situations," *Ecological Economics*, 2019, *166*, 106434.

Feiler, Lauren, "Testing models of information avoidance with binary choice dictator games," *Journal of Economic Psychology*, 2014, *45*, 253–267.

Ganguly, Ananda and Joshua Tasoff, "Fantasy and dread: The demand for information and the consumption utility of the future," *Management Science*, 2017, *63* (12), 4037–4060.

Grossman, Zachary, "Strategic ignorance and the robustness of social preferences," *Management Science*, 2014, *60* (11), 2659–2665.

_ **and Joël J Van Der Weele**, "Self-image and willful ignorance in social decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Hassan, Louise M, Edward Shiu, and Deirdre Shaw, "Who says there is an intention–behaviour gap? Assessing the empirical evidence of an intention–behaviour gap in ethical consumption," *Journal of Business Ethics*, 2016, *136* (2), 219–236.

Kahneman, Daniel, *Thinking, fast and slow*, Macmillan, 2011.

Kajackaite, Agne, "If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment," *Journal of Economic Behavior & Organization*, 2015, *116*, 518–524.

Knutsson, Mikael, Peter Martinsson, and Conny Wollbrant, "Do people avoid opportunities to donate?: A natural field experiment on recycling and charitable giving," *Journal of Economic Behavior & Organization*, 2013, *93*, 71–77.

Larson, Tara and C Monica Capra, "Exploiting moral wiggle room: Illusory preference for fairness? A comment," *Judgment and decision Making*, 2009, *4* (6), 467.

Li, Yufeng, Juanjuan Meng, Changcheng Song, and Kai Zheng, "Information avoidance and medical screening: a field experiment in China," *Management Science*, 2021, *67* (7), 4252–4272.

Matthey, Astrid and Tobias Regner, "Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior," *Games*, 2011, *2* (1), 114–135.

Rooney, Patrick M., Debra J. Mesch, William Chin, and Kathryn S. Steinberg, "The effects of race, gender, and survey methodologies on giving in the US," *Economics Letters*, 2005, *86* (2), 173–180.

Schnedler, Wendelin and Nina Lucia Stephan, "Revisiting a remedy against chains of unkindness," *Schmalenbach Business Review*, 2020, *72*, 347–364.

Serra-Garcia, Marta and Nora Szech, "The (in) elasticity of moral ignorance," *Management Science*, 2022, *68* (7), 4815–4834.

Spiekermann, Kai and Arne Weiss, "Objective and subjective compliance: A norm-based explanation of 'moral wiggle room'," *Games and Economic Behavior*, 2016, *96*, 170–183.

Stanovich, Keith E and Richard F West, "Individual differences in reasoning: Implications for the rationality debate?," *Behavioral and Brain Sciences*, 2000, *23* (5), 645–665.

Thunström, Linda, Madison Ashworth, Jason F Shogren, Stephen Newbold, and David Finnoff, "Testing for COVID-19: Willful ignorance or selfless behavior?," *Behavioural Public Policy*, 2021, *5* (2), 135–152.

Zizzo, Daniel John, "Experimenter demand effects in economic experiments," *Experimental Economics*, 2010, *13* (1), 75–98.

6 Appendix

Additional Tables

Table 4: The effect of the DESCRIBE treatment on whether the subject obtains information is even larger when controlling for demographics.

OLS Regression	Model 1		Model 2	
Subject obtains information	Coef.	Std. Err.	Coef.	Std. Err.
DESCRIBE	.128	(.053)	.142	(.052)
Female			.113	(.057)
Age			−.001	(.003)
Student			.180	(.079)
Graduated			−.202	(.068)
Econ. Student			−.065	(.077)
British			.028	(.071)
Italian			.034	(.069)
Constant	.460	(.037)	.453	(.141)
				N=358

Table 5: The effect of the DESCRIBE treatment on avoiding harmful choices is robust when controlling for demographics

OLS Regression	Model 1		Model 2	
Harm preventing choice	Coef.	Std. Err.	Coef.	Std. Err.
Informed	.453	(.077)	.488	(.079)
DESCRIBE	.172	(.085)	.163	(.086)
Female			−.043	(.065)
Age			.004	(.004)
Student			.095	(.094)
Graduated			.072	(.084)
Econ. Student			−.100	(.089)
British			−.128	(.085)
Italian			−.178	(.081)
Constant	.189	(.043)	.140	(.184)
				N=191

Table 6: Expected number of 'high' choices when 'high' is harmful

	'high'	'low'	total
CONTROL	82	55	137
DESCRIBE	61	69	130

Entries are computed using that 40% of subjects prevent harm in CONTROL and 53% in DESCRIBE—see p. 19.

Predictions using Self-image Model

The predictions for our experiment by the self-image model of Grossman and Van Der Weele (2017) depend on the exact interpretation of our DESCRIBE treatment and on whether the clicking to reveal information is seen as costly, neutral or even exciting (=benefit generating).

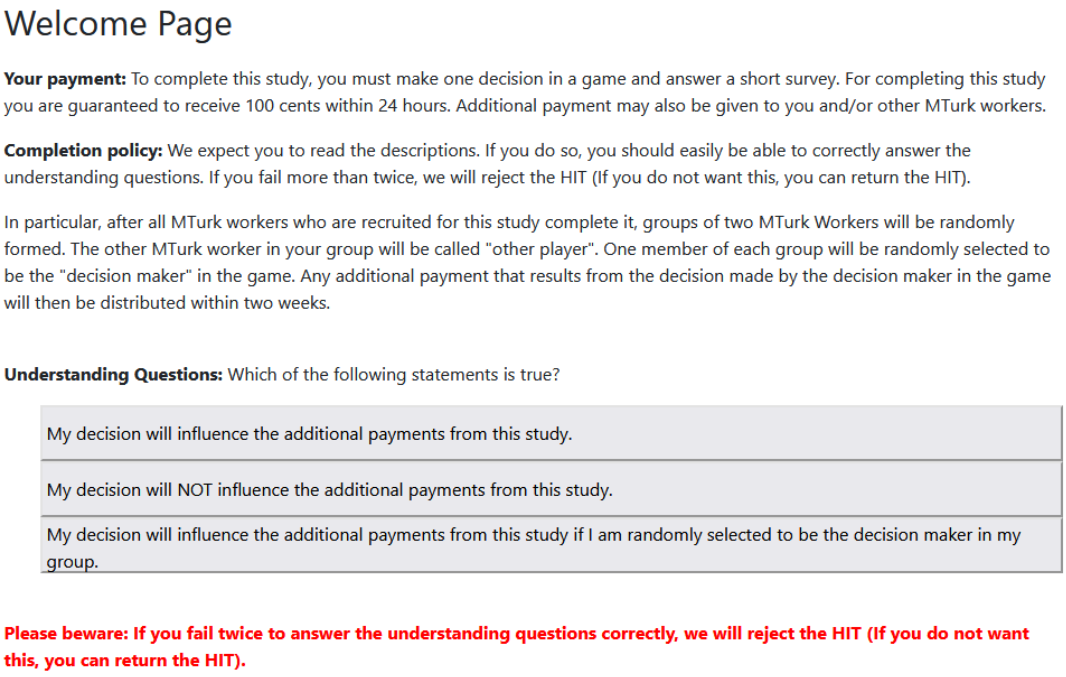
First, suppose that subjects in the DESCRIBE treatment upon reflection arrive at a contingent strategy. If subjects then come with a natural curiosity, then Proposition 4 in Grossman and Van Der Weele (2017) directly applies and fully in line with our own prediction, the share of subjects who inform themselves increases. Interestingly, subjects could still use a self-signaling mechanism very similar to that described by Grossman and Van Der Weele (2017) and deduce their type from the contingent strategies. More specifically, subjects who sufficiently care about others $\theta > \theta^*$ would pool and choose $a = 1$ whenever $W = w$ and obtain the image value from pooling with the other subjects in this group. At the same time subjects with $\theta < \theta^*$ would always choose $a = 0$ and likewise pool. Since the expected image gains from this behavior are the same as in the equilibrium, where subjects remember whether they have informed themselves or not, the cut-off value θ^* is exactly the same as in Proposition 1 in Grossman and Van Der Weele (2017). In this equilibrium, subjects engage in the same action choices in the CONTROL and DESCRIBE treatment. Since the share of those who inform themselves increases, the share of those who are informed and act pro-socially decreases. This prediction runs counter to our prediction.

Second, suppose that subjects in the DESCRIBE treatment arrive at a contingent strategy but revealing comes with (albeit small) costs. While informing is no longer necessary to signal the type, this signal can be sent as just described using the contingent actions. Subjects with $\theta > \theta^*$ in both treatments then need the information to execute their contingent plan and are willing to incur respective costs. Subjects with $\theta < \theta^*$ do not inform themselves. Informing oneself can thus still act as a signal, even though it is no longer required. In this interpretation, the share of those who inform themselves and those among the informed who act pro-socially is the same in both treatments. In contrast, we predict that the share increases in the DESCRIBE treatment.

Third, the DESCRIBE treatment may trigger a subject to reveal their own type perfectly to herself, which she is not able to do in the CONTROL treatment. For example, she may honestly admit to herself that she is selfish. In the CONTROL treatment, we would then have an equilibrium akin to that of Proposition 1 in Grossman and Van Der Weele (2017), where everyone with $\theta > \theta^*$ gets the information. In the DESCRIBE treatment, there would be no gains from signaling. Subjects will decide their action a based on their preferences. The subject defined by $W - \hat{\theta}c = 0$ will be indifferent and all above $\hat{\theta}$ will choose $a = 1$ whenever $W = w$ and all others $a = 0$. In order to implement their strategy, the subjects above $\hat{\theta}$ have to inform themselves. Notice that $\hat{\theta} > \theta^*$ because with self-signaling people have an additional incentive to inform themselves (pooling with the more attractive selfless type). In other words, the share of subjects who inform themselves should be lower in the DESCRIBE than in the CONTROL treatment, which is the opposite of our prediction.

Whichever of these interpretations we choose, the predictions differ from those made in our model.

Experimental Screens



Welcome Page

Your payment: To complete this study, you must make one decision in a game and answer a short survey. For completing this study you are guaranteed to receive 100 cents within 24 hours. Additional payment may also be given to you and/or other MTurk workers.

Completion policy: We expect you to read the descriptions. If you do so, you should easily be able to correctly answer the understanding questions. If you fail more than twice, we will reject the HIT (If you do not want this, you can return the HIT).

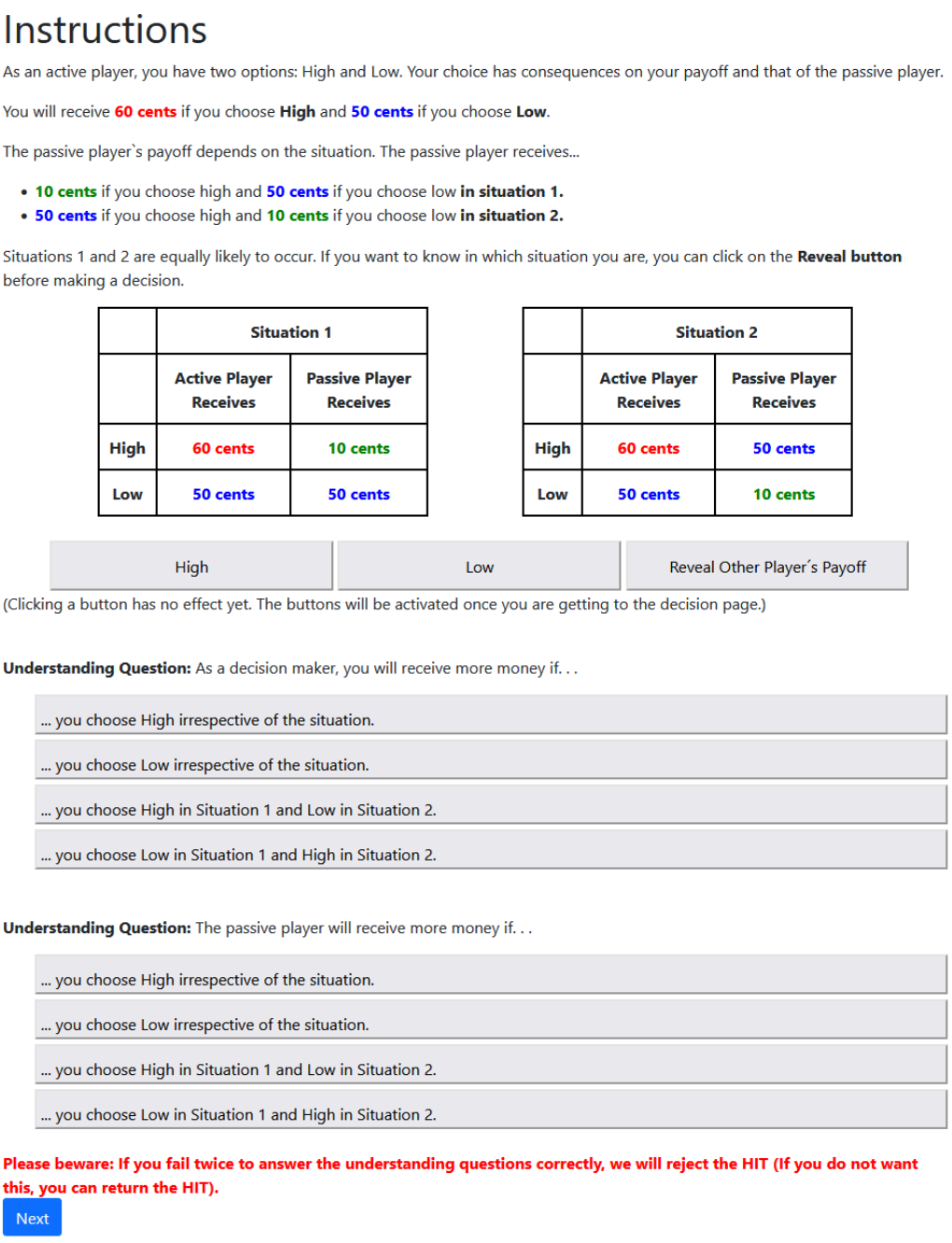
In particular, after all MTurk workers who are recruited for this study complete it, groups of two MTurk Workers will be randomly formed. The other MTurk worker in your group will be called "other player". One member of each group will be randomly selected to be the "decision maker" in the game. Any additional payment that results from the decision made by the decision maker in the game will then be distributed within two weeks.

Understanding Questions: Which of the following statements is true?

- My decision will influence the additional payments from this study.
- My decision will NOT influence the additional payments from this study.
- My decision will influence the additional payments from this study if I am randomly selected to be the decision maker in my group.

Please beware: If you fail twice to answer the understanding questions correctly, we will reject the HIT (If you do not want this, you can return the HIT).

Figure 7: First page of instructions including first understanding question



Instructions

As an active player, you have two options: High and Low. Your choice has consequences on your payoff and that of the passive player.

You will receive **60 cents** if you choose **High** and **50 cents** if you choose **Low**.

The passive player`s payoff depends on the situation. The passive player receives...

- **10 cents** if you choose high and **50 cents** if you choose low **in situation 1.**
- **50 cents** if you choose high and **10 cents** if you choose low **in situation 2.**

Situations 1 and 2 are equally likely to occur. If you want to know in which situation you are, you can click on the **Reveal button** before making a decision.

	Situation 1	
	Active Player Receives	Passive Player Receives
High	60 cents	10 cents
Low	50 cents	50 cents

	Situation 2	
	Active Player Receives	Passive Player Receives
High	60 cents	50 cents
Low	50 cents	10 cents

High | Low | Reveal Other Player´s Payoff

(Clicking a button has no effect yet. The buttons will be activated once you are getting to the decision page.)

Understanding Question: As a decision maker, you will receive more money if. . .

- ... you choose High irrespective of the situation.
- ... you choose Low irrespective of the situation.
- ... you choose High in Situation 1 and Low in Situation 2.
- ... you choose Low in Situation 1 and High in Situation 2.

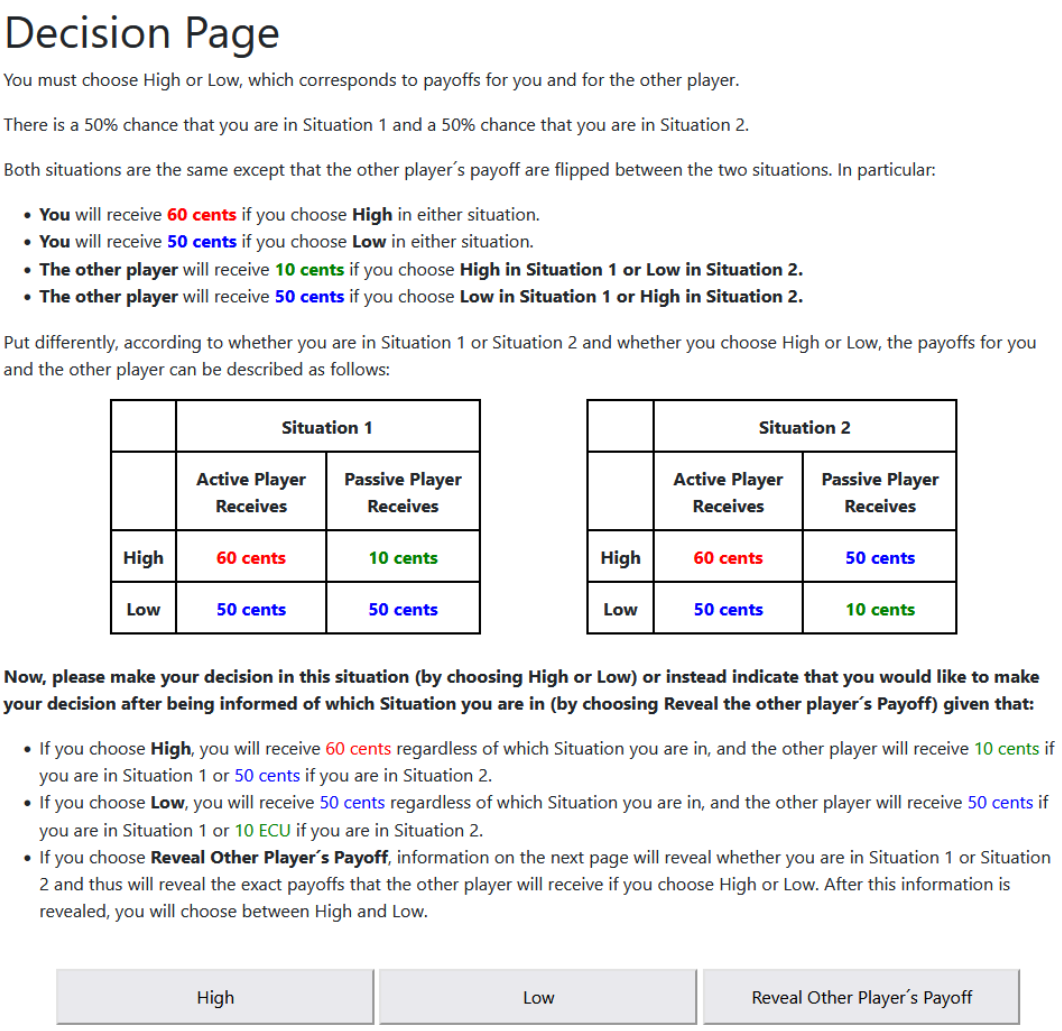
Understanding Question: The passive player will receive more money if. . .

- ... you choose High irrespective of the situation.
- ... you choose Low irrespective of the situation.
- ... you choose High in Situation 1 and Low in Situation 2.
- ... you choose Low in Situation 1 and High in Situation 2.

Please beware: If you fail twice to answer the understanding questions correctly, we will reject the HIT (If you do not want this, you can return the HIT).

Next

Figure 8: Second page of instructions including understanding questions about the decision problem



Decision Page

You must choose High or Low, which corresponds to payoffs for you and for the other player.

There is a 50% chance that you are in Situation 1 and a 50% chance that you are in Situation 2.

Both situations are the same except that the other player´s payoff are flipped between the two situations. In particular:

- **You** will receive **60 cents** if you choose **High** in either situation.
- **You** will receive **50 cents** if you choose **Low** in either situation.
- **The other player** will receive **10 cents** if you choose **High in Situation 1 or Low in Situation 2.**
- **The other player** will receive **50 cents** if you choose **Low in Situation 1 or High in Situation 2.**

Put differently, according to whether you are in Situation 1 or Situation 2 and whether you choose High or Low, the payoffs for you and the other player can be described as follows:

	Situation 1	
	Active Player Receives	**Passive Player Receives**
High	**60 cents**	**10 cents**
Low	**50 cents**	**50 cents**

	Situation 2	
	Active Player Receives	**Passive Player Receives**
High	**60 cents**	**50 cents**
Low	**50 cents**	**10 cents**

Now, please make your decision in this situation (by choosing High or Low) or instead indicate that you would like to make your decision after being informed of which Situation you are in (by choosing Reveal the other player´s Payoff) given that:

- If you choose **High**, you will receive 60 cents regardless of which Situation you are in, and the other player will receive 10 cents if you are in Situation 1 or 50 cents if you are in Situation 2.
- If you choose **Low**, you will receive 50 cents regardless of which Situation you are in, and the other player will receive 50 cents if you are in Situation 1 or 10 ECU if you are in Situation 2.
- If you choose **Reveal Other Player´s Payoff**, information on the next page will reveal whether you are in Situation 1 or Situation 2 and thus will reveal the exact payoffs that the other player will receive if you choose High or Low. After this information is revealed, you will choose between High and Low.

High | Low | Reveal Other Player´s Payoff

Figure 9: Initial decision page when subject does not know the consequences of 'High' and 'Low'

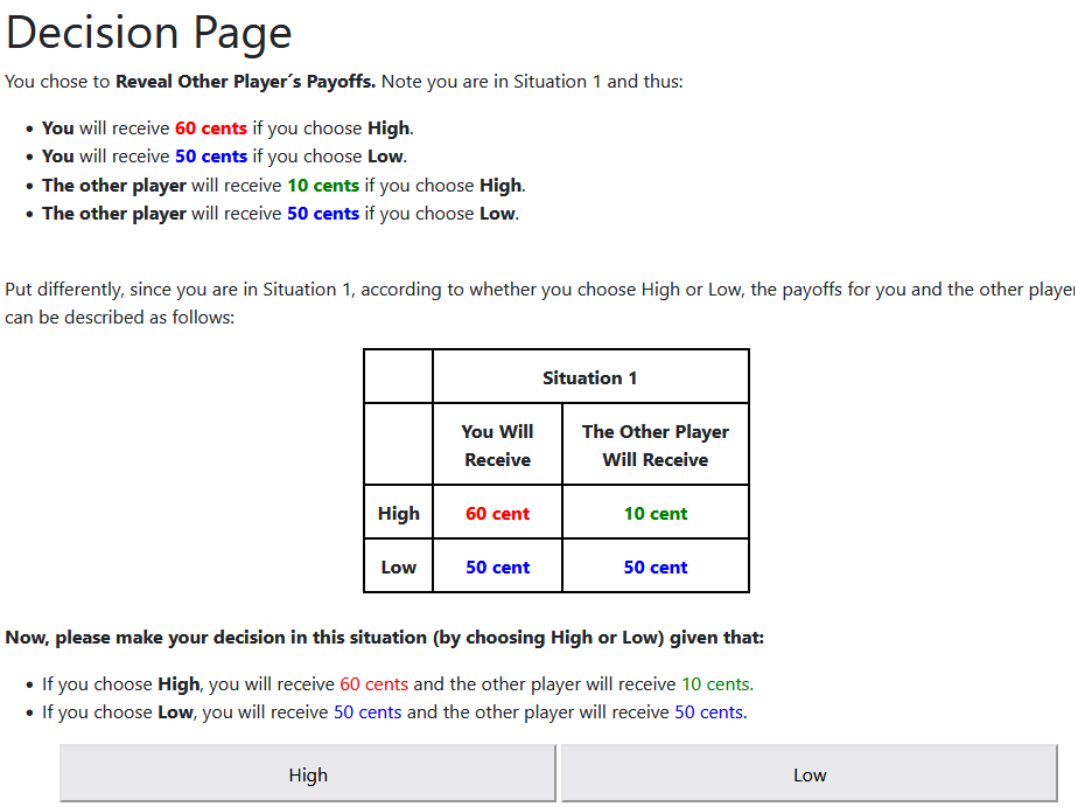


Figure 10: Decision screen when subject informed herself and 'High' leads to harm

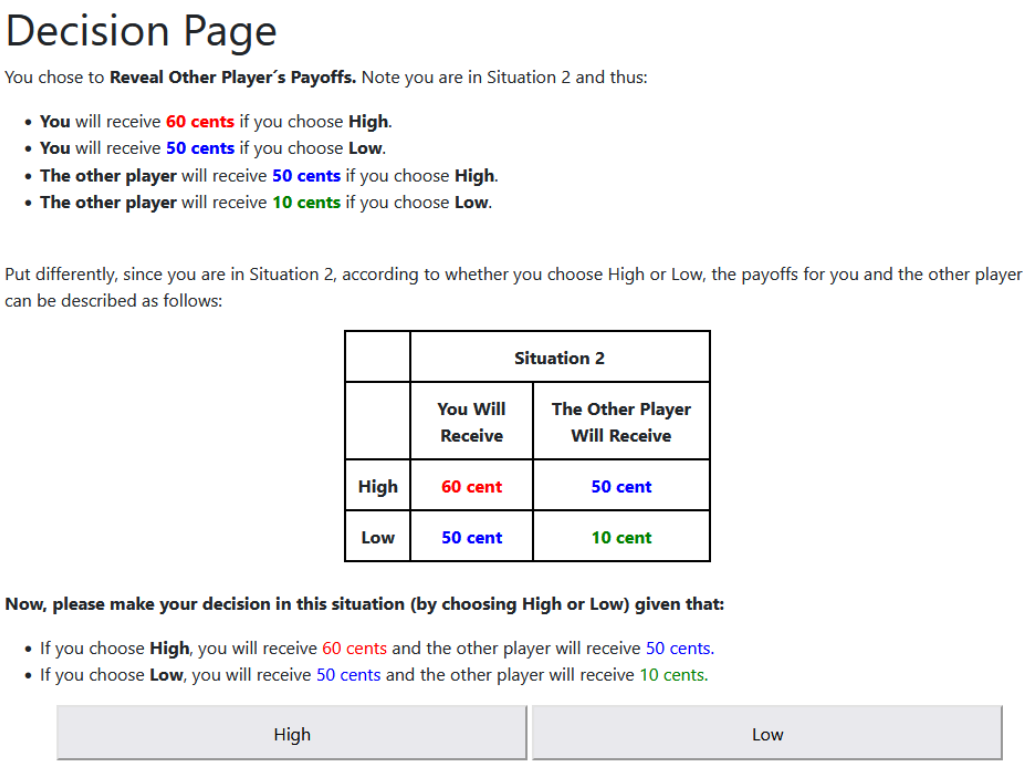


Figure 11: Decision screen when subject informed herself in and 'High' leads to no harm

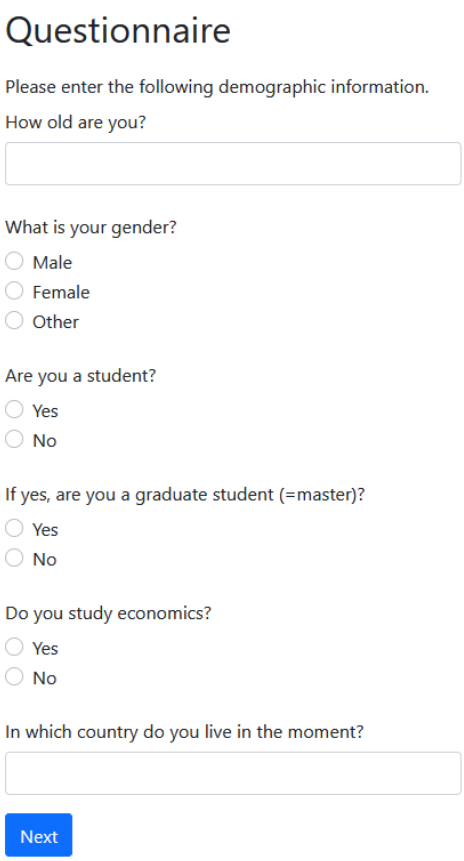


Figure 12: Demographic questionnaire

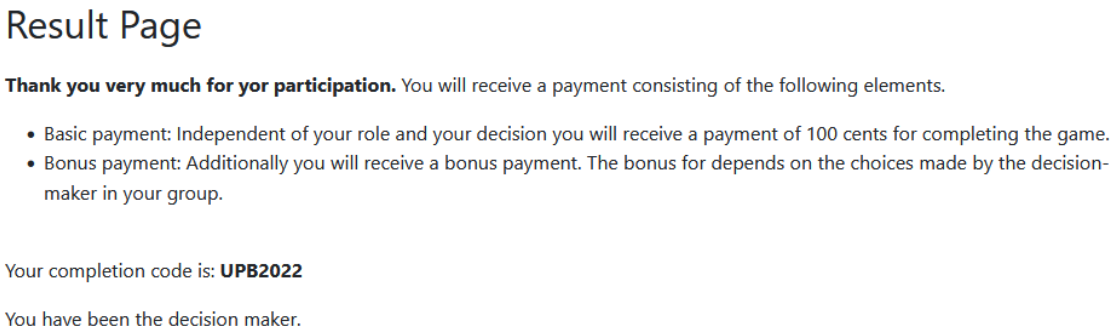


Figure 13: Landing page at the end of the experiment