TI 2025-023/I
Tinbergen Institute Discussion Paper

Giving as a self-control problem

Cristina Figueroa[1]
Jantsje Mol[2]
Ivan Soraperra[3]
Joël J. Van der Weele[4]

1 University of Amsterdam, Tinbergen Institute

2 University of Amsterdam

3 Max Planck Institute for Human Development

4 University of Amsterdam, Tinbergen Institute

Giving as a Self-Control Problem

Cristina Figueroa[*] Jantsje M. Mol[†] Ivan Soraperra[‡] Joël J. van der Weele[§]

March 24, 2025

Abstract

Social preferences depend on emotional states like compassion and anger. Since emotions are fleeting and subject to manipulation, they may generate demand for commitment. We investigate the use of commitment strategies in an online experiment ($n = 1,400$), where subjects decide to watch or avoid videos before engaging in a charitable giving task. We find that a video with emotional content increases giving, but is also avoided more than non-emotional videos. We estimate a structural model of state-dependent social preferences, and show evidence for sophisticated commitment to selfishness *and* altruism. We argue that giving can be fruitfully analyzed as a self-control problem.

JEL classification: D91, D64, C91

Keywords: Giving, Experiment, Empathy, Self-Control, Sophistication

[*]University of Amsterdam, Tinbergen Institute. Email: c.figueroa@uva.nl.

[†]University of Amsterdam, University of Applied Sciences Rotterdam. Email: j.m.mol@hr.nl.

[‡]Max Planck Institute for Human Development. Email: soraperra@mpib-berlin.mpg.de.

[§]University of Amsterdam, Tinbergen Institute. Email: j.j.vanderweele@uva.nl.

We thank Save the Children UK (especially Gemma) and Samuel Walsh, and seminar audiences at the LMU Munich, ICSD 2024 in Leiden, the Empathy and Emotion Lab at the University of California in San Diego, the EEA 2024 in Rotterdam, University of Heidelberg, University of Zürich and FAIR/NHH in Bergen. We particularly thank Jason Dana, Stefano Della Vigna, Ernst Fehr, David Levine, Clément Staner, Stefan Trautmann, Severine Toussaert and Bertil Tungodden for useful comments. We gratefully acknowledge financial support from the Amsterdam Center for Behavioral Change and A Sustainable Future at the University of Amsterdam. Ethical Approval was granted by the University of Amsterdam with reference number EB-778.

[A]nimum rege, qui nisi paret, imperat.
(Control your temper, because if it does not obey, it commands.)
– Horace

1 Introduction

Emotions can cause temporary shifts in social preferences. For instance, the emotions of anger and compassion affect social behavior in bargaining and social dilemma problems (Grimm and Mengel, 2011; Gneezy et al., 2014; Drouvelis and Grosskopf, 2016; Nguyen and Noussair, 2022). Moreover, others induce emotions strategically, like when opponents provoke us into anger (Gneezy and Imas, 2014) or charities manipulate our compassion (Small and Loewenstein, 2003). Emotions can thus create preference conflicts and self-control problems between emotionally charged ("hot") and neutral ("cold") states.

To mitigate such self-control problems, people may look for commitment strategies. Emotional commitment can serve to reinforce either altruism or selfishness. For instance, people may try to avoid requests for donations — which may explain "ask avoidance" observed in both the economic laboratory (Dana et al., 2006; Lazear et al., 2012) and in field studies (DellaVigna et al., 2012; Andreoni et al., 2017). But commitment may also favor altruistic behavior. Individuals may commit to acts of generosity in order to counteract materialistic impulses and "be a better person". This can explain pledges for periodic charitable giving and religious or spiritual practices that cultivate empathy.[1]

Commitment is a relatively unexplored topic in the domain of social preferences, contrasting with an extensive economic literature on commitment in contexts such as exercise, personal finance, and dietary choices. We address this gap, and study anticipated emotions and commitment seeking. In our experiments, 1400 UK residents on the online platform Prolific decide whether to make a donation to the charity Save the Children (StC). Before deciding, participants were asked to state their preference between watching either a neutral video or another video that depended on the experimental treatment. In the *Emotional* treatment, the video shows an emotional story of a victim in a warzone designed to stimulate empathy. In the *Direct Ask* treatment, the treatment video shows a direct fundraising ask. In a *Control* treatment, both videos are neutral and unrelated to charity. Before choosing which video to watch, participants read a short description of the video that allows them to anticipate its effect.

We identify sophisticated strategies of emotional self-regulation through several key design features. First, our treatments systematically vary the ability to regulate emotions and mitigate social pressure to give. Second, we disentangle the effect of video exposure from participants'

[1]For example, within Christianity, the "prayer of intercession" encourages believers to imagine the suffering of others and pray for their relief. Similarly, in Buddhist and Hindu traditions, the concept of Karuna emphasizes the intentional development of compassion as a means of fostering a more ethical and fulfilling life.

self-selection by occasionally overriding their video preference at random. This approach allows us to examine behavior in counterfactual scenarios that participants did not voluntarily choose. Third, we elicit participants' beliefs about their anticipated behavior across different emotional states, enabling us to assess their level of sophistication regarding potential self-control challenges. Finally, unlike field studies, our experimental design controls for both time costs (as all participants watch a video) and cognitive effort (as all participants make a donation decision), ensuring that these factors do not confound observed behavior.

The data show several patterns in line with the use of commitment, some of which replicate existing findings, while others are novel to our study. First, watching the emotional video increases giving by 18 percentage points compared to a neutral video, consistent with previous findings in the literature on empathy and giving (Long, 1976; Alpizar et al., 2008; Meer, 2011; Verhaert and Van den Poel, 2011; Smith et al., 2020; Andries et al., 2024). Second, we find evidence for sorting: more altruistic subjects are more likely to select the videos with emotional appeals or direct asks, in line with the literature (Lazear et al., 2012; Grossman and Van der Weele, 2017). Third, while the emotional video has a large effect on giving, it also results in the highest avoidance levels, as measured by the fraction of participants who choose to watch the alternative neutral video, compared to the neutral control or the direct ask video. Fourth, avoidance does not simply reflect aversion to videos with negative (anticipated) valence, but is closely linked to behavior. In particular, the video increases giving both among subjects who wanted to see it and who wanted to avoid it, and participants are at least partially aware of this.

To make quantitative assessments of the role of commitment and sophistication, we interpret our results in the context of a structural model of self-control based on Gul and Pesendorfer (2001) and Loewenstein et al. (2015). We assume that agents in a "cold" state decide which emotions to experience in a "hot" state, and that preferences for giving may differ across states. Sophisticated agents in the cold state may try to constrain or seek emotional impulses in order to affect their desired level of giving. Using our choice data as well as the participants' predictions about their own counterfactual giving, we quantify the distribution of giving preferences across states and the level of sophistication of the sample. We find that 26% of participants have "state-dependent preferences" (i.e. their choice depends on the video), and 60% of these agents chose the emotional video, in line with commitment to altruism. Depending on our precise structural assumptions, we estimate 40% to 60% of participants to be sophisticated.

Taken together, our results highlight the interplay between deliberation and affect, and show that giving can be fruitfully analyzed in a self-control framework. This opens new directions for research on social preferences, linking it more closely to research on emotions and emotion regulation in psychology, as well as theoretical and empirical approaches to self-control problems in economics. In the conclusion, we describe several detailed avenues for future research.

More specifically, our paper contributes to the literature in several ways. First, we add

to the literature on moral wiggle room and ask-avoidance. While previous literature has documented ask-avoidance, there has been little direct evidence about the mechanisms. Some papers highlight social pressure or self-image (DellaVigna et al., 2012; Cain et al., 2014; Grossman and Van der Weele, 2017), whereas other papers, based on similar giving situations, suggest emotion regulation as a mechanism (Andreoni et al., 2017). Moreover, ask-avoidance admits alternative explanations like minimizing cognitive load, or avoiding the time commitment associated with donation requests. Our paper contributes by showing that self-regulation of emotions and empathy plays an important role in ask avoidance. Moreover, it challenges the idea that people are mostly exploiting excuses for selfishness: a substantial fraction of subjects *seeks* to be emoted or convinced by the charity, behavior that is in line with a commitment to altruism but not selfishness. This is an understudied phenomenon, which relates to findings in Saccardo and Serra-Garcia (2023), who show that some people resist corrupting information in financial advice.

Second, we build a bridge between economic models of social preferences and the psychological literature on the motivated use of compassion (Cameron et al., 2022). Psychologists have shown that individuals use strategies to limit their exposure to empathy-inducing cues (Hodges and Biswas-Diener, 2007; Zaki, 2014) in order to avoid empathy's cognitive (Hodges and Klein, 2001; Cameron et al., 2019), emotional (Pancer, 1988; Davis et al., 1999), and material costs (Heider, 1982; Shaw et al., 1994). Shaw et al. (1994) relates most closely to our paper: in their experiment, individuals increase avoidance of descriptions labeled "high empathy" more if they know they will be asked to help the victims. Our paper differs in various ways: we use an incentivized giving task, we control for cognitive and time cost of giving, we avoid priming and experimenter demand effects as we do not use explicit labels for our intervention, we provide additional measures of counterfactual behavior to identify sophistication, and our sample sizes are approximately 30 times larger.

Third, we contribute to an understanding of the dual nature of human altruism and cooperation (Loewenstein and Small, 2007). A growing empirical literature, surveyed by Fromell et al. (2020), tries to identify whether cooperation and altruism are "intuitive" (as opposed to deliberative). Typical experimental interventions try to limit deliberative processes through the use of time pressure, ego depletion, or cognitive load, and study the resulting cooperation levels. This has yielded mixed results without clear conclusions. We provide an alternative empirical strategy, where we investigate deliberative influences through the use of commitment. We show clear heterogeneity, whereby some people prefer to experience emotions prior to giving, offering an explanation for the inconclusive results so far.

Finally, we add to the literature on sophistication about self-control and commitment demand. The measurement of sophistication has usually followed the modeling framework of time-inconsistency, in particular O'Donoghue and Rabin (1999). Studies classify individuals as time-consistent, sophisticated, or naive based on O'Donoghue and Rabin (1999)'s corre-

spondence between ideal, predicted, and actual behavior.[2] In the context of charitable giving, Andreoni and Serra-Garcia (2021) show that individuals are time inconsistent, while Kölle and Wenner (2023) find no evidence for time inconsistency. Dreber et al. (2016) show that delayed payments in dictator games lead to decreased giving, which is in line with predictions of the dual-self model in their paper. The findings on time inconsistency in giving are mixed, and none of these studies looks directly at commitment or sophistication. In our experiment, there is minimal delay across decisions. Thus, instead of time inconsistency, we concentrate on commitment across *emotional states*. Jia (2022) also investigates self-control in the face of empathy, and studies motivated beliefs as a commitment device. By contrast, we consider situations in which people choose their emotions directly. Our approach aligns with the temptation model of Gul and Pesendorfer (2001) and the experiment by Toussaert (2018). We provide a novel approach to quantify naiveté by directly contrasting people's predicted counterfactual behavior (validated by incentivized beliefs) with the actual behavior of similar subjects.

2 Experimental Design

We investigate the role of commitment in the face of emotions and empathy. To this end, we designed a simple online experiment. The experiment features a choice between earning money for oneself or giving it to a charity, Save the Children (StC). Before deciding, participants choose to watch a video, which, depending on the treatment, may induce emotions or social pressure relating to the charity.

2.1 Timeline

Figure 1 shows a timeline of the experiment. We now describe each stage in more detail. Full instructions are avaliable in Appendix G.

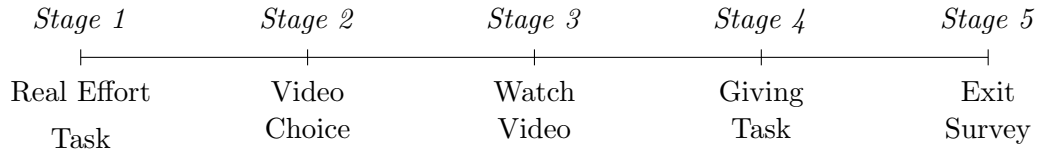


Figure 1: Timeline of the experiment

Stage 1: Real Effort Task. We implemented a short task to generate a sense of endowment over the money at stake in the participants. The task is based on the "Counting symbols on matrix" task in oTree (Chen et al., 2016). Participants need to count the number of right-pointing arrows in 4×5 arrow matrices for 2 minutes. Participants can only proceed if they answer correctly at least 7 matrices. This number was designed to make sure all participants could easily

[2]Examples include studies on university students' study habits (Wong, 2008; Mandel et al., 2017), weight loss (Cobb-Clark et al., 2024), and real-effort tasks (Cerrone and Lades, 2017).

make it. On average, participants tried out 11.5 matrices and solved correctly 10.7 of them. Participants were told that those who complete the task successfully would be entering a lottery to earn £5 with a probability of 20%. They would only learn at the end of the experiment whether they had won the £5. In particular, they had to make a giving choice conditional on earning the money. This design choice was made to strike a balance between our experimental budget and the potential stakes of the decision.

Stage 2: Video Choice. In this stage, participants are informed about how they can use the money earned in Stage 1, and we describe the giving task in Stage 4. Participants are informed that they will have to watch a video (2 minutes in length) and can choose between two videos. The videos depended on the treatment, as we describe in more detail below. Participants are informed that they will receive their preferred video most of the time but that it is possible that the computer overrides their preference, and they end up receiving the other video. To allow participants to anticipate the effect of the video, we briefly describe all videos (see descriptions in Table 1).

Stage 3: Watch Video. Individuals have to watch either their preferred (with 60% probability) or less preferred (with 40% probability) video in full screen and without the possibility of stopping it.

Stage 4: Giving Task. Subjects face a binary donation decision between themselves and StC. We chose StC as we wanted a well-known charity to avoid subjects choosing videos to seek additional information. StC was founded in the UK over a century ago and is one of the UK's biggest charitable organizations. The giving task has two options. Option A (you: £5, StC: £0) allows subjects to keep all money earned in Stage 1 for themselves. In Option B (you: £1, StC: £8), the subject keeps £1 and transfers £4, which is doubled by the experimenter and donated on the subject's behalf to StC. After the giving task, we ask individuals to explain their video preference and their donation decision in two separate open-ended questions.

Stage 5: Exit Survey. At the end of the experiment, subjects receive filled in a survey which includes (1) a questionnaire on donating behaviour, (2) the short Empathy Quotient Questionnaire (Wakabayashi et al., 2006) on emotions associated with the video, (3) two questions on the anticipated temptation and social pressure to donate associated with each video, (4) two questions on the experienced temptation and social pressure to donate under the received video, (5) a question on whether the random implementation / overriding rule mattered for the choice of preferred video, (6) an attention check for video visualization, and (7) two beliefs questions that served to better understand sophistication.

2.2 Treatments and videos

In each experimental treatment, participants could choose to watch one of two videos at the *Video Choice* stage. The first video was always the same across treatments, and was labeled "alternative video". It explains how grass grows and why it is green. This video was chosen to induce as little emotions as possible and be of roughly equal length as the treatment videos. It was described as "an alternative video of similar length unrelated to charity".

The treatments differ in the second video and its corresponding description, as summarized in Table 1.[3] Our choice of treatment videos was designed to vary the emotional content and the social pressure to donate. The short video descriptions are designed such that participants can form anticipation of these factors.

In the *Emotional* treatment, we used an existing promotional video from StC. This video shows a girl whose city slowly turns into a warzone. We chose this video for three reasons. First, as we will show below, the video induces strong feelings of compassion as well as a range of other emotions. Second, according to StC it has a proven record of increasing donations.[4] Third, it provides no information at all about the charity itself or its activities, as its focus is to create empathy with the protagonist.

Treatment	**Treatment Video**	**Treatment Video Description**
Control	Video about the mechanics of wave formation.	This video explains how waves are formed, the types of waves that exist, and how different factors affect the behaviour of water in the ocean.
Emotional	Existing StC promotional video, showing a girl in warzone.	This video is part of a campaign by Save the Children. It shows the struggles of a nine-year-old girl when her city becomes a warzone. As the political conflict escalates, the girl and her family experience increasingly traumatic hardships and perils.
Direct Ask	Direct ask from the Head of fundraising of StC UK.	This video shows Gemma, the Director of Fundraising at Save the Children UK, asking participants to donate their money to Save the Children in order to support their work providing education, food, and medicine to children.

Table 1: Overview of experimental treatment videos.

In the *Direct Ask* treatment, the treatment video was a recording by the Head of fundraising of StC UK. She explains that she is a representative of StC and asks participants to donate the

[3]Links to the videos: alternative video https://tinyurl.com/3sdbccxb, control video https://tinyurl.com/bdfb8cey, emotional video https://tinyurl.com/3mujbbut. Due to privacy reasons, the direct ask video is not publicly available. Please contact the authors for further information.

[4]Check the Results section on the video from the creative agency that produced it: Don't Panic at www.dontpaniclondon.com/work/most-shocking-second-a-day/

money they earned to the charity. To further mimic aspects of interpersonal communication with a sollicitor, participants in the *Direct Ask* treatment were asked to type a message explaining why they did or did not choose to give to the charity, which would be forwarded to StC. This allows participants to give a reason for their decision, like an excuse for not giving (Exley and Petrie, 2018). The treatment, therefore, changes both the video and the existence of a message. To control for this difference and understand whether this message matters, we also ran a *Direct Ask No Message* treatment that omitted the message. Since we do not find significant differences depending on the availability of the message, we do not analyze this treatment in the main text, but in Appendix D.

Our treatments are designed to isolate the impact of social pressure generally, and emotion regulation specifically. The comparison between the *Control* treatment and the other two treatments shows the effect of pressure to give (emotional or otherwise). The comparison between the *Emotional* and the *Direct Ask* treatment serves two purposes. First, we can compare the impact of two common fundraising practices to persuade potential donors: direct solicitations and emotional appeals. Second, it allows us to identify alternative mechanisms for avoidance given in the literature. While treatments generate pressure on the participants, the direct ask video does not include any affective content, so participants should anticipate less emotional intensity from the description. To evaluate these claims, we elicit various (anticipated) emotions from a separate sample — see Section 3.1.

2.3 Experimental Procedure and Sample Characteristics

The experiment was programmed in oTree (Chen et al., 2016) and conducted online in January 2024 on the Prolific platform. We recruited 1400 UK participants in total ($n_{Control} = 403$, $n_{Emotional} = 399$, $n_{DirectAsk} = 401$, and $n_{DirectAskNoMessage} = 199$). We chose a UK sample because StC is well-known in the UK, and the emotional video is tailored to UK residents. Our treatments are balanced on sociodemographic characteristics, with the average participant being 40 years old, employed, born and residing in the UK, and of white ethnicity (see our Balance Table A1). The experiment lasted 15'25" on average, and participants received £1.60 as a participation fee. As detailed above, individuals could potentially earn a £5 bonus by choosing not to donate the money, and 10 cents by providing an accurate guess in the incentivized beliefs question. On average, individuals earned 40 pence in addition to the participation fee with these bonuses.

2.4 Hypotheses

The hypotheses, main analyses, and sample sizes were preregistered and accessible via https://aspredicted.org/5HS_R5J. Our first hypothesis concerns the precondition to study self-control aspects of giving behavior, namely that social pressure manipulations should have an

impact on giving behavior. In Section 3.1, we report the results of a manipulation check to understand the difference in emotional reactions to the treatment. Our hypotheses concern the behavioral reaction to the video.

Hypothesis 1 (Exposure effect). The treatment video in the *Emotional* and the *Direct Ask* treatments increases donations relative to the *Control*.

Next, the results in the ask avoidance literature show that people avoid triggers that induce pressure or non-giving (Dana et al., 2006; Lazear et al., 2012; DellaVigna et al., 2012; Andreoni et al., 2017). For this reason, we hypothesized that both our *Emotional* and *Direct Ask* treatments induce higher avoidance of the treatment video compared to the *Control*. While this hypothesis is based on the literature on ask avoidance, commitment might also go in the other direction. Participants may want to overcome momentary temptations of greed or materialism, and therefore seek to be emoted or pressured. Thus, the video choice will reflect a mix of such commitment motives.

Hypothesis 2 (Avoidance). The alternative video will be chosen more often in the *Emotional* and the *Direct Ask* treatments relative to the *Control*.

In addition to these hypotheses, we perform a number of additional analyses. First, previous research on ask avoidance does not explicitly compare social pressure and emotional regulation as drivers of avoidance behavior. Here, we contrast the *Emotional* and *Direct Ask* treatments to see which generates more avoidance and how such avoidance relates to behavior. Ex-ante, it is not clear which of those treatments will have larger effects, which is why we did not pre-register a specific hypothesis.

Second, we assess the degree of sophistication underlying participants' strategies. Interpreting the video choice in terms of self-control and commitment requires that individuals understand the impact of emotions or social pressure on decisions, and act to steer their own behavior. To assess this kind of sophistication, we use the random implementation of the preferred video. This allows us to compare donating behavior in counterfactual scenarios while keeping video preferences fixed. We also assess subjects' predictions of counterfactual donating behavior. In theory, a sophisticated participant should display correct beliefs about giving after observing a counterfactual video. In Section 4, we combine these ideas to provide a theoretical framework and give quantitative estimates of different behavioral types and their sophistication.

3 Results

In this section, we present an overview of the main results. The next section will interpret our results in the context of a structural model of self-control. We report the p-values of two-sided

tests throughout, unless indicated otherwise.

3.1 Manipulation Check

The main manipulation between treatments was the treatment video and its description. Since our focus is on the self-regulation of emotions, we evaluated both the emotions that individuals anticipated and experienced. We conducted this check on a separate sample (N = 147, $n_{Control}$ = 51, $n_{Emotional}$ = 49, and $n_{DirectAsk}$ = 47). *Before* watching the video, individuals had to rate the intensity they anticipated experiencing of the following emotions: sadness, happiness, guilt, boredom, compassion, anger, disgust, and fear. All they knew about the video was the description we provided. *After* watching the video, we also asked them to rate the intensity with which they had actually experienced the same emotions. Both elicitations used a 5-point scale ranging from "Very weak" to "Very high" intensity.

Figure C1 illustrates our results. The video from the *Emotional* treatment produces higher anticipation of sadness, disgust, fear, anger, compassion, and guilt than the video from *Control* treatment ($p < 0.05$ on two-sided t-test in each case). Anticipated emotions are also statistically significantly higher compared to those anticipated in the video from the *Direct Ask* treatment for all emotions (sadness, disgust, fear, anger, and compassion, with $p < 0.05$ in each case) except anticipated guilt, which is statistically indistinguishable ($p = 0.47$). Emotions in the *Direct Ask* treatment are anticipated to be significantly higher than in the *Control* treatment ($p < 0.05$). Participants anticipate less boredom and happiness after both charity-related videos, compared to *Control*.

Figure C2 shows similar patterns for *experienced* emotion. Participants experienced higher sadness, disgust, fear, anger, compassion, and guilt in both charity-related videos compared to *Control*, and stronger emotions after the video from the *Emotional* treatment than from the *Direct Ask* treatment.

In summary, the video from the *Emotional* treatment raised anticipation and experience of a range of emotions compared to the other two videos. This implies that the treatment comparisons we perform below can identify the effect of this cluster of emotions, but not of any emotion individually. However, since the *Direct Ask* treatment also shows a high level of anticipated guilt, associated with increased social pressure, this is statistically indistinguishable across the two charity-related treatments. Thus, we can think of the comparison of the *Direct Ask* treatment and the *Emotional* treatment as identifying the role of other emotions, such as compassion.

3.2 Donations

We hypothesized that both the *Direct Ask* and the *Emotional* treatment group would donate more relative to the *Control* treatment group. Indeed, when we simply aggregate all donations by treatment, donation levels are 37.34%, 37.16%, 27.29% for the *Emotional*, the *Direct Ask* and the *Control* treatment, respectively. Thus, following a two-sided proportion test, we conclude that subjects in the *Emotional* treatment ($\chi^2 = 8.804$, df $= 1$, $p = 0.003$) and the *Direct Ask* treatment ($\chi^2 = 8.506$, df $= 1$, $p = 0.004$) donated more often compared to subjects in the *Control* treatment.

Of course, these comparisons are confounded by the selection effect of the video. To assess the causal impact of video exposure, we control for video preferences in a regression.[5] Table 2 shows how exposure to the treatment video affects the decision to donate for each of our main treatments, controlling for preference for the treatment video.

Table 2: Linear probability model of donation decision

	Dependent Variable: Donation decision		
	Control	Emotional	Direct Ask
Constant	0.230^{***}	0.221^{***}	0.178^{***}
	(0.040)	(0.034)	(0.032)
Preferred Treatment Video (ref=no)	0.072^{*}	0.158^{***}	0.267^{***}
	(0.044)	(0.049)	(0.047)
Exposed to Treatment Video (ref=no)	0.001	0.180^{***}	0.101^{**}
	(0.044)	(0.048)	(0.048)
Observations	403	399	401
R^2	0.01	0.07	0.10

Notes: Linear probability models for the donation decision for each of the treatments. "Preferred Treatment Video" takes the value of 1 if the participant indicated that she wanted to see the treatment video. "Exposed to Treatment Video" takes the value of 1 if the participant was assigned to see the treatment video. Ref=no means that the individual did not prefer/receive the treatment video (i.e., she preferred/received the alternative). Robust standard errors are in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

The coefficients for the Received Treatment dummy indicate that being exposed to the treatment video had no effect on donations in the *Control* treatment ($\beta = 0.001$), whereas the treatment video increased donations by 18.0 and 10.1 percentage points in the *Emotional* and the *Direct Ask* treatment, respectively. We use Clogg et al. (1995)'s Z-test to compare the coefficients of Received Treatment across regressions. The *Emotional* treatment video increased donations relative to the *Control* ($Z = 2.724$, $p = 0.003$, one-sided), but there is no statistically significant evidence that it increased donations with respect to the *Direct Ask* treatment ($Z =$

[5] We obtain similar results using inverse probability weighting where observations are weighted by the inverse of the probability of receiving the treatment video.

1.162, $p = 0.245$, two-sided). The latter does increase donations relative to the *Control*, with marginal statistical significance ($Z = 1.516$, $p = 0.064$, one-sided). Thus, both treatments increase donations relative to a baseline without emotions and social pressure.

In addition, we see a correlation between choosing the charity-related video and donations. This is in line with selection or sorting: the idea that individuals who are motivated to give expose themselves to emotions or social pressure (Lazear et al., 2012; Grossman and Van der Weele, 2017). There is also a weaker and marginally significant relation in the *Control* treatment, for which we do not have a good explanation.

3.3 Avoidance

We expected individuals to display higher levels of avoidance (preference for the alternative video) under the *Emotional* and *Direct Ask* treatments relative to the *Control*. Figure 2 shows the overall levels of avoidance for each treatment: 42%, 58%, and 46% for the *Control*, *Emotional*, and *Direct Ask* treatment, respectively. There is evidence of higher avoidance in the *Emotional* treatment relative to the *Control* ($\chi^2 = 21.076$, df $= 1$, $p < 0.001$), yet no differences in avoidance when comparing the *Direct Ask* and *Control* treatment ($\chi^2 = 0.973$, df $= 1$, $p = 0.324$). In addition, comparing the *Emotional* and the *Direct Ask* treatment suggests higher avoidance for the former ($\chi^2 = 12.54$, df $= 1$, $p < 0.001$). All tests are two-sided proportion tests.

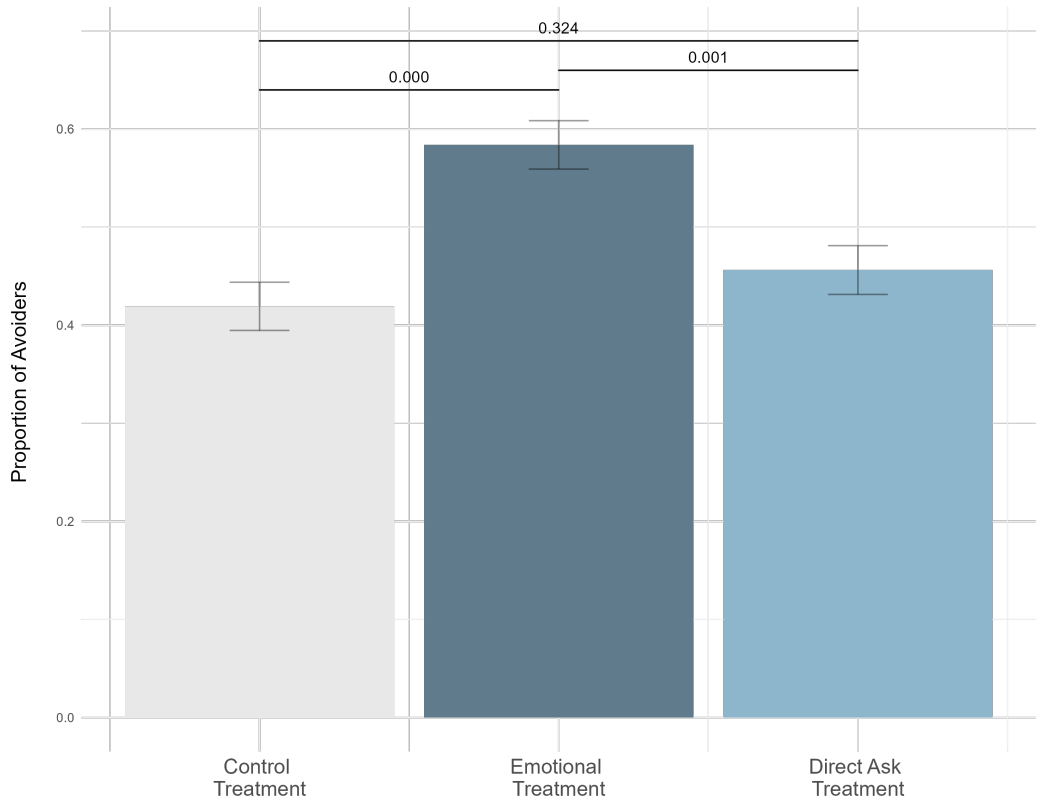


Figure 2: Proportion of individuals who chose the alternative video by treatment. p-values on top from proportions tests with Holm correction for multiple comparisons. Error bars denote +/- 1 SE.

3.4 Sophistication

Avoidance is often regarded as a behavior exhibited by "reluctant givers", who choose avoidance in order to reduce their giving or their discomfort. This interpretation presupposes some sophistication; agents need to anticipate the possible future states, their emotions in those states (guilt, compassion), and their behavioral response to these emotions. However, the results so far are also in line with other explanations, such as a wish to avoid (anticipated) negative emotions. To distinguish between these explanations, we investigate sophistication in more detail.

To define sophistication, we follow previous literature on self-control (O'Donoghue and Rabin, 1999; Gul and Pesendorfer, 2001): an individual is sophisticated when she can accurately predict her own choices in the future. In our context, this requires anticipating the effect of an altruistic trigger on one's own future utility and donating behavior.

So far, there has been little evidence on the role of sophistication in ask avoidance, or emotional regulation more broadly. Identifying sophistication is an empirical challenge, as the experimenter only observes how the subject behaved in a given situation, not in a counterfactual situation. Eliciting predictions of future behavior or emotional states before the behavior occurs is also problematic, as it may generate concerns for consistency or increase reflection that may

alter the very behavior intended to be measured. Incentives for accuracy may further magnify this problem.

Our approach is inspired by Toussaert (2018), who proposes to approximate the counterfactual behavior of the agent by looking at the behavior of "similar" individuals in the counterfactual situation. The design feature where we override the agent's preference with some probability creates an agent who is similar (i.e. has the same video preference), but who decided in a counterfactual emotional state, creating a proxy for the counterfactual behavior. Moreover, we assess sophistication by letting agents predict the counterfactual behavior of a similar other, as well as their own behavior. We discuss these measures of sophistication in turn.

Counterfactual donation behavior

The randomization of the video's assignment allows us to compare giving behavior after watching a neutral and a treatment video, *conditional* on the preference for the video. Figure 3 below shows the donation levels across treatments for individuals who wanted to see the alternative video (Panel A) and for those who wanted to see the treatment video (Panel B). The figure further distinguishes between those who received their preferred video and those who received the non-preferred video.

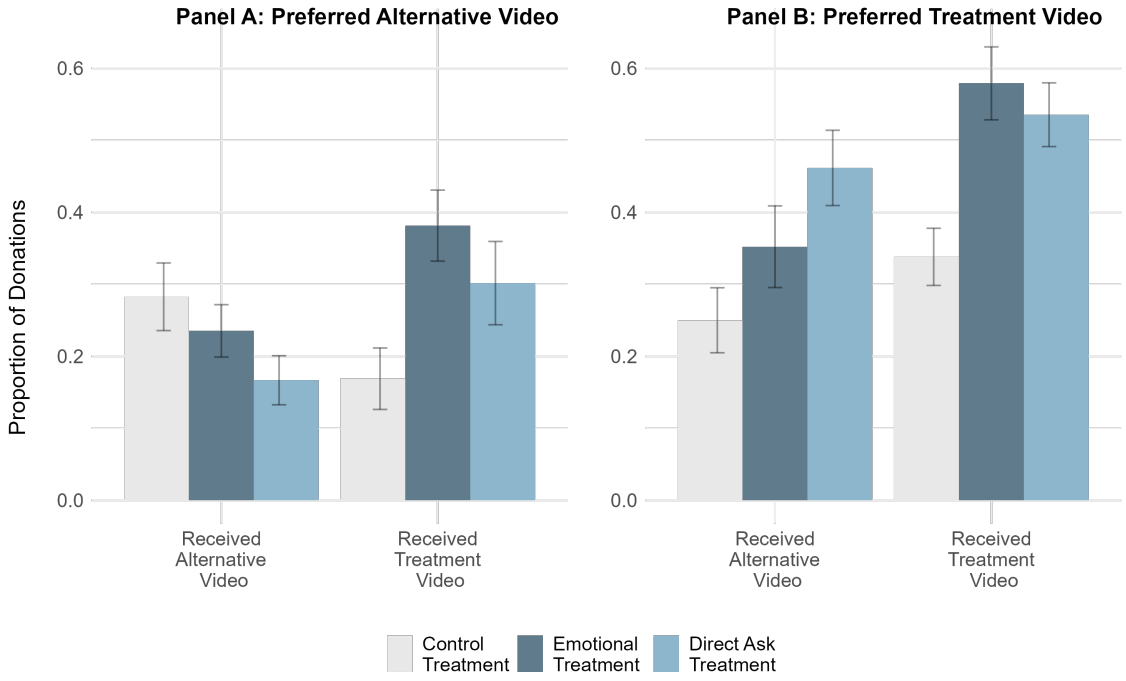


Figure 3: Proportion of donations across treatments (denoted in colors) split by preference (Alternative Video in Panel A; Treatment Video in Panel B) and exposure (within each panel: Receiving Alternative Video on the left versus Treatment Video on the right). Error bars denote +/- 1 SE.

Figure 3A shows that individuals who wanted to avoid the treatment video increased their donations if they were actually shown the video. This happens in both the *Emotional* treatment ($\chi^2 = 5.122, df = 1, p = 0.024$) and the *Direct Ask* treatment($\chi^2 = 3.716, df = 1, p = 0.054$) but not for the *Control* treatment, where we see a shift in the opposite direction ($\chi^2 = 2.450, df = 1, p = 0.118$). Although this pattern does not directly speak directly to participants' awareness, it is consistent with avoidance of the treatment video as a sophisticated commitment not to give. All tests are two-sided proportion tests.

Figure 3B shows donations from individuals who preferred to watch the treatment video. In this sample, it is not obvious ex-ante what the effect of the video should be. We observe increases in all treatments, which are statistically significant in the *Emotional* treatment ($\chi^2 = 7.490, df = 1, p = 0.006$) but not in the *Direct Ask* treatment ($\chi^2 = 0.881, df = 1, p = 0.348$) nor the *Control* treatment ($\chi^2 = 1.652, df = 1, p = 0.199$). It thus appears that preferences are less dependent on the video reception in the *Control* and *Direct Ask* treatment than in the *Emotional* treatment. This is compatible with individuals demanding the emotional video to "commit" to donating, something that has not been explored in the literature.[6] All tests are two-sided proportion tests.

In sum, the counterfactual donation is compatible with sophistication and demand for commitment. However, the behavioral patterns also admit other explanations. For instance, agents may simply be surprised by the effect of the treatment video and avoid it for other reasons. For more direct evidence that participants anticipate these effects, we turn to an analysis of beliefs.

Anticipation of counterfactual behavior

To measure anticipation of counterfactual behavior, we directly asked individuals about their donation in the hypothetical case where they saw the video they did not see in the actual experiment. The answers, elicited in the final questionnaire, allow us to see if people anticipate switching their behavior upon seeing a different video, and if the anticipation is consistent with the actual behavior.[7] To address the fact that this question is not incentive compatible, we also asked subjects to guess the proportion of participants donating (number of individuals out of 100) among those "similar others" who shared the same video preference but received the opposite video. They were incentivized by an award of 10 pence if the guessed proportion was +/- 0.1 of the actual proportion (see also the design section). The idea is that beliefs about others are a good proxy for what the agent herself would have done (Toussaert, 2018). Indeed, we find that both belief variables are highly and significantly correlated in all treatments, suggesting

[6] Note that in both panels of Figure 3, the behavior in the *Control* condition is consistent with *reactance*: subjects give less if they do not see their preferred video. We discuss this in more detail in Section 4.5.

[7] Note that both measures were elicited at the end of the experiment, after subjects watched the video and made the donation decision. This has the advantage that the belief elicitation does not distort the choices by causing additional reflection. However, it has the drawback that the answers may be influenced by the experience of watching the video.

that people took the unincentivized question about their own counterfactual behavior seriously.[8]

To assess subjects' anticipation, Figure 4 displays the proportion of "switchers", those who predict that they would have made a different choice in the counterfactual situation. We display proportions across treatments for individuals who preferred to watch the alternative video (Panel A) or the treatment video (Panel B), and further distinguish on the actual video that the participant saw and on their donation decision.

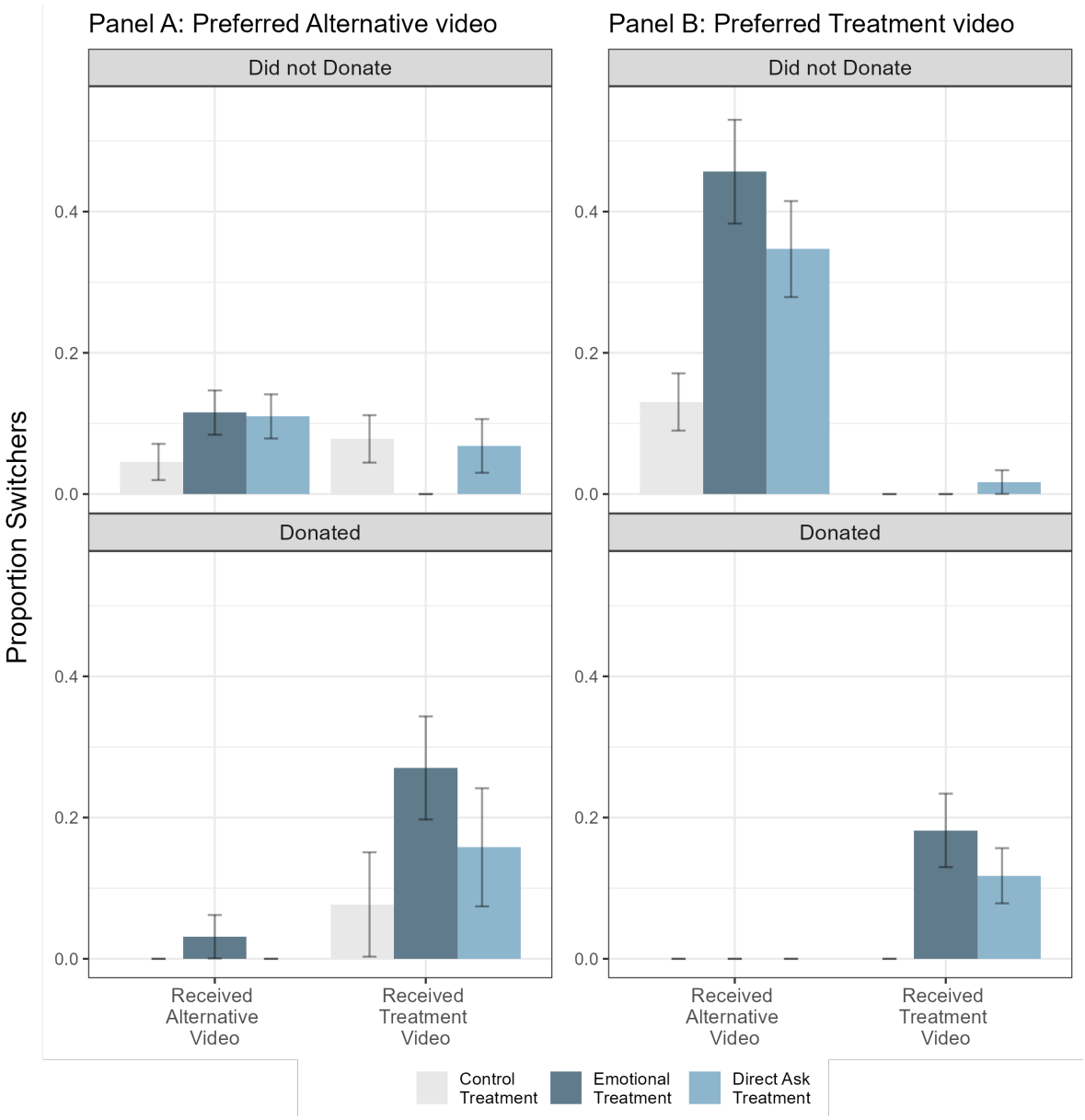


Figure 4: This figure displays the proportion of individuals who claim that they would have switched their donation decision had they received the opposite video. Panel A (B) displays this proportion for those who preferred the alternative (treatment) video. For both panels, we further distinguish on the video that the participant ended up seeing due to the random implementation and on whether they donated or not.

[8]Correlation coefficients are $r = 0.44$, $r = 0.40$, $r = 0.49$ and tests of the null where $r = 0$ yield always $p < 0.001$ for the *Control*, *Emotional* and *Direct Ask* treatments, respectively. For the full sample, we find $r = 0.471$, $p < 0.001$.

We find that participants in both the *Emotional* ($\chi^2 = 19.078$, df $= 1$, $p < 0.001$) and the *Direct Ask* ($\chi^2 = 10.348$, df $= 1$, $p = 0.001$) treatment predicted to switch more often than those in the *Control* treatment. This constitutes a sanity check on our results. Focusing on the *Emotional* and the *Direct Ask* treatments, we observe several patterns. First, most switching is predicted by individuals who i) did not donate after seeing the alternative video (top panels / left bars), and ii) individuals who donate after seeing the treatment video (bottom panels / right bars). All tests are two-sided proportion tests.

Second, switching is *not* predicted by individuals who i) already donate after seeing the alternative video: they believe that they would donate with the treatment video as well (bottom panels / left bars). Switching is also barely observed among individuals who ii) do not donate in treatment as they believe that they would also not donate after seeing the alternative video (top panels / right bars). In Appendix Table A3, we report statistical tests supporting the results of all these pairwise comparisons with the inclusion of a (Bonferroni) correction for multiple hypothesis testing.

These switching patterns are sensible and show that at least some of the individuals were aware that the charity-related videos increase the likelihood of donating. This complements the evidence in the previous subsection and suggests that video choice is used both as a commitment to give and a commitment not to give, depending on the subject's motivations. One might go one step further, and compare *predicted* switching levels to *actual* switching levels. We do this in the context of a structural model, where we can quantify the degree of sophistication and explore the different motivations behind video choices and their effects on giving behavior.

4 Modeling Giving as a Self-Control Problem

We have shown that emotional videos are avoided more than others, and that participants are at least partially aware that emotional videos can change their decisions. These facts suggest that the video choice functions as a commitment device. In this section, we interpret our results in the context of a structural model of self-control inspired by Loewenstein et al. (2015) and Gul and Pesendorfer (2001).[9] This exercise has two main objectives. First, to offer a structural framework that organizes the behavior in the experiment, and to make explicit the main assumptions behind an interpretation of self-control. Second, to produce quantitative estimates of the main behavioral profiles or "types" that we observe in the experiment, as well as the degree of naiveté and sophistication. Such estimates give a more precise picture of the direction of self-control and the degree of sophistication and can be compared to previous and future research.

[9] Bénabou and Pycia (2002) shows that the Gul and Pesendorfer (2001) model can be interpreted in terms of a multiple selves model where there is a costly intrapersonal conflict between a Planner with long-term preferences and a Doer with a myopic focus on current period utility, as in Thaler and Shefrin (1981).

Our model is most immediately applicable to the *Emotional* treatment, and we will connect it primarily to this treatment. However, it may apply to the *Direct Ask* treatment to the extent that this induces emotions of guilt associated with social pressure.[10] To keep the analysis transparent and tractable, our model leaves out motivations that may play a role in the experiment, such as a (non-strategic) desire to avoid negative emotions and curiosity about specific videos. We discuss these motivations in Section 4.5.

4.1 Model

In the model, each agent takes a decision $x \in \{A, S\}$, which reflects an altruistic (A) or selfish (S) act. Following Loewenstein et al. (2015), we assume the utility from action x consists of two components. $U(x)$ represents a "deliberative" component, i.e., a preference relation net of any strong or specific emotions, reflecting a reasoned consideration of arguments. Second, $E(x|\theta)$ is an "affective" component that reflects the emotional payoffs from giving in an emotional state θ. In our experiment, $\theta \in \{\theta_T, \theta_{Alt}\}$ represents the emotional state triggered by watching different videos, with θ_T being the emotional state triggered by the treatment video and θ_{Alt} the emotional state triggered by the alternative video.

We assume that the deliberative component $U(x)$ is not affected by the video. Since we do not have an explicit measure of U, we cannot test this assumption directly. However, by design, the videos do not provide information about specific activities of StC, nor do they contain any "rational" arguments for giving, thus making this assumption plausible. We also assume that the deliberative and affective components are additive. This yields the utility function

$$V(x|\theta) = U(x) + E(x|\theta). \tag{1}$$

We make several assumptions on the relation between θ and x (the choice of giving), as implemented by the videos and tasks in our experiment. First, we assume that the treatment video in the *Emotional* treatment triggers an emotional state that makes the altruistic option A emotionally more appealing than the selfish option S, i.e., $E(A|\theta_T) > E(S|\theta_T)$. Second, we assume the converse for the case of the alternative video, i.e. $E(A|\theta_{Alt}) < E(S|\theta_{Alt})$. We do not claim that the alternative video induces selfishness but rather that it preserves the emotions related to self-interest, materialism, or greed that arise from the experimental context in itself. Indeed, platforms like MTurk ("Make money in your spare time") and Prolific ("Get paid to change the world") try to recruit participants by advertising the opportunities for making money, so subjects are likely to be motivated by personal gain.

Third, we assume that donating after the treatment video yields higher emotional payoffs than donating after the alternative video ($E(A|\theta_T) > E(A|\theta_{Alt})$), and fourth, that not donating

[10] One might even see emotions in the model as a reduced form way to model (self-)image concerns that are associated with exposure to requests.

after the alternative video trumps not donating after the treatment video, $E(S|\theta_{Alt}) > E(S|\theta_T)$. These assumptions reflect the idea that once induced, emotions like compassion and guilt need to be "quenched" by an act of giving. Together, our assumptions boil down to positing a complementarity between emotions like compassion and the act of giving, and between the emotions of greed or materialism and non-giving.

Agents sequentially decide on θ and x. First, an agent chooses an emotional state θ in a "cold" stage where they do not experience strong emotions. They then choose x in a "hot" stage where agents experience the emotions associated with state θ. This mirrors the timing of decisions in our experiment, where the agent first chooses the video $\theta \in \{\theta_T, \theta_{Alt}\}$ and then chooses the option $x \in \{A, S\}$ right after having seen the video. We assume that the choice of x in the hot stage maximizes $V(x)$, the sum of the deliberative and emotional utility. Thus, in our setup, individuals evaluate the convex combination of the two types of utility for each alternative.

To describe the choice of θ, two concepts will be useful. First, in line with Loewenstein et al. (2015), we define the "Deliberative Optimum" x_U^* as the optimal choice according to the deliberative preferences, i.e. $x_U^* := \arg \max_{x \in \{A,S\}} U(x)$. Second, we define the "Aligned Set" Θ, as the set of states in which the deliberative optimum also maximizes total utility V, i.e.,

$$\Theta := \left\{ \theta | \arg \max_{x \in \{A,S\}} V(x|\theta) = x_U^* \right\}.$$

Note that our assumptions on $E(x|\theta)$ imply that Θ is non-empty.

We assume that in the cold stage, the agent chooses to maximize θ from the aligned set, i.e. $\theta^* = \arg \max_{\theta \in \Theta} V(x|\theta)$. Thus, agents apply a lexicographical criterion to the video choice: the first goal is to choose a video that leads to the implementation of the deliberative optimum in the hot state by making sure emotions are in the aligned set. Given this, they maximize $V(x|\theta)$ and reap the highest possible amount of emotional utility without deviating from the deliberative optimum.

This two-stage decision parallels the model by Gul and Pesendorfer (2001). Indeed, both models consider a two-stage decision where the choice in the second stage is based on the sum of two components reflecting a trade-off between two sets of preferences. Moreover, both models allow the decision-makers to manipulate their choices in the second stage, but in a different way. While in Gul and Pesendorfer's model, the decision maker manipulates choices by restricting the set of available alternatives, in our setup, the decision maker manipulates choices by selecting the emotional state they will be in when making their choices.[11]

[11] To further highlight the parallel, note that under our assumptions, the first stage choice of theta can also be expressed in terms of the Gul and Pesendorfer's first-stage utility, which is defined over sets. Let the utility of the set $\{A, S\}$ be $W(\{A, S\}|\theta) = \max_{x \in \{A,S\}} (U(x) + E(x|\theta)) - \max_{y \in \{A,S\}} E(y|\theta)$, then an agent picking $\theta \in \{\theta_T, \theta_{Alt}\}$ to maximize $W(\{A, S\}|\theta)$ would make the same choice of theta described in our lexicographic procedure. In this formulation, one can interpret the difference $E(y|\theta) - E(x|\theta))$ as the anticipated cost of the affective push towards y when preferences are not aligned. In a robustness check, we consider an alternative optimization procedure in

Finally, we consider the issue of naiveté. The choice of video in the cold state crucially depends on the ability of the agent to anticipate the emotions experienced in each state, i.e. after watching the videos in the experiment. We allow for the possibility of both sophisticated types, with perfect emotional foresight, as well as naive types with inaccurate beliefs about $E(x|\theta)$. Specifically, we assume that naive types believe videos have no effect on their choices, i.e. they believe that $E(S|\theta_{Alt}) = E(A|\theta_{Alt}) = E(S|\theta_T) = E(A|\theta_T)$.

4.2 Analysis

Our model produces four possible choice profiles, or "types", depending on the functions $U(\cdot)$ and $E(\cdot)$. First, with respect to $U(\cdot)$, we can distinguish "Altruistic" agents whose deliberative optimum is the altruistic action ($x^*_U = A$), from "Selfish" agents with $x^*_U = S$. Second, we can distinguish individuals according to the importance of their emotional utility E relative to deliberative utility U. If emotional utility E is relatively weak, the choice that optimizes V is invariant to the emotional state and will coincide with the deliberative optimum. We call these individuals "State-Independent" (SI). Such individuals will simply choose the emotional state that maximizes the emotional utility given their deliberative optimum. In the context of the experiment, SI Altruistic types will choose the emotional video, while SI Selfish types will choose the neutral video.

By contrast, if emotional utility is important relative to the deliberative component, the choice x that optimizes V will vary across emotional states θ, reducing the number of emotional states in the alignment set. Such "State-Dependent" agents will face a self-control problem. Their video choice thus serves to guide their choices in the hot stage toward the deliberative optimum.

Note that these conceptualizations align with the concept of commitment demand in Gul and Pesendorfer (2001), where individuals may seek commitment not only to avoid giving in to temptation (or, in terms of O'Donoghue and Rabin (1999), to address dynamic inconsistency) but also to minimize anticipated self-control costs, even when such costs do not alter their choice. In our case, subjects strictly prefer some videos even if it does not change their action, as they anticipate that they will result in additional emotional payoffs from their choice.

Finally, for each of the four types sketched here, there is a further dimension of sophistication versus naiveté. Following our discussion so far, sophisticated agents make their choices for θ based on its correct emotional impact. By contrast, the naive type has the wrong belief about the impact of θ. Thus, naiveté is identifiable by contrasting the predictions that agents make about behavior in different states with the actual behavior in those states.

In our experiment, beliefs are measured after the choice, i.e., after agents have experienced

which agents in the cold state care only about implementing the deliberative optimum and do not consider the emotional utility of giving.

one of the videos. We assume naiveté takes the form of projection bias, whereby agents underestimate the effect of the video or emotional state. Thus, when predicting counterfactual behavior, naive agents may mistakenly think that the video does not affect their choice. Moreover, as they do not anticipate changes in emotional payoffs, they are indifferent between choosing the different videos.

Table 3: Type classifications

Type (1)	Del. Optimum (2)	Video Preference (3)	Assigned Video (4)	Giving Choice (5)	Counterf. Prediction (6)
Panel A. Sophisticated Types					
Altruist SI	A	θ_T	θ_T	A	A
			θ_{Alt}	A	A
Altruist SD	A	θ_T	θ_T	A	S
			θ_{Alt}	S	A
Selfish SI	S	θ_{Alt}	θ_T	S	S
			θ_{Alt}	S	S
Selfish SD	S	θ_{Alt}	θ_T	A	S
			θ_{Alt}	S	A
Panel B. Naive Types					
Altruist SI	A	θ_T or θ_{Alt}	θ_T	A	A
			θ_{Alt}	A	A
Altruist SD	A	θ_T or θ_{Alt}	θ_T	A	A
			θ_{Alt}	S	S
Selfish SI	S	θ_T or θ_{Alt}	θ_T	S	S
			θ_{Alt}	S	S
Selfish SD	S	θ_T or θ_{Alt}	θ_T	A	A
			θ_{Alt}	S	S

Notes: Column (1) displays the type classification based on deliberative preference and state dependency (SD) or not (SI), column (2) presents the deliberatively favored alternative, column (3) presents the video choice, column (4) presents the two possible videos that the individual can be (randomly) assigned to and columns (5) an (6) show the choices and the predictions that individuals would make after watching either video. Panels A and B show this information for Sophisticated and Naive Types, respectively.

We summarize our type classification in Table 3, which shows the choices of each type as well as their predictions about counterfactual behavior. For example, consider a self-interested individual with state-dependent preferences (Selfish SD). This individual deliberatively wants to not donate. If sophisticated, he knows that the treatment video will make him donate and thus prefers to watch the alternative video. In addition, he is also able to accurately predict what choice he would make after watching each video, even if we ask him after watching a video and irrespective of what that video was. In contrast, the naive, self-interested, inconsistent type thinks that the videos have no effect on emotions and thus randomizes between them. Moreover,

when asked about behavior in counterfactual videos, his response depends on the video that he received: after watching the alternative video and not donating, he predicts that he also would not have donated had he watched the treatment video, and vice versa. The counterfactual predictions of the naive thus exhibit projection bias.

4.3 Estimation procedure

Our data allow us to estimate the fractions of these types in the population. To this end, we employ an approach similar to Costa-Gomes et al. (2001). Roughly speaking, identification occurs as follows. The comparison of giving choices after the different videos identifies the degree of state dependence, while the choice of video further identifies the deliberative optimum of sophisticated types. Naiveté is identified from the difference between predicted and actual behavior. We also make two further simplifying assumptions. First, we assume that agents generally act according to their type but occasionally make mistakes, that is, in each decision, there is a probability ϵ that they make the wrong choice. Second, we assume that the degree of sophistication is independent of the type; in other words, no type is inherently more likely to exhibit naive behavior than another.

With these assumptions, we calculate the probability of each profile $p(V, R, C, P|t, n, \epsilon)$, i.e., the probability of observing someone who prefers video V, chooses option C (donates or not) after being assigned video R, and predicts that she would have chosen option P had she received the counterfactual video to R, conditional on being type t with sophistication level n ($n = 1$ if naive, 0 if sophisticated), and an error rate ϵ. See Appendix E for the formal expression of this conditional probability.

We define w as the (constant across types) probability of being naive, and λ_t as the probability of being of type $t \in T$, where T is the type space containing our 4 types, i.e., $T = \{\text{Altruist SI}, \text{Altruist SD}, \text{Selfish SI}, \text{Selfish SD}\}$. This yields the unconditional probability of observing the sequence of choices (V, R, C, P) in our data:

$$p(V, R, C, P) = \sum_{t \in T} \sum_{n \in \{0,1\}} \lambda_t (1 - w)^{\mathbb{1}(n=0)} w^{\mathbb{1}(n=1)} p(V, R, C, P|t, n, \epsilon) \tag{2}$$

With equation 2 we can specify a log-likelihood function[12] and maximize it to obtain an estimate of the proportion of each type (the parameters λ_t), the probability of naiveté (w), and the error rate in our data (ϵ).

[12]The log-likelihood function is $logLik(\epsilon, w, \boldsymbol{\lambda}) = \sum_{k=1}^{K} \log(\sum_{t \in T} \sum_{n=0}^{1} \lambda_t (1 - w)^{(1-n)} w^n \, p(V_k, R_k, C_k, P_k))$ where $\boldsymbol{\lambda}$ is the vector of the λ_t with t in T.

4.4 Results

The results for the *Emotional*, *Direct Ask*, and the *Control* treatments are presented in Figure 5. Appendix Table C1 contains the tables with the estimated parameters and the associated log-likelihood. Some patterns stand out. First, note that the *Control* treatment provides a sanity check on our results. It does not have any state-dependent types, which makes sense because its treatment video is assumed not to generate any emotions or social pressure. Moreover, naiveté has no meaning in this setting. Since there is no treatment effect of the video and hence nothing to anticipate, everyone should be "naive", which is indeed what we empirically observe.

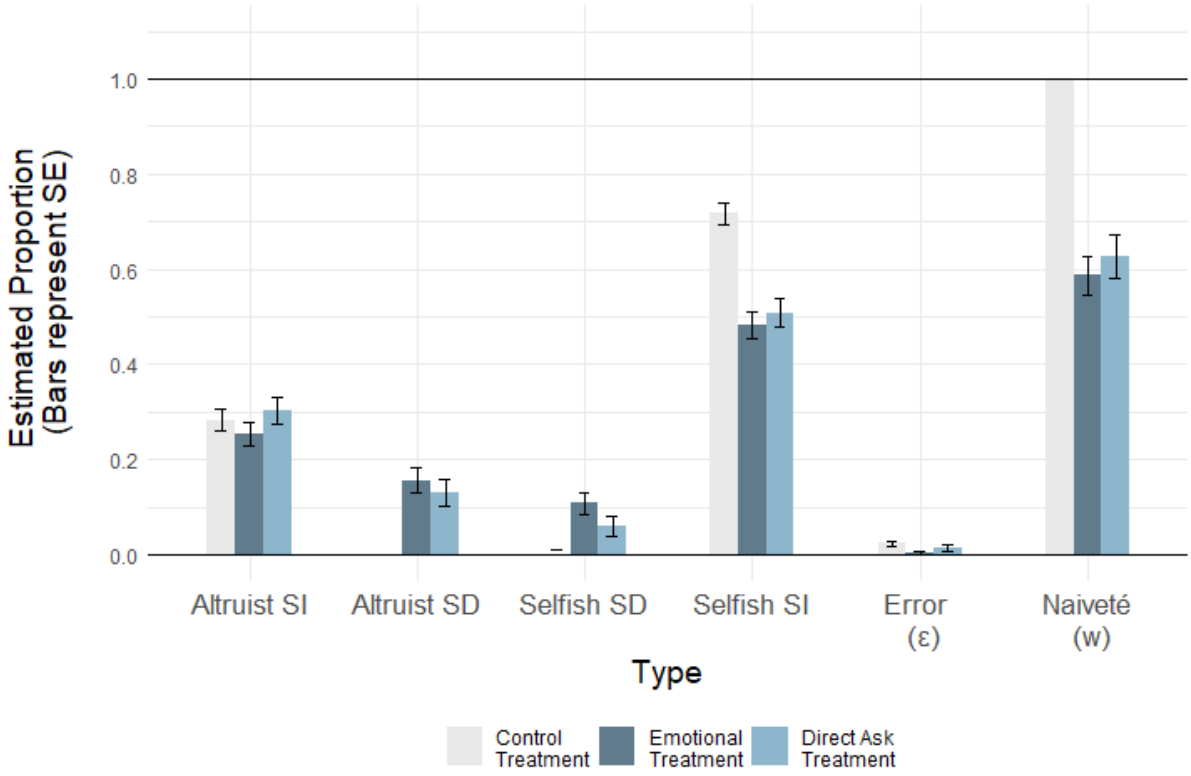


Figure 5: Estimated proportion of types, Naiveté and Error Rates

Comparing the *Emotional* and the *Direct Ask* treatments, we observe a similar distribution of types. Indeed, a likelihood ratio test does not reject a model that pools both treatments ($p = 0.38$). However, in the *Emotional* treatment, we do see somewhat higher fractions of SD types (both altruist and selfish) compared to the *Direct Ask* treatment. This difference is marginally significant at the 10% level ($Z = 1.56$, $p = 0.059$, one-tailed). Emotional appeals thus generate slightly higher commitment demand than a direct ask. In the discussion section, we analyze open-ended questionnaire responses and show that avoidance of emotions is indeed a strong motive in the *Emotional* treatment. However, since the difference is small, it is likely that the *Direct Ask* treatment generates pressure in different ways, for instance, by increasing normative expectations.

Focusing on the *Emotional* treatment, Figure 5 shows that selfishness is more prevalent than altruism: around 59% are (either state-dependent or independent) selfish. We find that a minority of about 26% of the participants is state-dependent. Approximately 11% are state-dependent selfish types who prefer not to donate and, to avoid doing so, choose the alternative video. However, if they receive the treatment video, they end up donating. This behavioral profile is sometimes called "reluctant altruists" in the literature (Lazear et al., 2012). About 16% are state-dependent altruists, whose self-control problem goes in the opposite direction: they prefer to donate and demand the treatment video to do so since if they receive the alternative video, they do not donate. It is noteworthy that among state-dependent types, a majority (60%) acts in line with commitment to altruism rather than selfishness, and might thus be called "willing altruists".

Figure 5 also shows that 60% of our sample is classified as naive and 40% as sophisticated. We consider this as a lower bound on the share of sophisticated behavior, because our assumptions impose a high bar for sophistication: sophisticated individuals not only perfectly anticipate the emotional payoffs in each video condition, they also optimize these payoffs. In the next section, we consider a relaxation of the sophistication assumption and show that it leads to higher estimates of the fraction of sophisticated types.

4.5 Discussion

Here, we discuss the strengths and weaknesses of our model and interpretation of giving as a self-control problem, as well as various extensions and robustness checks to the model.

Falsification. While our model is primarily a way to organize and interpret the results, it does rule out certain behaviors. More specifically, are 16 possible individual outcomes in the experiment[13], of which four are not rationalizable by the model. Table C.3 in Appendix C.2 provides a detailed breakdown of the number of participants for each outcome. It shows that only a very small number of individuals (10/1203) display one of the four patterns that would falsify the model. Thus, the behavioral profiles captured by the model are the ones that matter.

Sensitivity of sophistication. The estimated degree of naiveté and sophistication is sensitive to modeling assumptions about naive and sophisticated types. Our model implies that individuals are sophisticated only if the chosen video matches their donation choice in the cold state (i.e., those who donate want to watch the treatment video, while those who do not donate want to watch the alternative video), as this maximizes the emotional payoffs. In Appendix C.2, we relax this assumption by allowing sophisticated types to only care about implementing the deliberative optimum and not about maximizing emotional payoffs. Thus, if individuals do

[13] {choose alternative video, choose treatment video} × {see alternative video, see treatment video} × {donate, do not donate} × {predict counterfactual donation, predict counterfactual no donation}

not face self-control problems (i.e., they are state-independent), they simply randomize between the videos. This relaxation increases the estimated fraction of sophistication by approximately 48% (from 41.3% to 61.4%).

In addition, we analyze the effect of imposing full or zero sophistication on our estimates. The results are also available in Appendix C.2. Both models of either full or zero sophistication lead to a substantially lower likelihood than the model with a free parameter for naivité. Taken together, we think this is further evidence that a substantial portion of individuals are sophisticated about the impact of their emotional states on altruism. In sum, while the degree of sophistication is sizable, the estimated amount depends on untested assumptions that could be subject to further research.

Alternative motives. Our model focuses on a limited set of motives behind the video choice, and ignores drivers like curiosity, aversion to negatively valenced videos, or seeking further information about the charity. We designed the experiment to reduce some of these motives. For instance, we discouraged information seeking by using a well-known charity and giving video descriptions upfront. However, we cannot fully rule out the possibility that participants expect additional information from watching the video.

To better understand the relative strength of various motivations, we analyze open-ended responses to the question "Why did you choose the video?". We used OpenAI's GPT-4o model to classify responses into four categories, including avoidance of emotions, seeking of emotions, curiosity, random choice, and other motives. Moreover, we provide an overview of the 20 most used words in the Emotional and Direct Ask treatments. Appendix F provides a detailed overview of the methodology (e.g., AI prompts).

While we relegate detailed results to Appendix F, three patterns stand out. First, avoidance of emotions is a common reason for choosing the alternative video, especially in the *Emotional* treatment, where 65% of the answers by the 233 participants who prefer the alternative video fall into this category. The corresponding percentage in the *Direct Ask* treatment is 37%. Second, seeking emotions is frequently cited as a reason for choosing the treatment video, mostly so in the *Emotional* treatment. More generally, emotion-related words (feel, guilty, sad, emotional) appeared at least twice as often in the *Emotional* Treatment as in the *Direct Ask* Treatment. These pieces of evidence are in line with the idea that emotional self-regulation was a significant factor in decision-making. It also validates our experimental design as emotional motives are more present in the *Emotional* treatment and least in the *Control*. Note, however, that most responses were too generic to distinguish the exact reason why emotions were avoided or sought.

Third, curiosity and interest in the video content is a common motive across all treatments. This is especially true for the *Control* treatment, where more than 80% of responses fall into this category, but also the *Emotional* and *Direct Ask* treatments, where the majority of subjects who selected the treatment video gave an answer in this category. Thus, our attempts to suppress

information-seeking were, at most, partially successful.

While curiosity indicates a cognitive rather than emotion-driven mode of decision-making, information-seeking may nevertheless reflect a form of commitment, to the extent that one wants to be convinced to give. Indeed, it is the exact opposite of "information avoidance" that has often been documented in the literature on moral wiggle room (Dana et al., 2007; Vu et al., 2023). In general, the fact that a non-negligible fraction *seeks* to be convinced or emoted by the charity has been underexplored in the literature on altruism (though see Saccardo and Serra-Garcia, 2023). The rather brief nature of most responses does not allow for a more fine-grained inference about motives. Moreover, the distinction between emotions and information is conceptually fraught, but future research can seek to delineate the line between being emoted and being informed more clearly.

Randomization and reactance. A potential challenge regards the randomization by which individuals may have their choices overridden. While this randomization provides us with rich data to test for sophistication, it also entails some potential new issues. First, it could generate reactance, as individuals dislike having their choice overridden. Indeed, the *Control* treatment displays some evidence that is consistent with reactance, as people whose video choice was overridden donate less (even though both videos are unrelated to the charity) – see Figure 3. We can estimate the effect of the video on donations in the *Emotional* and the *Direct Ask* net of reactance, by subtracting the reactance effect in the *Control* treatment using a difference in difference regression (see Table A5 and Figure A1 in Appendix A). This exercise shows that the main results are robust to reactance.

Second, our randomization creates a stochastic decision environment where temptations are uncertain, while our model and analysis assume a deterministic one. This means that models of random indulgence, like Dekel and Lipman (2012), could potentially fit our data better. In order to assess the importance that individuals gave to the randomization, we asked them: "Did the randomization matter for deciding between the videos?" and they could give an answer of "Yes", "No" or "Unsure". Overall, 80.64 % of the total sample answered "No".[14] Thus, it seems that the randomization did not affect our results substantially.

5 Conclusion

Do individuals display sophisticated strategies to regulate their emotions when confronted with a giving task? We tested this question in an online experiment, where subjects could watch or avoid videos containing emotional triggers, before donating to a charity. We find that emotional videos have the largest effect on behavior compared to a direct ask or a neutral video, but are also

[14]The shares of "No" answers for the *Emotional*, *Direct Ask* and *Control* treatments are 84.37 %, 76.69 % , 80.05%, respectively.

avoided the most. We show that a substantial fraction of participants behave differently across the emotional states induced by our videos. Moreover, they partially predict the treatment effect on their own behavior and on that of others.

We interpret these observations in the context of a structural model, where agents use videos as commitment in the face of state-dependent social preferences. We estimate levels of naiveté and the direction of commitment. Approximately 26% of the participants make decisions that depend on their emotional state. Perhaps surprisingly, a majority of these decision-makers choose in line with commitment to altruism rather than to selfishness. That is, they choose to be exposed to emotional triggers, and predict they would not give otherwise.

This evidence suggests a type of person who wants to be altruistic but seeks to be more convinced, and therefore commits *towards* giving. Our study suggests that this "willing altruist" is at least as common as the "reluctant altruist", her well-studied counterpart in the literature on moral wiggle room (Dana et al., 2006; Lazear et al., 2012; Cain et al., 2014; Exley, 2020). By comparison, the existence of a willing altruist paints a more flattering picture of human nature. It also highlights the importance of commitment devices that promote empathy and compassion, such as giving pledges, automated donations, and (religious) rituals that serve to enhance charity and empathy.

We also provide evidence about the role of sophistication in emotional regulation. Previous literature has suggested that ask avoidance is a strategy to avoid empathy or social pressure, implicitly assuming sophistication (e.g. Grossman and Van der Weele, 2017). We provide precise evidence on this assumption and its drivers, using data on predicted and actual counterfactual behavior. We find that 40 to 60% of participants can be classified as sophisticated, depending on the assumptions of the structural model. Thus, both naiveté and sophistication are widespread. Since our structural analysis has left out some motives like curiosity, further research should confirm and refine these estimates.

Our focus on commitment provides new impetus to dual-self approaches to altruism (Loewenstein and Small, 2007; Loewenstein et al., 2015; Fromell et al., 2020), and opens up new avenues for research on social preferences. For instance, we can try to understand how emotional commitment relates to time-inconsistency. Our model suggests that time inconsistency can occur if emotions vary between the two decision-making episodes. Hence, self-control problems do not arise from the passing of time perse, but from the transient nature of emotions. This may explain why some studies do and others do not find time-inconsistency in giving (Kölle and Wenner, 2023; Andreoni and Serra-Garcia, 2021). Another area is welfare analysis. Our study underscores the importance of internal welfare accounting, but does not make quantitative welfare assessments (see also DellaVigna et al., 2012). Future research could explore whether individuals are ultimately better off when they can commit to either altruism or selfishness, and design commitment opportunities accordingly.

Our findings also open new avenues for studying market interactions and their regulation.

For example, Jain (2024) examines street vendors in India and finds that child vendors earn twice as much as adult vendors, as customers struggle to refuse them. A better understanding of the impact of emotional sales tactics might produce more informed regulation to protect both children and consumers. Similarly, our research might inform codes for responsible fundraising by charitable organizations, helping consumers pursue deliberative giving goals as opposed to ad-hoc responses to emotional appeals.

References

Alpizar, F., Carlsson, F., and Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in costa rica. *Journal of Public Economics*, 92(5):1047–1060.

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

Andreoni, J. and Serra-Garcia, M. (2021). Time inconsistent charitable giving. *Journal of Public Economics*, 198:104391.

Andries, M., Bursztyn, L., Chaney, T., and Djourelova, M. (2024). In their shoes. Technical report, National Bureau of Economic Research.

Bénabou, R. and Pycia, M. (2002). Dynamic inconsistency and self-control: a planner–doer interpretation. *Economics Letters*, 77(3):419–424.

Cain, D. M., Dana, J., and Newman, G. E. (2014). Giving versus giving in. *The Academy of Management Annals*, 8(1):505–533.

Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., and Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, 148(6):962–976.

Cameron, C. D., Scheffer, J. A., Hadjiandreou, E., and Anderson, S. (2022). Motivated empathic choices. *Advances in Experimental Social Psychology*, 66:191–279.

Cerrone, C. and Lades, L. (2017). Sophisticated and naïve procrastination: an experimental study. *MPI Collective Goods Preprint*, 08(2017).

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(C):88–97.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100(5):1261–1293.

Cobb-Clark, D. A., Dahmann, S. C., Kamhöfer, D., and Schildberg-Hörisch, H. (2024). Sophistication about self-control. *Journal of Public Economics*, 238:105196.

Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201.

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.

Davis, M. H., Mitchell, K. V., Hall, J. A., Lothert, J., Snapp, T., and Meyer, M. (1999). Empathy, expectations, and situational preferences: Personality influences on the decision to participate in volunteer helping behaviors. *Journal of Personality*, 67(3):469–503.

Dekel, E. and Lipman, B. L. (2012). Costly self-control and random self-indulgence. *Econometrica*, 80(3):1271–1302.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127(1):1–56.

Dreber, A., Fudenberg, D., Levine, D. K., and Rand, D. G. (2016). Self-control, social preferences and the effect of delayed payments. *Available at Social Science Research Network (SSRN): http://ssrn. com/abstract*, 2477454.

Drouvelis, M. and Grosskopf, B. (2016). The effects of induced emotions on pro-social behaviour. *Journal of Public Economics*, 134:1–8.

Exley, C. L. (2020). Using charity performance metrics as an excuse not to give. *Management Science*, 66(2):553–563.

Exley, C. L. and Petrie, R. (2018). The impact of a surprise donation ask. *Journal of Public Economics*, 158:152–167.

Fromell, H., Nosenzo, D., and Owens, T. (2020). Altruism, fast and slow? evidence from a meta-analysis and a new experiment. *Experimental Economics*, 23(4):979–1001.

Gneezy, U. and Imas, A. (2014). Materazzi effect and the strategic use of anger in competitive interactions. *Proceedings of the National Academy of Sciences*, 111(4):1334–1337.

Gneezy, U., Imas, A., and Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, 60(11):2645–2658.

Grimm, V. and Mengel, F. (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters*, 111(2):113–115.

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.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Heider, F. (1982). *The psychology of interpersonal relations*. Psychology Press.

Hodges, S. D. and Biswas-Diener, R. (2007). Balancing the empathy expense account: Strategies for regulating empathic response. In Farrow, T. and Woodruff, P., editors, *Empathy in Mental Illness*, pages 389–407.

Hodges, S. D. and Klein, K. J. (2001). Regulating the costs of empathy: the price of being human. *The Journal*

of Socio-Economics, 30(5):437–452.

Jain, R. (2024). Entrepreneurs of emotions: evidence from street vending in india. *Zurich University Working Paper 451*.

Jia, T. (2022). Empathy, motivated reasoning, and redistribution. *Mimeo*.

Kölle, F. and Wenner, L. (2023). Is generosity time-inconsistent? present bias across individual and social contexts. *Review of Economics and Statistics*, 105(3):683–699.

Lazear, E. P., Malmendier, U., and Weber, R. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Econometrics*, 4(1):136–163.

Loewenstein, G., O'Donoghue, T., and Bhatia, S. (2015). Modeling the interplay between affect and deliberation. *Decision*, 2(2):55–81.

Loewenstein, G. and Small, D. A. (2007). The scarecrow and the tin man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, 11(2):112–126.

Long, S. H. (1976). Social pressure and contributions to health charities. *Public Choice*, 28(1):55–66.

Mandel, N., Scott, M. L., Kim, S., and Sinha, R. K. (2017). Strategies for improving self-control among naïve, sophisticated, and time-consistent consumers. *Journal of Economic Psychology*, 60(C):109–125.

Meer, J. (2011). Brother, can you spare a dime? peer pressure in charitable solicitation. *Journal of Public Economics*, 95(7):926–941.

Nguyen, Y. and Noussair, C. N. (2022). Incidental emotions and cooperation in a public goods game. *Frontiers in Psychology*, 13:800701.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Pancer, S. M. (1988). Salience of appeal and avoidance of helping situations. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 20(2):133–139.

Saccardo, S. and Serra-Garcia, M. (2023). Enabling or limiting cognitive flexibility? evidence of demand for moral commitment. *American Economic Review*, 113(2):396–429.

Shaw, L. L., Batson, C. D., and Todd, R. M. (1994). Empathy avoidance: Forestalling feeling for another in order to escape the motivational consequences. *Journal of Personality and Social Psychology*, 67(5):879–887.

Small, D. A. and Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, 26(1):5–16.

Smith, K. E., Norman, G. J., and Decety, J. (2020). Medical students' empathy positively predicts charitable donation behavior. *The Journal of Positive Psychology*, 15(6):734–742.

Thaler, R. H. and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, 89(2):392–406.

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, 86(3):859–889.

Verhaert, G. A. and Van den Poel, D. (2011). Empathy as added value in predicting donation behavior. *Journal of Business Research*, 64(12):1288–1295.

Vu, L., Soraperra, I., Leib, M., van der Weele, J., 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):611–635.

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., and Weil, L. (2006). Development of short forms of the empathy quotient (eq-short) and the systemizing quotient (sq-short). *Personality and Individual Differences*, 41(5):929–940.

Wong, W.-K. (2008). How much time-inconsistency is there and does it matter? evidence on self-awareness, size, and effects. *Journal of Economic Behavior & Organization*, 68(3-4):645–656.

Zaki, J. (2014). Empathy: a motivated account. *Psychological Bulletin*, 140(6):1608–1647.

Appendix

A Tables

Table A1: Balance Table

	Control	Emotional	Direct Ask	Direct Ask No Message	p-value
Age	40.73	40.69	40.29	40.52	0.960
Sex (%)					0.645
Other	0.25	0.25	0.25	1.01	
Female	52.36	52.38	53.37	56.78	
Male	47.39	47.37	46.38	42.21	
Employment					0.531
Unemployed	13.65	13.28	12.72	11.56	
Employed	64.02	60.15	60.10	58.29	
Other	22.33	26.57	27.18	30.15	
Student					0.252
Other	14.64	19.80	15.21	20.60	
Yes	8.19	8.52	10.22	9.55	
No	77.17	71.68	74.56	69.85	
Time taken (seconds)	878.23	860.10	921.46	1042.33	0.020
Total approvals	771.17	832.22	766.39	811.11	0.426
Country of birth (% UK)	84.62	85.21	83.29	81.41	0.639
Country of residence (% UK)	100.00	99.75	100.00	99.50	0.350
Ethnicity					0.173
Asian	6.20	4.51	5.74	5.53	
Black	2.48	5.01	4.24	5.03	
No data	1.49	0.50	0.75	2.01	
Mixed	2.48	3.26	3.24	3.52	
Other	0.99	0.00	0.75	2.51	
White	86.35	86.72	85.29	81.41	
Language (% other than English)	9.43	7.77	9.23	8.54	0.841
Nationality (% UK)	89.83	88.72	86.28	88.94	0.450

Notes: This table presents descriptive statistics for demographic variables across treatment groups to assess balance. p-values report comparisons between treatment groups using two-sided tests: an ANOVA test for continuous variables and the Chi-square test for categorical variables. Demographics were provided by Prolific.

Table A2: Regression of donation decision on interaction of Preferred and Received Treatment

	Dependent variable: Donation decision					
	Control	Control	Emotional	Emotional	Direct Ask	Direct Ask
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.23^{***}	0.28^{***}	0.22^{***}	0.24^{***}	0.18^{***}	0.17^{***}
	(0.04)	(0.05)	(0.03)	(0.04)	(0.03)	(0.03)
Preferred Treatment	0.07^{*}	-0.03	0.16^{***}	0.12^{*}	0.27^{***}	0.29^{***}
	(0.04)	(0.07)	(0.05)	(0.07)	(0.05)	(0.06)
Received Treatment	0.001	-0.11^{*}	0.18^{***}	0.15^{**}	0.10^{**}	0.13^{**}
	(0.04)	(0.06)	(0.05)	(0.06)	(0.05)	(0.07)
Pref × Rec Treatment		0.20^{**}		0.08		-0.06
		(0.09)		(0.10)		(0.10)
Observations	403	403	399	399	401	401
R^2	0.01	0.02	0.07	0.07	0.10	0.10
Adjusted R^2	0.002	0.01	0.06	0.06	0.10	0.09
Residual Std. Error	0.45	0.44	0.47	0.47	0.46	0.46
	(df = 400)	(df = 399)	(df = 396)	(df = 395)	(df = 398)	(df = 397)
F Statistic	1.30	2.53^{*}	14.80^{***}	10.09^{***}	22.23^{***}	14.93^{***}
	(df = 2; 400)	(df = 3; 399)	(df = 2; 396)	(df = 3; 395)	(df = 2; 398)	(df = 2; 397)

Notes: This table presents the results of regressing the donation decision on whether the individual preferred the treatment video (Preferred Treatment), on whether (s)he received it (Received Treatment) and on the interaction between both for our two main treatments and the Control Treatment. Standard errors in parentheses (*p<0.1; **p<0.05; ***p<0.01).

Table A3: Corrected group comparisons of proportions of switchers

Emotional Treatment

Preferred Alternative

Not corrected				Bonferroni Corrected			
	A-A	A-S	T-A		A-A	A-S	T-A
A-S	0.284	NA	NA	A-S	1.000	NA	NA
T-A	0.018	0.049	NA	T-A	0.105	0.296	NA
T-S	0.748	0.015	0.000	T-S	1.000	0.093	0.001

Preferred Treatment

Not corrected				Bonferroni Corrected			
	A-A	A-S	T-A		A-A	A-S	T-A
A-S	0.000	NA	NA	A-S	0.00	NA	NA
T-A	0.056	0.006	NA	T-A	0.28	0.03	NA
T-S	*	0.000	0.012	T-S	*	15.53	0.00

Direct Ask Treatment

Preferred Alternative

Not corrected				Bonferroni Corrected			
	A-A	A-S	T-A		A-A	A-S	T-A
A-S	0.26	NA	NA	A-S	1.00	NA	NA
T-A	0.21	0.84	NA	T-A	0.11	0.30	NA
T-S	0.58	0.63	0.52	T-S	1.00	0.09	0.00

Preferred Treatment

Not corrected				Bonferroni Corrected			
	A-A	A-S	T-A		A-A	A-S	T-A
A-S	0.00	NA	NA	A-S	0.00	NA	NA
T-A	0.05	0.01	NA	T-A	0.32	0.035	NA
T-S	1.00	0.00	0.06	T-S	1.00	0.00	0.38

Notes: We conduct multiple proportion comparisons for the *Emotional Treatment* and *Direct Ask Treatment* proportions displayed in Panels A and B of Figure 4. We present both corrected and uncorrected group comparisons. We label the received video by A(T) if the alternative (treatment) video was received and the donation decision by A (S) if a donation was (not) given. *=both groups were zero.

Table A4: IPTW Linear probability model of donation decision

	Dependent Variable: Donation		
	Control	Emotional	Direct Ask
	(1)	(2)	(3)
Received Treatment	0.26***	0.46***	0.44***
	(0.03)	(0.04)	(0.04)
Did not Receive Treatment	0.26***	0.29***	0.32***
	(0.03)	(0.03)	(0.03)
Observations	403	399	401
R^2	0.26	0.39	0.38
Adjusted R^2	0.26	0.39	0.38
Residual Std. Error	0.63 (df = 401)	0.68 (df = 397)	0.68 (df = 399)
F Statistic	71.39*** (df = 2; 401)	128.37*** (df = 2; 397)	124.28*** (df = 2; 399)

Notes: This table presents the results of a IPTW weighted regression (observations are weighted according to the inverse of the probability to obtain the treatment video) with robust standard errors. The dependent variable is whether the subject donated or not (1=Yes, 0=No). Received Treatment is whether the subject received the treatment video (1=Yes, 0=No) and Did not Receive Treatment is its complementary. Standard errors in parentheses ($^{*}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

Table A5: Linear probability model of donation decisions net of reactance

	Dependent Variable: Donation	
	Emotional	Direct Ask
	(1)	(2)
Preferred Alternative	0.28^{***}	0.28^{***}
	(0.05)	(0.05)
Preferred Treatment	0.25^{***}	0.25^{***}
	(0.05)	(0.05)
Preferred Alternative × Received Treatment	-0.11^{*}	-0.11^{*}
	(0.06)	(0.06)
Preferred Treatment× Received Treatment	0.09	0.09
	(0.06)	(0.06)
Treatment × Preferred Alternative	−0.05	-0.12^{**}
	(0.06)	(0.06)
Treatment × Preferred Treatment	0.10	0.21^{***}
	(0.07)	(0.07)
Treatment × Preferred Alternative × Received Treatment	0.26^{***}	0.25^{***}
	(0.09)	(0.09)
Treatment × Preferred Treatment × Received Treatment	0.14	−0.01
	(0.10)	(0.09)
Observations	802	804
R^2	0.36	0.37
Adjusted R^2	0.36	0.37
Residual Std. Error	0.46 (df = 794)	0.45 (df = 796)
F Statistic	56.36^{***} (df = 8; 794)	58.97^{***} (df = 8; 796)

Notes: This table presents the results of a linear probability model regression with robust standard errors. The dependent variable is whether the subject donated or not (1=Yes, 0=No). Preferred Treatment(Alternative) is whether the subject preferred the treatment (alternative) video (1=Yes, 0=No), Received Treatment is whether the subject received the associated treatment video (1=Yes, 0=No) and Treatment is a dummy that takes value 1 for each respective column treatment relative to the control (baseline). Standard errors in parentheses ($^{*}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}\ p < 0.001$).

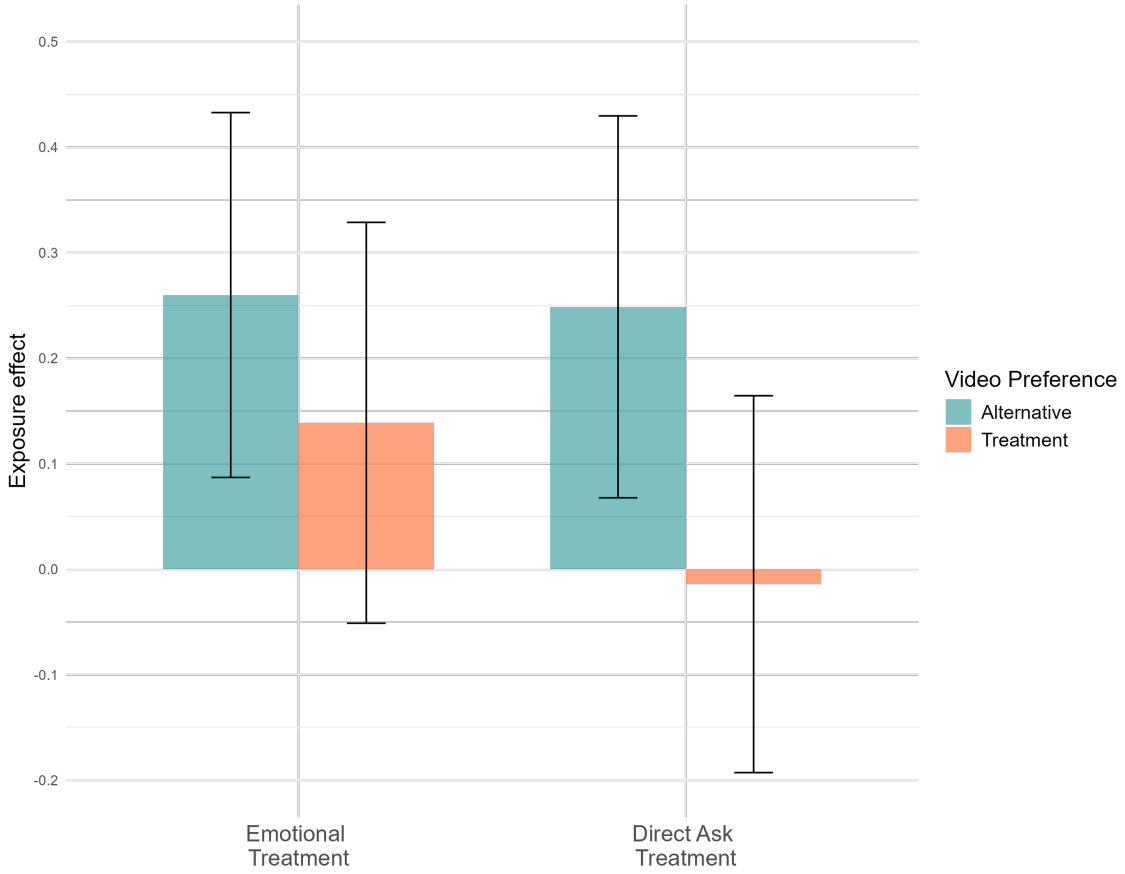


Figure A1: Plot of coefficients for (Treatment × Preferred Alternative × Received Treatment) and (Treatment × Preferred Treatment × Received Treatment) for each the *Emotional* vs *Control* and the *Direct Ask* vs *Control* models in Table A5 above. These coefficients capture the effect of exposure to the treatment video taking into account selection (distinguishes both individuals who preferred the treatment video and those who preferred the alternative) and reactance (the coefficients are net of the reactance identified in the *Control* Treatment.)

B Manipulation check

B.1 Pretest data

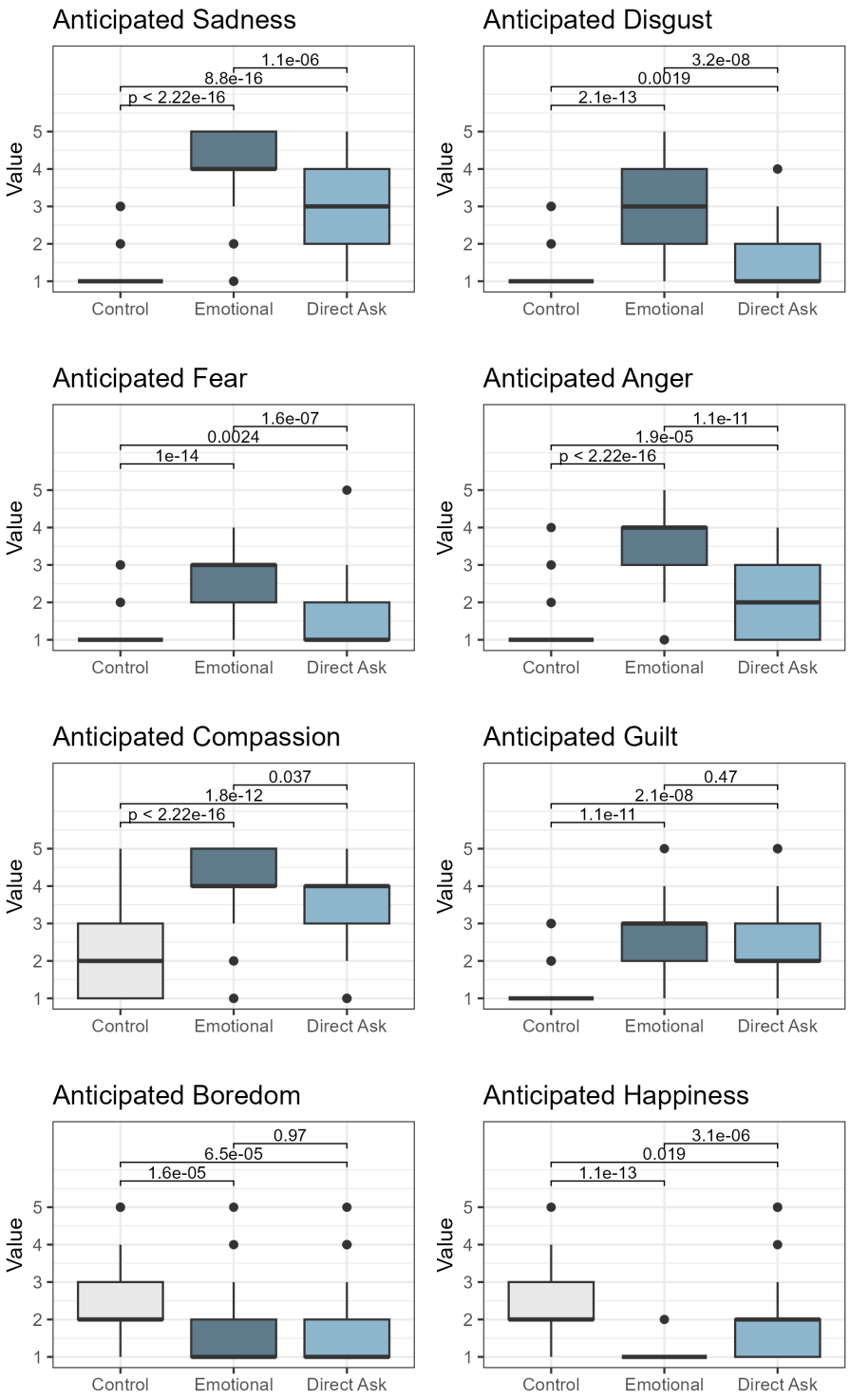


Figure C1: This figure presents the answers to evaluations on the anticipated intensity of eight emotions across videos. The emotion intensity scale ranges from 1: Very weak, 2:Weak, 3:Moderate, 4:High, 5:Very High. P-values for two sided t-test mean comparison provided.

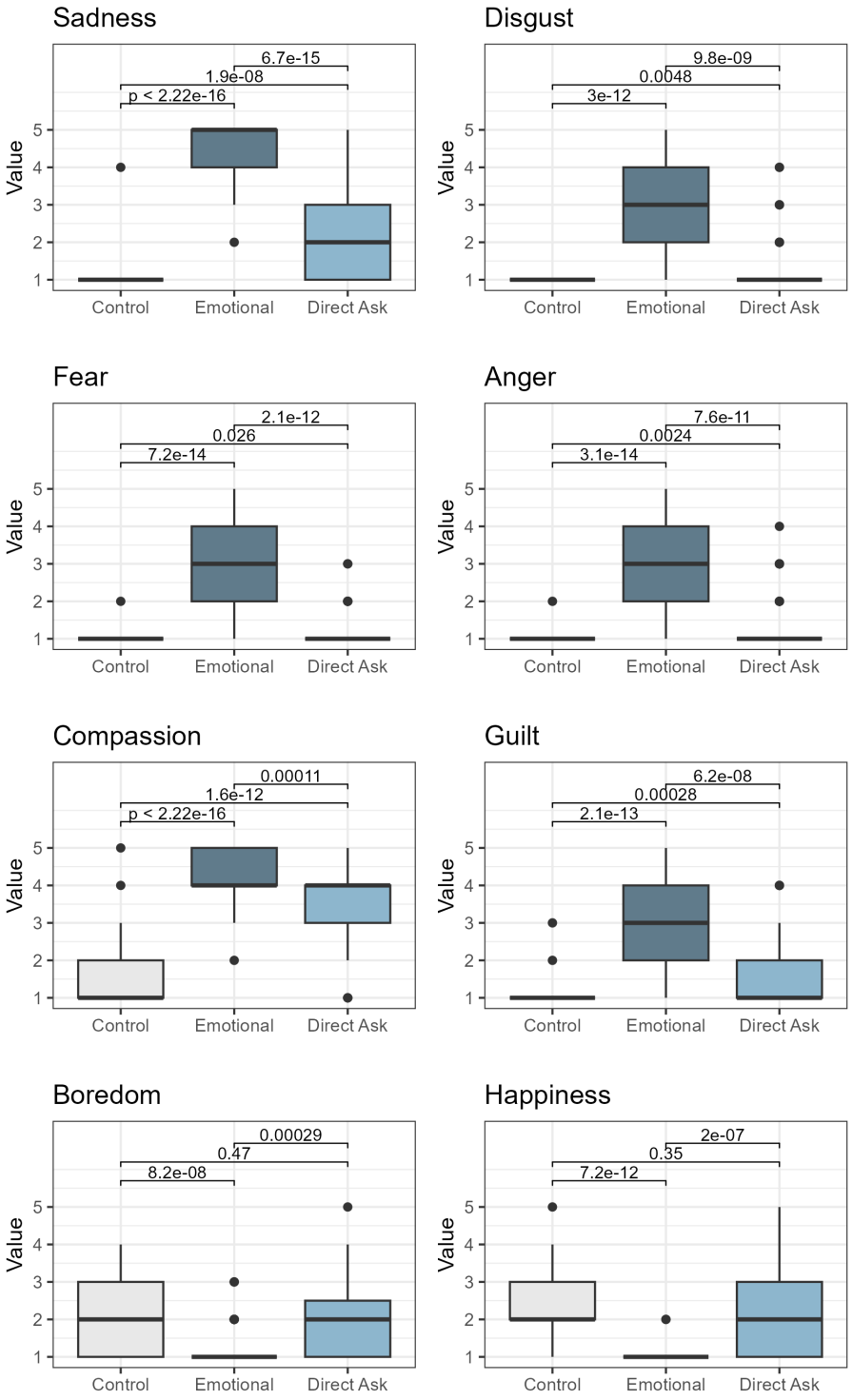


Figure C2: This figure presents the answers to evaluations on the experienced intensity of eight emotions across videos. The emotion intensity scale ranges from 1: Very weak, 2:Weak, 3:Moderate, 4:High, 5:Very High. P-values for two sided t-test mean comparison provided.

C Model

C.1 Estimated models results

Table C1: Proportion of estimated types across conditions.

Panel A. Emotional Treatment

	Estimate	SE	2.5 %	97.5 %
Altruistic SI	0.253	0.026	0.203	0.304
Altruistic SD	0.156	0.027	0.103	0.208
Selfish SD	0.108	0.023	0.063	0.154
Selfish SI	0.483	0.030	0.425	0.540
Error (ϵ)	0.004	0.004	-0.004	0.012
Naiveté (w)	0.587	0.040	0.508	0.666
logLik	-929.112			

Panel B. Direct Ask Treatment

	Estimate	SE	2.5 %	97.5 %
Altruistic SI	0.302	0.028	0.248	0.357
Altruistic SD	0.129	0.028	0.074	0.184
Selfish SD	0.060	0.023	0.015	0.104
Selfish SI	0.509	0.031	0.448	0.569
Error (ϵ)	0.014	0.007	0.001	0.028
Naiveté (w)	0.626	0.046	0.537	0.716
logLik	-919.856			

Panel C. Control

	Estimate	SE	2.5 %	97.5 %
Altruistic SI	0.283	0.024	0.237	0.329
Altruistic SD	0.000	0.001	-0.003	0.003
Selfish SD	0.000	0.009	-0.018	0.018
Selfish SI	0.717	0.023	0.671	0.763
Error (ϵ)	0.023	0.005	0.012	0.033
Naiveté (w)	1.000			
logLik	-869.159			

C.2 Robustness checks for the model

C.2.1 Alternative model: Relaxing assumption on video choice

We introduce a small change to the way agents choose videos in the cold state. In the main text, we assumed that agents applied a lexicographical criterion to the video choice: 1. they choose a video that leads to the implementation of the deliberative optimum in the hot state and 2. given this, they maximize $V(x|\theta)$ to reap the highest possible amount of emotional utility without deviating from the deliberative optimum. The modification to this is that we assume here that individuals do not engage in point 2, once they have a video that guarantees the deliberative optimum, they do not further care about maximization of emotional utility. This means that:

(Modified video choice in the cold state) The agent in the cold state chooses the video in

the following way:

(a) They choose θ s.t. $\text{argmax}_x V(x|\theta) = \text{argmax}_x U(x)$

The agent is then happy to choose a video that aligns the choice in the hot state with U, but if both videos do this, agents do not try to maximize the emotional utility.

The only predictions that change compared to Panel A in Table C1 are the ones relative to the choice of the video for the sophisticated state-independent altruists and selfish types. Under the new assumption, these types are indifferent between the videos.

The Table C2 below shows the new estimated parameters under the relaxed assumption on video choice for the *Emotional Treatment*.

Table C2: Estimates proportion of types for the *Emotional Treatment* relaxing the assumption of video choice in the cold state

	Estimate	SE	2.5 %	97.5 %
Total Altruist SI	0.281	0.031	0.220	0.343
Total Altruist SD	0.125	0.032	0.063	0.187
Total Selfish SI	0.505	0.034	0.438	0.572
Total Selfish SD	0.088	0.025	0.040	0.137
Error(ϵ)	0.003	0.003	-0.003	0.008
Naiveté (w)	0.386	0.134	0.123	0.649
logLik	-936.748			

C.2.2 Naiveté

We compare the fit of each model estimated in the tables below (the first one with full sophistication $w = 0$ and the second one with no sophistication $w = 1$) relative to the main model with w as a free parameter. A likelihood ratio test shows that the model fit is significantly worst for the restricted models ($\chi^2 = 148.568, df = 1, p = 0$ and $\chi^2 = 81.692, df = 1, p = 0$, for the model full and no sophistication, respectively).

Table C3: Estimates proportion of types Emotional Treatment imposing no Naiveté ($w = 0$)

	Estimate	SE	2.5%	97.5%
Altruistic SI	0.323	0.032	0.261	0.386
Altruistic SD	0.059	0.029	0.002	0.117
Selfish SD	0.000	0.006	-0.011	0.012
Selfish SI	0.617	0.032	0.555	0.679
Error(ϵ)	0.171	0.014	0.144	0.198
logLik	-1003.397			

Table C4: Estimates proportion of types Emotional Treatment imposing full Naiveté ($w = 1$)

	Estimate	SE	2.5%	97.5%
Altruistic SI	0.322	0.036	0.251	0.392
Altruistic SD	0.003	0.002	-0.000	0.006
Selfish SD	0.093	0.051	-0.006	0.193
Selfish SI	0.582	0.038	0.508	0.657
Error(ϵ)	0.073	0.010	0.053	0.093
logLik	-969.958			

C.3 Raw Data of Type Profiles

Table C5 presents the distribution of observations for each combination (V, R, C, P) across treatments (columns 2–5). As before, V denotes the video that is chosen, R represents the video that is received, C indicates the choice made after watching R, and P is the prediction of behavior in a counterfactual scenario. Columns 6 and 7 display the types that are compatible with each combination (V, R, C, P) for both the main model, as detailed in the main text, and the alternative model that relaxes the assumption about video choice in the cold state presented in this section. We define the types as $t_{i,n} = \{t_{1,n}, t_{2,n}, t_{3,n}, t_{4,n}\} = \{\text{Altruist SI}, \text{Altruist SD}, \text{Selfish SI}, \text{Selfish SD}\}$, where $n = 0$ denotes full sophistication and $n = 1$ indicates naiveté.

Table C5 allows us to assess which types' behaviors are predicted to change when the assumption about video choice in the cold state is relaxed. As explained before, only the predictions related to video choice for sophisticated state-independent altruists and selfish types differ between the main and alternative models. In the alternative model, these types randomize between videos instead of choosing the video that yields the highest emotional utility. For instance, consider the sophisticated state-dependent altruist $t_{1,n=0}$. In the main model, this type should only display the combinations $(A, A, \theta_T, \theta_{Alt})$ and $(A, A, \theta_T, \theta_T)$. However, in the alternative model, due to the relaxed assumption, the following combinations also become possible: $(A, A, \theta_{Alt}, \theta_T)$ and $(A, A, \theta_{Alt}, \theta_{Alt})$. This expansion of possible combinations illustrates how relaxing the cold state video choice assumption impacts the alignment of observed data with model predictions.

Table C5: Raw Data — N Observations for Each Sequence of Choices by Treatment.

V	R	C	P	Control	Emotional	Direct Ask	Types Main Model	Types Alt. Model
A	A	θ_T	θ_T	48	45	60	$t_{1,n=0} + t_{1,n=1} + t_{2,n=1} + t_{3,n=1}$	$t_{1,n=0} + t_{1,n=1} + t_{2,n=1} + t_{3,n=1}$
A	S	θ_T	θ_T	0	10	8	$t_{2,n=0}$	$t_{2,n=0}$
S	A	θ_T	θ_T	0	0	1	—	—
S	S	θ_T	θ_T	94	40	58	$t_{4,n=1}$	$t_{4,n=0} + t_{4,n=1}$
A	A	θ_T	θ_{Alt}	23	25	42	$t_{1,n=0} + t_{1,n=1}$	$t_{1,n=0} + t_{1,n=1}$
A	S	θ_T	θ_{Alt}	0	0	0	—	—
S	A	θ_T	θ_{Alt}	9	21	17	$t_{2,n=0}$	$t_{2,n=0}$
S	S	θ_T	θ_{Alt}	60	25	32	$t_{2,n=1} + t_{4,n=1} + t_{3,n=1}$	$t_{4,n=0} + t_{2,n=1} + t_{4,n=1} + t_{3,n=1}$
A	A	θ_{Alt}	θ_T	12	27	16	$t_{1,n=1} + t_{2,n=1} + t_{3,n=1}$	$t_{1,n=0} + t_{1,n=1} + t_{2,n=1} + t_{3,n=1}$
A	S	θ_{Alt}	θ_T	1	10	3	$t_{3,n=0}$	$t_{3,n=0}$
S	A	θ_{Alt}	θ_T	5	0	3	—	—
S	S	θ_{Alt}	θ_T	59	60	41	$t_{4,n=0} + t_{4,n=1}$	$t_{4,n=0} + t_{4,n=1}$
A	A	θ_{Alt}	θ_{Alt}	26	31	20	$t_{1,n=1}$	$t_{1,n=0} + t_{1,n=1}$
A	S	θ_{Alt}	θ_{Alt}	0	1	0	—	—
S	A	θ_{Alt}	θ_{Alt}	3	12	11	$t_{3,n=0}$	$t_{3,n=0}$
S	S	θ_{Alt}	θ_{Alt}	63	92	89	$t_{4,n=0} + t_{2,n=1} + t_{4,n=1} + t_{3,n=1}$	$t_{4,n=0} + t_{2,n=1} + t_{4,n=1} + t_{3,n=1}$

D Disentangling Social Pressure

Notice that our *Direct Ask* Treatment added an extra layer of manipulation: we requested individuals to write a message justifying their donation decision, which is shared with StC. In this section we explore how this requirement influenced the effect of social pressure on donations and strategic avoidance.

We first look at the effect on donations. Total donations in the *Direct Ask* and *Direct Ask No Message* treatments are approximately 37% and 43%, respectively, and are not significantly different following a two-sided proportion test ($\chi^2 = 1.500, df = 1, p = 0.221$).

In order to assess the effect of each treatment on donations while controlling for potential differences in selection, we use the regressions displayed in Table D1 below.

Table D1: Linear probability model of donation decision

	Dependent Variable: Donation	
	Direct Ask	Direct Ask No Message
	(1)	(2)
Constant	0.18^{***}	0.25^{***}
	(0.03)	(0.05)
Preferred Treatment (ref=no)	0.27^{***}	0.19^{***}
	(0.05)	(0.07)
Received Treatment (ref=no)	0.10^{**}	0.13^{*}
	(0.05)	(0.07)
Observations	401	199
R^2	0.10	0.06

Notes: This table presents the results of a linear probability regression of the donation decision on whether the individual preferred the treatment video (Preferred Treatment) and on whether (s)he received it (Received Treatment) for the *Direct Ask* and *Direct Ask No Message* treatments. Standard errors in parentheses ($^{\circ}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; *** $p < 0.001$).

Table D1 shows that the effect of receiving the treatment video (controlling for selection) was similar independently of whether a message was requested along with the video visualization. A Z-test to compare the coefficients of Received Treatment across regressions reveals that there is no differential effect on donations for both treatments ($Z = 0.316, p = 0.376$, two-sided).

Finally, avoidance levels were approximately 46% and 39% for the *Direct Ask* and the *Direct Ask No Message* treatment, respectively. We cannot conclude that these proportions are significantly different from each other using a two-sided proportion test ($\chi^2 = 1.990, df = 1, p = 0.158$).

E Formulas

E.1 Conditional probability

The following equation shows the probability of observing the sequence of outcomes (V, R, C, P) in the experiment given the type $t \in \{\text{Altruist SI}, \text{Altruist SD},$
$\text{Selfish SI}, \text{Selfish SD}\}$, the naiveté $n \in \{0, 1\}$, and the probability to make a mistake when choosing ϵ. Note that, in the equation, we assume that naive agents randomly choose the video when they are indifferent. Moreover, we assume that mistakes in different choices are independent and have the same probability.

$$
\begin{aligned}
p(V,R,C,P|t,n,\epsilon) = \Big[&(1-\epsilon)^{\mathbb{1}(V=V_t^*)} \cdot \epsilon^{1-\mathbb{1}(V=V_t^*)} \cdot 0.6^{\mathbb{1}(R=V)} \cdot 0.4^{1-\mathbb{1}(R=V)} \cdot \\
&\cdot (1-\epsilon)^{\mathbb{1}(C=C_{t,R}^*)} \cdot \epsilon^{1-\mathbb{1}(C=C_{t,R}^*)} \cdot (1-\epsilon)^{\mathbb{1}(P=P_{t,R,s}^*)} \cdot \epsilon^{1-\mathbb{1}(P=P_{t,R,s}^*)}\Big]^{1-n} \cdot \\
&\cdot \Big[\frac{1}{2} \cdot 0.6^{\mathbb{1}(R=V)} \cdot 0.4^{1-\mathbb{1}(R=V)} \cdot \\
&\cdot (1-\epsilon)^{\mathbb{1}(C=C_{t,R}^*)} \cdot \epsilon^{1-\mathbb{1}(C=C_{t,R}^*)} \cdot (1-\epsilon)^{\mathbb{1}(P=P_{t,R,n}^*)} \cdot \epsilon^{1-\mathbb{1}(P=P_{t,R,n}^*)}\Big]^{n}
\end{aligned} \tag{3}
$$

With V_t^*, $C_{t,R}^*$, $P_{t,R,s}^*$, $P_{t,R,n}^*$ denoting the correct choices for type t (as per Table 3 above). Specifically, V_t^* is the correct choice of video; $C_{t,R}^*$ is the correct choice between A and S after seeing R; $P_{t,R,s}^*$ is the correct prediction of a sophisticated agent t after watching R; and $P_{t,R,n}^*$ is the correct prediction of a naive agent t after watching R.

F Text Analysis

F.1 AI classification

We used the OpenAI API with the GPT-4O model to classify participants' responses to the question: "Why did you choose the video?" To guide the AI in this task, we provided a detailed prompt explaining the context of the experiment (see prompt below) and a set of five clear categories with examples for each. The categories were as follows:

- Avoidance of Emotional Influence: For responses indicating that participants avoided emotional impact. Our analysis is not able to distinguish whether such avoidance occurs for strategic reasons (not wanting to give) or because people simply dislike emotions. While we tried to make further classifications, most responses were too generic to allow such fine-grained classifications.
- Seeking Emotional Influence: For responses showing that participants wanted to feel more emotionally connected or motivated to donate.
- Curiosity or Interest in Content: For responses motivated by an interest in learning or the topic itself, without emotional reasons.
- Random: For responses indicating that participants chose randomly or without much thought.
- Other: For responses that did not clearly fit into any of the above categories.

The AI was explicitly instructed to assign each response to only one category, defaulting to "Other" if the appropriate category was unclear.

F.2 Results

The results for this exercise are shown below in Table D2.

Table D2: Reasons for Video Choice by Treatment

Treatment	Video Choice	Avoidance of Emotions	Seeking Emotions	Curiosity/Interest in Content	Other	Random	n
Emotional	Alternative	64.38%	0.43%	24.03%	6.44%	4.72%	233
	Treatment	2.41%	24.70%	65.06%	7.23%	0.60%	166
Direct Ask	Alternative	37.16%	1.09%	43.17%	11.48%	7.10%	183
	Treatment	3.67%	12.84%	75.69%	5.50%	2.29%	218
Control	Alternative	2.96%	0.00%	82.25%	1.18%	13.61%	169
	Treatment	0.85%	1.28%	88.89%	1.28%	7.69%	234

Table D2 evidences a series of patterns. First, and consistent with our previous findings, individuals frequently cited *Avoidance of Emotions* as a reason for choosing the alternative video, particularly in the *Emotional* treatment. While the *Direct Ask Treatment* displays some level of this type of avoidance it is almost half of what we observe in the *Emotional* treatment. Figure D1 shows the list of most frequent 20 words for both treatments. Highlighted in red are words that can be associated to emotions: feel, guilty, sad and emotional. Notice that the frequency with which these words are mentioned in the *Emotional* treatment is, at minimum, more than double than the frequency in the *Direct Ask* treatment.

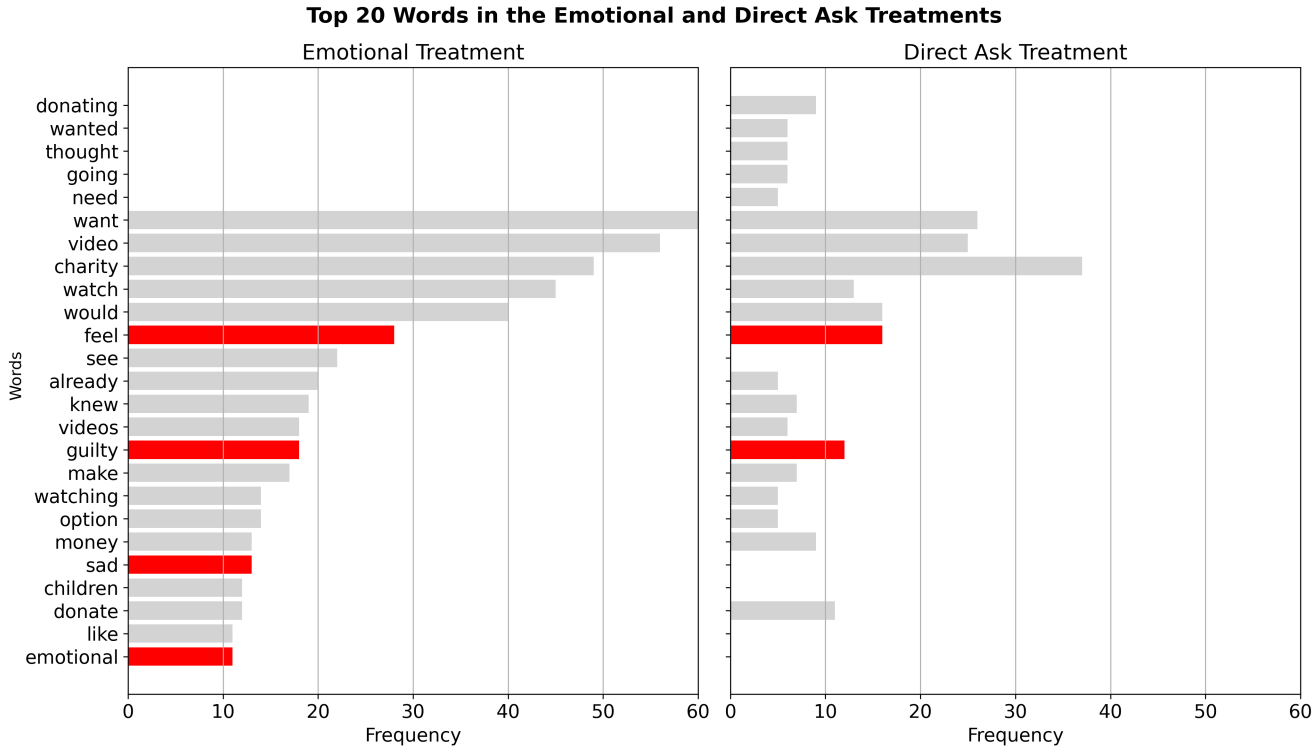


Figure D1: Top 20 most frequent words in the *Emotional* and the *Direct Ask* treatment. Basic stopwords from a default list of stopwords were removed.

Second, and in line with the idea of commitment towards altruism, participants also mentioned *Seeking Emotions* as a reason for choosing the Treatment video in the *Emotional Treatment*. These reason for choosing the Treatment video is again the highest for the *Emotional Treatment*. The fact that individuals who chose the Alternative video gave reasons in line with

Avoidance of Emotions and that those who chose the treatment video gave reasons in line with *Seeking Emotions* further supports our interpretation that emotional self-regulation is a strategy that individuals commonly used in the *Emotional* treatment.

In addition, and as discussed in the text, participants also cited *Curiosity and Interest in Content* as a reason for choosing the Treatment video. This was a significant motive across all treatments, though it was most prominent in the *Control* treatment. For the *Emotional* and the *Direct Ask* treatment, where participants knew the video was described as being related to the charity, it is challenging to distinguish between information seeking as a form of commitment to altruism and information-seeking as a form of pure information demand. We cannot clarify this with the existing data and further research would be needed to understand whether the first or the second, or mixture of both motives, was at play. Nevertheless and taking the open ended answers at face value, this does suggest that non-emotional reasons where most important when choosing the Treatment video.

F.3 Prompt

The prompt varied the description of the treatment video for each treatment. We provide the prompt for the *Emotional* treatment below.

Prompt for AI-Based Classification Task

prompt = f""" I am analyzing responses from an online experiment where participants were asked to choose between two videos before making a donation decision.

The experiment was designed to investigate emotional self-regulation and commitment in altruistic decision-making.

Video Options Participants Could Choose From:

1. **Charity Video (Emotional/Empathy-Inducing Video)**
Description (provided to participants): This video is part of a campaign by Save the Children. It shows the struggles of a nine-year-old girl when her city becomes a warzone. As the political conflict escalates, the girl and her family experience increasingly traumatic hardships and perils.
Purpose: Designed to evoke empathy and emotional engagement, increasing the likelihood of donating.

2. **Alternative Video (Neutral Video)**
Description (provided to participants): This is an alternative video of similar length, unrelated to charity.
Purpose: Serves as a neutral, non-emotional control, minimizing emotional engagement.

Participants made their video decision knowing that video allocation was randomized, meaning they could end up watching their less preferred video.

After watching the assigned video, participants made their donation decision.

Response Classification Task

After this, participants were asked an open-ended question:
"Why did you choose this video¿'

Their answers need to be categorized into one of the following five classifications:

Classification Categories & Examples

Category 1: **Avoidance of Emotional Influence**
Participants explicitly state they chose the alternative video to avoid emotional influence, social pressure, feelings of guilt/compassion or emotions in general.
Example responses:

"I didn't want to feel pressured to donate."
'I knew the charity video would make me feel guilty, so I avoided it."
"I don't like being manipulated emotionally."
"I don't like watching sad or emotional videos."
"I find these kinds of videos too distressing."
"I avoid war-related content because it's upsetting."

Category 2: **Seeking Emotional Influence**
Participants explicitly state they chose the charity video to increase their emotional engagement or to ensure they would donate.
Example responses:
"I wanted to feel more connected to the cause and be more motivated to donate."
"The charity video helps me remember why donating is important."
"I thought the charity video would make me care more about the issue."

Category 3: **Curiosity or Interest in Content**
Participants choose a video due to general interest in the topic, rather than emotional reasons.
Example responses:
"I wanted to learn more about how grass grows."
"I was curious about how the charity presents its appeal."
"I just wanted to see what the videos were about."

Category 4: **Random**
Participants explicitly indicate that they chose randomly.
Example responses:
"I just picked one at random."
"I didn't really think about it."
"It didn't really matter to me which one I watched."

Category 5: **Other**
Responses that do not fit any of the above categories.

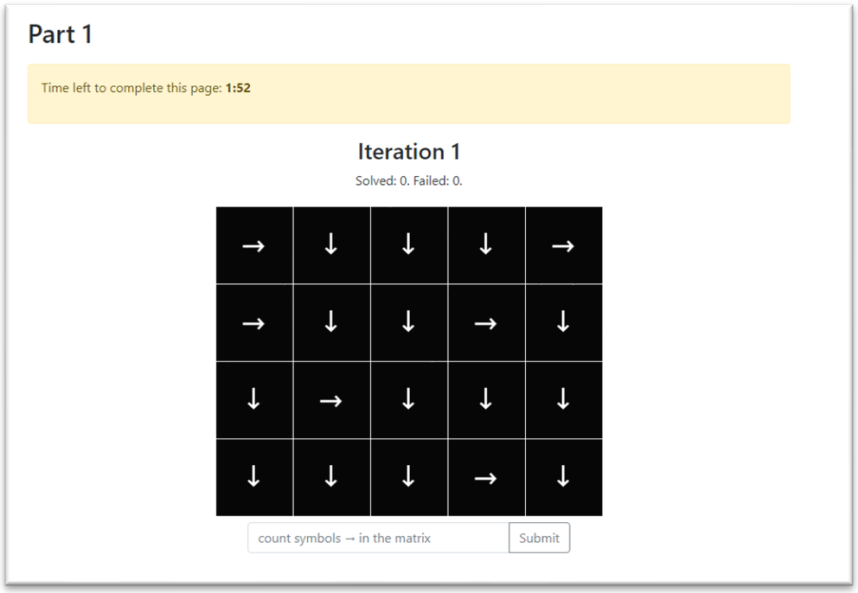
—

Instructions for Classification Task:

1. Read the response carefully.
2. Assign the response to only the most relevant category.
3. Ensure consistency in classification.

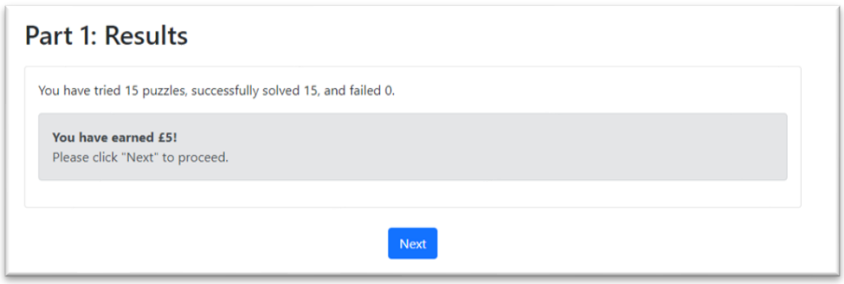
Final Step:
Now, classify the following participant response based on these categories:
Participant's Response: {*narrative*}
Assigned Category: [Provide category number from 1–5, provide just the number] """

G Experimental protocol (Emotional Treatment)



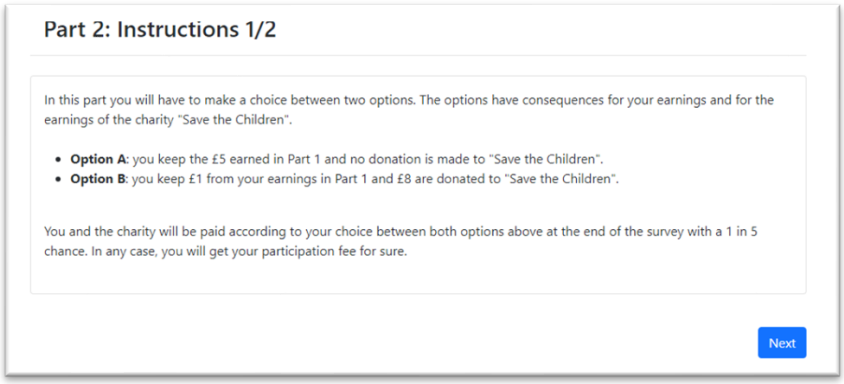
Real Effort Task

Participants count the arrows in the matrix for 2 minutes.



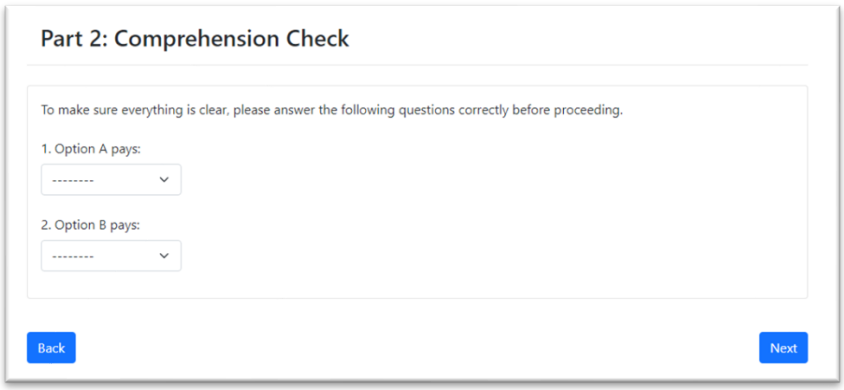
Real Effort Task Bonus

We award the 5 pound endowment.

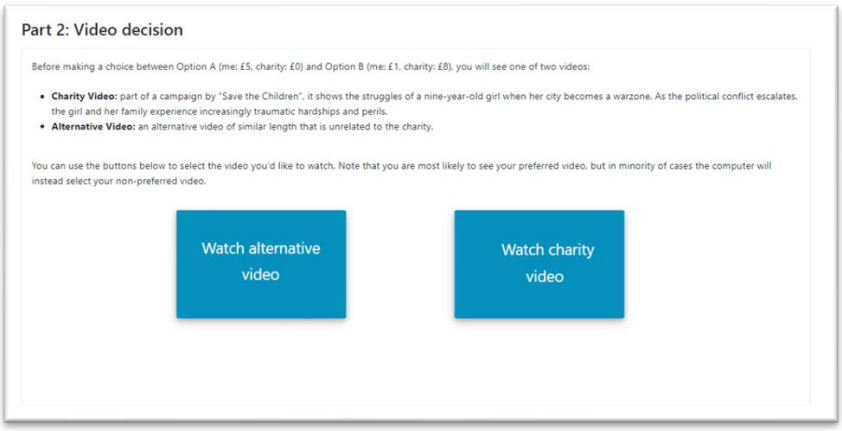


Instructions about Donation Possibility

Participants are informed of the donation opportunity and the available payoffs.

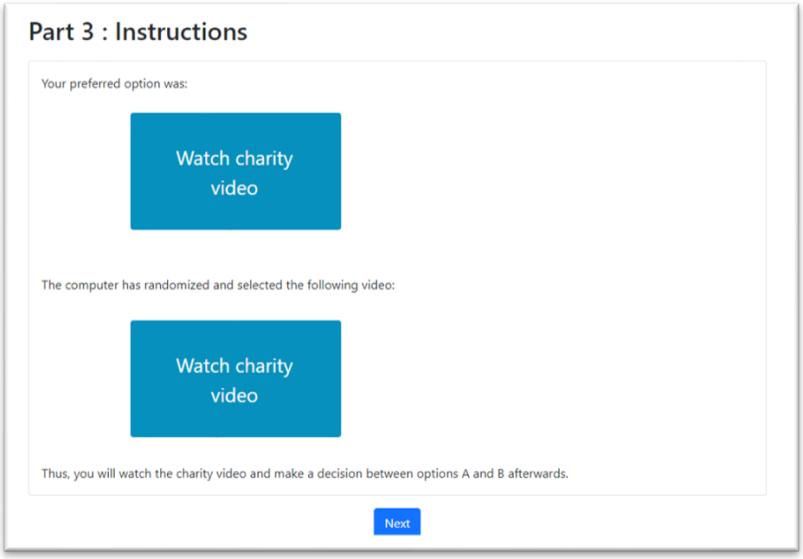


Comprehension Check



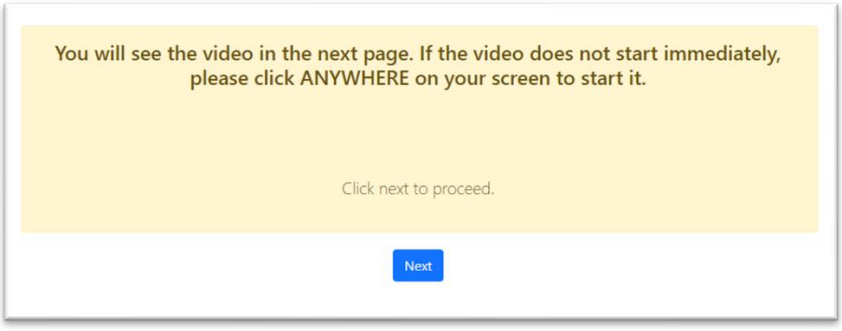
Instructions of Video

Individuals are informed about the two videos. The description of the treatment video (in this case the Charity video) varies per treatment. Individuals need to select one of the two videos to proceed.



Implemented Video Result

Individuals are informed about the result of the random implementation.

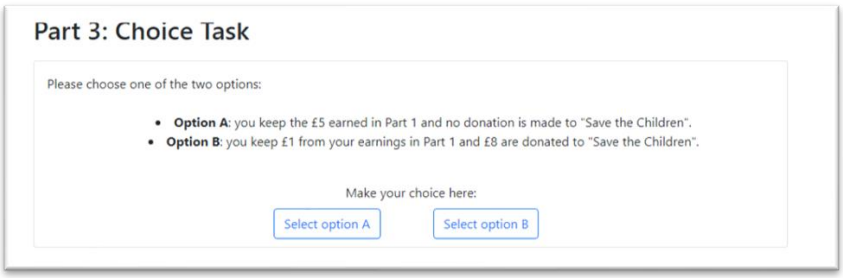


Video Alert

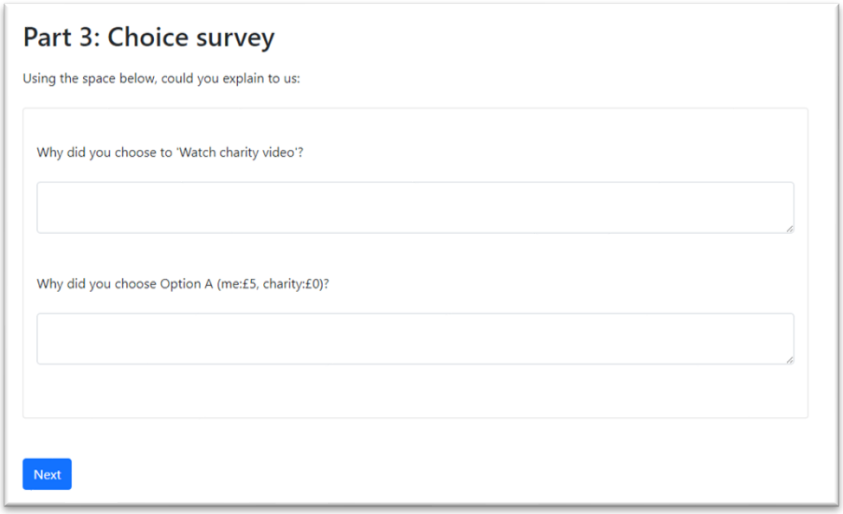
Some browsers block the video, so we provided instructions on what to do to avoid that.

Video Visualization

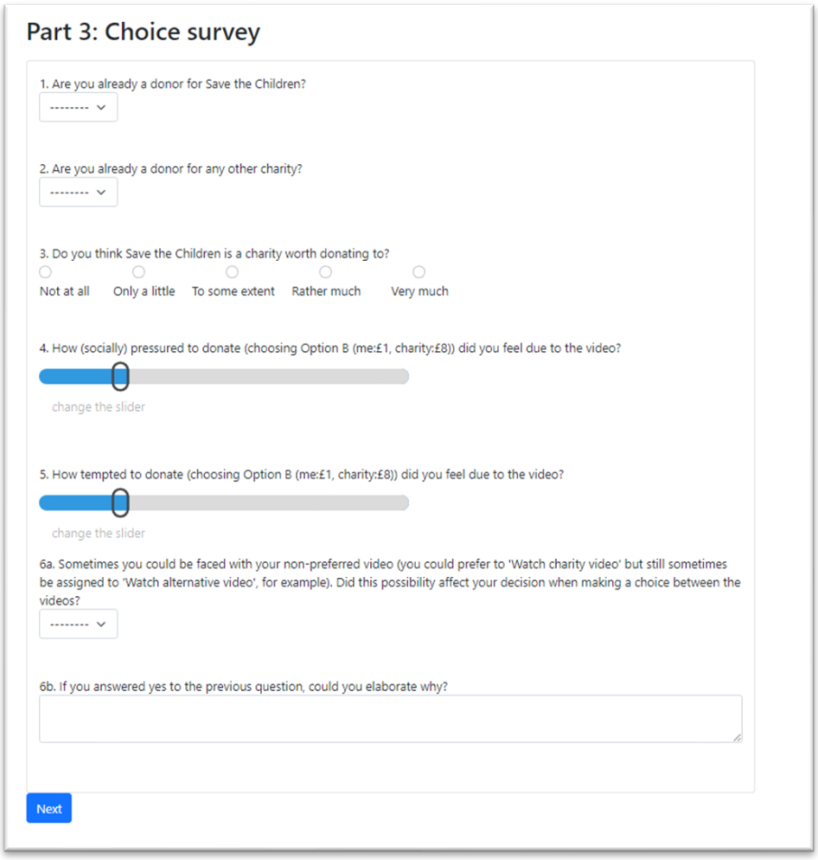
Individuals watch the video, full screen and without the possibility to stop. When the video stops, a “Next” button to navigate further is activated.



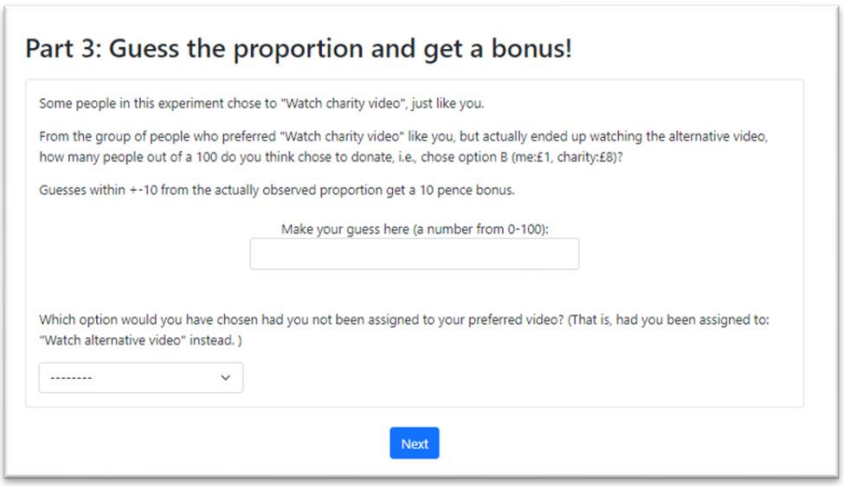
Donation Decision



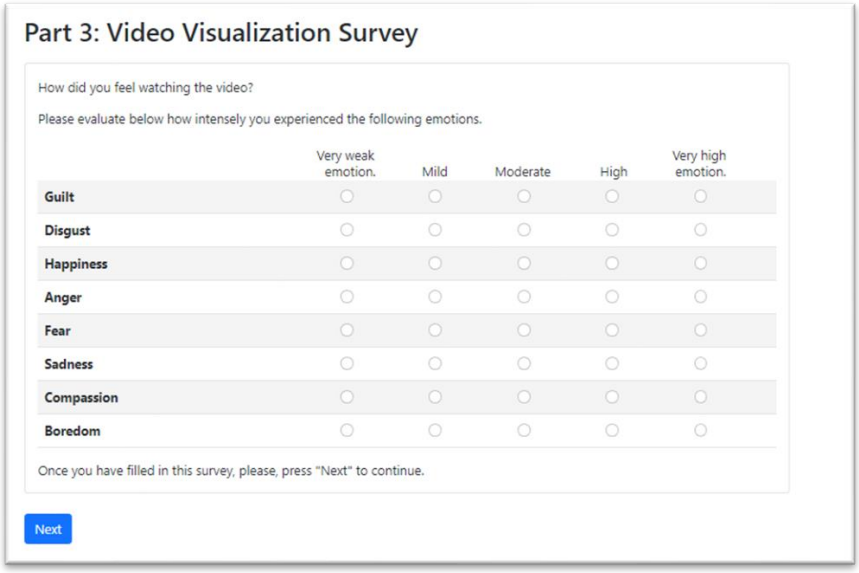
Choice Survey (1)



Choice Survey (2)



Choice Survey (3)
Incentivized and unincentivized belief questions.



Part 3: Video Visualization Survey

How did you feel watching the video?

Please evaluate below how intensely you experienced the following emotions.

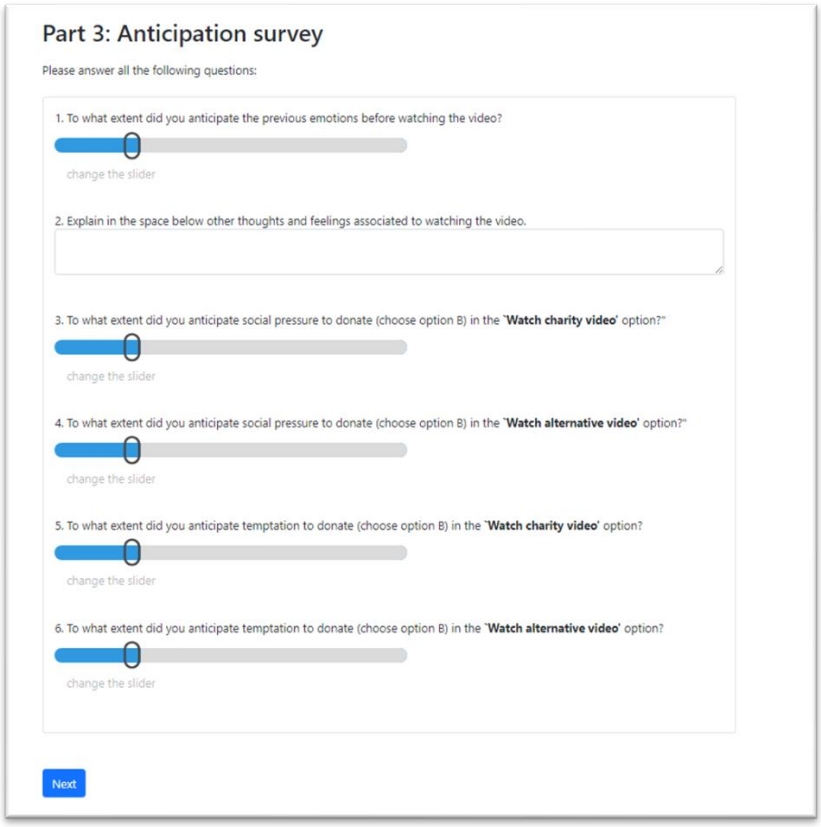
	Very weak emotion.	Mild	Moderate	High	Very high emotion.
Guilt	○	○	○	○	○
Disgust	○	○	○	○	○
Happiness	○	○	○	○	○
Anger	○	○	○	○	○
Fear	○	○	○	○	○
Sadness	○	○	○	○	○
Compassion	○	○	○	○	○
Boredom	○	○	○	○	○

Once you have filled in this survey, please, press "Next" to continue.

Next

Choice survey (4)

Emotions questionnaire



Part 3: Anticipation survey

Please answer all the following questions:

1. To what extent did you anticipate the previous emotions before watching the video?

change the slider

2. Explain in the space below other thoughts and feelings associated to watching the video.

3. To what extent did you anticipate social pressure to donate (choose option B) in the '**Watch charity video**' option?"

change the slider

4. To what extent did you anticipate social pressure to donate (choose option B) in the '**Watch alternative video**' option?"

change the slider

5. To what extent did you anticipate temptation to donate (choose option B) in the '**Watch charity video**' option?

change the slider

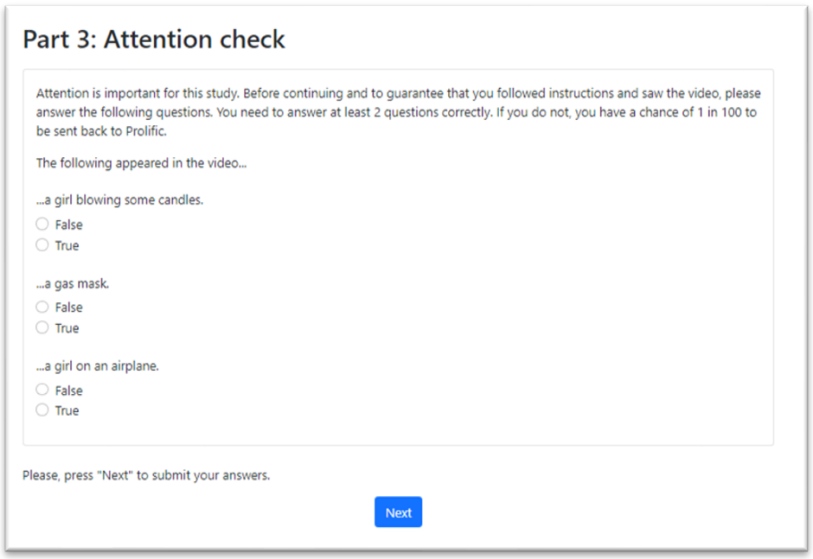
6. To what extent did you anticipate temptation to donate (choose option B) in the '**Watch alternative video**' option?

change the slider

Next

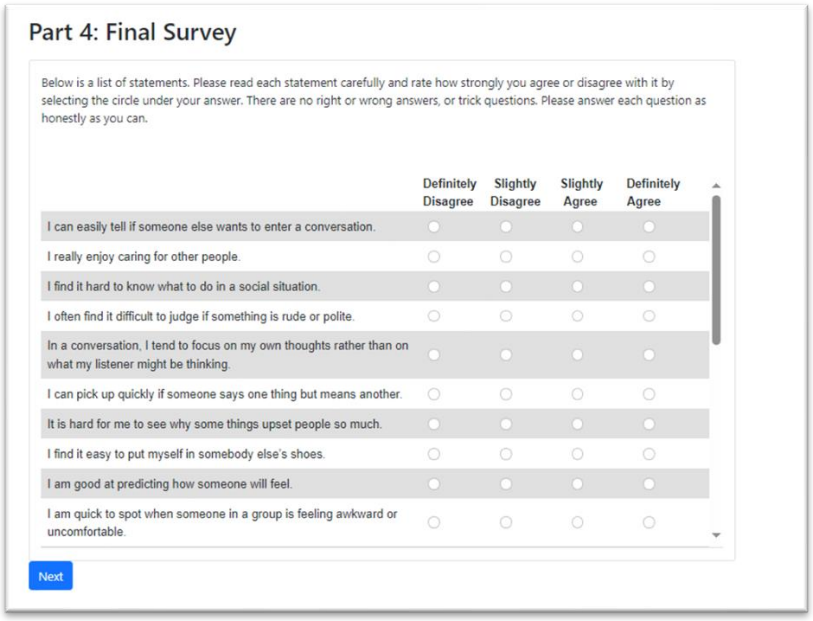
Choice survey (4)

Anticipation of emotions, pressure and temptation.



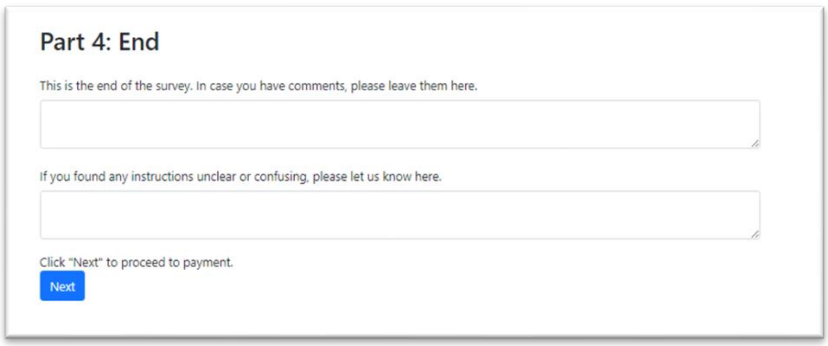
Attention Check for Video Visualization

Three questions associated to the implemented video.

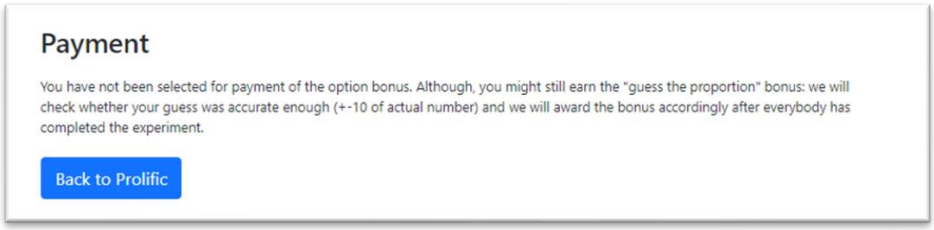


Final Survey

Short version of the Empathy Quotient Questionnaire cited in the text. Participants can scroll down to answer a total of 22 questions.



Feedback for experiment



Bonus Payment

Participants are informed whether they were selected to receive the bonus.

End of experiment