Diffusion of Being Pivotal and Immoral Outcomes*

Armin Falk[†] Thomas Neuber[‡] Nora Szech[§]

May 26, 2019

Abstract

We study how the diffusion of being pivotal affects immoral outcomes. In our main experiment, subjects decide about agreeing to kill mice and receiving money versus objecting to kill mice and foregoing the monetary amount. In a baseline condition, subjects decide individually about the life of one mouse. In the main treatment, subjects are organized into groups of eight and decide simultaneously. Eight mice are killed if at least one subject opts for killing. The fraction of subjects agreeing to kill is significantly higher in the main condition compared with the baseline condition. In a second experiment, we run the same baseline and main conditions but use a charity context and additionally study sequential decision making. We replicate our finding from the mouse paradigm. We further show that the observed effects increase with experience, i.e., when we repeat the experiment for a second time. For both experiments, we elicit beliefs about being pivotal, which we validate in a treatment with non-involved observers. We show that beliefs are a main driver of our results.

Keywords: Diffusion of being pivotal, group decisions, morality, replacement logic
JEL Codes: C91, C92, D01, D02, D23, D63, D71

*We thank K. Albrecht, S. Altmann, R. Bénabou, T. Dohmen, J. Engel, D. Engelmann, S. Gächter, P. Heidhues, D. Huffman, S. Jäger, F. Kosse, B. Kőszegi, F. Krämer, G. Loewenstein, F. Rosar, J. Sobel, N. Schweizer, F. Zimmermann and participants at various seminars for helpful comments. We thank M. Antony, T. Graeber, and T. Wölk for excellent support. Falk acknowledges financial support by the German Science Foundation (DFG) through the Leibniz Program and by the European Research Council (ERC Advanced Grant 340950). The study was approved by Ethical Committee of the University of Bonn, Reference Number: 066/12.
[†]Institute on Behavior & Inequality (briq) and University of Bonn; armin.falk@uni-bonn.de
[‡]University of Bonn; thomas.neuber@uni-bonn.de
[§]Karlsruhe Institute of Technology (KIT); nora.szech@kit.edu

1 Introduction

This paper studies how groups favor moral transgression in diffusing responsibility and notions of being pivotal. Intuitively, acting in groups provides an excuse for acting immorally simply because an individual may perceive himself as not or only partly responsible for an outcome. To investigate the consequences of group settings that diffuse being pivotal, we ran two sets of experiments, varying the choice environment and contrast environments where subjects are fully pivotal with contexts where being pivotal is diffused by an exogenous change in organizational design. In the latter, subjects are organized into groups and individual decisions are aggregated such that the individual can easily believe that his decision is unlikely to be pivotal. Organizing people into groups and implementing a decision rule that does not require the support of all members for immoral action enables a simple "replacement logic" (see Sobel, 2010), which denotes the procedural phenomenon whereby people can mutually excuse their own immoral behavior with their individual powerlessness in the face of others' immoral behavior.

In our main experiment, the paradigm involves the trade-off between life and money. Subjects decide between receiving money and agreeing to kill mice versus not receiving money and objecting to the killing.[1] Importantly, mice used in the experiment are so-called "surplus" mice, all of which would have been killed without our intervention (see Section 2). Subjects learn about this default in a post-experimental debriefing. The paradigm is informed by the widely-held view that harming others in an unjustified and intentional way is considered immoral.[2] We contrast two treatments: the *Baseline* treatment implements a simple binary choice where subjects either receive €0 for saving a mouse (Option A) or €10 for killing the mouse (Option B). In Baseline, subjects are hence fully pivotal. This condition serves as a comparison benchmark for the main *Simultaneous* treatment, in which eight subjects simultaneously decide between Option A and Option B. As in Baseline, a subject receives no money for choosing Option A

[1] The study was approved by the ethics committee of the University of Bonn.

[2] See, e.g., Gert (2012, Section 1) on "The Definition of Morality", The Stanford Encyclopedia of Philosophy: "In this descriptive sense, although avoiding and preventing harm is common to all, 'morality' can refer to codes of conduct of different societies with widely differing content, and still be used unambiguously."

and €10 for choosing Option B, irrespective of the other subjects' choices. However, if at least one subject chooses Option B, eight mice are killed. Thus, if a subject believes that at least one other subject is likely to choose Option B, he may no longer consider himself as being pivotal. In line with our argument, we find that the fraction of subjects choosing Option B is significantly higher in Simultaneous than in Baseline, despite the fact that—upon being pivotal—killing causes the death of eight mice rather than one. Moreover, the likelihood that a subject chooses to kill mice decreases with his belief of being pivotal. At the aggregate level, all mice are killed in Simultaneous. The result shows how a simple change in organizational design can causally promote immoral behavior and outcomes. Our findings thus contribute to understanding the sources of malleability of moral behavior and why "ordinary" people endowed with given moral values may engage in activities to which they would generally object.

Our second choice paradigm involves the binary decision between receiving €10 for oneself or donating €15 to a charity that supports children suffering from cancer. We replicate the two treatments from the first experiment as closely as possible (BaselineC and SimultaneousC). For completeness, we also study a dynamic setting of diffusion of responsibility which mimics SimulatenousC but in which decisions are made sequentially (SequentialC). Finally, we investigate experience effects, i.e., whether the observed effects become larger if subjects repeat the same experiment one more time. The charity experiment replicates the main effect from the mouse conditions. The share of subjects choosing the selfish Option B is significantly higher in the simultaneous condition than in the baseline treatment. Moreover, choosing a second time in BaselineC on average does not affect the likelihood of donations but—as expected—induces more selfish choices in both SimultaneousC and SequentialC. In the latter, we additionally find that previous history matters for behavior. In particular, learning that Option B has already been chosen essentially eradicates the choice of Option A further down the line.

Perceptions of being pivotal are central to the mechanism under study and they critically hinge on beliefs about the behavior of others. This is why, in both experiments, we elicit beliefs and confirm that choices are strongly associated with the perceived likelihood of being pivotal. Given the critical role of beliefs, we ran a further treatment with non-involved subjects. In this condition, subjects read the instructions of all three treatments implemented with the charity paradigm and were asked to predict the results from the experiment. These independently elicited beliefs of spectators corroborate our above-mentioned findings. In particular, we find that beliefs of spectators are very similar to those of subjects actually making a decision.

Organizational contexts that generate replacement arguments are pervasive at various levels of social interaction. They range from state-organized violence and corrupt bureaucracies to cheating in sports, morally dubious market transactions, and malpractice within corporations. We discuss a few examples below. Some examples are more closely related to our simultaneous condition, others to the sequential choice context. Most real-world examples, however, share features of both. In this sense, our experimental group treatments represent limiting cases, where subjects decide either in isolation and complete uncertainty about other individuals' behaviors (Simultaneous and SimultaneousC), or with perfect information about previous choices and the exact timing and order of moves (SequentialC).

A striking example that closely corresponds to our simultaneous conditions is the practice of firing squads, which comprise of a group of executors rather than a single person. For all members, shooting entails the personal advantage of avoiding disciplinary measures, and "technologically" one person who shoots his gun is sufficient to bring about the killing. From an individual member's perspective, being pivotal is diffused, as many people shooting at the same time implies that the killing is likely to happen, regardless of whether a particular member fires his gun or not. Moreover, members of firing squads are often randomly issued a gun containing a blank cartridge, which additionally diffuses being pivotal: even if a member of the squad shoots his gun, he remains uncertain whether or not he can effectively cause the killing at all. Apparently, these features reduce feelings of responsibility and facilitate participating in executions.

Corruption is another example that closely resembles the simultaneous decision making context. Suppose that a citizen wants to gain illegitimate access to a public permit and therefore intends to bribe an official. He may approach different officials, but he only needs to find one single official who accepts the bribe. Since any official taking the bribe would do so secretly, there is no way to credibly signal honesty. If a given official is sufficiently certain that at least one of his colleagues is corrupt, he may now feel tempted to accept the bribe himself. This logic can give rise to an equilibrium where a large proportion of officials actually act corruptly. Doping in sports provides a similar example. Most athletes publicly state that they detest doping. However, many are later found guilty, with the road cyclist Lance Armstrong being an infamous example. This places athletes in a dilemma. They might actually object to cheating—at least because it jeopardizes the credibility of their discipline—but believe that others are doped anyway, which makes it seem more acceptable or even necessary to engage in doping themselves.

Reasoning about not being pivotal also helps to explain outcomes in markets that violate traders' own moral or fairness preferences. Here, a replacement argument prevails if traders prefer concluding a trade themselves over letting another trader perform the same transaction, even if trading creates unfair outcomes for traders themselves or imposes negative externalities on others. In cases where buying decisions create negative externalities, a frequent "excuse" is that "if I don't buy, another buyer will." On the opposite side of the market, suppliers of potentially harmful goods are in a similar situation, arguing that market demand would be met with or without their involvement. British Secretary of State Boris Johnson invoked an argument along these lines in October 2016 after allegations about weapons exported to Saudi Arabia being used for war crimes in Yemen. Faced with a motion in the House of Commons to suspend sales, he retorted that the respective members of parliament should "be in no doubt that we would be vacating a space that would rapidly be filled by other Western countries who would happily supply arms with nothing like the same compunctions or criteria or respect for humanitarian law" (Peck, 2016).[3]

The replacement logic also contributes to corporate crime. For example, Andrew

[3]This is a refined version of the discussed argument in pointing at positive "side effects" associated with Britain taking an active role (see Glover and Scott-Taggart, 1975, pp. 177). Yet, the latter might often represent mere excuses rather than sound justifications.

Fastow—chief financial officer (CFO) of Enron from 1998 until 2001, who played a central role in concealing massive losses prior to the firm's bankruptcy in 2001 and served a six-year sentence in prison—himself drew the following parallel: "But the reality is, if at any point in my career I said 'time out, this is bullshit, I can't do it'... they would have just found another CFO, but that doesn't excuse it. It would be like saying it's OK to murder someone because if I didn't do it someone else would have" (Soltes, 2016, p. 225). The above quote underscores our main hypothesis, while it also highlights that behavior in response to uncertainty about being pivotal may nevertheless be perceived as morally repulsive.[4] Note that the replacement logic draws on consequentialist moral thinking. By contrast, deontological moral reasoning would dictate doing the "right" thing regardless of being pivotal or not. The extent to which groups are vulnerable to transgression therefore crucially depends on the share of individuals following consequentialist vs. deontological moral reasoning, respectively. We discuss the relative shares in the context of our experiments in Sections 3 and 4.

Our paper is related to work on contextual factors affecting fair outcomes in the context of simple dictator, bargaining, or allocation games. While we focus on the role of beliefs about being pivotal, other mechanisms that have been identified to favor "unfair" outcomes are delegation or exploiting moral "wriggle rooms," as discussed, e.g., in Bartling and Fischbacher (2012), Hamman, Loewenstein, and Weber (2010), Dana, Weber, and Kuang (2007), and Serra-Garcia and Szech (2018).[5] Falk and Szech (2013) analyze the malleability of moral outcomes in bilateral and multilateral market situations and Falk (2017) studies the role of status inequality.

The diffusion of being pivotal can be interpreted in terms of higher costs of acting

[4]Another example in this vein is the role of the replacement logic in the organization of the Holocaust (Arendt, 1963; Darley, 1992; Lifton, 1986). Lifton (1986) interviewed German doctors stationed in Auschwitz. They were operating in a nightmarish environment, with one of their objectives being to "select" prisoners who would be allowed to live while others would be immediately gassed. Being ordinary doctors, this activity was likely to be morally terrible and self-contradictory to them. Nevertheless, they engaged in the selection procedures. One of the frequently made justifications was that the "horrible machinery would go on", regardless of whether a particular doctor continued to participate. Replacement arguments suggesting the impossibility to stop ongoing moral crime were also used in the Nuremberg Trials as excuses for having participated in various kinds of atrocity under the Nazi Regime (see, e.g., Crawford, 2007 and references therein).

[5]On the effects of institutions on values, see also Bowles (1998). On the role of authority, see Milgram [1974] (2009).

morally, because in our group treatments the probability of reaching a moral outcome when acting morally and foregoing the additional payment is smaller than one. In this sense, our findings are related to work by Andreoni and Miller (2002) and Fisman, Kariv, and Markovits (2007), who show that when exogenously varying the price of giving in simple dictator games, the observed willingness to share varies accordingly. Two important features differentiate our setup from this literature. First, we contrast a monetary benefit for oneself with a *moral* good. It remains to be shown whether people readily engage in trade-offs here as well. Second, we do not set the probability of being pivotal exogenously but it is determined endogenously by the behavior of others, rendering equilibrium considerations.[6] Another related strand of literature in social psychology concerns the so-called bystander effect (see, e.g., Latané and Darley, 1968; and for a recent overview Fischer et al., 2011). Typical bystander experiments study helping behavior in response to a staged emergency (e.g., the experimenter becomes injured). What sets our simultaneous treatments apart is that even if a subject opts for the moral outcome, he remains uncertain about whether the moral outcome is implemented or not, similar, e.g., to firing squads. By contrast, in typical bystander experiments this uncertainty does not exist. If a subject opts for helping, the person in need receives help. Furthermore, in a bystander experiment, while deliberating whether to help or not, subjects often observe that others do not help either. In our simultaneous-move setup, this type of social learning is ruled out. When deciding to kill a mouse or not to donate, respectively, subjects do not know whether other subjects also opt for the selfish option. The dynamic properties of observing others, however, are explicitly studied in our sequential treatment. In addition, in a bystander experiment, participants need to realize that their help is required (and that it is better to step in than to hope that some other, say, more able helper will step in), while in our setup the consequences of decisions are straightforward. We also note that in our experiment, consequences are real, incentives are exactly specified, and the mechanism (beliefs about being pivotal) is explicitly measured.

The remainder of the paper is organized as follows. In Section 2, we describe the

[6] For an equilibrium analysis of group decisions in morally relevant contexts, see Rothenhäusler, Schweizer, and Szech (2018).

design and implementation of the main experiment and develop our hypotheses. The results are presented in Section 3. Section 4 covers the charity experiment. We first present a replication of our main results and provide evidence for the validity of elicited beliefs. We then proceed by investigating an additional sequential condition. Finally, Section 5 concludes by summarizing the paper and discussing additional observations.

2 Experiment

Avoiding and preventing unjustified harm is central to most notions of morality. It is this notion that informs the "mouse paradigm" used in our main experiment, which involves the trade-off between killing a mouse and receiving money versus saving a mouse life and receiving no money (Falk and Szech, 2013).[7] Subjects are explicitly informed that each mouse is a young and healthy mouse that will live for about two years if saved. For illustrative purposes, we present subjects the picture of a mouse on an instruction screen. We guarantee subjects that mice—if saved—live in an appropriate, enriched environment, jointly with a few other mice. Hence, in case subjects decided to save mice, these mice were kept alive in an enriched environment, with good feed and comfortable nesting material, precisely as stated in the instructions.

2.1 Design

Subjects are also informed in detail about the killing process. In the instructions, they read the following passage: "[T]he mouse is gassed. The gas flows slowly into the hermetically sealed cage. The gas leads to breathing arrest. At the point at which the mouse is not visibly breathing anymore, it remains in the cage for another 10 minutes. It will then be removed." To further rule out uncertainty about the decision context, subjects

[7] Deckers et al. (2016) provide convergent and discriminatory validity of the mouse paradigm as a measure for morality. Killing is negatively related to agreeableness—one of the Big Five facets—which describes a tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others, and positively related to Machiavellianism, measuring a person's tendency to be unemotional, and detached from conventional morality. Moreover, killing is not related to disposable income, whether students are professionally involved with animal research or animal experiments or have a simple preference for animals, as expressed by having a pet at home.

are shown a short demonstration video of the killing process. In the video, four mice first move vividly in the cage, then they successively slow down as more and more gas enters the cage. Eventually they die, with their hearts visibly beating heavily and slowly.

It is important to stress that the mice used in the experiment were so-called "surplus" mice: these mice were bred for animal experiments, but proved to be unsuited for scientific research. They were perfectly healthy, but keeping them alive would have been costly. It is common practice in laboratories conducting animal experiments to gas such mice. Thus, as a consequence of our experiment, many mice that would have otherwise all died were saved. Subjects were informed about this default in a post-experimental debriefing.[8]

Treatments. We study the role of diffusion of being pivotal in contrasting two decision environments, one where subjects are fully pivotal (Baseline) and one where being pivotal is diffused by organizing subjects into groups (Simultaneous). The two decision contexts differ in terms of how likely it is that any given subject is pivotal, keeping overall moral and financial consequences identical. In Baseline, each subject decides about the life of one mouse. Subjects face a simple binary choice between Option A and Option B: Option A implies that the mouse will survive and that the subject receives no money, while Option B implies the killing of the mouse and receiving €10. The Baseline treatment informs us about the share of subjects who are willing to kill the mouse for €10 when obviously being pivotal.

In Simultaneous, subjects decide in groups of eight and are endowed with eight mice. As in Baseline, each subject faces an individual binary choice between Option A and Option B: Option A implies that a subject receives no money. If a subject choses Option B, he receives €10. Individual monetary consequences are independent of other subjects' decisions. All subjects choose simultaneously. They know that if at least one subject chooses Option B, all eight mice are killed. Furthermore, they know that they will not receive feedback on whether the mice are ultimately killed or not (although it is obvious

[8] While perceptions of the situation may have changed due to this information, the consequences were exactly the same and as stated in the instructions. In future research, it would be interesting to explore whether using an alternative framing would affect decisions in response to institutional changes differently (compare evidence on the so-called omission–commission bias, e.g., in Spranca, Minsk, and Baron, 1991).

for a subject that the mice die if he chooses Option B). Note that we chose to endow a group with eight mice to keep the number of mice at the aggregate level identical to Baseline. Of course we do not know whether the valuation of mice lives is proportional to the number of saved mice, but keeping numbers identical at the aggregate level allows for a clean comparison of the overall impact of group vs. individual decision making.

In Simultaneous, right after subjects have made their decision, we elicit beliefs about being pivotal. Subjects are asked to indicate the probability that all other seven group members have chosen Option A (*belief_pivotal*). We also ask subjects to estimate how many other subjects in their group have chosen Option B. They can enter any number from 0 to 7 and are paid €1 for a correct estimate (*belief_B*).

Procedure. 252 subjects—mainly undergraduate university students from all majors—took part in the experiment, 124 subjects in Baseline and 128 in Simultaneous. Each subject participated only in one treatment condition. We used z-Tree as the experimental software (Fischbacher, 2007). Subjects were recruited using the software ORSEE (Greiner, 2004). At the beginning of an experimental session, participants received detailed information about the rules and the structure of the experiment. In all treatments, the experiment started only after all participants had answered several control questions correctly.

To reduce possible communication between subjects across sessions, the experiment was run on two consecutive days in six different rooms at the *Beethovenhalle*, the largest concert hall in Bonn. We set up six parallel, computerized labs in these rooms. Subjects received payments according to the rules of the experiment and an additional show-up fee of €20 to compensate for the remote location. In both treatments, subjects received their payments in a sealed envelope outside the room where the experiment had taken place. This way, neither other subjects nor the experimenters handing over the envelopes knew what a particular subject had earned. This procedure was explained in the instructions.

To ensure credibility, we stated right at the beginning that all statements made in the instructions were true—as is standard in economic experiments—and that all consequences of subjects' decisions would be implemented exactly as described in the instructions. We emphasized orally that the experimenters would personally guarantee the truthfulness of the instructions. Subjects were also invited to send us an email if they wanted to discuss the study.

2.2 Hypotheses

Our predictions start from the premise that most subjects follow consequentialist reasoning rather than exhibiting rule-based deontological behavior. We expect that subjects in the Simultaneous treatment will engage in strategic considerations, thinking about how other subjects will decide. If they come to the conclusion that the likelihood of being pivotal is sufficiently small, subjects will find it justifiable to opt for the morally problematic Option B. Consequently, we would expect a higher share of subjects opting to kill in the group treatment compared with Baseline, in which subjects know that they are pivotal for certain.

To fix ideas, we normalize the utility from receiving €10 to one and the utility from receiving €0 to zero. There is a subjective moral cost c_n associated with the death of $n = 1$ or 8 mice, respectively. Furthermore, we denote by $p_i \in [0, 1]$ the probability of subject i being pivotal. Clearly, if $p_i = 0$ it is optimal to select Option B and so there exists an equilibrium in which all players select Option B in case $n = 8$. From now on, we focus on the case of $p_i > 0$. If a subject chooses Option A and proves to be pivotal—i.e., killing is averted—utility is given by 0. Otherwise, the resulting level of utility is $-c_n$. The expected utility from choosing Option A therefore amounts to $-(1 - p_i)\, c_n$. The utility from choosing Option B is always given by $1 - c_n$. In making their decisions, deontological subjects disregard cost–benefit considerations and always choose Option A.[9] Any consequentialist subject chooses Option B if and only if the respective utility is at least as large as the expected utility from Option A or—equivalently—if $c_n \leq p_i^{-1}$.

[9] Alternatively, one could assume that deontologists take into account moral costs but always act as if they were deciding alone, i.e., they deliberately abstain from equilibrium considerations. Indeed, Kant's categorical imperative requires people to "[a]ct only in accordance with that maxim through which you can at the same time will that it become a universal law" (cited from Kant, 1996, p. 73). Deontologists would then choose Option B if $c_n \leq 1$ and Option A otherwise. Qualitative predictions are the same and changes for numerical results are minor. For a discussion of the differences between consequentialist vs. deontological reasoning, see Bénabou, Falk, and Tirole (2018a, 2018b).

Obviously, in the individual decision context, it holds that $p_i = 1$. By contrast, the chance of being pivotal in the simultaneous condition depends on the behavior of the other subjects in the same group.

This recursive relationship between subjects' decisions in Simultaneous can be understood as a strategic game between eight players whose types are characterized by their subjective moral costs c_8 and their respective moral conceptions, i.e., whether they are deontologists or consequentialists. Types are independently drawn, with the distribution F of moral costs c_8 having convex support and d denoting the probability of a subject following deontological ethics. Observe that if a consequentialist subject with subjective moral cost $\tilde{c}_8$ is indifferent between Options A and B (but—by assumption—chooses Option B), then any other consequentialist for whom $c_8 > \tilde{c}_8$ strictly prefers Option A. Conversely, any consequentialist subject with moral cost $c_8 < \tilde{c}_8$ strictly prefers Option B. Hence, any symmetric Bayesian equilibrium must have an equilibrium cut-off $k^\star \geq 1$ such that consequentialist subjects choose Option B if $c_8 \leq k^\star$ and Option A otherwise.[10] Consider a candidate k for such a cut-off. In conjunction with the distribution of types, it implies a probability of being pivotal, which is given by $p(k) = \{d + (1-d)[1 - F(k)]\}^7$. For any equilibrium cut-off that lies in the interior of the support of F, there exists a marginal subject for whom $c_8 = k^\star$ who needs to be indifferent between the two choice options. An interior equilibrium cut-off is thus a fixed point for which $k^\star = p(k^\star)^{-1}$, i.e.,

$$k^\star = \{d + (1-d)[1 - F(k^\star)]\}^{-7} . \quad (1)$$

As can be seen from the above equation, the equilibrium cut-off $k^\star$ is always weakly larger than one, the latter being the cut-off under individual decision making. The equilibrium cut-off is *strictly* larger than one as long as there exist any consequentialists ($d < 1$) of whom some have moral costs smaller or equal than one ($F(1) > 0$). Intuitively, some subjects choosing Option B even when fully pivotal reduce the likelihood of being pivotal for others, causing subjects with moral costs just above one to also choose Option B.

[10]It is straightforward to establish that any equilibrium must be symmetric. We sketch the proof next. Suppose otherwise, i.e., that $k_j > k_i$ for some pair of subjects i, j. Using that $p_i = \prod_{j \neq i} \{d + (1-d)[1 - F(k_j)]\}$, implies that $p_i < p_j$ and therefore $k_i = p_i^{-1} > p_j^{-1} = k_j$, a contradiction.

Depending on the precise distribution of moral costs and the prevalence of deontologists, this then also leads other subjects with still higher moral costs to adjust their behavior. The described moral unraveling is our first hypothesis.

Hypothesis 1. *The share of subjects choosing Option B—thereby taking €10 and agreeing to kill—will be higher in Simultaneous than in Baseline.*

It is worth noting that as long as—for each individual—moral costs c_8 of killing eight mice are higher than moral costs c_1 of killing just one mouse, we tend to underestimate the role of being less pivotal in groups relative to Baseline. We could have endowed groups only with one mouse. In this case, we would expect even larger treatment effects. We opted for eight mice, however, to keep the maximum possible extent of harm fixed at the aggregate level when comparing treatments.

In deriving Hypothesis 1, we have assumed that subjects hold perfectly rational and therefore identical beliefs, which constitutes a useful benchmark case. In reality, however, we expect that subjects have heterogeneous beliefs about the likelihood of being pivotal, which we elicit as part of our experimental procedure. According to the decision rule for consequentialists, this heterogeneity in beliefs should also translate into differences in decisions. This is our second hypothesis.

Hypothesis 2. *In the Simultaneous treatment, the likelihood that a given subject opts for taking €10 and killing the mice decreases with the subjective probability assigned to being pivotal.*

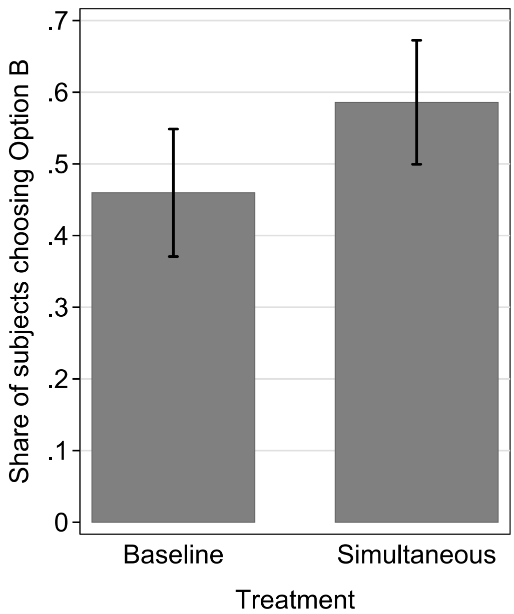
In sum, diffusion of being pivotal in groups leads consequentialist subjects to adjust their behavior. In equilibrium, p is small, making immoral behavior more attractive than when deciding individually. In addition, individual heterogeneity in the belief about p should translate into respective propensities to choose Option B. Hence, we expect that, on average, Option B is chosen more often in Simultaneous than in Baseline, and that—at the individual level—the likelihood of choosing Option B is inversely related to perceptions of being pivotal.

3 Results

We now present the results of our main experiment. We start with a treatment comparison and then provide evidence for the underlying mechanism at the level of both individual decision making and equilibrium behavior. For the latter, we estimate the subjective moral costs of killing eight mice, which we then use to derive welfare implications.

3.1 Choices and Beliefs

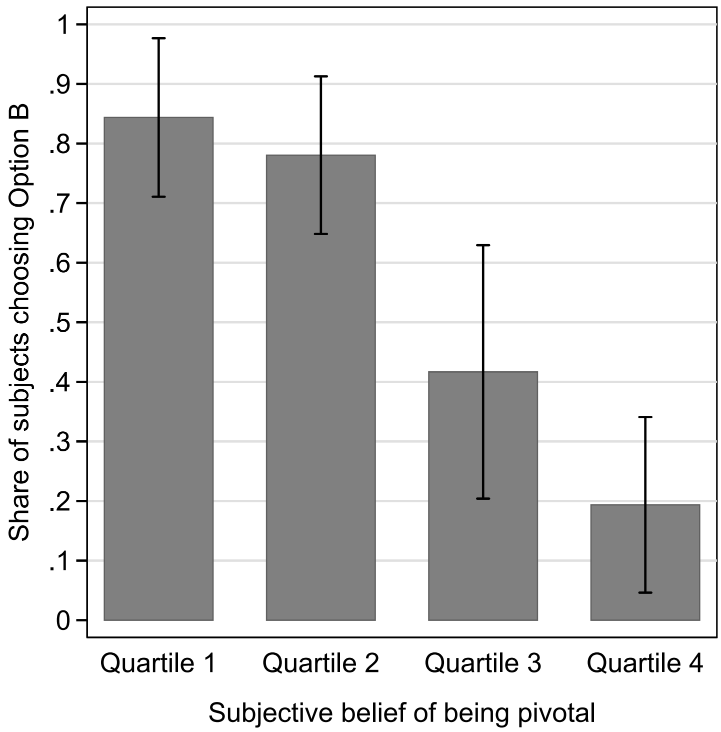
Our main result from the mouse experiment is shown in Figure 1, where we compare the shares of subjects choosing to kill in Baseline and Simultaneous, respectively. In Baseline, 46.0% of subjects choose Option B. In Simultaneous, the respective share is 58.6%, marking a difference of about 27%. This difference is significant ($p = 0.04$, two-sample test of proportions, two-sided) and confirms Hypothesis 1. At the aggregate level, the group impact is striking. While 46% of mice are killed in Baseline, *all* mice are killed in *all* groups in Simultaneous.



Note: Share of subjects choosing Option B in Baseline and Simultaneous. Error bars show 95% confidence intervals from OLS estimations using heteroscedasticity-consistent standard errors.

Figure 1: Treatment comparison

We have argued above that individual perceptions of being pivotal are critical in driving the increase in selfish behavior in Simultaneous. Accordingly, we should observe that an individual's willingness to choose Option B decreases with his belief of being pivotal. This is indeed what we find. Recall that we asked subjects about the probability that all other group members had chosen Option A (*belief_pivotal*). Figure 2 displays the fraction of subjects choosing Option B depending on this belief. The four categories in Figure 2 are based on quartiles of the belief distribution with respective percentage values of 0–3.5, 3.5–10, 10–35, and 35–100. In line with Hypothesis 2, the figure shows a clear negative relation between subjective perceptions of being pivotal and the likelihood of choosing Option B (Spearman rank correlation: -0.54, $p < 0.001$).[11]



Note: Share of subjects in Simultaneous choosing Option B depending on their belief of being pivotal. Error bars show 95% confidence intervals from OLS estimations using heteroscedasticity-consistent standard errors.

Figure 2: Belief quartiles (Simultaneous)

To integrate the above patterns in a structured way, we now apply the data to our formal equilibrium framework introduced in Section 2.2. Recall that with rational beliefs, the share of subjects choosing Option B would be given by $F(p^{-1})$ for consequentialists

[11] The values of *belief_pivotal*—which we use here—and those of the incentivized *belief_B* are strongly and significantly correlated (Spearman rank correlation: -0.63, $p < 0.001$).

and by zero for deontologists. To proceed, we need to make two assumptions about the distribution of moral costs.

Assumption 1. *The subjective moral costs of consequentialist subjects follow a log-normal distribution F, with log-costs having mean μ and standard deviation σ.*

Assumption 2. *Moral costs are independent of the perceived likelihood of being pivotal.*

In addition, we use the fact that subjects' beliefs about being pivotal are in fact heterogeneous. We can now write the probability of a given subject choosing Option B in terms of the cumulative distribution function of the standard normal distribution.

$$\mathbb{P}(\text{Option B}) = \begin{cases} \Phi\left(\frac{\ln\left(\textit{belief\_pivotal}^{-1}\right)-\mu}{\sigma}\right) & \text{for consequentialists} \\ 0 & \text{for deontologists} \end{cases} \tag{2}$$

Now, consider a finite mixture model with two latent classes, one capturing consequentialists and the other deontologists. For consequentialists, a probit model is estimated which regresses the likelihood of choosing Option B on the log of the inverse probability of being pivotal and a constant. For consequentialists, the probability of choosing Option B is always zero.

$$\mathbb{P}(\text{Option B}) = \begin{cases} \Phi\left[\beta_0 + \beta_1 \ln\left(\textit{belief\_pivotal}^{-1}\right)\right] & \text{for consequentialists} \\ 0 & \text{for deontologists} \end{cases} \tag{3}$$

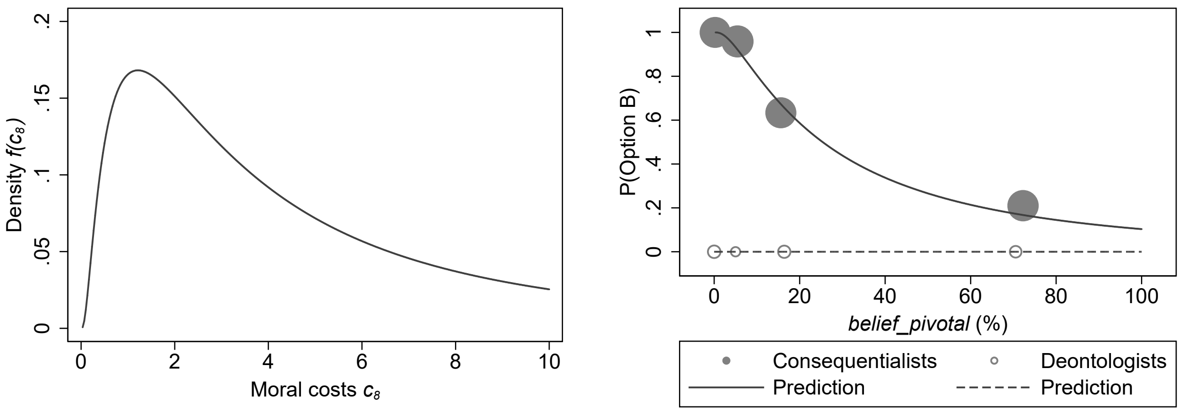
From Equations 2 and 3, it follows that

$$\sigma = \beta_1^{-1} \qquad \text{and} \qquad \mu = -\frac{\beta_0}{\beta_1}\,.$$

We estimate the finite mixture model given by Equation 3 using the expectation–maximization (EM) algorithm, assigning subjects to latent classes in terms of probabilities.[12] The invariance property of maximum likelihood estimates then allows us to

[12] If *belief_pivotal* is reported as 0%, we treat it as 0.1%.

convert the point estimates for coefficients into estimates for the parameters of F, as described above.



(A) Estimated moral costs **(B)** Observed vs. predicted decisions

Note: The left panel shows the estimated probability density function for moral costs c_8 in the Simultaneous treatment, denoted in multiples of the utility from receiving €10. The right panel plots the implied probabilities of choosing Option B for different beliefs about the chance of being pivotal against observed shares in the experiment. Subjects are partitioned into quartiles of the belief distribution and for each group weighted average beliefs and shares of Option B are calculated, separately for consequentialists and deontologists. Sizes of bubbles are proportional to estimates for the expected numbers of subjects. The solid line depicts the predictions for consequentialists, which are given by $F\big(belief\_pivotal^{-1}\big)$. Predictions for deontologists are given by zero and shown as the dashed line.

Figure 3: Moral costs (Simultaneous)

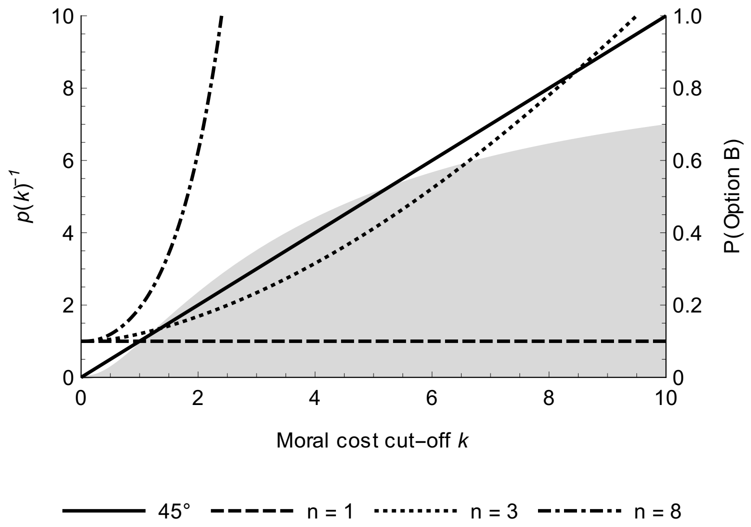
Figure 3 visualizes the results. The left panel shows the density function of moral costs c_8. The underlying estimates for the distributional parameters are $\hat{\mu} = 1.37$ and $\hat{\sigma} = 1.09$, corresponding to the mean and the standard deviation of log-costs, respectively. The expected value of moral costs is given by 7.098. We further estimate that the share of deontologists within our population of subjects is 13.6%, which is quite close to 18% of subjects choosing Option A despite being certain that they will not be pivotal.[13] The right panel uses these estimates to predict subjects' choices depending on their subjective beliefs about being pivotal for consequentialists (solid line) and deontologists (dashed line). Bubbles show observed choice probabilities by quartiles of the belief distribution, again separately for consequentialists and deontologists (solid and hollow, respectively). Deontologists never choose Option B. Consequentialists always choose Option B if they are certainly not pivotal but this probability decreases to 10.3% if they believe that they

[13]Of course, it may also be the case that some subjects made mistakes. In this sense, deontologists comprise all people whose choice behavior is unresponsive with respect to beliefs about being pivotal.

are pivotal for sure. If subjects were deciding individually—as in Baseline—but about the lives of eight mice rather than just one, these estimates imply that 8.9% of them would choose Option B, which is much lower than the observed 46.0% opting to kill in Baseline.

3.2 Equilibrium Implications and Welfare

We now use the estimated distribution of moral costs and the share of deontologists to inspect equilibrium implications. Figure 4 is a visual inspection of the equilibrium



Note: Dashed lines visualize the function $p(k)^{-1}$, which maps candidate cut-off values k into corresponding probabilities of being pivotal. Considered group sizes are one and three (both counterfactual), as well as eight. Values are given on the left axis. Fixed points, for which $p(k^\star)^{-1} = k^\star$, are intersections of the respective dashed line with the solid 45° line and mark Bayesian equilibria. The figure is calibrated using the estimates from Section 3.1. Numbers on the horizontal axis denote multiples of utility from receiving €10. The shaded area represents the cumulative distribution function $(1-d)\,F(k)$ of subjects choosing Option B. Values are given on the right axis.

Figure 4: Equilibria (Simultaneous)

condition given by Equation 1 in Section 2.2, calibrated according to the estimates from Section 3.1. The dashed lines show the inverse probabilities of being pivotal as functions of the cut-off value k for moral costs at which subjects switch from choosing Option B to Option A. To gain a better intuition, we show them for counterfactual group sizes of one and three, as well as for the actual group size of eight. Equilibria are intersections of the dashed lines with the solid 45° line. Thus, e.g., for $n = 3$ (and still assuming the life of eight mice being at stake), there would be two equilibria on the interval from zero to ten, one at 1.35 in which still only 14.0% of subjects would choose Option B and one at 8.46 in which 65.6% would choose Option B. For our actual case of $n = 8$, no equilibria exist on the considered interval.

Despite being individually rational, the behavior of subjects in Simultaneous is not welfare-optimal even when applying their own preferences. For quantifying welfare implications, we assume that consequentialists and deontologists share the same distribution of moral costs and treat utility from money as linear.[14] Then, the average moral costs of killing eight mice across all subjects is equivalent to €70.98. Nonetheless, all mice are killed. All of those subjects who choose Option B secure a monetary payoff of €10, so that the average utility in Simultaneous is equivalent to a loss of €65.12. If all subjects had chosen Option B, it would have been equivalent to a loss of €60.98. By contrast, if subjects were deciding alone, the utility would be weakly positive for everybody: all deontologists as well as those consequentialist subjects with moral costs above one would choose Option A and receive a utility of zero, while consequentialists with moral costs between zero and one would opt for killing and receive utility corresponding to the subjective excess of utility from €10 over the cost of killing. The average level of utility would thus be given by $(1 - d) \int_0^1 (1 - c)\, f(c)\, \mathrm{d}c \times €10$, which—according to our estimates—equals €0.31. Interestingly, a utility level of zero could also have been achieved in Simultaneous, had all subjects behaved as deontologists and saved the mice. This increased efficiency captures the intuition regarding why—from an evolutionary perspective—some degree of rule-based moral behavior could actually be expected (Alger and Weibull, 2013). However, our results point to a dominant role of consequentialist reasoning and question the relatively high fractions of Kantian types in survey data such as the trolley problem, where consequences are hypothetical rather than real.

[14] Moral costs for consequentialists can be interpreted as a lower bound for deontologists' moral costs.

4 Replication and Extensions

In this section, we employ a different setup. This second choice paradigm involves the binary decision between receiving €10 for oneself or donating €15 to a charity that supports children suffering from cancer. The charity treatments are essentially the same as in the mouse experiment, except that we use a different choice paradigm and study the role of experience as well as an additional sequential condition. As far as possible, we use the same design features, stake sizes (€10 for the selfish option), wording and framing of choice options. At the beginning of the experiment, subjects are made familiar with the charity, which is devoted to supporting children who suffer from cancer. In particular, the charity is engaged in psychological assistance and organizing leisure activities for children and their families, it helps with follow-up care and school-related issues, and supports parents and siblings as well as clinical research on cancer.

Charity treatments. To check the replicability of our experimental results from the mouse paradigm, we study a baseline (*BaselineC*, "C" for "charity") and a simultaneous group condition (*SimultaneousC*), analogous to the mouse conditions. In BaselineC, subjects make the binary decision to either donate €15 (Option A) or keep €10 for themselves (Option B).[15] In SimultaneousC, subjects are in groups of eight and simultaneously choose either Option A or Option B, respectively. Choosing Option B implies receiving €10 and choosing Option A receiving no money, irrespective of the choices of other group members. A donation of €120 (8 $\times$ €15) for the charity is only initiated if all group members choose Option A. If one group member or more choose(s) Option B, the donation of €120 is destroyed. To study how a dynamic setting affects the diffusion of responsibility, we further run treatment *SequentialC*. This treatment is identical to SimultaneousC (including payments, donation, wording, etc.), except that subjects choose sequentially. It is randomly determined at which position a subject is asked to decide, one subject being first, another second, up to position 8. Prior to making the binary decision (Option A or Option B), subjects are informed about their position (1 to 8) and about the previous

[15] Note that the design choice to donate €15 limits the plausibility of the argument that the €10 kept are spent on an alternative good cause.

choice history, i.e., how many subjects have previously chosen A and how many have opted for B. In both SimultaneousC and SequentialC, we also elicit beliefs analogous to Simultaneous in the mouse condition. Subjects are asked to indicate the probability that all other seven group members have chosen Option A. Responses are given in percent using a slider, with higher percentages reflecting a higher perceived likelihood of being pivotal for the respective subject (*belief_pivotal*).[16] We also ask subjects to estimate how many other subjects in their group have chosen Option B, with possible responses from 0 to 7 (*belief_B*). Correct answers are remunerated with €2.

To measure potential experience effects, all three conditions include a second round, which came to subjects as a surprise.[17] Subjects were told that they will make one more and final decision. In SimultaneousC and SequentialC, subjects learn whether at least one subject in their group has chosen Option B and thereby destroyed the donation, and that they will make the same decision in the same group of eight, as in the first round. In SequentialC, they also know that they act in the exact same order, i.e., each subject chooses at the same position as before. Payoffs and consequences are identical to the first round.

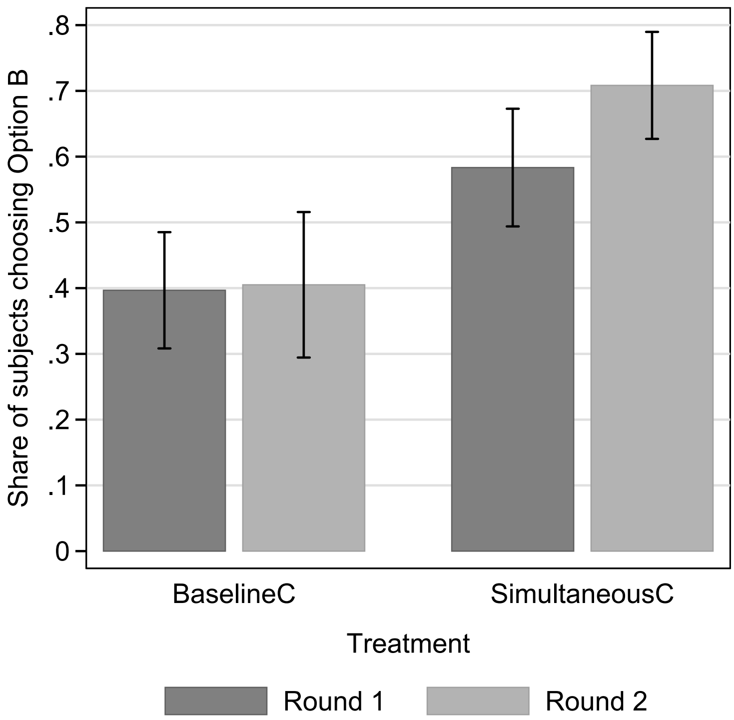
Charity procedures. 481 subjects—mainly undergraduate university students from all majors—took part in the experiments, 121 subjects in BaselineC, 120 in SimultaneousC and 240 in SequentialC (30 groups). Each subject participated in only one treatment condition. We used oTree as experimental software (Chen, Schonger, and Wickens, 2016). Subjects were recruited using the software ORSEE (Greiner, 2004). At the beginning of an experimental session, participants received detailed information about the rules and structure of the experiment. In all treatments, the experiment only started after all participants had answered several control questions correctly. The experiments were run at the BonnEconLab in March 2017. Subjects received a show-up fee of €10.

[16]Beliefs are elicited in the same way in SimultaneousC and SequentialC, but we note that in the latter, beliefs will depend on position and responses are affected by previous play, e.g., getting to know that Option B has already been chosen.

[17]Of our 121 subjects who took part in BaselineC, only 79 took part in an experience condition, i.e., in a second round. For the first two sessions (with 42 subjects) we only ran one round. In the analysis, we therefore either use 121 observations (Round 1) or 79 observations (Round 2), respectively.

4.1 Replication and Experience Effects

We begin by presenting the results for the two treatments that correspond to the ones in our main experiment. The main findings are summarized in Figure 5, which displays the share of subjects choosing Option B (not to donate) in conditions BaselineC and SimultaneousC, respectively. The dark bars show results from the first round, the light bars those of the second round (which was unexpected for subjects). Two observations can be made. First, we replicate the main result from the mouse experiment using

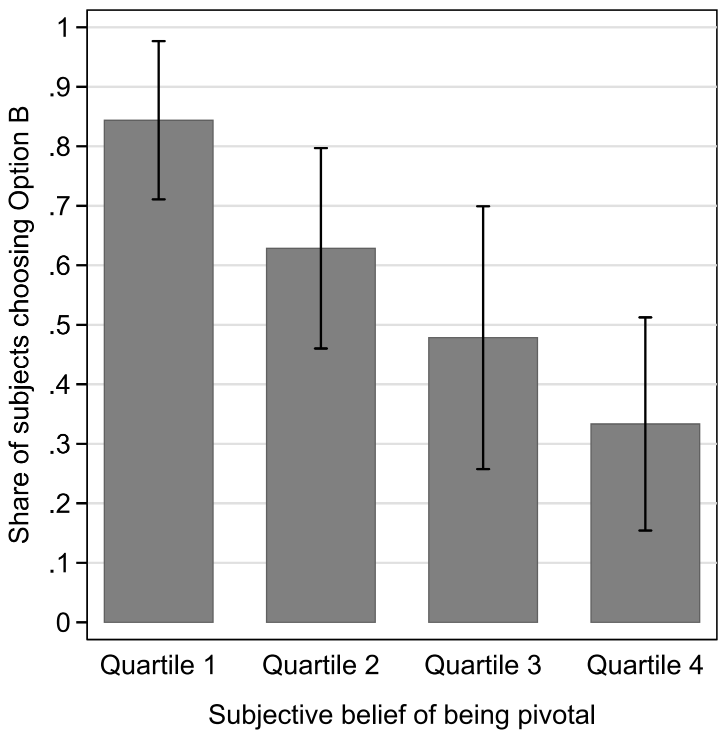


Note: Share of subjects choosing Option B in BaselineC and SimultaneousC, per round. Error bars show 95% confidence intervals from OLS estimations, where standard errors are clustered at the group level for the second round of SimultaneousC.

Figure 5: Comparison between BaselineC and SimultaneousC

a different choice paradigm. The share of subjects choosing Option B is significantly higher in SimultaneousC than in BaselineC, with means of 58.3% and 39.7%, respectively ($p = 0.004$, two-sample test of proportions, two-sided). The increase in selfish behavior amounts to 47.0%, which is higher than the respective increase in the mouse condition. At the aggregate level, no single group in SimultaneousC effectively donated. Second, the detrimental effect of group decision making on prosocial outcomes seems to increase with experience. Comparing the results between periods one and two reveals an increase in the likelihood of immoral choices upon learning the previous outcome of 12.5 percentage points ($p = 0.03$, see OLS estimates shown in Column 1 of Table 3 in Appendix A.1). In sharp contrast, moral behavior is not vulnerable to repetition in BaselineC, with an increase of Option B below one percentage point.

Analogous to the mouse experiment, we find that the association between the belief about being pivotal and choosing Option B is negative and statistically significant for SimultaneousC.[18] This relationship is shown in Figure 6, where we display the share of



Note: Share of subjects choosing Option B in the first round of SimultaneousC depending on the belief of being pivotal. Error bars show 95% confidence intervals from OLS estimations using heteroscedasticity-consistent standard errors.

Figure 6: Belief quartiles (SimultaneousC)

subjects choosing Option B depending on *belief_pivotal*. In SimultaneousC, among those who believe that they are not pivotal (estimated likelihood of being pivotal of 0%), again 18% (three out of 17) of subjects choose Option A, presumably reflecting a Kantian kind of moral reasoning.

[18]Again, both types of beliefs (*belief_B* and *belief_pivotal*) are significantly correlated (Spearman rank correlation: -0.35, $p < 0.001$).

In both treatments, we observe some subjects switching from one choice option to the other between rounds. In the case of SimultaneousC, this switching is asymmetric, as reflected by the higher share of subjects choosing Option B in the second round. If beliefs about being pivotal are important drivers of behavior, changes in beliefs should have predictive power for switching. In Table 1, we regress the choice in Round 2 on the

	Dependent variable: Option B in Round 2		
	OLS (1)	Probit (2)	Logit (3)
Option B in Round 1	0.324***	0.333***	0.322***
	(0.0989)	(0.0962)	(0.0973)
Decrease in *belief_pivotal*	0.00398**	0.00385**	0.00389**
	(0.00169)	(0.00162)	(0.00175)
Constant	0.498***		
	(0.0880)		
Observations	120	120	120
Clusters	15	15	15
R^2	0.131		

Note: Columns 2 and 3 report average marginal effects and average discrete changes due the binary choice in Round 1. Standard errors are clustered at the group level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 1: Switching behavior

choice in the first period and the change in the belief about being pivotal. There is a significant effect in the expected direction: subjects who consider themselves less pivotal in the second period than in Round 1 indeed become more likely to choose Option B in Round 2.

A possible concern in interpreting beliefs is the potential endogeneity of beliefs due to motivated reasoning (Epley and Gilovich, 2016; Gino, Norton, and Weber, 2016). To limit the problem, we incentivized beliefs about the number of other participants choosing Option B in the mouse and charity treatments, such that subjects could earn additional money for accurate estimates. However, we also ran an additional belief experiment with non-involved observers. In the belief experiment, participants read the original instructions of treatments in the charity experiment (avoiding textual redundancies). We then ask them for the probability that a subject is in a group in which all other seven group members choose Option A (*belief_pivotal*). If the percentage answer (*belief_pivotal*) is correct within an interval of plus/minus five percentage points, they receive €2. 87 subjects participated in this condition, which was programmed with oTree (Chen, Schonger, and Wickens, 2016) and run at the BonnEconLab in March 2017.[19] The beliefs of the spectators are not significantly different from those of active subjects in the first round of SimultaneousC ($p = 0.59$, Mann–Whitney U test, two sided). Active subjects' average beliefs are even slightly higher, suggesting that self-serving belief distortions do not play a dominant role in our main conditions. For details, see Appendix A.2.

To summarize, we replicate the main findings from the mouse condition. Subjects are less likely to choose the morally desired action in SimultaneousC than in BaselineC (pertaining to Hypothesis 1 from Section 2.2) and beliefs about being pivotal seem to be critical (pertaining to Hypothesis 2). In addition, we document that selfish outcomes in groups tend to increase with experience in contrast to individual decisions, further supporting the crucial role of beliefs about being pivotal.

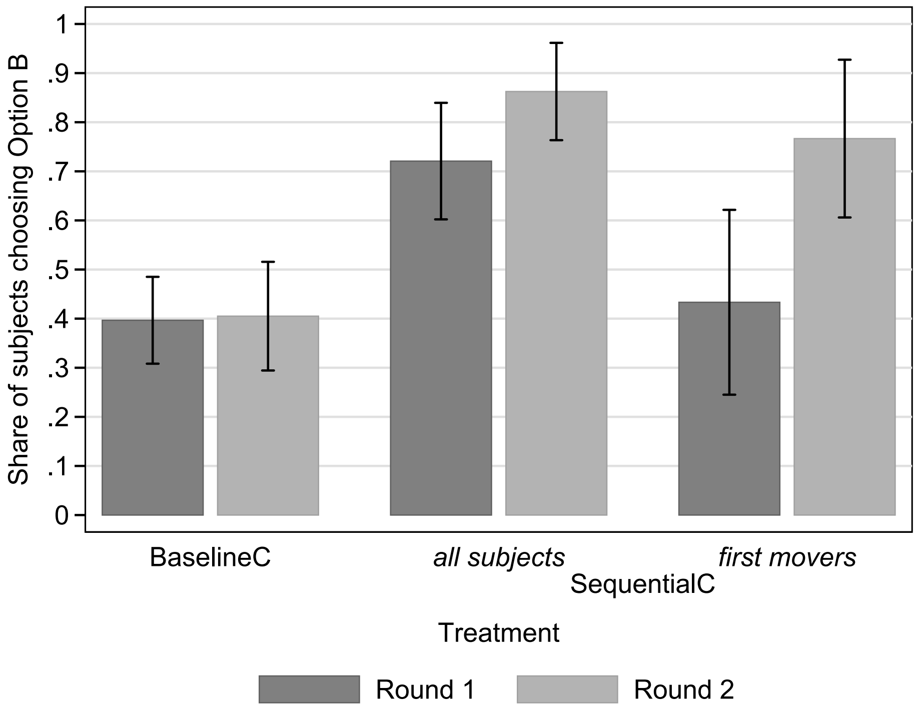
4.2 Sequential Decision Making

We now turn to the sequential decision making setup SequentialC. The central findings for this treatment are summarized in Figure 7. The overall share of participants choosing Option B in the first round of SequentialC is 72.1%, an increase of 81.7% relative to BaselineC. The difference between the two treatments is statistically significant ($p < 0.001$, see OLS estimates shown in Column 1 of Table 3 in Appendix A.1). This share increases by another 14.2 percentage points towards the second round ($p = 0.06$, see the same table).[20] At the aggregate level, in both rounds only two out of the 30 groups in

[19] Another interesting extension would be to investigate how behavior depends on different sources of being pivotal. In fact, there is evidence that endogenously determined probabilities resulting from choices of other group members ("social risk") can give rise to different behavior than probabilities determined by a correspondingly calibrated random device (see, e.g., Bohnet et al., 2008). In particular, if subjects in our experiment cared about fairness in relation to their fellow group members, they would potentially have additional reasons to act selfishly in the respective treatments: they could either wish to equalize their own monetary payoffs with the ones of other group members who are selfish, or they could feel "betrayed" if others did not cooperate in implementing the moral outcome. By contrast, if the probability of being pivotal was exogenously determined, social motives should be less relevant.

[20] Again, the two types of beliefs (*belief_B* and *belief_pivotal*) are significantly correlated (Spearman rank correlation: -0.65, $p < 0.001$).

SequentialC do not destroy the donation of €120.



Note: Share choosing Option B among subjects in BaselineC, all subjects in SequentialC and subjects in SequentialC who decide first in their groups, per round. Error bars show 95% confidence intervals from OLS estimations, where standard errors are clustered at the group level for both rounds of SequentialC.

Figure 7: Comparison between BaselineC and SequentialC

Acting in a chain renders the specific position within the decision process relevant. Subjects deciding first in their group are of particular interest, since in a certain sense they are in a similar situation as subjects in SimultaneousC. They have no information about others' behavior in the given round and the consequences of the moral choice Option A for them depend on the behavior of seven other subjects. In the first round, 43.3% of first movers choose Option B. Interestingly, this share is not significantly larger than in BaselineC ($p = 0.71$, two-sample test of proportions, two-sided). One can think of a number of plausible mechanisms contributing to this finding. First, the chance of being pivotal indeed seems to be higher for first movers in SequentialC than for subjects in SimultaneousC. Of the 17 cases in the first round where first movers choose Option A, two result in an actual donation. In light of the simple logic employed in Section 2.2, this might even be expected. Conditional on the donation not having been destroyed yet, choosing Option A becomes increasingly attractive the further down the line that a given subject decides, because the donation has to "survive" fewer remaining decisions. Subjects deciding at earlier positions should anticipate this recovery of moral behavior over positions, incentivizing them to preserve the donation themselves. Second, first movers overestimate their chance of being pivotal. The two surviving donations out of 17 cases where first movers choose Option A correspond to a likelihood of 11.8%, but first movers on average believe that it is 31.3%.[21] This could hint at exaggerated optimism regarding the possibility of acting as a prosocial role model (Gächter et al., 2012; Gächter, Nosenzo, and Sefton, 2013). Third, subjects in SequentialC who choose Option B first in their groups are strongly identified with destroying the donation even in a constellation where, in fact, the decision would not have altered the final outcome. This is because subjects in positions 2 to 8 make state-contingent choices, such that their counterfactual behavior remains unknown. In particular, a subject who has chosen Option B will almost certainly observe that all subjects deciding at later positions will do the same but will not know what they *would have done* otherwise. The last two points lose most of their power in the second round. The large majority of first movers who had chosen Option A in the first round will learn that the donation was destroyed, meaning that they have not been pivotal. If they were hoping to be role models, they will feel frustrated. If they did not want to take the blame for choosing Option B first, they will now have a good excuse. Indeed, in the second round the fraction of subjects who choose Option B increases to 76.7%, which is now significantly different to the second round of BaselineC ($p < 0.001$, two-sample test of proportions, two-sided). It thus seems that with experience, the diffusion of being pivotal erodes moral behavior also in the context with sequential decision making.

Of course, the points discussed above may to some degree also apply to subjects deciding on other positions. It is therefore informative to consider the dynamics of choice behavior in this treatment more broadly. In Table 2 we explore the role of position and choice history in a simple panel regression framework using both rounds. In Columns

[21] Note that for all 30 movers in the first round we also find that the belief of being pivotal and Option A are significantly correlated in the expected direction (Spearman rank correlation: -0.68, $p < 0.001$) for *belief_pivotal* and 0.80, $p < 0.001$ for *belief_B*).

	Dependent variable: Option B					
	Round 1			Round 2		
	(1)	(2)	(3)	(4)	(5)	(6)
Position (1–8)	0.0552***		-0.0121	0.0210*		0.00108
	(0.0139)		(0.0110)	(0.0115)		(0.00963)
Not destroyed		-0.626***	-0.607***		-0.458***	-0.154
		(0.0626)	(0.115)		(0.122)	(0.126)
Interaction			-0.0170			-0.126***
			(0.0278)			(0.0229)
Constant	0.473***	0.948***	1.013***	0.768***	0.968***	0.962***
	(0.0988)	(0.0228)	(0.0467)	(0.0798)	(0.0161)	(0.0572)
Observations	240	240	240	240	240	240
Clusters	30	30	30	30	30	30
R^2	0.0794	0.450	0.458	0.0196	0.313	0.435
Adj. R^2	0.0755	0.448	0.451	0.0155	0.310	0.428

Note: OLS regression coefficient estimates, with binary choice option (Option B: destroy donation vs. Option A: donate) as dependent variable. Data come from the SequentialC treatment. *Position* is the position in the move order from 1–8, *Not destroyed* is a dummy that is 1 if all subjects in the respective group have chosen Option A thus far, and *Interaction* is the interaction of the two above variables. Standard errors in parentheses are clustered at the group level (30 groups). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 2: Choice dynamics

1 and 4, we regress a participant's choice of Option B on his position. Descriptively, subjects are more likely to choose Option B the further down the line that they decide. In Columns 2 and 5, we regress Option B on a dummy indicating that no other group member has chosen Option B yet ("not destroyed"). The respective intercepts are close to one and show that, conditional on the donation having already been destroyed, almost all subjects choose Option B. The remaining subjects' decisions could reflect either a lack of attention or understanding (which is unlikely given the control questions and the prominent display of previous play on the decision screen) or a non-consequentialist notion of rule-based decision making. More importantly, in the first as well as in the second round, subjects react strongly to being potentially pivotal, as reflected in the negative and significant coefficients (Columns 2 and 5, respectively). In Columns 3 and 6, we combine position and history and also include the interaction of the two. Turning to Round 1 (Column 3), the coefficients for both position as well as the interaction are

insignificant, and the coefficient indicating that Option B has not yet been chosen is essentially identical to the one in Column 2. This suggests that subjects largely ignore their positions. Turning to the second round, this is no longer true. Now, the conditional probability of choosing Option B is generally high but decreases over positions. This can be interpreted as evidence of successful learning about moral types of subsequent subjects in the same group as well as about the imposed choice mechanism itself.

5 Conclusion

This paper has documented the deterioration of moral outcomes in response to diffusion of being pivotal. Simple organizational changes from an individual decision context to group conditions increase moral transgression at the individual and even more so at the aggregate level.

In our main experiment, subjects decide between taking €10 for themselves or killing mice. In Baseline, subjects decide individually about the life of one mouse. In Simultaneous, subjects decide in groups of eight about the lives of eight mice. A single subject is enough to bring about the killing. We observe a statistically significant increase from 46.0% choosing to kill in Baseline to 58.6% in Simultaneous. In the group setting, all mice are killed. Our second paradigm closely resembles that of the first experiment but replaces killing of mice with destroying charitable donations of €15 and €120 in the individual and group contexts, respectively. Analogously to the above comparison, we find a significant increase from 39.7% choosing the selfish option in BaselineC to 58.3% in SimultaneousC. To test for experience effects, we repeat the experiment in an unexpected second round. Repetition leaves the share in BaselineC virtually unchanged, while the share increases by another 12.5 percentage points in SimultaneousC. Using the charity paradigm, we also study a sequential context, in which eight subjects decide in a line and know whether the donation has already been destroyed. On average, 72.1% of subjects opt for destroying the donation and the share rises by another 14.2 percentage points towards the second round. Among subjects deciding first in their groups, 43.3% destroy the donation in the first round and 76.7% do so in the second round. Thus, with experience, immoral behavior also deteriorates for first movers in the sequential choice context.

Consequentialism and deontological ethics have been center stage in occidental moral philosophy for the last centuries. Empirical studies using the so-called trolley problem put forward by Philippa Foot (see also, e.g., Greene et al., 2004; Thomson, 1976)[22] have provided support for the relevance of both. However, the evidence highlights the importance of situational and emotional factors. In contrast to the trolley evidence—which uses hypothetical outcomes—subjects in our experiment face real consequences. In all of our group treatments, we elicit beliefs about being pivotal. Subjects consistently respond to notions of being pivotal and only few subjects appear to follow a Kantian conception. In Simultaneous, 18% of subjects who hold the belief that the chance of being pivotal is exactly zero choose Option A. In SimultaneousC, the respective share is again 18%. Finally, in SequentialC, of the 153 individuals for whom the group donation was already destroyed before, eight subjects (5.2%) nevertheless choose Option A. These numbers suggest the existence of deontological reasoning but they are quite low. Our findings thus question the relatively high fractions of Kantian types in survey data.

Using incentivized answers from non-involved observers, we show that there is no indication that subjects form or report self-servingly biased beliefs in an attempt to justify selfish behavior in our context.

Generally, we find that beliefs about being pivotal are too high. Had they been more realistic, the willingness to engage in selfish behavior may have been even more pronounced. In this sense, it is conceivable that repeated interactions with learning possibilities even further increase the likelihood of immoral outcomes, as we observe in the second round of our experiment using the charity paradigm. Overestimating one's sense of being pivotal could point to a human tendency to overestimate one's impact in general. This may well extend to other (non-moral) contexts and seems worth further investigating, e.g., in voting contexts (Duffy and Tavits, 2008). In this context, Quattrone and Tversky

[22]The quandary to be resolved in this problem is to follow either the deontologically warranted option (and not to throw a switch that will divert a trolley and kill one person) or the option preferred from a consequentialist perspective (killing the person to save five others).

(1984) argue and provide evidence that people use their own actions as prognostic for the behavior of others, therefore trying to "induce" others to behave in a desired way even when no causal impact can exist. Another possible reason for overestimating one's impact could come from a desire for meaning, self-attribution and determination, as well as for motivating action in general. Such a desire for self-efficacy is already known in the context of the so-called IKEA effect (e.g., Norton, Mochon, and Ariely, 2012).

While the focus of this paper is to highlight possible negative consequences of organizational design on moral behavior, the reverse inference is of course our main interest. Our findings suggest that organizations aiming to promote morality should reduce diffusion of being pivotal, and instead attribute individual responsibility to their members.

References

Alger, Ingela, and Jörgen W. Weibull. 2013. "Homo Moralis—Preference Evolution Under Incomplete Information and Assortative Matching". *Econometrica* 81 (6): 2269–2302.

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism". *Econometrica* 70 (2): 737–753.

Arendt, Hannah. 1963. *Eichmann in Jerusalem.* New York: Viking Press.

Bartling, Björn, and Urs Fischbacher. 2012. "Shifting the Blame: On Delegation and Responsibility". *Review of Economic Studies* 79 (1): 67–87.

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018a. *Eliciting Moral Preferences.* Mimeo.

— . 2018b. *Narratives, Imperatives, and Moral Reasoning.* NBER Working Paper 24798. Cambridge, MA: National Bureau of Economic Research.

Bohnet, Iris, Fiona Greig, Benedikt Herrmann, and Richard Zeckhauser. 2008. "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States". *American Economic Review* 98 (1): 294–310.

Bowles, Samuel. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions". *Journal of Economic Literature* 36 (1): 75–111.

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

Crawford, Neta C. 2007. "Individual and Collective Moral Responsibility for Systemic Military Atrocity". *Journal of Political Philosophy* 15 (2): 187–212.

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

Darley, John M. 1992. "Social Organization for the Production of Evil". *Psychological Inquiry* 3 (2): 199–218.

Deckers, Thomas, Armin Falk, Fabian Kosse, and Nora Szech. 2016. *Homo Moralis: Personal Characteristics, Institutions, and Moral Decision-Making.* Mimeo.

Duffy, John, and Margit Tavits. 2008. "Beliefs and Voting Decisions: A Test of the Pivotal Voter Model". *American Journal of Political Science* 52 (3): 603–618.

Epley, Nicholas, and Thomas Gilovich. 2016. "The Mechanics of Motivated Reasoning". *Journal of Economic Perspectives* 30 (3): 133–140.

Falk, Armin. 2017. *Status Inequality, Moral Disengagement and Violence.* CESifo Working Paper 6588. Munich: CESifo.

Falk, Armin, and Nora Szech. 2013. "Morals and Markets". *Science* 340 (6133): 707–711.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments". *Experimental Economics* 10 (2): 171–178.

Fischer, Peter, Joachim I. Krueger, Tobias Greitemeyer, Claudia Vogrincic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalena Wicher, and Martina Kainbacher. 2011. "The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies". *Psychological Bulletin* 137 (4): 517–537.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. "Individual Preferences for Giving". *American Economic Review* 97 (5): 1858–1876.

Gächter, Simon, Daniele Nosenzo, Elke Renner, and Martin Sefton. 2012. "Who Makes a Good Leader? Cooperativeness, Optimism and Leading-by-Example". *Economic Inquiry* 50 (4): 953–967.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2013. "Peer Effects in Pro-Social Behavior". *Journal of the European Economic Association* 11 (3): 548–573.

Gert, Bernard. 2012. "The Definition of Morality". In *Stanford Encyclopedia of Philosophy*, Fall 2012, ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically". *Journal of Economic Perspectives* 30 (3): 189–212.

Glover, Jonathan, and M. Scott-Taggart. 1975. "It Makes no Difference Whether or Not I Do It". *Proceedings of the Aristotelian Society, Supplementary Volumes* 49:171–209.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment". *Neuron* 44 (2): 389–400.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments". In *Forschung und wissenschaftliches Rechnen*, ed. by Kurt Kremer and Volker Macho, 63:79–93. GWDG-Bericht.

Hamman, John R., George Loewenstein, and Roberto A. Weber. 2010. "Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship". *American Economic Review* 100 (4): 1826–1846.

Kant, Immanuel. 1996. *Practical philosophy.* Ed. by Mary J. Gregor. Cambridge: Cambridge University Press.

Latané, Bibb, and John M. Darley. 1968. "Group inhibition of bystander intervention in emergencies". *Journal of Personality and Social Psychology* 10 (3): 215–221.

Lifton, Robert Jay. 1986. *The Nazi Doctors: Medical Killing and the Psychology of Genocide.* New York: Basic Books.

Milgram, Stanley. 2009. *Obedience to Authority: An Experimental View.* Reprint. New York: Harper Perennial Modern Classics.

Norton, Michael I., Daniel Mochon, and Dan Ariely. 2012. "The IKEA effect: When labor leads to love". *Journal of Consumer Psychology* 22 (3): 453–460.

Peck, Tom. 2016. "If we don't sell arms to Saudi Arabia, someone else will, says Boris Johnson". *The Independent* October 26.

Quattrone, George A., and Amos Tversky. 1984. "Causal Versus Diagnostic Contingencies: On Self-Deception and on the Voter's Illusion." *Journal of Personality and Social Psychology* 46 (2): 237–248.

Rothenhäusler, Dominik, Nikolaus Schweizer, and Nora Szech. 2018. "Guilt in voting and public good games". *European Economic Review* 101:664–681.

Serra-Garcia, Marta, and Nora Szech. 2018. *The (In)Elasticity of Moral Ignorance.* Working Paper Series in Economics 120. Karlsruhe: Karlsruher Institut für Technologie (KIT).

Sobel, Joel. 2010. *Do Markets Make People Selfish?* Mimeo.

Soltes, Eugene. 2016. *Why They Do It: Inside the Mind of the White-Collar Criminal.* New York: PublicAffairs.

Spranca, Mark, Elisa Minsk, and Jonathan Baron. 1991. "Omission and commission in judgment and choice". *Journal of Experimental Social Psychology* 27 (1): 76–105.

Thomson, Judith Jarvis. 1976. "Killing, Letting Die, and The Trolley Problem". *The Monist* 59 (2): 204–217.

Appendix A Further Details

A.1 Comparison of Charity Treatments and Rounds

	Dependent variable: Option B		
	OLS	Probit	Logit
	(1)	(2)	(3)
BaselineC, Round 2	0.00837	0.0217	0.0348
	(0.0712)	(0.184)	(0.295)
SimultaneousC, Round 2	0.125**	0.338**	0.551**
	(0.0583)	(0.157)	(0.257)
SequentialC, Round 2	0.142*	0.506*	0.888*
	(0.0746)	(0.276)	(0.492)
SimultaneousC	0.187***	0.472***	0.756***
	(0.0635)	(0.163)	(0.263)
SequentialC	0.324***	0.847***	1.368***
	(0.0726)	(0.205)	(0.339)
Constant	0.397***	-0.262**	-0.419**
	(0.0447)	(0.116)	(0.186)
Observations	920	920	920
Clusters	395	395	395
R^2	0.123		
AIC	1115.3	1065.5	1065.5

Note: Difference estimates for treatments in the charity experiment and for second-round effects within each treatment. BaselineC is the omitted category. Estimates refer to coefficients. Standard errors are clustered at the group level for the second round of SimultaneousC and for both rounds of SequentialC. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

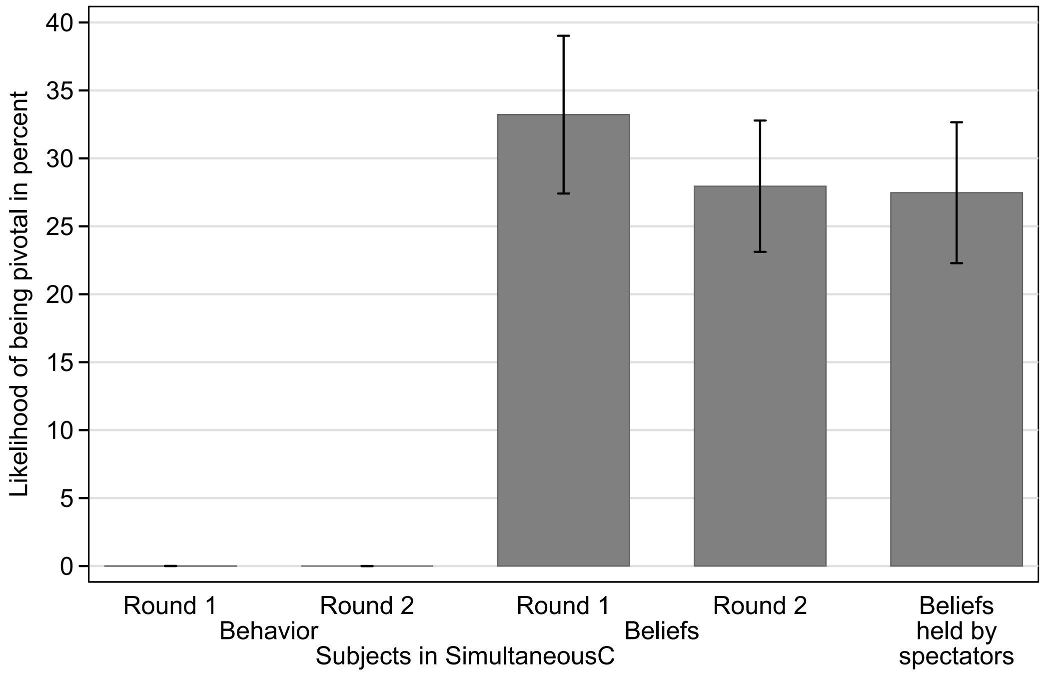
Table 3: Difference estimates

A.2 Validity of Beliefs

We report results from the belief experiment in which participants read the original instructions of treatments in the charity experiment (avoiding textual redundancies) and reported incentivized estimates corresponding to *belief_pivotal* and *belief_B* in SimulatenousC.

Figure 8 shows results concerning *belief_pivotal*. The actual probability for a subject to be in a group with all other seven group members choosing Option A was 0%, both in

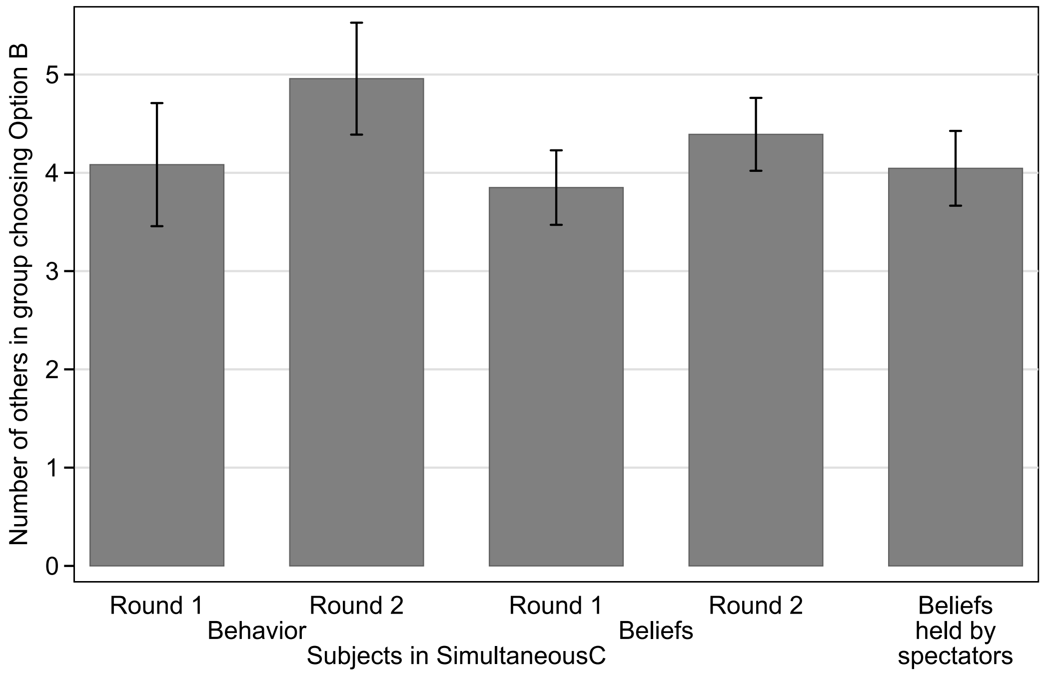
Rounds 1 and 2. In no single group, there were more than six subjects choosing Option A. A different way to estimate the actual probability of being pivotal is to use the whole distribution of choices and to calculate the likelihood—given the probability for Option A (41.7%)—of randomly being matched with seven group members who all choose Option A, which is 0.22%. This value is shown in the first bar and the analogous value of 0.02% for Round 2 in the second bar (the probability of Option A in the latter round is 29.2%). Bars 3 and 4 show subjects' beliefs for Rounds 1 and 2, respectively. It is obvious that subjects heavily overestimate how likely it is that they are pivotal. While the shown average beliefs hide a substantial amount of heterogeneity, almost all subjects perceive themselves as being pivotal with a higher likelihood than they actually are. Moving from Round 1 to Round 2, subjects adjust in the correct direction but still heavily overestimate their impact.



Note: Likelihood of being pivotal, i.e. the probability that all other seven members of a given subject's group choose Option A (in percent). Error bars show 95% confidence intervals from OLS estimations, where standard errors are clustered for the second round.

Figure 8: Belief comparison (*belief_pivotal*)

For *belief_B*, results are qualitatively quite similar to those concerning *belief_pivotal*. Figure 9 shows the actual behavior of subjects in Rounds 1 and 2 (first two bars), beliefs



Note: Number of other group members choosing Option B (0–7). Error bars show 95% confidence intervals from OLS estimations, where standard errors are clustered for the second round.

Figure 9: Belief comparison (*belief_B*)

in Rounds 1 and 2 (bars 3 and 4), as well as beliefs of spectators (fifth bar). The number of subjects choosing Option B increases from Round 1 to 2, which is reflected in changes in the beliefs of subjects. In contrast to *belief_pivotal*, however, subjects are overall much more accurate about actual outcomes.[23] Importantly, as for *belief_pivotal*, the beliefs of active subjects and spectators are not statistically significantly different (comparison of bars 3 and 5 in Figure 9; $p = 0.59$, Mann–Whitney U test, two sided).

Appendix B Instructions of the Mouse Experiment

Instructions have been translated from German.

[23]A possible explanation is that subjects found estimating absolute numbers easier than estimating a probability.

B.1 Baseline

Thank you very much for your participation!

For your participation you will in any case receive 20 euros. In the following you can earn an additional amount of money. At the end of the experiment you will receive your money in an envelope. Neither the other participants of the experiment nor the experimenter will be able to see how much money you have earned.

Please note: Throughout the whole experiment **communication between the participants is not allowed**. On the computer, please only use the functions intended to be used. If you have questions please raise your hand. Your question will then be answered at your cubicle!

Please note: **All statements made in these instructions are true.** This holds for all experiments carried out by the BonnEconLab, and also for this experiment. **In particular, all actions to be taken will be implemented exactly in the way they are described.** If you want to, you will be able to verify the correctness of all statements made in these instructions after the experiment.

In this experiment, there is a **Quiz A** and a **Quiz B**. Both, Quiz A and Quiz B, are simple trivia quizzes with questions from history, geography, sports, and so on. One example question could be: "Capital of Belgium?" There will, respectively, be four possible answers out of which one answer is correct. The posed questions in Quiz A and Quiz B are identical, that means, they are exactly the same regarding their difficulty. You will get three minutes to solve the quiz. The more questions you solve correctly, the more you can earn. **For each question that is answered correctly, you receive 5 cents.**

B.1.1 Details on the mouse

In this study, the life of a mouse is entrusted to your care. It is a healthy, young mouse, living with some other mice together in a small group. The expected lifetime of this mouse is approximately two years.

B.1.2 What is the difference between Quiz A and Quiz B?

Quiz A: In Quiz A, at the end of the experiment, you earn no additional money besides the **20 euros** for participation and the mouse stays alive.

Quiz B: In Quiz B, at the end of the experiment, you **get 10 euros in addition**. As another consequence, **the mouse will get killed**.

B.1.3 Details on the killing process

If you opt for the death of the mouse, the mouse is gassed. The gas flows slowly into the hermetically sealed cage. The gas leads to breathing arrest. At the point at which the mouse is not visibly breathing anymore, it remains in the cage for another 10 minutes. It will then be removed.

B.1.4 Summary

In Quiz A you earn no additional money, and the mouse does not get killed. In Quiz B, you earn additionally 10 euros, and the mouse gets killed. The decision is yours. You make your decision on a decision screen that will be shown as soon as you have answered the control questions on the following screen.

B.1.5 Control questions

In case of Quiz A: How many euros do you receive in addition? ____

Will a mouse be killed? ☐ Yes ☐ No

In case of Quiz B: How many euros do you receive in addition? ____

Will a mouse be killed? ☐ Yes ☐ No

B.2 Simultaneous

Introduction as in Baseline

In this study, the life of eight mice is entrusted to your group's care. These are healthy, young mice, living with some other mice together in a small group. The expected lifetime of these mice is approximately two years.

B.2.1 What is the difference between Quiz A and Quiz B?

In the following we describe the consequences of choosing Quiz A and Quiz B. The choice options and consequences are identical for all eight group members.

Quiz A: In Quiz A, at the end of the experiment, you earn no additional money besides the **20 euros** for participation. This holds for all group members. Each group member who chooses Quiz A receives no additional money.

Quiz B: In Quiz B, at the end of the experiment, you **get 10 euros in addition**. This holds for all group members. Each group member who chooses Quiz B receives 10 euros in addition.

Another consequence is that **eight mice get killed if at least one member of your group of eight chooses Quiz B**. Thus if in total one member of the group, or two, three, four, five, six, seven or eight group members choose Quiz B, eight mice get killed.

Only if no member in your group of eight chooses Quiz B, the mice will not get killed.

B.2.2 Details on the killing process

If your group opts for the death of the mice, these will be gassed. The gas flows slowly into the hermetically sealed cage. The gas leads to breathing arrest. At the point at which the mice are not visibly breathing anymore, they remain in the cage for another 10 minutes. They will then be removed.

B.2.3 Summary

In Quiz A you earn no additional money. In Quiz B, you earn additionally 10 euros. Whether the mice get killed depends on whether at least one member of your group of eight has chosen Quiz B. You make your decision on a decision screen, which will be shown as soon as you have answered the control questions on the following screen.

Control questions and video

Appendix C Instructions of the Charity Paradigm

Instructions have been translated from German.

C.1 BaselineC

Welcome and thank you very much for your interest in today's experiment!

This experiment is part of a research project of the *Bonner Laboratorium für experimentelle Wirtschaftsforschung (BonnEconLab).*

For your participation you will in any case receive €10.00, which will be handed to you in cash today at the end of the experiment. During the experiment, you will make decisions on the computer. Depending on how you decide, you can earn additional money.

During the experiment, it is not allowed to communicate with other participants. Also note that the curtain of your cubicle has to be shut throughout

the entire experiment. Please now switch off your mobile phone, to make sure that other participants are not being disturbed. On the computer, please only use the functions intended to be used and make all inputs using either the mouse or the keyboard. If you have questions, please contact the conductor of the experiment. To do so, please stick your hand out of the cubicle.

All statements made in this experiment are true. This holds for all experiments carried out by the BonnEconLab, and also for this experiment. **In particular, all actions to be taken will be implemented exactly in the way they are described.** If you want to, you will be able to verify the correctness of all statements made in these instructions after the experiment.

In what follows, we will first ask you to answer a question regarding your mood. Subsequently, the decisions you will have to make will be explained in detail.

C.1.1 How is your current mood?

Please give an answer to this on the following scale from 0 to 10.

0 means that your mood is very bad.

10 means that your mood is very good.

You can choose any integer number on the scale from 0 to 10 to express your current mood.

C.1.2 The donation

This experiment is about a donation to the *Förderkreis für krebskranke Kinder und Jugendliche e. V.*, a regional charity from Bonn.

Every participant, that means also you, will first be entrusted with a donation which will be made to the *Förderkreis für krebskranke Kinder und Jugendliche e. V.* after today's experiment.

During the experiment, you will make decisions which affect this donation. Moreover, the information which follow are also relevant for your personal payoff from this experiment.

Therefore, please carefully read the following instructions. In particular, make sure that you understand all decisions you can make as well as their potential consequences.

C.1.3 Information about the ***Förderkreis***

The *Förderkreis*. The *Förderkreis für krebskranke Kinder und Jugendliche e.V.* **supports young people suffering from cancer and their families comprehensively in dealing with the disease**. The society is committed to psychological support, to organizing free time activities, as well as to aftercare and to supporting children and adolescents with school. Moreover, indirectly affected individuals like parents and siblings are extensively supported. This takes, for example, the form of a specifically established home for parents and of pedagogic support. Moreover, the *Förderkreis* supports clinical research on cancer.

Projects and tasks of the *Förderkreis*.

- *Klassissimo* school project: offers participation in school lessons using Skype
- *Bärenstark*: support of families at home
- Psychosocial and psychooncological counseling of patients and relatives
- Pedagogic support at the hospital department
- Start-up financing for new positions and financing of specific training of departments' staff.
- Financing of hospital clowns and music therapy
- Aftercare
- Support of clinical research on cancer

C.1.4 Your decision

The donation. You are entrusted with a donation of €15.00, which is supposed to be made to the *Förderkreis für krebskranke Kinder und Jugendliche*

***e. V.* following today's experiment. Whether this amount will in fact be transferred to the *Förderkreis* at the end of the experiment depends on the decisions that you will make.**

Anonymity. No other participant in this experiment can see your decisions. The subsequent analysis of all data is done anonymously, such that all your decisions cannot be linked to your identity anymore.

You can choose between two options: **Option A** and **Option B**. **Depending on which of both options you choose, you can earn different amounts of money. Additionally, depending on which option you choose, consequences differ for the donation of €15.00 which was described above.**

In what follows, the consequences associated with choices of **Option A** and **Option B**, respectively, will be described.

Option A. If you choose **Option A**, besides €10.00 for participation you will receive **no** additional money at the end of the experiment.

Option B. If you choose **Option B**, you will **additionally** receive €10.00 at the end of the experiment.

As a further consequence, **the previously described donation of €15.00 will be destroyed**.

Summary. If you choose **Option A**, you do not receive an additional payment and the donation will not be destroyed. If you choose **Option B**, you additionally receive €10.00 and the donation is destroyed. The decision rests with you.

You make your decision on a decision screen, which will be shown as soon as you have answered the control questions on the following screen.

C.1.5 Control questions

In case of Option A. How many euros do you receive in addition? ____

Will the donation be destroyed? ☐ **Yes** ☐ **No**

In case of Option B. How many euros do you receive in addition? ____

Will the donation be destroyed? ☐ **Yes** ☐ **No**

C.1.6 Your decision

Please now choose between **Option A** and **Option B**.

I choose: ☐ **Option A** ☐ **Option B**

C.1.7 Result

If Option A was chosen: You have decided **not to destroy** the donation.

Therefore, a donation of €15.00 to the *Förderkreis für krebskranke Kinder und Jugendliche e. V.* will be made for you by the BonnEconLab.

If Option B was chosen: You have decided **to destroy** the donation.

Therefore, **no donation will be made**.

C.1.8 Experiment 2

Now follows a second experiment. This experiment is the last experiment. Your final payoff comprises of €10.00 for participation in the experiment, your decision in the first experiment, and, independently, on how you decide in the second experiment.

The decision in the second experiment is the same as in the first experiment. Thus, you can again choose between **Option A** and **Option B**, i.e., you can decide whether a donation will be destroyed or not. The donation is again a donation to the *Förderkreis für krebskranke Kinder und Jugendliche e. V.*

C.2 SimultaneousC

Introduction as in BaselineC

C.2.1 Your decision

Your group. You are together with 7 other participants of today's experiment in a group of 8 people. Your group members have been allotted to you at the beginning of the experiment. You will at no point learn which participant is in your group.

Note: **You are making all decisions within this experiment autonomously and independent of the other members of the group.** The consequences of your decisions can depend on decisions of other group members. On the following screens, all decisions, alternatives, and consequences will be introduced and explained in detail.

The donation. Your group is entrusted with a donation totaling €120.00, which is supposed to be made to the *Förderkreis für krebskranke Kinder und Jugendliche e. V.* following today's experiment. Whether this amount will in fact be transferred to the *Förderkreis* at the end of the experiment, depends on the decisions that you and the other members of your group will make.

Anonymity. No other participant in this experiment can see your decisions. This is also true for the other members of your group. The subsequent analysis of all data is done anonymously, such that all your decisions cannot be linked to your identity anymore.

You can choose between two options: **Option A** and **Option B**. **Depending on which of both options you choose, you can earn different amounts of money. Additionally, depending on which option you choose and which options the other participants of your group choose independently, consequences differ for the donation of €120.00 which was described above.**

In what follows, the consequences associated with choices of **Option A** and **Option B**, respectively, will be described. The choices and the consequences are the same for all 8 participants in your group.

Option A. If you choose **Option A**, besides €10.00 for participation you will receive **no** additional money at the end of the experiment.

This holds for all group members: Each group member who chooses **Option A** receives no additional money.

Option B. If you choose **Option B**, you will **additionally** receive €10.00 at the end of the experiment.

This holds for all group members: Each group member who chooses **Option B** additionally receives €10.00.

As a further consequence, **the previously described donation of €120.00 will be destroyed if at least one of the 8 members of your group chooses Option B**. Thus, if one group member, or if two, three, four, five, six, seven, or eight group members decide for **Option B**, the donation is destroyed. **Only if none of the 8 members of your group chooses Option B, the donation will not be destroyed.**

Summary. If you choose **Option A**, you do not receive an additional payment. If you choose **Option B**, you additionally receive €10.00. Whether the donation to the *Förderkreis* is destroyed depends on whether at least one of the 8 members of your group has chosen **Option B**.

C.2.2 Decisions of participants in your group

Note: The consequences of your choice do not just depend on you but also on the decisions of the other 7 members of your group. This holds in particular for the execution of the donation to the *Förderkreis für krebskranke Kinder und Jugendliche e.V.*: **Only if none of the members of your group has chosen Option B, the donation of €120.00 is made.**

You and the other 7 members of your group decide **simultaneously**. After all group members have made their decision, you learn whether the donation will be made.

At the end of today's experiment, you will also learn how many members of your group have in total chosen **Option A** and how many members of your group have in total chosen **Option B**.

You make your decision on a decision screen, which will be shown as soon as you have answered the control questions on the following screen.

C.2.3 Control questions

Suppose, two/one/no/six group member(s) choose(s) Option B.

You choose Option A: How many euros do you receive in addition? ____

Will the donation be destroyed? ☐ **Yes** ☐ **No**

You choose Option B: How many euros do you receive in addition? ____

Will the donation be destroyed? ☐ **Yes** ☐ **No**

C.2.4 Your decision

Please now choose between **Option A** and **Option B**.

I choose: ☐ **Option A** ☐ **Option B**

C.2.5 What do you estimate?

How likely is it in your opinion that all other group members have chosen **Option A**? *Please enter a probability (from 0 to 100 percent):* [Slider]

What do you think, how many of the other 7 group members have chosen **Option B**? If you estimate the correct number, you will additionally receive €2.00. *Enter a number between 0 and 7:* ____

C.2.6 Result

If Option A was chosen: You have decided **not to destroy** the donation.

In your group, at least one participant has decided **to destroy** the donation. The donation over €120.00 from you and the other members of your group will therefore **not be made**.

You have not made a correct estimation and therefore do not receive any additional payoff.

C.2.7 Experiment 2

Now follows a second experiment. This experiment is the last experiment. Your final payoff comprises of €10.00 for participation in the experiment, your decision in the first experiment, and, independently, on how you decide in the second experiment.

The decision in the second experiment is the same as in the first experiment. Thus, you can again choose between **Option A** and **Option B**, i.e., you can decide whether a donation will be destroyed or not. The donation is again a donation to the *Förderkreis für krebskranke Kinder und Jugendliche e.V.*

Please note: You are in the same group of 8 participants as in the first experiment.

C.3 SequentialC

Introduction as in SimultaneousC

C.3.1 Decisions of participants in your group

Note: The consequences of your choice do not just depend on you but also on the decisions of the other 7 members of your group. This holds in particular for the execution of the donation to the *Förderkreis für krebskranke Kinder und Jugendliche e.V.*: **Only if none of the members of your group has chosen Option B, the donation of €120.00 is made.**

You and the other 7 members of your group decide **one after the other**. Your position is randomly determined by a computer.

When it is your turn, you will learn whether among the people who have decided before you, someone has already chosen Option B. You will also learn your position within the sequence. Moreover, you will learn how many members of your group have already chosen **Option A** and how many members of your group have already chosen **Option B**. At the end of today's experiment, you will also learn how many members of your group have in total chosen **Option A** and how many members of your group have in total chosen **Option B**.

Please note: If another participant in your group has already decided for **Option B** before it was your task, this means that the donation has already been destroyed. Thus in this case, your decision has **no** effect any more on whether the donation is made.

Control questions as in SimultaneousC.

C.3.2 Your decision

You are on **position 1** in the order of your group. Consequently, no other member in your group has made a decision yet.

Or:

You are on **position 2** in the order of your group. Consequently, 1 group member has already made a decision.
Of the 1 group members who have decided before you, 1 has decided for **Option A** and 0 for **Option B**.

Or:

You are on **position 3** in the order of your group. Consequently, 2 group members have already made a decision.
Of the 2 group members who have decided before you, 1 has decided for **Option A** and 1 for **Option B**.
Thus, the donation has already been destroyed.

Please now choose between **Option A** and **Option B**.

I choose: ☐ **Option A** ☐ **Option B**

Remaining instructions as in SimultaneousC.

C.4 Belief Experiment

In the belief experiment, participants read the original instructions (avoiding redundancies, however), learn how many subjects have taken part in the respective treatment, and *are then asked to answer the following questions.*

BaselineCB:

- How likely do you think it is that a randomly chosen participant of the just described experiment decides for Option A, i.e., not to destroy the donation?

SimultaneousCB:

- How likely was it for a participant in the experiment to be in a group in which all other 7 group members choose Option A? (answer in percent)
- How likely do you think it is that the donation is not destroyed in such a group in the end, i.e., that all 8 group members choose Option A. (answer in percent)
- Please imagine you are in the new situation at the BonnEconlab which was just described. What do you think: How many of the other 7 members of your group have decided for Option B, i.e., to destroy the donation?

SequentialCB:

- How likely was it for a participant in this experiment to be in a group in which all other 7 group members choose Option A? (answer in percent)
- How likely do you think it is that the donation is not destroyed in such a group, i.e., that all 8 group members choose Option A? (answer in percent)

Please now imagine yourself being in the situation of a participant in the described experiment at the BonnEconLab.

- Imagine, you decide first and choose Option A. How many of the other 7 group members do you think also choose Option A, such that the donation is not destroyed? (answer in percent)
- Imagine, the member at position 1 in your group chooses Option A. You decide second and also choose Option A. How likely do you think it is that all further 6

people in the group also choose Option A, such that the donation is not destroyed? (answer in percent)

- Imagine, the members at positions 1 to 3 in your group all choose Option A. You decide as the fourth and also choose Option A. How likely do you think it is that all further 4 people in the group also choose Option A, such that the donation is not destroyed? (answer in percent)

- You decide last, i.e., as the eight. How likely do you think it is that all 7 before you have chosen Option A? (answer in percent)

- Please again imagine yourself being in the situation of the described experiment at the BonnEconLab. You decide first. What do you think: How many of the 7 other members of your group decide for Option B, i.e., for destroying the donation?

- Now, please imagine that you decide last in your group, i.e., as the eighth. All 7 group members before you have chosen Option A. Would you then choose Option A or Option B? *(unincentivized)*

- How likely do you think it is that a participant in the just described situation – decided last, all group members before have chosen Option A – also has chosen Option A? (answer in percent)