Human Capital and Economic Opportunity Global Working Group

Working Paper Series

Working Paper 2016-013

Diffusion of Being Pivotal and Immoral Outcomes

Armin Falk
Nora Szech

May, 2016

Human Capital and Economic Opportunity Global Working Group
Economics Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
www.hceconomics.org

Diffusion of Being Pivotal and Immoral Outcomes

Armin Falk[1] and Nora Szech[2]

September 2014

Abstract. We study how diffusing being pivotal affects the willingness to support immoral outcomes. Subjects decide about agreeing to kill mice and receiving money versus objecting to kill mice and foregoing the monetary amount. We investigate an exogenous diffusion of being pivotal imposed by organizational design as well as self-imposed, endogenous diffusion of being pivotal. Regarding exogenous diffusion, we compare two treatments. We keep overall financial incentives and overall payoff consequences identical, yet vary the decision rule: In Baseline subjects decide individually about the life of one mouse. In the Exogenous Diffusion treatment, subjects are organized into groups of eight. Eight mice are killed if at least one subject supports the killing. The fraction of subjects agreeing to kill is significantly higher in Exogenous Diffusion than in Baseline. Moreover, in Exogenous Diffusion, the likelihood to agree to the killing decreases in subjective perceptions of being pivotal. We then show that many subjects actually have a preference to actively create a situation where being pivotal is diffused. In the Endogenous Diffusion treatment, each subject chooses the probability of killing a mouse. The monetary amount a subject receives is proportional to the killing probability. More than 30 percent of subjects opt for intermediate killing probabilities, thereby actively diffusing being pivotal at a proportional reduction of money. Response times and feelings of remorse and bad conscience suggest that it is in particular subjects experiencing moral conflict who prefer diffusing being pivotal. Presumably, this serves as a means to keep a positive self-image while behaving selfishly.

Keywords. Diffusion of being pivotal, morality, replacement logic, self-image, response times
JEL. C91, D01, D03, D23, D63

We thank K. Albrecht, S. Altmann, B. Köszegi, R. Bénabou, T. Dohmen, J. Engel, D. Engelmann, S. Gächter, P. Heidhues, D. Huffman, S. Jäger, F. Kosse, F. Krämer, G. Loewenstein, F, Rosar, J. Sobel, F. Zimmermann and participants at various seminars for helpful comments. We acknowledge financial support by the German Science Foundation (DFG) through the Leibniz Program. The study was approved by Ethical Committee of the University of Bonn, Reference Number: 066/12.

[1]Center for Economics and Neuroscience, University of Bonn, armin.falk@uni-bonn.de
[2]Chair of Political Economy, Institute of Economics, Karlsruhe Institute of Technology, and Berlin Social Science Center (WZB), nora.szech@kit.edu

1. Introduction

This paper studies how diffusing whether an agent is pivotal causes moral transgression. Intuitively, diffusing being pivotal provides an excuse for engaging in immoral activities simply because an actor may perceive himself as irresponsible or as only partly responsible for the outcome. A striking example how institutions facilitate engaging in a morally difficult activity is the practice of firing squads, which typically consist of a group of executors rather than a single person. From an individual member's perspective being pivotal is diffused as many people shoot at the same time implying that the killing is likely to happen, regardless of whether a particular member does or does not fire his gun. Moreover, members of firing squads are often randomly issued a gun containing a blank cartridge. This diffuses being pivotal in another respect. Even if a member of the squad shoots his gun, he remains uncertain whether he can effectively cause the killing at all. Apparently, these features facilitate participating in executions.

We first investigate how being pivotal is diffused due to the specific details of an exogenously given institutional environment. To this end we vary organizational design and contrast a situation in which subjects are fully pivotal with one in which being pivotal is exogenously diffused. In the latter, subjects are organized into groups of eight and individual decisions are aggregated such that the individual can easily believe that his or her decision is not pivotal. In a second step, we analyze whether subjects exhibit a preference for actively diffusing being pivotal. To this end, we study behavior in an individual decision context in which subjects can determine how likely their decision becomes pivotal, at a proportional reduction of payment. We study the impact of exogenous and endogenous diffusion in isolation. It is likely, however, that both effects reinforce each other, e.g., if people deliberately self-select into environments that dilute being pivotal.

Organizing people into groups and implementing a decision rule that does not require the support of all members for immoral action to happen enables a simple "replacement logic" (Sobel, 2010). It allows each single actor to believe that even if he or she does not agree to engage in a morally questionable activity, others will, such that the immoral outcome happens anyway. This diffusion of being pivotal is pervasive at all levels of social interaction, such as organizations and markets.

A particularly striking, and sad, example for the potential to enforce evil outcomes is the organization of the Holocaust (Darley, 1992; Arendt, 1963; 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" between prisoners who would be allowed to live and those who would be gassed right away. 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 for the obvious evil was that the "horrible machinery would go on", regardless of whether or not a particular doctor would continue to participate. Replacement arguments suggesting the impossibility to stop ongoing moral crime were also used as excuses in the Nuremberg Trials for having participated various kinds of atrocity under the Nazi Regime (see, e.g., Crawford, 2007 and the references therein).

Replacement arguments also help explaining outcomes in markets that are violating traders' own moral or fairness preferences (Sobel, 2010). Here replacement prevails if traders prefer concluding a trade themselves to letting another trader perform the same transaction, even if trading creates unfair outcomes for traders themselves, or negative externalities on others. In cases where buying decisions create negative externalities, a frequently made "excuse" is that "if I don't buy, another buyer will".

The exogenous variation of being pivotal through organizational design or market interactions is potentially complemented and reinforced by a preference for self-imposed, i.e., endogenous, diffusion of being pivotal. This is the case if decision makers actively self-select into organizations that allow hiding behind uncertain outcomes. To study both channels, exogenous and endogenous diffusion, and their impact on immoral outcomes in isolation, we ran a series of laboratory experiments. In all treatments, our experimental paradigm to model moral decision-making is the trade-off between life and money. Subjects decided between receiving money and agreeing to kill mice versus not receiving money and objecting to the killing.[1] Importantly, mice used in the experiment were so-called "surplus" mice, which would have all been killed without our intervention (see Section 2). Subjects learned about this default in a post-experimental debriefing. The paradigm is informed by the

[1] The study was approved by the Ethics Committees of the University of Bonn and of the Karlsruhe Institute of Technology.

widely held consensus that harming others in an unjustified and intentional way is considered as immoral. While there exists no universal consensus about how to define the content of morality, avoiding and preventing harm is a central element according to most notions of morality.[2]

To study the role of exogenous diffusion of being pivotal, we contrast two treatments: The Baseline treatment implements a simple binary choice according to which subjects either receive zero euros for saving a mouse (option A) or 10 euros for killing the mouse (option B). In Baseline subjects are fully pivotal. This condition therefore serves as a comparison benchmark for the Exogenous Diffusion treatment. In the latter, eight subjects simultaneously decide between option A and option B. As in Baseline a subject receives zero euros for choosing option A and 10 euros for choosing option B, irrespective of the other subjects' choices. If at least one subject chooses option B, however, *eight* mice are killed. Thus, if a subject believes that it is likely that at least one other subject chooses option B, he may no longer consider himself as pivotal. A low chance of being pivotal provides a potential excuse to choose option B, as choosing B is unlikely changing the outcome but guarantees a payoff of 10 euros.[3]

We find that the fraction of subjects choosing option B is significantly higher in Exogenous Diffusion than in Baseline, despite the fact that killing implies the death of eight mice rather than one mouse. Moreover, the likelihood that a subject chooses to kill mice is decreasing in his or her belief of being pivotal. At the aggregate level all mice were killed in Exogenous Diffusion. From an organizational "efficiency" point of view, killing mice in the latter treatment was also "cheaper" than in Baseline. The cost for killing one mouse was ten euros in Baseline, but less than six euros in Exogenous Diffusion. These results demonstrate the malleability of moral behavior. In particular they show the power of organizational design to causally promote immoral behavior and outcomes. The "same" people produce different levels of moral transgression when acting in different institutions. Our findings thus contribute to

[2] See, e.g., Bernard (2012) 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."

[3] This excuse is more plausible according to a utilitarian rather than a deontological moral conception, see the next section for a discussion.

understanding why "ordinary" people endowed with given moral values may engage in activities they would generally object to.

To study the role of self-imposed or endogenous diffusion of being pivotal we ran a further individual treatment. Using the same choice paradigm, in the Endogenous Diffusion treatment each subject chooses death-probabilities between 0 percent and 100 percent for one mouse. If a subject opts for zero percent, the mouse is saved with certainty. Likewise, if he or she opts for 100 percent, the mouse gets killed for sure. If an intermediate percentage value is selected, the mouse gets killed with that likelihood. Subjects receive their chosen percentage-level times 20 euros. While in the Exogenous Diffusion condition the presence of other decision makers potentially dilutes being pivotal by organizational design, subjects in the Endogenous Diffusion condition can actively implement non certain outcomes themselves – yet at a proportional reduction of payment. Our results show that more than 30 percent of subjects nevertheless choose intermediate probability values. We hypothesized that the intermediate option is attractive in particular for subjects who experience a need to resolve the moral conflict between receiving money and killing their mouse. Choosing uncertain outcomes may allow cashing in some money while maintaining a rather positive self-image (Bénabou and Tirole, 2006). To test this conjecture we use response times, which are indicative of (moral) conflict, as well as survey responses concerning feelings of remorse and bad conscience. We find that response times are significantly higher for subjects choosing uncertain rather than certain outcomes. Furthermore, when equipped with the possibility to "hide" behind uncertain outcomes, the fraction of subjects who save their mouse is significantly lower in comparison to an individual set-up in which subjects can only decide between the two certain outcomes: saving versus killing. Thus, even if subjects can save a mouse with certainty, less do so if they have the option to diffuse being pivotal; and this holds despite the fact that choosing an intermediate death probability proportionally lowers the received monetary amount.

In sum our findings show that an institutionally imposed diffusion of being pivotal facilitates moral transgression. In addition we observe a preference for self-imposed uncertainty. In various real life situations, both mechanisms, exogenous and endogenous diffusion, are likely to reinforce each other, by creating a demand for institutions that dilute being pivotal.

Our paper is related to work showing malleability of fair outcomes in the context of simple dictator, bargaining or allocation games. We focus on the role of exogenous and endogenous variation in being pivotal. Other mechanisms that have been identified to produce "unfair" outcomes are delegation or exploiting moral "wriggle rooms", as discussed, e.g., in Bartling and Fischbacher (2012), Hamman et al. (2010) and Dana et al. (2007).[4] Falk and Szech (2013) analyze malleability of moral outcomes in bilateral and multilateral market situations. Another related 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 Exogenous Diffusion treatment apart is that even if a subject opts for the moral outcome, he or she remains uncertain whether the moral outcome is implemented or not. This is closer to the situation of execution teams than to the situation in a bystander experiment where, clearly, if a subject opts for helping, the person in need receives help. Furthermore, in a typical bystander experiment, while deliberating about helping or not, subjects observe that others do not help either. Feelings of guilt can hence be shared. In our simultaneous-move set-up, this type of social learning is ruled out: When deciding to kill a mouse, subjects do not know whether other subjects opt for killing as well. 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 more able helper will step in), while in our set-up consequences of decisions are straightforward. Note also 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. Section 2 describes the research paradigm, treatments, procedural details and hypotheses. Results are presented in Section 3, and Section 4 concludes.

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

2. Design and Hypotheses

In this section we first discuss our paradigm, before we present treatments, procedural details and hypotheses.

Paradigm. A key challenge in studying immoral outcomes is to define a choice paradigm that involves harming others in an unjustified way. For this purpose, we used the "mouse paradigm", 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).[5] In all treatments subjects were explicitly informed that each mouse is a young and healthy mouse, which will live for about two years if saved. For illustrative purposes, we presented subjects the picture of a mouse on an instruction screen (Figure 1). We guaranteed subjects that mice, if saved, live in an appropriate, enriched environment, jointly with a few other mice. To this end, we purchased the mice saved by subjects' decisions. These mice are now kept in an enriched environment, with good feed and comfortable nesting material, precisely as stated in the instructions.

Subjects were also informed in detail about the killing process. In the instructions they read the following passage: "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." To further rule out uncertainty about the decision context, subjects saw 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 beating visibly heavy and slow.

[5] Deckers et al. (2014) 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 with Machiavellianism, measuring a person's tendency to be unemotional, and detached from conventional morality. Moreover, killing is not related at all 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.

Figure 1: Picture of a mouse as presented in the instructions of the experiment.

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 turned out to be unsuited for study. They were perfectly healthy, but keeping them alive would have been costly. It is standard to gas such mice. This is common practice in laboratories conducting animal experiments. Thus, as a consequence of our experiment, many mice that would otherwise all have died were saved. Subjects were informed about this default in a post-experimental debriefing.[6]

Treatments. To study the role of diffusion of being pivotal, we explore three decisions contexts: One in which subjects are fully pivotal (Baseline), one in which being pivotal is diffused by organizing subjects into groups (Exogenous Diffusion), and one in which subjects have the possibility to individually diffuse being pivotal, if they want to (Endogenous Diffusion).

To study the effects of *exogenous* diffusion of being pivotal, we contrast Baseline with Exogenous Diffusion. The two decision contexts differ in how pivotal a single subject is, keeping *overall financial incentives and payoff consequences* identical. In Baseline, each subject decided about the life of one mouse. Subjects faced a simple binary choice, labeled option A and option B. Option A implied that the mouse would survive and that the subject would receive no money. Option B implied the killing of the mouse and receiving 10 euros. The Baseline treatment

[6] While *perceptions* of the situation may have changed due to this information, *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., Spranca et al. (1991)).

informs us about the fraction of subjects who are willing to kill the mouse for 10 euros if they are fully pivotal.

In Exogenous Diffusion, subjects decided in groups of eight. As in Baseline, each subject faced an individual, private and binary choice option, again labeled option A and option B. Option A implied that a subject would receive no money. If a subject chose option B, he or she would receive 10 euros. Individual payoff consequences were independent of other subjects' decisions. All subjects chose simultaneously. They knew that if at least one subject chose option B, all eight mice were killed. Furthermore, they knew that they would not receive feedback on whether the mice were finally killed or not (though it is clear that the mice die if they chose B). On the aggregate level, the group of eight subjects was facing the same financial incentives and was endowed with the same number of mice as in Baseline. The difference between treatments only concerns the organizational design: While in Baseline each subject was pivotal, this was not the case in the Exogenous Diffusion condition if more than one subject supported the killing.

In both treatments we elicited beliefs of being pivotal right after subjects had taken their decision. Subjects were asked to assume a group of eight subjects and to estimate how many other subjects they think had chosen option B. Subjects could enter any number from 0 to 7 and were paid one euro for a correct estimate.[7] In Exogenous Diffusion we also asked subjects to indicate the probability that all other seven group members had chosen option A. In this case their decision would be pivotal for the survival of the mice. Subjects were asked to enter an integer percentage number, i.e., higher percentages indicate a higher likelihood that the subject perceived himself as pivotal.

In order to study whether subjects have a preference for actively diffusing being pivotal themselves in a very controlled, individual set-up, we ran the Endogenous Diffusion treatment. In this treatment, subjects chose the percentage level with which their mouse dies. They entered an integer probability level between 0 and 100. As payoffs, subjects received the percentage level chosen times 20 euros. If they chose zero percent, the mouse was guaranteed to survive and subjects received zero euros. Likewise, if subjects opted for 100 percent, the mouse was killed with certainty and subjects received 20 euros. If, e.g., a subject opted for 30 percent, he

[7] To calculate accuracy of beliefs and payments, in Baseline seven other subjects were randomly drawn for each subject in a given session.

received 30 percent of 20 euros, which amounts to 6 euros, and his mouse was killed with a probability of 30 percent. Subjects knew that they would not receive feedback on whether their mouse was finally killed or not if they opted for an intermediate percentage level.

As another control treatment, we furthermore use data from an individual Price List treatment, in which subjects chose individually for 20 different decision rows about survival or death of one mouse. In each decision row, option A guaranteed the survival of the mouse. Option B implied the killing of the mouse and receiving a monetary amount. The monetary amount increased over decision rows from 2.50 euros up to 50 euros in steps of 2.50 euros. Subjects knew that one of the rows would be randomly selected and implemented with all consequences. Data from this treatment informs us about the willingness to agree to kill for various different monetary amounts and helps us to compare decisions to save mice when being pivotal can be endogenously diffused or not.

Procedures. A total of 458 subjects, mainly undergraduate university students from all majors, took part in the experiments. 124 subjects participated in Baseline, 128 in Exogenous Diffusion, 110 in Endogenous Diffusion, and 96 in Price List, respectively.[8] Each subject participated only in one treatment condition. We used z-Tree as experimental software (Fischbacher, 2007); subjects were recruited using the software ORSEE (Greiner, 2003). At the beginning of an experimental session, participants received detailed information about the rules and 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 experiments were run between May 3 and 4 in 2012, and were conducted 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 euros to compensate for the remote location. In all treatments subjects received their payments in a sealed envelope outside the room where the experiment had taken place. This

[8] The two control treatments, Baseline and Price List, have also been used in Falk and Szech (2013).

way, neither other subjects nor the experimenter 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 personally guarantee the truthfulness of the instructions. Subjects were also invited to send us an email if they wanted to discuss the study.

Hypotheses. Our main hypothesis was that exogenous as well as endogenous diffusion of being pivotal would tempt subjects to behave selfishly rather than morally. We hence expected a higher killing rate in Exogenous Diffusion than in Baseline. If subjects in the former treatment believe that other group members engage in the immoral activity and choose option B, they are no longer pivotal. This should make it easier to choose option B, applying some replacement logic that the decision of other group members would have lead to the killing anyway (Sobel, 2010). In line with this hypothesis, we expected that in Exogenous Diffusion, agreeing to the killing should be inversely related to perceptions of being pivotal.

While these hypotheses seem intuitive, two important qualifications should be made. First, even if being pivotal is diluted in Exogenous Diffusion, the consequences from an individual perspective are much more severe as well. First, in Baseline, subjects decide about the life of one mouse, while subjects in Exogenous Diffusion potentially cause the death of eight mice. Second, replacement arguments corresponding to a low perception of being pivotal provide a moral excuse only from an outcome-based or utilitarian moral perspective, as advocated for example by Jeremy Bentham and John Stuart Mill. If subjects follow a deontological moral principle (e.g., the Kantian Categorical Imperative), they may rather stick to their morally preferred option, regardless whether being pivotal is diffused or not. The empirical relevance of these two moral conceptions, which have been the main combatants in occidental moral philosophy for the last centuries, has been empirically studied using the so-called trolley problem put forward by Philippa Foot (see also,

e.g., Thomson, 1976; Greene et al. 2004).[9] Evidence from this literature suggests that both moral approaches are empirically relevant and that the extent to which people follow the one or the other largely depends on situational and emotional factors (e.g., framing the trolley problem as footbridge problem, which assigns actors a more active role, leads to more rule based behaviors). In sum, while the diffusion of being pivotal would suggest a higher killing rate in Exogenous Diffusion than in Baseline, the effect may also depend on how severe subjects perceive the killing (eight mice vs. one mouse in case one is pivotal), on the extent to which subjects follow a utilitarian moral approach, and whether they believe that other subjects are utilitarians.

We hypothesized that the diffusion of being pivotal would render the pursuit of immoral activity less damaging for self-image (Bénabou and Tirole, 2006). With regard to the Endogenous Diffusion treatment, we therefore expected a non-trivial fraction of subjects, in particular those who experienced a conflict between money and morality, to choose intermediate probabilities in an attempt to resolve that conflict. As a consequence, we further expected the fraction of subjects who save their mouse with certainty and thereby forego all monetary payments to be lower in Endogenous Diffusion than in an individual condition where subjects had to choose a deterministic outcome, i.e., either killing a mouse with certainty and receiving 20 euros or saving a mouse with certainty and receiving 0 euros.

3. Results

Exogenous diffusion of being pivotal. Our first result concerns the comparison of individual killing rates between Baseline and Exogenous Diffusion. As Figure 2 shows, in Baseline 45.9 percent of subjects chose option B. In Exogenous Diffusion this fraction was 58.6 percent, an increase of about 30 percent. This difference is significant ($p<0.05$, Two sample test of proportions, two-sided). Thus, a simple organizational change increases the likelihood of killing in a causal and significant way.

[9] The quandary to be resolved in this problem is to either follow 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).

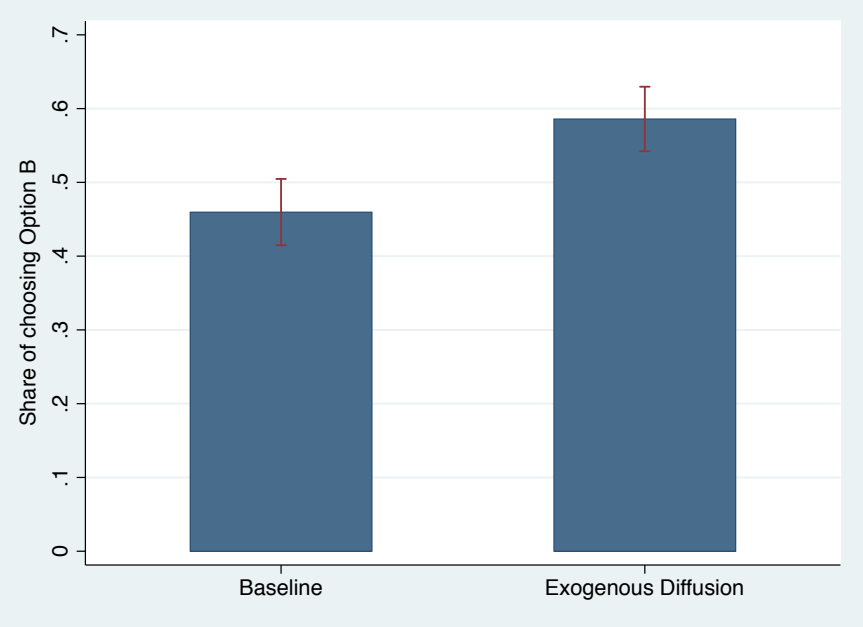

Figure 2: Fraction of subjects choosing option B in Baseline and Exogenous Diffusion. Error bars show standard deviations at the means.

If individual perceptions of being pivotal are responsible for immoral behavior in Exogenous Diffusion, we should observe that the likelihood of subjects choosing to kill their mouse decreases in the belief of being pivotal. This is what we find. Recall that we asked subjects about the probability (in percentage points) that all other group members had chosen option A. Figure 3 displays the fraction of subjects choosing option B depending on this belief. The four categories in Figure 3 are based on quartiles of the belief distribution with respective percentage values of 0-3.5; 3.5-10; 10-35 and 35-100. The figure shows a clear negative relation between subjective perception of being pivotal and the likelihood of choosing option B (Spearman rank correlation: -0.5409, $p<0.001$).

Figure 4 displays the full distribution of beliefs about being pivotal. It shows the cumulative distribution of beliefs separately for subjects choosing option A and option B, respectively. The figure shows that beliefs vary strongly between subjects, suggesting considerable uncertainty about being pivotal. Moreover, beliefs differ vastly depending on the option chosen. For example, while among those who choose option B, 31 percent of subjects believe that the likelihood of being pivotal is zero, the corresponding value for those who choose option A is only 9.4 percent. Likewise, subjective estimates of below or equal 20 percent hold for 90.7 percent of subjects

who decided to kill, while the respective value is only 39.6 percent for those who decided to save mice.

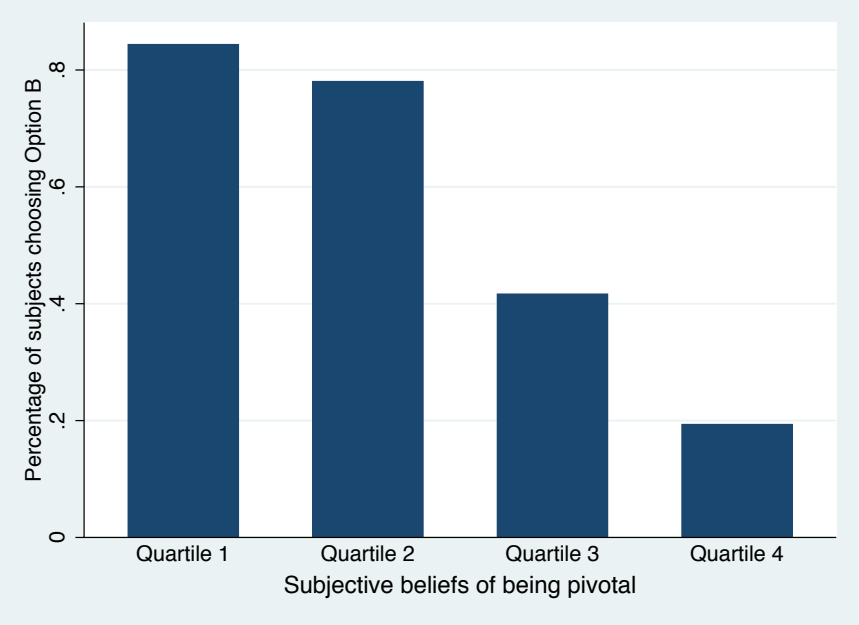

Figure 3: Percentage of subjects choosing option B depending on the belief of being pivotal.

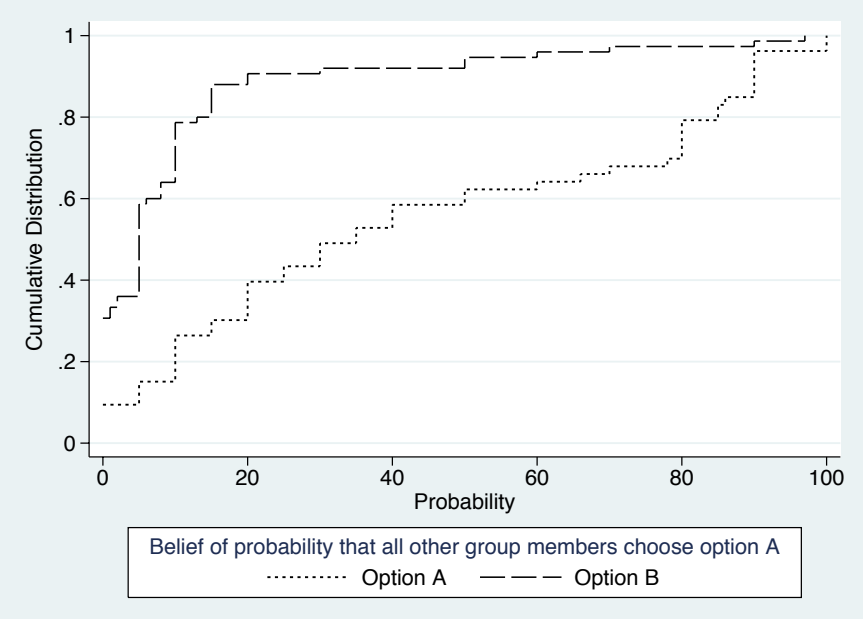

Figure 4: Cumulative distribution of beliefs of being pivotal in the Exogenous Diffusion treatment, separately for subjects choosing option A or B, respectively. *Probability* refers to the estimated probability that all other group members have chosen option A. Only in this case, the own decision becomes outcome-relevant. n = 128.

In light of Figures 2 and 3 it is not surprising that the likelihood of killing depends statistically significantly on perceptions of being pivotal. In Table 1 we show OLS regression results in which we regress the likelihood of choosing option B on this belief.[10] The respective coefficient is negative and significant. Assuming a linear relationship, the coefficient implies that a change in perceptions of being pivotal of 10 percent decreases the likelihood of killing mice by 8.4 percentage points (see Table 1, column 1). Figure 4 suggests that especially very low beliefs coincide with choosing option B. Therefore we tested for convexity: if we include a squared term of the percentage value, both coefficients are significant at the 1-percent level (see Table 1, column 2). Thus, the likelihood of killing depends strongly on the belief of being pivotal, and option B is chosen particularly frequently if subjects perceive themselves as very unlikely to be pivotal. This finding suggests that in particular large groups should give way to immoral behavior, enabling many subjects to believe that they are most likely not pivotal anyway. The importance of subjective perceptions of being pivotal can also be derived from results reported in column 3. In this specification we use subjects' responses to the question how many other subjects they think had chosen option B.[11] We constructed a dummy, which takes value 1 if a subject stated the point belief that no other subject has chosen option B, which would render the subject's decision pivotal. 17 subjects (13.3 percent) stated this belief. Among these 17 subjects only one subject chose option B. The negative and significant coefficient of the dummy implies that relative to subjects with other point beliefs the likelihood of choosing option B is about 61 percent lower for subjects who believe that they are pivotal.

[10] We obtain the same results using Probit estimates.

[11] We did not elicit a complete belief distribution. Especially if a subject stated a low number of participants choosing B, there may have been considerable uncertainty about being pivotal when opting for B.

	(1)	(2)	(3)
Probability of being pivotal	-.0084***	-.0224***	
	(.0012)	(.0049)	
Probability of being pivotal squared		. 00016***	
		(.00005)	
Belief that nobody chooses B			-.6078***
			(.1174)
Constant	.7910***	.8871***	.6667***
	(.0478)	(.05667)	(.0428)
Observations	128	128	128
R-squared	0.2733	0.3207	0.1754

Table 1: OLS regression coefficient estimates, with binary choice option (option B: agree to kill mice vs. option A: do not agree to kill mice) as dependent variable and standard errors in parentheses. Data come from the Exogenous Diffusion treatment. *Probability of being pivotal* is the belief that all other group members have chosen option A (in percent). *Belief that nobody chooses B* is a dummy, which takes value 1 for the point belief that all other subjects have chosen option A. *** indicate significance at the 1-percent-level.

As briefly discussed above, notions of being pivotal provide a moral excuse only from an outcome-based or utilitarian moral perspective. In this respect it is noteworthy that in Exogenous Diffusion about 10 percent of subjects chose option A, despite holding a belief that the chance of being pivotal is exactly zero. Thus from a utilitarian moral perspective, these subjects had no reason not to choose option B and cashing in 10 euros. Instead, these subjects seem to have followed a deontological moral principle, sticking to their morally preferred option regardless of outcomes. In line with survey-based evidence on the prevalence of utilitarian and deontological moral conceptions (Trolley problem), our findings suggest the coexistence of both moral conceptions using an incentivized choice task.

A possible concern in interpreting beliefs is the potential endogeneity of beliefs, e.g., due to reasons of false-consensus effects (Marks and Miller, 1987)[12], or

[12] A finding that suggests a limited role of false consensus as source of beliefs is the following. Suppose that due to false consensus subjects who choose option A believe all other subjects choose option A as well. If this were the case, the fraction of subjects who choose A and think all others choose option A should be identical in Baseline and in Exogenous Diffusion. If, however, beliefs of being pivotal have an independent impact on subjects' decisions, the fraction of subjects who choose option B while believing that everybody else chooses option A should be lower in

driven by subjects trying to justify behavior ex post in an attempt to maintain a positive self-image. Eliciting and interpreting beliefs is notoriously difficult in this respect. To limit this problem we incentivized beliefs about the number of other participants choosing option B in both treatments. We did not incentivize the likelihood of being pivotal belief for the following reasons. A possible remuneration would have depended on the distance to a binary outcome; either the subject turned out to be pivotal in his group or not. We felt such a remuneration scheme would have been difficult to understand for many subjects. Furthermore, many subjects may have chosen intermediate percentages to diversify against the only two possible outcomes of 0 percent and 100 percent. Reassuringly, however, both types of beliefs (number of other participants choosing option B and probability of being pivotal) are significantly correlated (Spearman rank correlation: -0.5214, $p<0.001$). Moreover, all our results are qualitatively the same (also in terms of significance) if we use incentivized beliefs about the number of other participants choosing option B instead of using the (more appropriate) belief about being pivotal (see results in column (3) of Table 1).

We conclude this section in briefly discussing aggregate moral consequences. While in Baseline 45.9 percent of mice were killed, all mice were killed in all groups of the Exogenous Diffusion treatment. Moreover, killing is not only much more pronounced in Exogenous Diffusion, it is also more "efficient" in the following sense: the money needed to kill one mouse is 10 euros in Baseline but only 5.86 euros in Exogenous Diffusion. Thus if the goal of an organization is to promote immoral activities it can achieve this by organizing people into groups and aggregating decisions such that the support of few group members is sufficient for the immoral to happen. Such an organization will be more vulnerable to moral transgression than an organization that attributes individual responsibility to its members.

Endogenous diffusion of being pivotal. In the previous section we have shown how an institutionally imposed diffusion of being pivotal affects behavior. Using the same research paradigm we now explore whether subjects also have a preference to actively *create* uncertainty about outcomes. Intuitively, self-imposed uncertainty facilitates maintaining a positive self-image since people do not cause morally questionable

---

Exogenous Diffusion. This is what we find. 11.1 percent of subjects who hold this belief choose option B in Baseline, but only 5.9 percent do so in Exogenous Diffusion.

outcomes in a deterministic way. Maintaining a positive self-image should be particularly easy with ex-post uncertainty about the final outcome, as in our Endogenous Diffusion treatment. Here, subjects never learned the outcome unless they chose a killing probability of either zero or 100 percent, respectively.

Figure 5 displays the cumulative probability of chosen probabilities. The figure shows that 31.8 percent of subjects chose a probability strictly above 0 and below 100 percent. It is hence attractive for many subjects to "hide" behind uncertain outcomes, even though it implies earning less money, presumably as it allows maintaining a positive self-image.

If the possibility to diffuse responsibility tempts subjects into choosing intermediate probabilities we should further find that the fractions of subjects who save their mouse with certainty is lower in the Endogenous Diffusion treatment in comparison to the Price List treatment evaluated at 20 euros. This is what we find. The fraction of subjects who save their mouse with certainty when 20 euros are at stake is 42.7 percent in Price List, compared to 30.9 percent in Endogenous Diffusion, which is weakly significant (Two-sample test of proportion, $p<0.1$, two-sided). Thus in comparison to Price List, fewer subjects save mice with certainty in the Endogenous Diffusion condition.

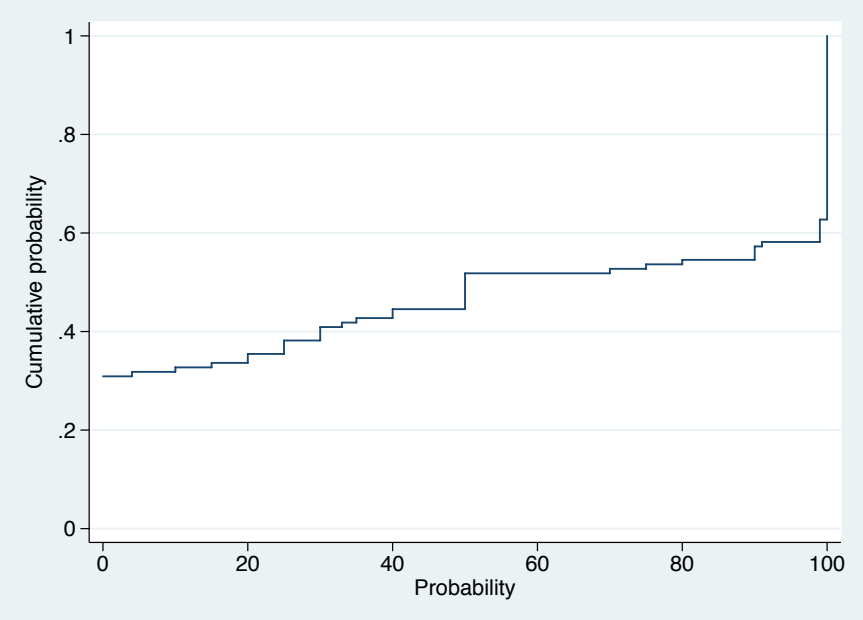

Figure 5: Cumulative distribution of chosen probabilities.

Who is tempted into choosing intermediate probabilities? If hiding behind uncertainty is a means to resolve the conflict between selfishness and morality, it is plausible to assume that intermediate probabilities are chosen by subjects who experience such a conflict. To test this intuition we rely on two measures of normative conflict, response times and expressed feelings of remorse and bad conscience. Response times are generally associated with the processing of conflict, including moral or normative conflict (Davelaar, 2008, Kerns et al., 2004, Greene et al., 2001). In addition we asked all subjects who chose a probability strictly above 0 percent to express possible feelings of remorse and bad conscience on a seven-point Likert scale. Higher values imply stronger moral concerns. Reassuringly, both types of measures are significantly positively correlated. The Spearman rank correlations between response times are 0.3486 ($p<0.01$, two-sided) for bad conscience, and 0.3386 ($p<0.01$, two-sided) for remorse, respectively.

We hypothesized that subjects with either very high or very low moral concerns should not experience moral conflict and therefore readily choose a probability of either 0 or 100 percent, respectively. In contrast, subjects who are torn between killing and receiving money use intermediate probabilities in an attempt to "resolve" their moral conflict. As a consequence, we would expect response times to be longer for subjects choosing intermediate probabilities in comparison to subjects choosing probabilities of either 0 or 100 percent. As Figure 6 shows, this is what we find. Mean response times for 0 and 100 percent are 25.5 seconds (std. dev. 14.1) and 29.7 seconds (std. dev. 18.8), respectively. For intermediate probabilities the mean response time is 42.0 seconds (std. dev. 33.1), significantly longer than for zero percent ($p<0.01$) and 100 percent ($p<0.05$), respectively (t-tests, two-sided)[13]. Note that differences in response times cannot be explained by trivial differences in "transaction costs". Subjects had always to type in their desired probability. In addition, stated levels of remorse and bad conscience are higher for subjects choosing intermediate probabilities than for those killing their mouse with certainty ($p<0.05$ for bad conscience and ($p<0.01$) for remorse, t-tests, two-sided)[14]. These findings suggest that subjects who experience more pronounced normative conflict use the intermediate probability option in an attempt to diffuse responsibility.

[13] The difference between zero and 100 percent is not statistically significant.
[14] A comparison with subjects choosing a probability of zero is impossible as those subjects were not asked about feelings of remorse and bad conscience.

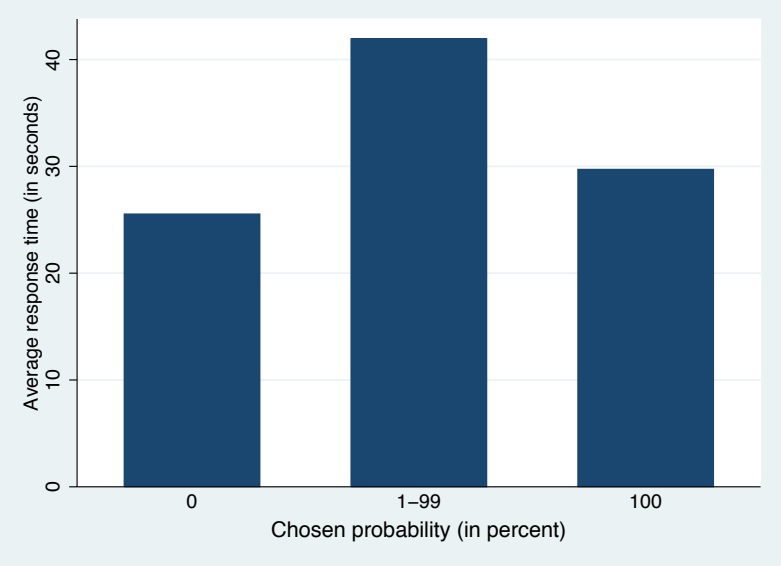

Fig. 6. Mean response times for subjects choosing either certain outcomes (probabilities of either 0 or 100 percent) or uncertain outcomes (probabilities strictly larger than 0 and strictly smaller than 100 percent).

Subjective statements about moral sentiments further display an interesting pattern: Both, feelings of bad conscience and remorse, decrease in chosen killing probabilities (for killing probabilities above zero). The Spearman rank correlation between remorse and chosen probability is -0.4490 ($p<0.01$, two-sided). Similarly, the correlation between bad conscience and chosen probability is -0.3032 ($p<0.01$, two-sided). The respective correlations for subjects choosing intermediate probabilities only, i.e., excluding those who chose a probability level of 100, are -0.4059 ($p<0.05$, two-sided) for remorse, and -0.3714 ($p<0.05$, two-sided) for bad conscience, respectively. This pattern suggests that subjects with relatively strong moral concerns are tempted into choosing low probabilities. This way they are almost certain that the mouse will survive, yet they earn some money. Likewise, subjects with rather low standards opt for high probabilities in an attempt to earn as much money as possible, without causing the death of the mouse with certainty, and thereby leaving room for a positive self-interpretation.

4. Concluding Remarks

This paper has documented the malleability of moral outcomes in response to both an exogenous as well as an endogenous diffusion of being pivotal. The simple

organizational change from Baseline to Exogenous Diffusion increased moral transgression on the individual and the aggregate level. Individual effects are systematically affected by beliefs about being pivotal. On an aggregate level the effects are fatal in the sense that diluting being pivotal leads to a killing rate of 100 percent, rendering the killing more "efficient" from an organizational point of view. This happens even though many subjects still overestimated their potential impact on outcomes: The overall willingness to agree to the killing was almost 60 percent in Exogenous Diffusion. If beliefs about being pivotal had been more realistic, the willingness to agree to kill may have been even more pronounced. In this sense we would expect that repeated interactions with learning possibilities would increase the likelihood of immoral outcomes.

Additional suggestive evidence for the importance of exogenous diffusion of being pivotal comes from an explorative control treatment. In this treatment everything was identical to Exogenous Diffusion except that in the control treatment, at least *three* subjects had to choose option B in order to kill eight mice. Thus in this treatment a subject was pivotal if exactly two other subjects had chosen option B, rather than no other subject as in Exogenous Diffusion. In this control condition the killing rate is 53.1 percent (n=64), i.e., right in between Baseline and Exogenous Diffusion. The difference is neither statistically significant with respect to Baseline nor with respect to Exogenous Diffusion. However, the fact that the killing rate is higher than in Baseline but lower than in Exogenous Diffusion is in line with notions of being pivotal: Perceived being pivotal is lower than 100 percent, but higher than in Exogenous Diffusion. While the average subjective assessment of being pivotal in Exogenous Diffusion is 24.4 percent, the respective value is 32.5 percent in the control condition ($p<0.001$, Wilcoxon rank sum test, two-sided).

The effects of diffusion of being pivotal induced by organizations are potentially reinforced by a tendency to self-select into environments where outcomes do not materialize with certainty. Such a preference for self-imposed uncertainty is shown in our Endogenous Diffusion treatment where more than 30 percent of subjects chose intermediate probabilities, at a proportional cost. Our data suggest that the uncertainty option is preferred in particular by subjects who experience moral conflict. In future research, it would be interesting to explore the role of feedback about outcomes or authorship. In both of our treatments that allowed for a diffusion of being pivotal, subjects had the possibility to hide behind uncertainty. This allowed

blaming others, or hoping for a "good outcome" and thereby keeping a rather positive self-image. We speculate that resolving uncertainty ex post will affect the attractiveness of uncertainty as a moral excuse.

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 at promoting morality should reduce diffusion of being pivotal, and instead attribute individual responsibility to its members.

References

Arendt, H. (1963) Eichmann in Jerusalem: A Report on the Banality of Evil. Viking.

Bartling, B., Fischbacher, U. (2012), Shifting the Blame: On Delegation and Responsibility. Review of Economic Studies 79, 67-87.

Bénabou, R., Tirole, J. (2006), Incentives and prosocial behavior. American Economic Review 96, 1652-78.

Bowles, S. (1998), Endogenous Preferences: The cultural consequences of markets and other economic institutions. Journal of Economic Literature 36, 75-111.

Crawford, N. (2007), Individual and Collective Moral Responsibility for Systemic Military Atrocity. The Journal of Political Philosophy, 187 – 212.

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

Darley, J.M. (1992), Social Organization for the Production of Evil. Psychological Inquiry 3, 199-218.

Davelaar E.J. (2008), A computational study of conflict-monitoring at two levels of processing: Reaction time distributional analyses and hemodynamic responses. Brain Research 1202, 109-119.

Deckers, T., Falk, A., Kosse, F. and Szech, N. (2014), Individual Characteristics as Source of Variation in Moral Outcomes, University of Bonn (mimeo).

Falk, A., Szech, N. (2013), Morals and Markets. Science, 340, 707-711.

Fischbacher, U. (2007) z-Tree: Zurich Toolbox for ready-made economic experiments. Experimental Economics 10, 171.

Fischer P., Krueger J.I., Greitemeyer T., Vogrincic C., Kastenmüller A., Frey D., Heene M., Wicher M., Kainbacher M. (2011), The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies, Psychological Bulletin, 137(4), 517-37.

Foot, P. (1967), The Problem of Abortion and the Doctrine of Double Effect. Oxford Review 5, 5-15.

Gert, B. (2012), The Definition of Morality, The Stanford Encyclopedia of Philosophy (Fall 2012 Edition), Edward N. Zalta (ed.).

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001), An fMRI Investigation of Emotional Engagement in Moral Judgment. Science, 293, 2105-2108.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004), The Neural Bases of Cognitive Conflict and Control in Moral Judgment. Neuron 44, 389–400.

Greiner, B. (2003), An online recruitment system for economic experiments (in Forschung und wissenschaftliches Rechnen. GWDG Bericht 63, Ges. für Wiss. Datenverarbeitung, Göttingen, pp. 79–93).

Hamman, J., Loewenstein, G., Weber, R.A. (2010), Self-interest through delegation: An additional rationale for the principal-agent relationship. The American Economic Review 100, 1826-46.

Kerns, J.G., Cohen, J.D., MacDonald, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S., (2004), Anterior cingulate conflict monitoring and adjustments in control. Science 303(5660): 1023-1026.

Latané, B., Darley, J.M. (1968), Group inhibition of bystander intervention in emergencies. Journal of Personality and Social Psychology 10, 215-21.

Lifton, R.J. (1986), The Nazi Doctors: Medical Killing and the Psychology of Genocide. Basic.
Marks, G., Miller, N. (1987), Ten years of research on the false-consensus effect: An empirical and theoretical review. Psychological Bulletin 102 (1): 72-90.
Milgram, S. (2009), Obedience to Authority: An Experimental View, Harper Perennial Modern Classics; Reprint edition.
Sobel, J. (2010), Do markets make people selfish? Discussion Paper, Economics Department, University of California.
Spranca, M., Minsk, E., Baron, J. (1991), Omission and Commission in Judgment and Choice. Journal of Experimental Social Psychology 27 (1), 76-105.
Thomson, J. (1976), Killing, Letting Die, and the Trolley Problem. The Monist 59, 204-17.

Appendix: Instructions of the experiment

In the following we present the English translation of the instructions for Baseline, Exogenous Diffusion and Endogenous Diffusion. The introduction to the experiment as well as the description of the mouse paradigm, including the video etc., were identical in all treatments. The Price List treatment was identical to the Baseline treatment except that instead of offering a single binary choice, subjects were presented a table with an increasing price-list, ranging from 2.50 euros to 50 euros. In all treatments, the choice between Options A and B was associated with taking part in an identical trivia quiz.

1. Instructions: Baseline treatment

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 Bonn Econ Lab, 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 und 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.**

Depending on which quiz you choose, you may earn different amounts of money in addition. Additionally, depending on which quiz you choose, there will be different consequences for a mouse.

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.

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**.

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. As soon as the mouse is not visibly breathing anymore, it remains in the cage for another 10 minutes. It will then be removed.

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 take your decision on a decision screen that will be shown as soon as you have answered the control questions on the following screen.

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

Video
To visualize the killing of mice by gas, you will in the following see an excerpt of a documentation video (30 seconds). The mouse will be killed in an identical way.

2. Instructions Exogenous Diffusion treatment

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.

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**.

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. As soon as the mice are not visibly breathing anymore, they remain in the cage for another 10 minutes. They will then be removed.

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 take your decision on a decision screen that will be shown as soon as you have answered the control questions on the following screen.

Control questions and video

3. Instructions Endogenous Diffusion treatment

Introduction as in Baseline

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.

You choose the probability with which the mouse will get killed. The probability which you can choose is in steps of 1-Percent from 0 percent to 100 percent. Thus if you insert, e.g., a **zero**, the mouse will definitely not get killed, and be **saved with certainty**. If you insert a **7**, the probability that a mouse gets killed is **7 percent**. If you insert a 83, the probability that a mouse gets killed is **83 percent**. If you insert a **100**, the probability is **100 percent**, i.e., the mouse gets killed with certainty. Please note: **the higher the chosen probability, the higher is your payment**. For each percentage point you receive 20 cents.

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. As soon as the mouse is not visibly breathing anymore, it remains in the cage for another 10 minutes. It will then be removed.

Summary:

You are solving a quiz. In addition you choose the probability with which a mouse gets killed. The higher the chosen probability the higher is your payment. You receive 20 cents for each percentage point.
You take your decision on a decision screen that will be shown as soon as you have answered the control questions on the following screen.

Control questions and video