Making Up for Harming Others — An Experiment on Voluntary Compensation Behavior

Frauke Stehr[*] Peter Werner[†]

July 3, 2023

Abstract

We investigate in a controlled laboratory setting to what extent buyers are willing to offset negative consumption externalities. In one treatment dimension, we vary whether the externality associated with a purchase is irreversible, or can be reduced ex post by a voluntary payment. In a second treatment dimension, we induce diffusion of harm among harmed subjects and diffusion of responsibility among buyers by separately varying the matching of buyers and harmed subjects. We find that subjects are on average willing to compensate for their negative externalities, and that this willingness is sensitive to the surplus from buying. Yet, experimental buyers are highly heterogeneous, with some never compensating. While the introduction of voluntary compensation significantly reduces externalities, the net externality still remains high across all treatments. Diffusion of responsibility tends to reduce the size of compensation and to increase overall net externalities in the main experiment. An additional control treatment reveals that under diffusion of responsibility among buyers, patterns of conditional cooperation seem to drive compensation: The amount paid for compensation increases with higher beliefs about the compensation by other buyers.

Keywords: voluntary compensation, moral behavior, socially-responsible consumption, diffused responsibility, diffused harm

JEL Codes: D91, D62, H41, Q58

[*]Frauke Stehr, Department of Microeconomics and Public Economics, Maastricht University, Tongersestraat 53, 6211 LM Maastricht, The Netherlands (e-mail: f.stehr at maastrichtuniversity.nl).

[†]Peter Werner, Department of Microeconomics and Public Economics, Maastricht University, Tongersestraat 53, 6211 LM Maastricht, The Netherlands (e-mail: p.werner at maastrichtuniversity.nl).

1 Introduction

Voluntary offsetting tools can be an efficient means to internalize externalities for consumers who wish to diminish their contribution to a public bad (Kotchen, 2009), for example to environmental pollution. These tools rely on individuals' willingness to pay to reduce externalities they have caused, for instance by financing emission reductions elsewhere in the world. As such, offsetting tools may be welfare enhancing, allowing individuals who otherwise would have abstained from the consumption of certain goods due to concerns for the externality to consume because they can now offset the externality. The relevance of voluntary (carbon) offsetting tools is evidenced by their global market volume, which was estimated to be $748 million in the first half of 2021 alone, offsetting around 239 Mt CO_2 equivalents (Forest Trends' Ecosystem Marketplace, 2021). In recent years, a large number of organizations have started to offer the possibility to compensate negative environmental externalities themselves, for instance through donations to climate change mitigation projects.[1]

Given the popularity of such offsetting tools, it is vital to understand the determinants of the willingness to engage in compensation. While recent field experiments have provided insights into consumers' responses to measures aimed at promoting offsetting, such as matching or rebate schemes (Kesternich et al., 2016), or changes in the choice architecture (Kesternich et al., 2019), further evidence is required to assess the circumstances in which voluntary offsetting may be an effective means to mitigate climate change. Insights into how contextual factors in the decision environment affect individuals' willingness to compensate are the basis for these assessments. Laboratory experiments can complement field data analysis by designing treatments with counterfactual decision environments to cleanly identify the causal effect of different contextual characteristics on behavior. The controlled environment of the laboratory additionally allows us to control subjects' beliefs of the size and severity of the externality, which might vary widely in field settings and create substantial noise in compensation decisions.

We therefore employ a laboratory experiment to investigate to what extent participants in the role of buyers are willing to offset the harm their purchasing decisions create and how this

[1] Apart from a large number of private organizations (e.g., Carbon Footprint Ltd., Atmosfair, myclimate), also publicly funded institutions (e.g., climateneutralnow.org by the UN) and organizations linked to the church (e.g., German Klimakollekte) offer this service.

willingness depends on the tangibility of the externality. Subjects in our experiment can buy a virtual good, where upon buying a negative real-world externality is imposed on another subject in the lab. In one set of treatments, the externality associated with buying the good is irreversible whereas in another set of treatments buyers are offered the possibility to voluntarily compensate for the externality, thus reducing the harm on the harmed subject ex-post. The other set of treatment variations targets the effect of the externality's tangibility. In our baseline condition, one buyer is matched with one harmed subject. In two treatments, we independently manipulate both diffusion of the harm created and diffusion of responsibility in creating the externality by changing the matching structure of buyers and harmed subjects. In addition to individual patterns of voluntary compensation, we investigate buyers' purchasing decisions and the resulting net externalities with and without the possibility to compensate.

Our study makes two contributions to the literature: First, we elicit an incentivized laboratory measure for the willingness of buyers to voluntarily compensate for the negative externalities they are responsible for. The previous literature in the field of environmental economics mostly relies on survey measures on the willingness to compensate or focuses on field experiments conducted in specific contexts. In contrast to these approaches we elicit compensation behavior in a controlled and abstract setting, also testing to what extent this behavior responds to diffusion. Diffusion may be a decisive factor for individuals' real-world compensation decisions, given, for example, the low tangibility of the damages caused by consumption. In particular, we investigate to what extent compensation behavior responds to the diffusion of the externality itself (by varying the concentration of induced harm) and to the diffusion of responsibility for the externality (by introducing free-riding incentives between buyers). To assess the role that voluntariness can play in achieving significant reductions of negative real-world externalities, it is important to understand better how these factors shape compensation behavior.

Second, we extend the literature on moral behavior in market settings by focusing on a specific type of moral decisions that has received little attention before: the motivation to repair the damage caused by one's actions. In our experiment, we give buyers the possibility to compensate negative externalities ex-post. This option is relevant in practice, because abstaining from consumption entirely or purchasing a good with a lower or no externality might often not be possible (for example, long-distance flights). For a decision-maker who trades off personal benefits of consumption against a loss of utility due to the creation of harm (sometimes conceptualized as "moral costs" in the literature), ex-post compensation allows to reduce the externality (and hence the moral cost) at their own discretion. This is different from the case where externalities from trading are fixed and irreversible, which has been the focus of most of the previous literature. Moreover, given that buyers in our experiment can behave morally in two dimensions – either in the ex-ante abstention from buying or in choosing to compensate ex-post – it is a priori unclear whether total externalities will be more or less pronounced than in settings where externalities cannot be offset. Comparing treatments with and without the possibility to compensate allows us to analyze moral behavior and the resulting externalities in both dimensions.

Our results are as follows: First, subjects are on average willing to compensate for the externalities they create, and compensation decisions are sensitive to the surplus at stake. Yet, experimental buyers are highly heterogeneous, and some of them never compensate. Furthermore, the introduction of the compensation technology tends to increase the frequency of buying in the treatment without diffusion. Overall, the introduction of voluntary compensation significantly reduces the net externality in all treatments by between 13 and 18%. Average net externalities, however, remain on a high level, still accounting for 64 to 79% of the maximum possible externality, and externalities tend to be higher in the treatment with diffusion of responsibility as compared to the baseline condition.

Second, we find that in the main experiment buyers tend to compensate less under diffusion of responsibility than in the baseline condition. In an additional control treatment, we find that under diffusion of responsibility for the externality, beliefs about the behavior of other buyers on average matter for compensation, in line with patterns of conditional cooperation: The higher one's expectation of the other buyer's compensation, the larger one's own compensation.

The remainder of our paper is organized as follows: we review the literature related to our study in Section 2. In Section 3, we describe our experimental design and the hypotheses. We present the experimental results in Section 4; Section 5 details an additional treatment with belief data in the diffused responsibility setting. Section 6 briefly discusses the results and concludes.

2 Related Literature

First, our study links to previous research in the field of environmental economics that investigates individual motivations to invest resources in order to mitigate climate change. Some studies in this area collect incentivized measures for individuals' willingness to pay for climate protection per se: For instance, the studies by Löschel et al. (2013) and Diederich and Goeschl (2014) based on data gathered with non-laboratory subjects from Germany indicate a generally low willingness to pay for climate change mitigation as well as a large heterogeneity in preferences among subjects. A recent study by Andre et al. (2021) analyzes the willingness to donate in order to fight climate change in a representative sample from the US, showing that information interventions highlighting a norm for climate protection can significantly increase donations. Feldhaus et al. (2022) find in an artefactual field experiment that highlighting individual responsibility increases donations to climate change mitigation projects. Most closely related to our study, research focusing on the nature of individual motivations behind ex-post compensation of environmental externalities typically use evidence from surveys and vignette studies (Blasch and Farsi, 2014, Blasch and Ohndorf, 2015, Schwirplies and Ziegler, 2016, Lange and Ziegler, 2017).[2] In contrast to these studies, we elicit an incentivized measure for the voluntary ex-post compensation of negative externalities, thereby obtaining insights into revealed preferences for reducing the harm generated through one's own actions. Measuring revealed preferences for lower levels of externalities is important as there might be a discrepancy between stated and revealed preferences in the sense that answers from survey questions may exaggerate the true willingness to bear costs in order to reduce one's own negative externalities (see, for instance, related evidence from dictator games reported by Schier et al. 2016). Berger et al. (2021) investigate the choices of customers to offset CO_2 emissions from flights with observational field data. Their results indicate this potential discrepancy, as less than 5% of the flight passengers in their data set choose to offset the CO_2 emissions.

To the best of our knowledge, there are only two studies so far that investigate the individual motivation to remove negative externalities ex-post in controlled laboratory settings, and their designs differ markedly from the present study: The study by Jakob et al. (2017) tests concerns for moral responsibility in a real-effort experiment in which some participants work on a task that

[2] In particular, three factors are commonly found to correlate with stated demand for carbon offsets: environmental preferences, warm glow and the perceived responsibility in the creation of the externality.

creates a real-world externality. The authors show that a significant share of participants exhibit moral responsibility in the sense that they have a preference for removing this externality themselves rather than delegating it to another participant, despite efficiency losses associated with this choice. Moreover, Kuhn and Uler (2019) conduct a laboratory experiment in which subjects act as buyers and sellers in double auction markets, and realized trades impose a negative externality on all market participants. After the double auction markets, former traders can purchase carbon offsets that reduce the externality equally for all participants, thus transforming the choice to offset into a public goods game. Among other things, Kuhn and Uler find that the demand for offsets is price-sensitive, and individual choices to offset the externality are linked to a survey measure for personal responsibility. In contrast, our setting focuses on buyers who stay unaffected by the externality themselves, resembling situations in which the negative effects are not (directly) experienced by those who cause the externality. In addition, unlike the previous studies, voluntary compensation in our baseline condition directly reduces the harm for the third party and is not subject to potential free-riding. Moreover, by comparing compensation behavior within subject between our baseline condition and our treatment with responsibility diffusion, we can directly investigate the effect of free-riding incentives on the willingness to compensate and the resulting externalities.

Second, our study is related to research on moral concerns related to economic behavior. The study by Andre et al. (2021) mentioned above highlights the moral dimension of choices related to the reduction of environmental externalities; a higher degree of universal moral values is significantly correlated with donations to fight climate change. Moreover, several studies have shown that pro-environmental behavior and altruism as well as moral universalism are strongly correlated (Lades et al., 2021, Enke et al., 2022). Previous literature focusing on moral behavior investigates the determinants of moral behavior when (potential) traders on markets have to weigh personal payoffs against negative externalities, investigating for example the role of market institutions and market structure in fostering moral behavior (see, for example, Falk and Szech, 2013, Bartling et al., 2015, Kirchler et al., 2016, Pigors and Rockenbach, 2016b, Ockenfels et al., 2020, Sutter et al., 2020, Bartling et al., 2023, Riehm et al., 2022).[3] In these studies, participants have to choose whether or

[3]Several studies focus on buyers' selection processes when there are both goods with a negative externality and fair goods without a negative externality. In general, the fair good has a significant positive market share, despite coming at a higher price (Bartling et al., 2015, Pigors and Rockenbach, 2016a, Friedrichsen, 2017, Bartling et al., 2022). Moreover, Pace and van der Weele (2020) investigate the willingness to buy a product with a negative CO_2 externality, focusing on the interaction with prices and beliefs about the externality. In particular, the authors find

not to engage in trading activities or choose between products that vary in their degree of social costs. In contrast to the present experiment, however, choices then determine whether or not a negative externality is created which is irreversible afterwards. Hence, previous studies do not consider the option to reduce the damage created by one's actions and thus cannot investigate to what extent this option influences moral consumption behavior in the first place.

The option to offset harm ex-post in our setting enables decision-makers to "scale" moral behavior in the sense that compensation can be chosen in a way that balances private benefits and moral costs, as the degree of their externality can be adjusted endogenously by compensating either fully or only partially. Moreover, previous studies on altruistic and ethical behavior would suggest that the separation of buying and compensation decisions central to the present setting might make a difference from the buyer perspective compared to the situation in which buying triggers an irreversible externality. In particular, Gneezy et al. (2014) formalize the idea of "conscience accounting" when subjects face a sequence of choices that have ethical and prosocial components and balance own consumption, the consumption of others and feelings of guilt when violating moral constraints. As a result of conscience accounting, the guilt that arises after an unethical action increases the likelihood of a subsequent prosocial action. In the context of ex-post compensation, this would suggest that buying the good and causing the harm for the third party might then eventually trigger the willingness to compensate. If buyers behave in line with conscience accounting, the anticipation of the possibility to behave in a prosocial way through compensating might lead to more frequent buying, and thus, to higher (initial) externalities.[4]

Third, our study builds on previous experimental literature focusing on the behavioral effects of diffusion of cause and effect. Diffusion of responsibility has been found to significantly reduce prosocial behavior in dictator games (see, for instance, Dana et al., 2007, Hamman et al., 2010, Cryder and Loewenstein, 2012). Related to our setting, Branas-Garza et al. (2009) find that reduced responsibility makes dictators less likely to equalize highly unequal payoff distributions ex-post. In the context of moral behavior on markets, some studies have suggested that diffusion might have a detrimental effect but the evidence is mixed so far: Falk et al. (2020) test whether diffusion of responsibility in simultaneous and sequential decision making in groups leads to a higher

that beliefs about the CO_2 externality are significantly negatively related to the decision to buy the product.

[4]Gneezy et al. (2014) find evidence similar to this conjecture: If experimental subjects foresee that the possibility to donate to charity is available at a later stage, they become more likely to behave in a dishonest way.

willingness to impose negative externalities. They observe that individuals who decide in groups are significantly more likely to choose in a selfish way regardless of the nature of the externality. Irlenbusch and Saxler (2019) find no significant difference in the willingness to accept a negative externality when an individual's decision is compared to the case when two decision-makers share responsibility. Relatedly, Bartling and Özdemir (2023) observe that subjects do not make use of a "replacement excuse" (i.e., choosing an immoral action because someone else would have done so otherwise) in situations where a clear social norm exists that classifies the behavior causing an externality as immoral. Finally, the only study we are aware of that focuses on the diffusion of harm is the study by Bartling et al. (2019) that investigates the effect of diffused harm on the outcomes of markets in which goods with and without externalities can be traded. Holding the absolute magnitude of the externality constant, the authors vary how many subjects are harmed by the externality (one versus six subjects) and find that diffused harm only weakly affects the market share of fair goods.[5] Given the differences in the nature of moral behavior that previous studies focus on (either the choice to directly impose externalities or the choice between socially responsible products and products with externalities), it is unclear to what extent the diffusion of harm and the diffusion of responsibility change voluntary ex-post compensation in the present setting. Moreover, unlike previous research in this area, we investigate the impact of the two potential types of diffusion on both purchasing and offsetting decisions within the same setting.

3 Experimental Design and Hypotheses

3.1 Basic Setup

We employed a 2×3 experimental design. To assess the effect of the introduction of compensation, we varied the availability of compensation within subjects, so that every subject participated in the two conditions *NoCompensation* and *Compensation*. Between subjects, our three main treatments then varied the tangibility of the negative externality: *NoDiffusion*, *DiffusedHarm* and *DiffusedResponsibility*.

[5] In addition, in a series of distributional games by Schumacher et al. (2017), participants decide about whether or not a good is provided that imposes a certain cost for each payer in a group of payers. Here, a large fraction of subjects are insensitive to the size of the group who is bearing costs, which results in ignorance of large provision costs when these are spread among many payers.

We first describe our baseline condition *NoDiffusion* before we move to the diffusion treatments in the next subsection. The structure of the basic decision situation in the *NoCompensation* × *NoDiffusion* condition was as follows: There were two players, one buyer and one harmed party. Roles were randomly assigned. Subjects kept their role throughout the entire experiment.

In every round, both the buyer and the harmed party received an endowment of 75 Experimental Currency Units (ECU). The buyer could then decide whether she wanted to buy a fictitious good at price p from the experimenter. This price was randomly drawn from a uniform distribution with $p \in \{1, 2, ..., 100\}$ ECU. If the buyer decided not to buy the good, both players kept their endowment and no externality was created. When buying the good, on the other hand, the buyer received an extra payoff of 100 ECU reflecting her induced valuation. Given that the price could lie between 1 and 100, the gains from buying were between 0 and 99 ECU. Thus, buyers could always afford to buy the good; in fact, given the parameters of the task it was optimal for a buyer not concerned about the externality to buy the good at all price realizations.

At the same time, buying imposed a real-world negative externality on the harmed party. In our setting, the externality was implemented by obliging the harmed party to work on a tedious effort task after the end of the experiment. Specifically, a harmed subject had to correctly place 240 sliders in the slider task (Gill and Prowse, 2012)[6] in order to receive their payment after all other subjects who did not have to work on any sliders had left. As mentioned above, a profit-maximizing buyer not concerned about the externality would always buy and only be indifferent between buying and not buying at the highest possible price (100 ECU). By varying the surplus available to buyers from their buying decisions, we can therefore gain insights to what extent buyers trade-off material gains and negative externalities.

In the *Compensation* × *NoDiffusion* treatment, we added a second stage after the buying decision in each round. If a buyer decided not to buy the product, the consequences stayed the same as in the *NoCompensation* condition, i.e., both buyer and harmed party kept their endowments and no externality was created. Yet, once a buyer bought the good, she moved to the second stage and had the opportunity to use between 0 and 30 ECU of her experimental payoff to reduce the

[6] In this task, subjects were confronted with a number of movable sliders which could be dragged to all possible positions between 0 and 100. The computer randomly allocated the starting position of the slider. The subjects' task was then to drag all sliders to the middle position (in our case 50). Solving 240 sliders took between 9 and 18 minutes in our experiment.

externality on the harmed player. This decision mimics the choice of a buyer to compensate for the negative externality created by her purchase. In particular, each ECU spent by the buyer reduced the workload for the harmed player by eight sliders.[7] We chose the parameters to ensure that at least for price realizations below 70 ECU buying and fully compensating was a Pareto improvement compared to not buying at all, as for these prices buyers would increase their payoff beyond their endowment even after completely offsetting the externality (and thus, entirely avoiding extra working time for the third party).[8] This allows us to interpret a buyer who never buys even when compensation is available as displaying deontological preferences to not create harm, as it comes at a cost of a foregone increase in own surplus.

All subjects played 12 rounds of the game in the *NoCompensation* condition. To investigate the effect of the introduction of a compensation technology on buying behavior and net externalities, participants in all treatments then played the game for another 12 rounds in the *Compensation* condition.[9] In each round, prices were randomly and independently drawn for each buyer. To rule out the possibility that accumulated gains of trade would affect buying and compensating decisions, one round was randomly drawn to determine final payments and sliders at the end of a session. It was made clear in the instructions that harmed parties were other subjects who participated in the same session. To reduce reputation building, subjects knew that buyers and harmed parties would be re-matched in every round, ensuring that no subject would be matched to the same person in two consecutive rounds. We used a market frame labeling the roles as Person A and Person B, with Person A being described as a buyer who had to decide whether to buy a good. Example instructions can be found in Appendix B.

3.2 Treatments

To manipulate diffusion of the externality, we changed the matching structure in the treatments *DiffusedHarm* and *DiffusedResponsibility* holding the total externality following purchases (i.e.,

[7] The maximal compensation amount was set to 30 ECU in the experimental software, ensuring that net externalities could not be negative.

[8] For prices between 70 and 100 ECU the evaluation of the efficiency of buying and compensating depends on the (unknown) social costs of the externality, in our setting the individual disutility of working on the slider task. We allowed buyers to spend money from their endowment to compensate the harmed party for the negative externality when the surplus created from buying was not sufficient to fully compensate. Full compensation (reducing the externality to zero) was therefore always feasible for any realization of the prize.

[9] This also reflects more closely the real world setting, where offsetting tools have been introduced only relatively recently.

number of sliders) constant. As in *NoDiffusion*, buyers took the decisions both without and with the possibility to compensate. While in *NoDiffusion* one buyer was matched with one harmed party, we varied the number of buyers and harmed parties in *DiffusedResponsibility* and *DiffusedHarm*, respectively. In *DiffusedHarm*, one buyer was matched with two harmed parties who had to solve 120 sliders each if the good was bought. In the *Compensation* condition, each point spent as compensation would then reduce the number of sliders for each harmed party by four so that the efficiency of the compensation technology was identical to the other treatments. To keep the decision environment comparable across treatments, buyers were not allowed to allocate compensation freely between the two harmed parties. In *DiffusedResponsibility*, two buyers were matched with one harmed party who had to solve 240 sliders whenever at least one of the two buyers bought the good, which is similar to the implementation rule in Falk et al. (2020). If one buyer decided to purchase the good, the full externality was imposed on the third party; whether or not the second buyer bought was then irrelevant for the externality.[10] The fact that immoral behavior of one buyer is already sufficient to generate the full externality resembles many examples related to voluntary compensation in field settings. For instance, the bulk of CO_2 emission of a flight or a bus trip is already generated when only one passenger takes the trip.

In our *DiffusedResponsibility* treatment, each point spent as compensation reduced the number of sliders by eight. If the two buyers together spent more than 30 ECU in total, in principle yielding a negative net externality, excess points expired and the externality was simply reduced to zero. To avoid that the experience over rounds would inform beliefs about the compensation behavior of other buyers and thereby affect behavior, we did not give any feedback in between rounds about the other buyer's buying and compensation decisions and the resulting externalities.[11] Again, we made sure that in any consecutive round buyers were not matched with the same subjects (harmed parties and/or buyers). Therefore, it was not possible for buyers in the *DiffusedResponsibility* treatment to coordinate; for instance, a pattern where two buyers would always buy and take turns

[10] We acknowledge that by modeling diffusion of responsibility in this way, the ratio between potential gains to buyers and externality increases relative to the other treatments.

[11] In real offsetting markets, overcompensation of one's environmental externalities are possible, simply financing additional climate change mitigation projects. However, in our experiment, to keep the size of the maximum compensation fixed from the perspective of buyers, only the harm created by the decisions of the buyers matched in a particular round could be compensated by the buyers. Repaying amounts for compensation that exceed the total externality would automatically provide feedback on the compensation behavior of the other buyer, and might in turn affect behavior. We therefore decided to let excess compensation expire.

in compensating the third party was ruled out by the matching protocol.

3.3 Hypotheses

Our first hypothesis concerns buying and compensation choices. If buyers experience moral costs as a result of creating the externality on the third party, compensation provides a possibility to assuage this moral cost. Hence, buyers who experience disutility due to their moral concerns can be expected to choose positive compensation levels to reduce the moral costs of consumption. This pattern would be in line with the positive voluntary compensation levels observed in the field and in the patterns of moral behavior found in market settings in the presence of negative externalities.

At the same time, the possibility of compensating the externality (at least partially) may have the effect of increasing buying frequencies. We assume buyers trade off their own material gain against the moral cost of inducing harm on another subject. Anticipating the compensation stage, buyers can then be expected to become more likely to purchase the good in the first stage, thus making it more likely that the externality is created. Conscience accounting in line with Gneezy et al. (2014) would amplify this effect. A related example from the field is the study by Harding and Rapson (2019), who found an increase in energy consumption by 1-3% among buyers after signing up for a green energy program that would offset the CO_2 effects.[12] Overall, we pose a two-part hypothesis:

Hypothesis 1

a) Buyers choose positive compensation levels on average.

b) The introduction of compensation leads to a higher buying frequency compared to the case when compensation is not possible.

It is important to note that it is a priori unclear whether an increased frequency of buying after the introduction of the compensation technology will lead to higher or lower net externalities compared to the case in which compensation is not possible, as the resulting externality depends on buyers' moral costs. As described in the previous subsection, given the parameters of our

[12] A related phenomenon is the rebound effect reported in the field of energy economics which is defined as an increase in energy consumption after an increase in energy efficiency, which may even offset efficiency driven energy savings (Sorrell and Dimitropoulos, 2008). While the size of the rebound effect is still debated in the literature, its existence is well-documented (Gillingham et al., 2016).

experiment, buying and fully compensating the externality would be a Pareto improvement over abstaining from buying for prices below 70 ECU.

In our setting, a voluntary reduction in the harm created can be achieved either by not buying or by compensating. We use the term social responsibility to capture both of these behaviors. Our treatment variations focus on two additional factors, which may hamper this social responsibility by reducing the tangibility of the externality: diffusion of harm and diffusion of responsibility. First, many negative externalities share the characteristic that their harm is borne by many people at the same time, e.g., carbon emissions, water pollution or noise pollution all affect many people simultaneously. That many are harmed at once reduces the identifiability of victims and hence may lead to "compassion fade", the psychological phenomenon that the likelihood to help decreases when the number of victims increases. This effect has been found in many studies in humanitarian contexts (see Butts et al. 2019 for a meta-analysis), and was also confirmed in a study which comprised vignette and behavioral results for a sub-group of non-environmentalist subjects in various environmental donation contexts (Markowitz et al., 2013). Transferred to our setting, this leads us to hypothesize that diffusion of harm reduces social responsibility relative to the baseline condition.

Similarly, diffusion of responsibility is a frequent characteristic of transactions where negative externalities are created, for example in the transport sector where voluntary offsetting schemes are currently often implemented. Diffusion of responsibility may increase the moral wiggle room for buyers at both stages of the decision situation: First, if both buyers choose to buy the good and the externality is imposed, the link between one's individual action and the harm for the third party becomes weaker. Second, and closely related, the (potentially self-serving) belief that another buyer purchases the good and triggers the externality can diminish the perceived guilt associated with consumption, as the harm is generated regardless of one's individual decision. Third, if both buyers have decided to buy the good, the compensation decision becomes a free-riding problem, and the belief of one buyer that the other buyer will compensate may lead to a reduction of one's perceived necessity for compensation relative to a situation in which the buyer decides on the compensation alone. Indeed, diffusion of responsibility has been shown to reduce prosocial and moral behavior for example in dictator games and in the study by Falk et al. (2020). Taken together, these findings lead us to hypothesize that diffusion of responsibility reduces the reaction to the negative externality.

Importantly, as described above, decision-makers in our setting can respond to the reduction of the externality's tangibility along two dimensions. They can either buy the good more frequently or reduce voluntary compensation. As it is difficult to predict ex-ante on which dimension the behavioral response should be stronger pronounced, we formulate our hypothesis about the effect of the treatments in terms of social responsibility, integrating both buying and compensation choices.

Hypothesis 2

Diffusion of harm/responsibility reduces social responsibility: It increases buying and decreases compensation (conditional on buying), relative to the situation without diffusion.

3.4 Experimental Procedures

The experiment was run in June 2019 at the BEELab at Maastricht University, after we conducted a pilot in April 2019 to calibrate experimental parameters. In total, we conducted seven experimental sessions with altogether 150 subjects. Subjects were recruited using ORSEE (Greiner, 2015) and the experiment was programmed in oTree (Chen et al., 2016). The experiment lasted on average for 1 hour and 45 minutes and payments were made in Euro with the exchange rate 10 ECU = EUR 2.50. Average earnings were EUR 24.14 (SD 7.73), with buyers earning on average EUR 29.48 (SD 7.9) and harmed parties earning EUR 18.80 (SD 0). When all subjects who were registered for a session showed up, all treatments were conducted in this session to mitigate confounds between the treatment effects and session-specific effects. Upon arrival, subjects were randomly assigned to their seats in the lab.

During the sessions, subjects were seated in cubicles with closed doors and received paper instructions. If they had questions before and during the experiment, these questions were answered privately by the experimenter. Buyers and third parties were not able to see each other as cubicles were closed. Moreover, buyers were not able to detect in which cubicles the third parties were seated either, as participants in the latter role were also active in each round – they had to answer non-incentivized belief questions about buyers' behavior.

In a session with full participation, we had 6 participants in the *NoCompensation* treatment (3 buyers, 3 harmed parties) and 9 participants in either of the diffusion treatments (3 buyers and 6 harmed parties in *DiffusedHarm*, 6 buyers and 3 harmed parties in *DiffusedResponsibility*). We call

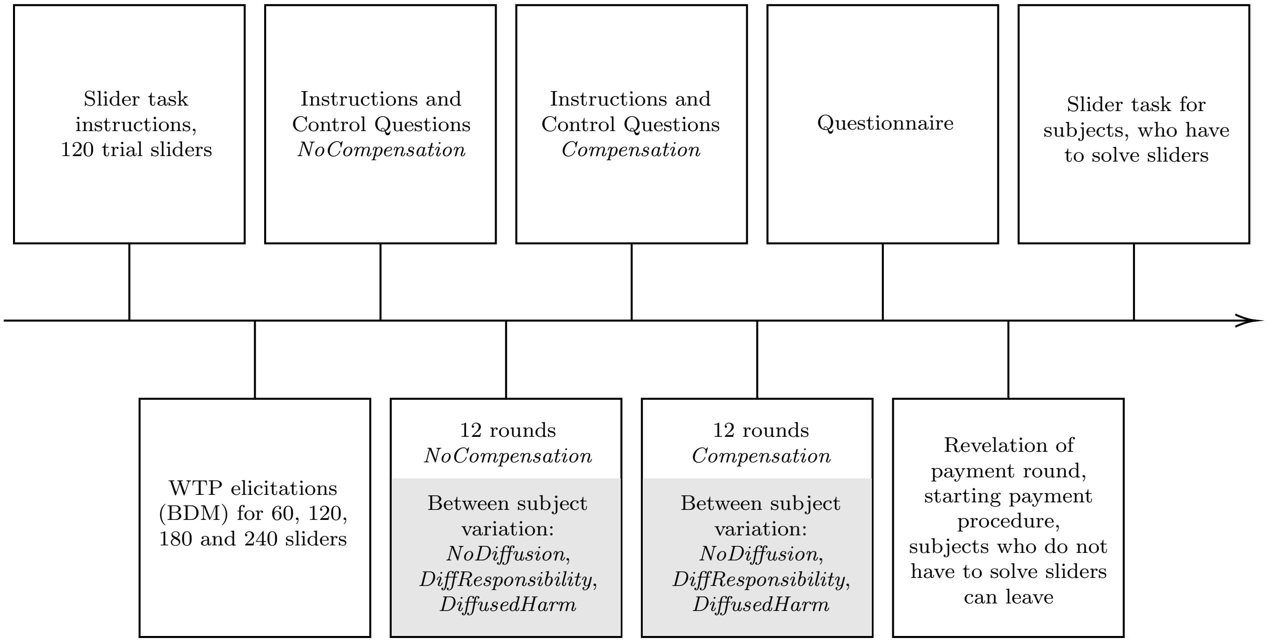

Figure 1 – Structure of an experimental session, which was the same for all subjects. Differences stemming from between subject treatments are indicated in shaded boxes. WTP refers to the willingness to pay to avoid having to solve sliders.

each of these treatment groups a cohort. Subjects were re-matched within their cohort, according to a pre-defined protocol ensuring that no one interacted repeatedly in two consecutive rounds. When not all registered subjects showed up for a session, we dropped one of these cohorts and consequentially conducted the session only with the remaining two treatment cohorts.

The structure of a session is depicted in Figure 1. At the beginning of the experiment, subjects received general instructions about the procedure of the session. We then explained the slider task and subjects had to solve 120 trial sliders to make sure they were familiar with what would later become the externality. For working on the slider task, they provisionally received 100 ECU which were used to elicit subjects' willingness to pay to avoid having to solve another set of sliders at the end of the session. Subjects had to state their willingness to pay for four different numbers of sliders (60, 120, 180, and 240), with each quantity constituting one "round"; they could bid up to 100 ECU (their proceeds from the work task) in each case. Incentive compatibility was ensured by implementing a Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964): For each number of sliders and each subject a random price between 0 and 100 ECU was drawn and compared to the subject's willingness to pay to avoid the respective number of sliders. If the

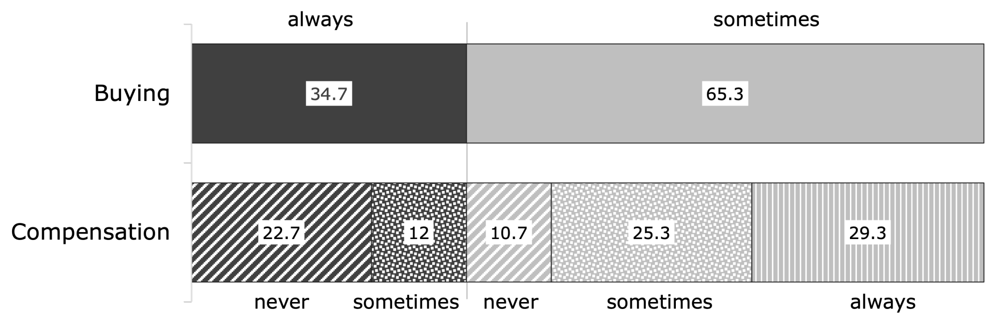

Figure 2 – Proportions of buyer types, based on buying and compensation behavior.

subject's willingness to pay was higher than or equal to this random price, the subject would have to pay the random price and not have to work on the task (in case this task was selected to determine final payments); otherwise the subject would keep the 100 ECU and work on the respective number of sliders.

After the willingness to pay elicitations, subjects received the instructions for *NoCompensation* and then had to answer control questions with example calculations. They could only proceed to the 12 rounds of the main experiment once they had correctly answered each of the control questions. Only then, subjects were informed of their role (buyer or harmed party). At the end of the *NoCompensation* part, this procedure was repeated with the instructions for the *Compensation* condition.

One round out of the 4 BDM rounds and the 24 rounds of the main experiment was randomly chosen to determine the payoffs for the subjects and possibly the externalities.[13] Before the payment round was revealed, subjects had to fill in a questionnaire with demographics and stated attitudes, for example related to general altruism and the attitudes towards ethical consumption. Subjects who had to work on the slider task then had to solve the sliders before they would receive their payments, while we already started handing out the payments of subjects who did not have to work.

4 Results

We start our analysis with aggregate buying and compensation patterns before we focus on potential effects of the experimental treatments and the determinants of individual choices. In the first step, we group participants into categories depending on their decisions in all rounds of the experiment; Figure 2 shows this classification and the corresponding shares of participants assigned to each category. The first observation is that only a minority of buyers (34.7%) buy the good in every round, although this would be optimal from a profit-maximizing perspective. The remaining subjects (65.3%) buy only in some of the rounds. No subject in our sample decides to avoid the externality entirely by never buying the product, suggesting that no subject in our sample displays strong deontological preferences to not cause any harm.

Moving to compensation behavior, we find that of those subjects who always buy, 65.4% never compensate. Hence, these subjects (22.7% of all buyers) do not seem to respond to the harm they cause to third parties. On the contrary, 44.9% of the subjects who buy only in some rounds (29.3% of the total buyer population) always compensate the third party. Overall, while we observe strong heterogeneity in buying and compensation decisions among subjects, the two diffusion treatments do not differ from the baseline condition in the distribution of buyer categories (all χ^2-tests yield significance levels of $p > 0.200$).

In the next step, we take a closer look at buying behavior. Table 1 depicts buying frequencies in the different treatment conditions. First, and in line with the previous analysis, we observe that in all treatments buyers are on average willing to forego private earnings by refraining from buying, i.e., the buying frequency significantly differs from 1 in all treatments ($p < 0.003$ in all two-sided Wilcoxon matched pair signed-rank tests comparing individual buying rates without or with compensation; in the following, all statistical tests are two sided). Once compensation is available, subjects use on average around 6 ECU to voluntarily reduce the harm they created, which is about 20% of the 30 ECU required to compensate for the full externality. In all treatments, this amount differs significantly from zero ($p < 0.001$, Wilcoxon signed-rank test comparing mean individual compensation amounts), confirming our Hypothesis 1.a).

[13] Given that subjects in the BDM tasks stated their willingness to pay prior to the main task and before the roles of buyers and sellers were assigned, these decisions were implemented irrespective of the experimental treatments *DiffResponsibility*, *DiffusedHarm*, or *NoDiffusion*, if one of the BDM tasks was selected for payments.

	Buying rate (SD)		
	NoCompensation	*Compensation*	Wilcoxon signed-rank test
NoDiffusion	73.81%	80.95%	0.059
	(0.31)	(0.21)	
DiffusedHarm	83.33%	88.43%	0.666
	(0.24)	(0.13)	
DiffusedResponsibility	87.04%	88.43%	0.198
	(0.21)	(0.19)	
Mann-Whitney U tests			
DiffusedHarm	0.396	0.398	
DiffusedResponsibility	0.105	0.121	

Table 1 – Average buying rates per treatment using subject mean buying rates across rounds, standard deviation in parentheses. The bottom two rows give the p-values of Mann-Whitney U tests, comparing buying rates in *NoDiffusion* with the respective treatments. The fourth column reports the p-values for a Wilcoxon signed-rank test of differences in buying rates between the *NoCompensation* and *Compensation* condition.

Abstaining from buying in the *NoCompensation* condition might also be an error of subjects who do not care about the externality rather than an act of social responsibility. We therefore look at the correlation between an individual's buying frequency in *NoCompensation* and an individual's compensation behavior in the second part of our experiment. We find that average compensation in the second part is highly negatively correlated with an individual's average frequency of buying (in % of rounds where the good was bought) in the first part when compensation is unavailable (Spearman's rho = -0.6504, $p < 0.001$). Hence, subjects who choose relatively high amounts for compensation are less likely to buy when the possibility to compensate does not exist, suggesting that these subjects indeed account for the externality in their buying decisions.

As described in subsection 3.3, we expected the introduction of compensation to increase the likelihood of buying (Hypothesis 1.b)). When we compare buying frequencies in *NoCompensation* and *Compensation*, we find some evidence for this hypothesis in our *NoDiffusion* condition. In this treatment, the availability of compensation increases the average frequency of buying by 7 percentage points, and (marginally) significantly so ($p = 0.059$, Wilcoxon signed-rank test comparing individual buying rates between *NoCompensation* and *Compensation*). At the same time, while the likelihood of buying also increases slightly on average in *Compensation* in the two diffusion conditions *DiffusedHarm* and *DiffusedResponsibility*, these increases are not significant, as the probability of buying is already very high in the first part of the experiment (around 85%).

	NoDiffusion (1)	*DiffusedHarm* (2)	*DiffResponsibility* (3)
Compensation	0.069*	0.048	0.021
	(0.039)	(0.047)	(0.016)
Price	-0.007***	-0.005***	-0.004***
	(0.001)	(0.001)	(0.001)
Constant	1.144***	1.152***	0.995***
	(0.081)	(0.055)	(0.151)
Observations	504	432	864
# buyers	21	18	36
Session dummies	Yes	Yes	Yes

Table 2 – Linear RE models using a binary dependent variable equal to one if the decision-maker chooses to buy the good in a particular round. *Compensation* treatment dummy takes the value of 1 in part 2 of the experiment, when compensation is available. Robust standard errors (clustered by subject) in parentheses, *** p=0.01, ** p=0.05, * p=0.10

Next, we conduct parametric analyses in which we also take into account the individual heterogeneity of experimental buyers (see Table 2). We estimate panel linear probability models with the decision to buy in a given round as the binary dependent variable, including buyer-level random effects, separately for each of the three experimental conditions. As independent variables, we include a dummy variable equal to 1 if the compensation technology was available to the buyers, and the price realized in the specific round. Also here, the introduction of compensation is associated with an increased likelihood of buying in *NoDiffusion* (Model 1), as the coefficient of the *Compensation* dummy is positive and weakly significant. In line with the results from non-parametric tests, we do not find a similar positive effect of the introduction of compensation for the two diffusion treatments (Models 2 and 3); the respective coefficients are positive but insignificant.

Moreover, across all models we observe a negative and significant price coefficient, suggesting that the lower the price, the higher the probability to buy. Note that given our experimental design, profit-maximizing subjects should buy, irrespective of the price realization. This is because the price can be at most 100 ECU, which is exactly equal to the induced valuation. Yet, the fact that subjects are price sensitive in all treatments suggests that buyers bear a moral cost of imposing a negative externality on the harmed party on top of their individual consumption utility. This leads to a trade-off between their own monetary benefits and the moral costs, which can explain why subjects are less likely to buy the good at a high price (low potential gain from buying).

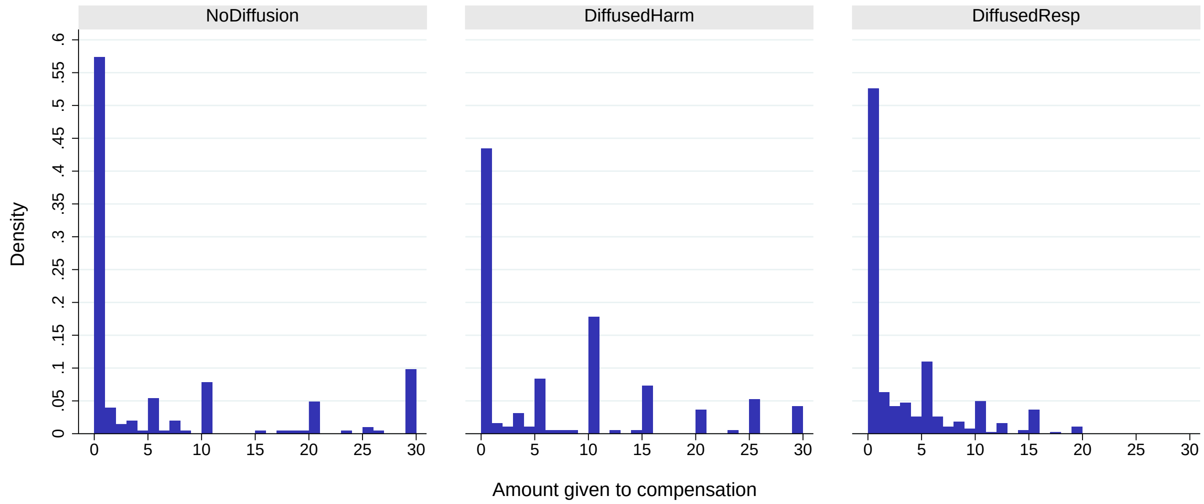

Figure 3 – Histograms of amounts given to compensation in all rounds by treatment conditional on having bought the good. Full compensation of the externality costs 30 ECU.

In the next step, we compare buying rates across treatments. As apparent from Table 1, both *DiffusedHarm* and *DiffusedResponsibility* lead to higher buying rates compared to *NoDiffusion* when compensation is not possible (83% or 87%, respectively vs. 73%) as well as when it becomes possible (both 88% vs. 81%). Yet, the differences between either of the diffusion treatments and *NoDiffusion* are not significant when comparing observations ($p > 0.100$, Mann-Whitney U tests comparing individual mean buying rates; also using regression analysis).

Overall, we can summarize our results concerning buying patterns as follows:

Result 1

The introduction of the compensation technology tends to increase buying frequency (weakly significantly) in the NoDiffusion treatment. In the treatments with diffusion, the introduction of the compensation technology does not significantly increase the probability of buying.

We now turn towards a more detailed analysis of compensation behavior. Comparing subject means of compensation across treatments, we find that subjects give on average 8.12 ECU (SD 10.84) in *NoDiffusion*, 7.78 ECU (SD 7.77) in *DiffusedHarm* and 3.71 ECU (SD 4.81) in *Diffused-Responsibility*. These differences are, however, not significant at conventional levels ($p > 0.200$, Mann-Whitney U tests comparing subject means of the diffusion treatments to *NoDiffusion*). When looking at the distributions of the amounts given to compensation across buyers and rounds, we

	(1)	(2)		(3)	
		Hurdle 1	Hurdle 2	Hurdle 1	Hurdle 2
DiffusedHarm	4.053	0.808*	-4.268	0.857	-5.024**
	(4.406)	(0.490)	(2.746)	(0.559)	(2.112)
DiffResponsibility	-2.956	0.436	-7.349***	0.454	-5.432**
	(3.626)	(0.428)	(2.130)	(0.460)	(2.471)
Price	-0.129***		-0.135***		-0.135***
	(0.009)		(0.010)		(0.010)
WTP(240 sliders)	0.217***	0.017*	0.050	0.019*	-0.021
	(0.070)	(0.009)	(0.045)	(0.011)	(0.042)
Altruism				0.051	1.875**
				(0.152)	(0.899)
Risk aversion				-0.123	2.151***
				(0.099)	(0.443)
Constant	4.084	-0.334	17.787***	-0.159	-1.280
	(5.122)	(0.390)	(2.905)	(1.215)	(7.611)
Session dummies	Y	N	Y	N	Y
Observations	777	777		777	
# buyers	75	75		75	
Model	Tobit	Hurdle		Hurdle	
Clustered SE	No	No		No	

Table 3 – RE regressions on compensation conditional on having bought the good. Model (1) describes the results of a RE-Tobit regression on compensation behavior. Models (2) and (3) give the results of a double hurdle model, where the first hurdle describes the probability of being the "compensator" type. The second hurdle in each case then uses again a panel specification on compensation, conditional on being the compensator type. *** p=0.01, ** p=0.05, * p=0.1

note that the modal compensation amount (conditional on buying) is zero (Figure 3). Furthermore, patterns seem to differ across treatments: In the *NoDiffusion* treatment, the externality is fully offset in 10% of the cases, while in the *DiffusedResponsibility* treatment the maximum amount that is ever given to compensation is 20 ECU (equivalent to two thirds of the externality).

In Table 3, we provide a parametric analysis of the determinants of compensation behavior. In the first step, we estimate a Random Effects Tobit model with the compensation in ECU chosen by the buyer as the dependent variable (Model 1). Thereby, we account for the censoring at zero of the compensation variable, and the zero-inflated nature of the data. As independent variables, we include treatment dummies as well as the price of the virtual good and a buyer's willingness to pay for not working on 240 sliders after the experiment (WTP 240 sliders; our measure for the perceived cost of the externality). In this model, we do not find significant treatment effects on overall compensation.

The zero-inflated data displayed in Figure 3 pools two different types of zeros. On the one hand, there are some subjects, who do not always compensate (for example because of the available surplus). Yet, it is also conceivable, that some subjects would never compensate, regardless of the treatment and decision environment (cf. also the heterogeneity of buyer behavior in Figure 2). Therefore, we estimate a panel double hurdle model (Model (2); Engel and Moffatt 2014), which accounts for the existence of never-compensators. In this model, the first hurdle determines whether a buyer is of the "compensator" type; that is, whether the buyer is in general willing to reduce the harm resulting from the buying decision. Conditional on passing the first hurdle, the second hurdle then analyses the determinants of the compensation amount.

Results from the first hurdle provide some indication that *DiffusedHarm* is associated with a higher likelihood of being the compensator-type, but this effect is only marginally significant ($p = 0.099$). Also, a higher perceived cost of the externality (WTP 240 sliders) tends to be associated with being compensator-type (i.e., pass the first hurdle), but only marginally significantly. Looking at compensation conditional on having passed the first hurdle, we find that the level of compensation falls both in *DiffusedHarm* and *DiffusedResponsibility*, in line with our Hypothesis 2, yet the effect is only significant for *DiffusedResponsibility*. Moreover, the model reveals a negative price effect on the amount given as compensation. This can be interpreted as an income effect, where a lower price implies a larger surplus to the buyer, which then leads to a larger amount of compensation. Finally, we find no significant effect of the willingness to pay to avoid 240 sliders for oneself on compensation.

As a further check for the robustness of the treatment effects, we add further controls for social and risk preferences to the hurdle model. First, given that in our setting buying induces a negative externality on another participant, social preferences should be a relevant motivation for the decision to compensate for the harm. Therefore, we add a survey measure for general altruism taken from Falk et al. (2018) as an independent variable to our model. Second, a buyer's risk preference might influence the decision to compensate in the *DiffusedResponsibility* treatment given that compensations exceeding the total harm expire in our setting. Therefore, we also include the general risk question from Dohmen et al. (2011) as independent variable. We re-scaled answers so that higher values indicate higher risk aversion.[14] Controlling for these preferences does not change our conclusions regarding the effect of *DiffusedResponsibility* on compensation behavior. Additionally, the treatment dummy of *DiffusedHarm* becomes now significant. Our conclusion regarding the effect of perceived harm stays the same as in Model (2) without preference controls.

We therefore arrive at the following result:

Result 2

Accounting for buyers who never compensate, diffused responsibility leads to lower compensation amounts. Diffusion of harm does not seem to robustly affect compensation.

To obtain further insights into compensation patterns in response to price variations, we look at how much compensation varies on the buyer level (as measured by the within subject standard deviation), both absolutely and as a share of the available surplus. For example, a standard deviation of zero implies that a buyer always chooses the same compensation amount. 28 buyers (37% of the buyers) do not vary compensation amounts across rounds. However, looking at their compensation amounts reveals though that 25 out of these are exactly the buyers who never compensate a positive amount. Out of the remaining 3, 2 subjects always give 30 ECU, and 1 subject always gives 25 ECU whenever reaching the compensation stage.[15] The only subjects which compensate a constant share of the available surplus are the 25 subjects who never compensate.[16] This provides further evidence that the majority of the buyers in our experiment are price-sensitive regarding their compensation decisions.

In the next step, we evaluate the effects of the treatments on the net externality (after compensation). We convert the total number of sliders harmed parties have to solve in a given round into the ECU amount necessary to offset them. In order to compare independent observations, we then calculate

[14]The altruism measure used the following wording: "How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are 'completely unwilling to share' and a 10 means you are 'very willing to share'. (...)" The risk preference question was phrased as follows "How do you see yourself: are you a person who is generally willing to take risks, or do you try to avoid taking risks? Please use a scale from 0 to 10, where a 0 means you are 'completely unwilling to take risks' and a 10 means you are 'very willing to take risks'. (...)" For the sake of interpretability as risk aversion, we inverted the scale for our analysis (0 meaning very willing to take risks, 10 meaning unwilling to take risk).

[15]Loosening this definition of "constant" giving to account for small errors, this result does not change much. For example, when we look at subjects with a standard deviation of compensation below 1, we need to account for 4 more subjects. 3 of them give on average less than 1 ECU to compensation (suggesting they mostly give nothing), the remaining buyer gives an average of 15.22 ECU across all rounds (i.e., mostly giving exactly 15 ECU).

[16]One additional buyer tends to give a constant share of the available surplus (SD < 1), compensating on average 0.56% of the available surplus.

	Net externality (SD)		Wilcoxon signed-rank test
	NoCompensation	*Compensation*	
NoDiffusion	22.14	19.31	0.022
	(4.33)	(3.86)	
DiffusedHarm	25.00	20.22	0.046
	(3.84)	(2.00)	
DiffResponsibility	29.03	23.78	0.028
	(0.82)	(2.89)	
Mann-Whitney U tests			
DiffusedHarm	0.196	0.721	
DiffResponsibility	0.012	0.046	

Table 4 – Cohort means of net externality (after compensation) in ECU equivalents by treatment, standard deviation in parentheses. The bottom two rows give the p-values of Mann-Whitney U tests, comparing net externalities in *NoDiffusion* with the respective treatments. The last column reports p-values of a Wilcoxon signed-rank test comparing net externalities between *NoCompensation* and *Compensation* for each treatment.

the mean net externality in ECU on cohort level (i.e., all the subjects in the same treatment within a session). Table 4 summarizes these cohort means.

The introduction of compensation reduces the net externality substantially and significantly in all treatments. In *DiffusedHarm*, this effect is descriptively the strongest, reducing the externality by 19% ($p = 0.046$, cohort level Wilcoxon signed-rank test), while in *DiffusedResponsibility* the externality is reduced by 18% ($p = 0.028$), and in *NoDiffusion* the externality is reduced by 13% ($p = 0.022$). Comparing net externalities across diffusion treatments, we find that *DiffusedResponsibility* leads to a significantly larger net externality both without and with the possibility of compensation ($p = 0.012$, and $p = 0.046$ respectively, cohort level Mann-Whitney U test) relative to the *NoDiffusion* treatment. The difference between *DiffusedHarm* and *NoDiffusion* is, however, not significant on conventional levels and descriptively small when compensation is possible.

Despite the non-negligible reduction induced by compensation, the externality remains on a high level in absolute terms. Across all treatments, two thirds or more of the maximum possible externality are still imposed on the harmed parties (64% in *NoDiffusion*, 67% in *DiffusedHarm* and finally, 79% in *DiffusedResponsibility*). Hence, the possibility to compensate on a voluntary basis does not remove the bulk of the negative externality in our setting.

Table 5 reports the results of linear panel regressions (with cohort level random effects) with the mean net externality in ECU per cohort and round as dependent variable. Model 1 refers to the

	NoCompensation (1)	*Compensation* (2)	Both (3)
DiffusedHarm	3.048**	1.687	3.352**
	(1.535)	(1.247)	(1.682)
DiffResponsibility	6.091***	4.447***	6.420***
	(1.167)	(1.112)	(1.265)
Compensation			-2.870***
			(0.578)
DiffHarm x *Comp*			-1.951
			(1.703)
DiffResp x *Comp*			-2.218*
			(1.206)
Price	-0.150***	-0.038*	-0.089***
	(0.035)	(0.021)	(0.025)
Constant	30.911***	21.813***	27.615***
	(1.906)	(1.775)	(1.638)
Observations	228	228	456
# cohorts	19	19	19
Session dummies	Y	Y	Y
Model	Linear	Linear	Linear

Table 5 – RE regressions on cohort level with robust standard errors (clustered by cohort) in parentheses. A cohort is the group of subjects within a session who were assigned to the same treatment and who interacted with each other repeatedly. The dependent variable is the mean net externality (after compensation) in ECU that is created in a cohort in a given round. *** p=0.01, ** p=0.05, * p=0.1

first part of the experiment, Model 2 to the second part where compensation was available. In these models, we find that the net externality in *DiffusedResponsibility* is larger than in *NoDiffusion*. We observe further that *DiffusedHarm* also tends to be associated with higher externalities when no compensation possibility is available (Models 1 and 2).

In the next step, we look into the effect of the compensation technology on the created externality in a model that includes interaction effects of compensation with both the *DiffusedHarm* and the *DiffusedResponsibility* treatment. This model includes the data for both experimental parts and, next to dummy variables for the *DiffusedHarm* and the *DiffusedResponsibility* treatments, a dummy variable for *Compensation* as well as its interaction terms with the diffusion treatments. The effects of *DiffusedHarm* and *DiffusedResponsibility* are positive and significant, indicating a higher negative externality than in *NoDiffusion* in the part without compensation. Moreover, the effect of the introduction of *Compensation* is negative and significant, highlighting that the introduction of voluntary compensation is associated with lower externalities. At the same time, the mitigating effect of compensation seems to be roughly similar across treatments: While the signs of the interaction effects are both negative, only the interaction *DiffusedResponsibility* x *Comp* is weakly significant, suggesting that the effect of the introduction of compensation is somewhat stronger in this treatment.

Finally, we observe an overall negative effect of prices on the net externality. In a framework where subjects trade-off material benefits with moral costs, the effect of prices is a priori unclear: While fewer goods are bought at higher prices (reducing the externality), higher prices also reduce compensation conditional on buying (increasing the externality). The analysis in Table 5 suggests that the effect of prices is stronger for the propensity to buy, leading to an overall negative effect of prices for the creation of the externality in our setting.

Result 3

In all treatments, compensation significantly reduces the total externalities.

Diffused responsibility leads to higher externalities both without and with compensation.

We try to disentangle what drives the effect of *DiffusedResponsibility* on the net externality. As we have seen in our previous analysis, buyers on average compensated lower amounts. At the same time, given our implementation rule in this treatment, the net externality is not purely determined by compensation behavior. Rather, the fact that the externality is created whenever at least one of the buyers buys the good already mechanically leads to a larger probability of the externality being created in the first place.[17] This can also be seen in the data: While the externality is created in the *NoDiffusion* treatment in 74% (SD 0.14; *Compensation*: 81%, SD 0.03) of the cases in the part without compensation, this share accounts for 97% (SD 0.03) in *DiffusedResponsibility* in both parts taking cohort level averages. These differences are also statistically significant ($p = 0.012$ in *NoCompensation*, $p = 0.002$ in *Compensation*, cohort-level Mann-Whitney U test).

To see whether reduced compensation in *DiffusedResponsibility* also (partially) drives the result of the larger net externality, we look at the amount of compensation that was received by a given harmed player in *NoDiffusion* and *DiffusedResponsibility*. Using a cohort-level Mann-Whitney U

[17]Even when assuming that there are no treatment differences in the propensity to buy, it is still more likely in *DiffusedResponsibility* that the externality is created. To see this note that $Pr(\text{externality created}) = Pr(\text{at least one player buys}) = 1 - Pr(\text{no player buys}) = 1 - (1 - Pr(\text{buying}))^2 \geq Pr(\text{buying})$.

test, we fail to find a significant difference between the amount of compensation received by a given harmed person ($p = 0.886$; 6.22 ECU (SD 4.38) on average in *NoDiffusion* and 5.48 ECU (SD 2.52) in *DiffusedResponsibility*). Hence, it seems that the individual reductions in compensation level out such that the overall compensation received by the harmed person is similar to the *NoDiffusion* treatment.

5 The Role of Beliefs under Diffused Responsibility

Under diffused responsibility, a buyer's belief about the behavior of other buyers may be a decisive factor for their own decision to buy and compensate for the externality. As we point out in Section 3 where we derive our hypotheses, there are various ways how diffusion of responsibility possibly affects the choice of a buyer to account for the harm created by the buying decision. In particular, moral costs (perceived guilt) of inducing the externality may be reduced by the (potentially self-serving) belief that the other buyer will also buy, as in this case harm would be created regardless of one's own decision. Also, if both subjects buy, they face a free-riding problem, and the belief that the other subject compensates might lower one's own willingness to compensate. As our original *DiffusedResponsibility* treatment cannot provide insights into the role of beliefs, we run an additional treatment *ResponsibilityBelief.* This treatment is identical to the original *DiffusedResponsibility* treatment; the only difference is that we also collect subjects' beliefs about the likelihood that the buyer they are matched with is going to buy. In the *Compensation* part of the experiment, we additionally elicit the expected compensation amount of the matched buyer.[18] We elicited subjects' beliefs in all rounds before their buying decision, which allows us to investigate whether and how own compensation behavior correlates with one's compensation beliefs.[19]

We conducted three sessions in the BEELab at Maastricht University in April 2023.[20] The

[18]The precise wording was "Assume now that the Person A you are matched with in this round has bought the good. (...) How many points (between 0 points and 30 points) will the other Person A use to reduce the number of sliders for Person B?".

[19]Incentivizing belief accuracy compared to the true realization in this setting is difficult. For probabilistic buying beliefs, it is not clear how to compare beliefs to the other's binary action (buy/not buy). Similarly, we elicited compensation beliefs conditional on the event the other buyer bought the good. Comparing this compensation belief to the realized compensation behavior of the opponent may therefore be difficult in rounds where the other buyer in fact did not buy. To keep the instructions as simple as possible, we therefore decided to not incentivize beliefs, also relying on results that incentives do not necessarily help make beliefs more accurate (Danz et al., 2022).

[20]We started one more session which had to be aborted because of technical difficulties before the main part of

experimental decision situation was implemented as in the main experiment. We collected data for 42 additional buyers linked to 21 harmed parties in 7 different matching groups. Based on the demographics we collected (age, gender) subjects in these followup sessions were not significantly different from the subjects in our main experiment. Average earnings of buyers were EUR 27.05 (SD 5.63) and for harmed parties EUR 22.19 (SD 3.01);[21] as in the previous wave of data collection, sessions lasted on average 1 hour and 45 minutes.

5.1 Results

		buy belief	buy	Wilcoxon signed-rank test
NoCompensation	Mean	63.51 %	72.82 %	< 0.001
	SD	(18.97)	(25.84)	
Compensation	Mean	66.73 %	80.75 %	< 0.001
	SD	(16.75)	(23.46)	
		compensation belief (ECU)	compensation (ECU)	
Compensation	Mean	7.05	6.64	0.577
	SD	(4.314)	(6.755)	

Table 6 – Buyers' average buying beliefs and true buying frequency in part 1 where no compensation was available, and in part 2 where compensation was available. The last row displays average compensation beliefs and true compensation (conditional on buying). The fifth column reports p-values of a two-sided Wilcoxon signed-rank test, comparing subject mean beliefs with subject mean actual behavior.

We start our analysis of the new treatments by looking descriptively at the beliefs of buyers. Table 6 displays the average believed probability that the other buyer buys the good, separately for the *NoCompensation* and the *Compensation* part. Additionally, it shows the observed relative frequency of buying, averaging over subject means across rounds. First, we observe that buyers on average underestimate the frequency of buying: For *NoCompensation* (*Compensation*) the average expected probability that the other buys accounts for 64% (67%); the actual frequency of buying is higher with 73% (81%). These differences in buyers' beliefs and actual buying propensity are significant for both *NoCompensation* and *Compensation* (both $p < 0.001$, Wilcoxon signed-rank

the experiment started.

[21] Average realized earnings are different from the main experiment because in some sessions one of the BDM rounds was drawn to determine final payments. Unlike in the main experiment, subjects were paid by bank transfer.

	Buying Behavior		Compensation Behavior	
	NoComp	*Compensation*		
	(1)	(2)	(3)	(4)
Buy_belief_prob	0.010***	0.011***		0.019
	(0.001)	(0.001)		(0.016)
Comp_belief			0.385***	0.371***
			(0.046)	(0.047)
Price	-0.000	0.001*	-0.040***	-0.031**
	(0.001)	(0.001)	(0.009)	(0.012)
Constant	0.060	0.030	1.038	-0.794
	(0.112)	(0.097)	(2.337)	(2.845)
Observations	504	504	407	407
# buyers	42	42	41	41
Model	Linear	Linear	Tobit	Tobit
Clustered SEs	Y	Y	N	N
Session dummies	Yes	Yes	Yes	Yes

Table 7 – RE regressions on buying (Model 1 and 2) and compensation behavior (Model 3 and 4). Buy_belief_prob is a buyer's probabilistic belief their matched buyer will buy the good in a given round. Comp_belief is a buyer's belief of the other's compensation amount assuming the other one has bought the good. Model 3 and 4 have fewer observations because not all subjects bought the good in every round. In addition, one subject decided to never buy the good, and hence never entered the compensation stage. *** p=0.01, ** p=0.05, * p=0.1

tests). On average, we thus have no evidence that buyers self-servingly inflate their beliefs about the buying decisions of others.

We next turn to the average belief about on the compensation amount (conditional on buying). We find that beliefs are relatively well calibrated with the true amounts – the average belief accounts for 7.05 ECU, whereas the true average compensation amount is 6.64 ECU ($p = 0.577$, Wilcoxon signed-rank test). We also calculate the expected compensation amount (multiplying the buying probability belief with the compensation belief). Naturally, this amount is lower with 5.39 ECU (SD 3.34), but also not significantly different from the true average compensation amount ($p = 0.262$, Wilcoxon signed-rank test).

In the next step, we analyze how beliefs about the behavior of others shape one's own behavior. Table 7 displays the results of models that test the correlation between buying and compensation beliefs and actual behavior of buyers. Models (1) and (2) are Random Effects linear probability models using a binary variable equal to 1 if the buyer bought the good in a particular round (zero otherwise) as the dependent variable . Model (1) refers to observations from *NoCompensation* and

Model (2) to the *Compensation* part. Besides controlling for the price, we include a subject's belief about the likelihood that the other buyer buys the good as an independent variable in the models. In both parts of the experiment, the likelihood of buying increases significantly with the expected likelihood that the other buyers also chooses to buy. This observation is in line with the notion of shared guilt between the buyers. The more likely the other subject buys, the more likely it becomes that the externality is triggered irrespective of one's own decision. We note, that unlike in the previous models, the price coefficient in Model (1) and (2) is insignificant and nearly zero. This seems to be driven by the fact that buying beliefs and prices are highly (negatively) correlated (Spearman's rho = -0.680, $p < 0.001$). Indeed, when running the regressions without the buying belief as dependent variable, the price coefficient is slightly larger and significant at the 1% level as in the previous analyses.

Models (3) and (4) consider the role of beliefs for compensation behavior. We estimate Tobit models with buyer level random effects. Besides the price, we include a buyer's belief about the compensation chosen by the other subject (conditional on the other subject having bought the good) as an independent variable (Model 3). Note that a priori, it is not clear how subjects respond to their beliefs – in principle they could see compensation as a game of strategic substitutes, where a higher contribution of other buyers means that subjects could reduce their own contributions. In contrast, we observe a positive and significant correlation between compensation belief and own compensation amount, even when controlling for prices.[22] This result speaks against the conjecture that subjects free-ride on others' compensation (which would imply a negative correlation between beliefs and compensation), but is instead in line with the interpretation that compensation in our setting relies on conditional cooperation of buyers (e.g., Fischbacher et al., 2001, Fischbacher and Gächter, 2010). A higher belief of the other's compensation amount is associated with a higher own compensation amount.[23] This conclusion remains the same if we additionally include buyer's beliefs about the other's buying probability (Model 4). Here, while the expected compensation still has a significant and positive effect on one's own compensation, the effect of one's buying belief is

[22] Beliefs as a function of prices are somewhat heterogeneous across subjects, with some believing in price sensitivity of their opponent and others having beliefs independent from prices. We describe this heterogeneity in Appendix A.1.

[23] We acknowledge that this positive correlation is also in line with egocentric belief formation, where subjects extrapolate from their own intended behavior on the other subject. It does however speak against self-serving belief formation, which would suggest that subjects strategically inflate their beliefs of the other's compensation to justify their own low compensation amounts.

not significant.

All in all, the additional treatment highlights the relevance of beliefs for buying and compensation behavior. On the aggregate, the positive correlation of beliefs about the other's buying probability with one's own buying behavior seems to be in line with shared guilt. At the same time, the decisions to compensate and thus to reduce the harm created can be interpreted in line with conditional cooperation: The higher the expected compensation of the other buyer, the higher one's own compensation on average.[24]

Finally, we note that, compared to our main treatment *DiffusedResponsibility*, the general probability of buying seems smaller in the new control treatment *ResponsibilityBelief* at least in the part without compensation ($p = 0.004$ and $p = 0.065$, respectively for *NoCompensation* and *Compensation*, two-sided Mann Whitney U test).[25] Buyers in the new *ResponsibilityBelief* treatment seem to compensate more than buyers in the main *DiffusedResponsibility* treatment ($p = 0.041$, two-sided Mann Whitney U-test). As a first test, we rule out that this is driven by a change in valuation of the externality over the time between the two experiments. In fact, subjects' average willingness to pay to avoid 240 sliders is not significantly different between the two treatments ($p = 0.924$, Mann-Whitney U test; the mean willingness to pay (ECU) was 20.50 (SD 22.88) in *DiffusedResponsibility* and 19.81 (SD 23.06)). To investigate further, we run parametric analyses focusing on the determinants of compensation, following the same approach as in the main experiment (Table A.1 in the Appendix shows the results).

As before, we use compensation as the dependent variable in the regression models and the treatment dummies as well as the price and the willingness to pay to avoid 240 sliders as dependent variables. Again, we use a Random Effects Tobit model (Model 1) and Random Effects double hurdle models (Models 2 and 3) as alternative specifications, with Model 3 additionally controlling for social and risk preferences. In the Tobit model, as before, none of the treatment dummies is significant. At the same time, all treatments, including the new *ResponsibilityBelief* treatment, are

[24] The aggregation of beliefs across buyers hides some heterogeneity that we show in Figure A.3 in the Appendix. For example, some subjects give a roughly constant amount, but believe the other one does not. For other subjects, the correlation between the expected compensation amount of the other buyer and their own compensation is positive. Notably, no subject in our sample displays a negative correlation between expected compensation and own compensation, which would be the result of self-servingly inflated beliefs on the other's compensation behavior, to justify lower own contributions.

[25] In the new treatment, the introduction of the compensation technology significantly increases the likelihood of buying, in line with Hypothesis 1 ($p < 0.001$, Wilcoxon signed-rank test).

negatively associated with compensation levels in the second part of the first hurdle model (Model 2). Once we control for social and risk preferences in Model 3, only the *DiffusedResponsibility* treatment from the main experiment remains significantly negatively correlated with the amount paid for compensation. In this model, while the sign of the coefficient for the *ResponsibilityBelief* treatment is still negative, it fails to reach significance. Hence, only in the main *DiffusedResponsibility* treatment, we find a robust negative effect on compensation behavior. Given our previous results on the link between beliefs about others' compensation level and one's own compensation, it might be possible that buyers in *ResponsibilityBelief* had on average more positive expectations on the behavior of other buyers which triggered somewhat higher compensation amounts in turn relative to the original *DiffusedResponsibility* treatment.

6 Discussion and Conclusion

We conducted an experiment to examine subjects' willingness to incur costs to compensate others for the negative externalities their purchasing decisions impose on them. While buyers are widely heterogeneous, positive compensation levels are chosen on average, suggesting that voluntary compensation can in principle be an effective means to reduce negative consumption externalities.[26] However, in our setting, buyers pay substantially less for compensation than what would be prescribed by the perceived harm. Subjects' willingness to pay to avoid working on the task reveal non-negligible social costs associated with the externality. In fact, the mean willingness to pay to avoid having to solve 240 sliders at the end of the experiment – equal to the negative externality imposed on the harmed party – is 21.73 ECU (SD 23.87), which is equivalent to EUR 5.43, a substantial share of the buyers' earnings in the main experiment (18%). Taking the perspective of a social planner, these social costs would make it optimal (conditional on a price of 70 ECU or less) that the good is bought in each round and a compensation of 22 ECU is paid as compensation to offset the externality. Yet, even in our baseline condition without diffusion, average compensation accounts for only 8.12 ECU, and thus for less than 40% of the perceived harm of the externality. Correspondingly, the net externality after compensation is still substantial

[26]Provided the compensation technology effectively reduces the externality.

in all treatments (between 64% and 79% of the maximum possible externality).[27]

Moreover, we find that compensation is price-sensitive – lower prices are associated with both higher probabilities to offset and higher amounts paid to the harmed party. That is, when there is more surplus to distribute, subjects are more willing to compensate. This price sensitivity suggests that buyers trade off material gains against the moral cost of causing the externality. At the same time, an alternative explanation for the correlation between prices and voluntary compensation might be possible. As the price in our setting is randomly determined and varies widely between periods, this might lead to moral licensing in the sense that having a "bad draw" in a previous round could be used as an excuse for the buyer for choosing not to compensate in a given round. While we cannot fully rule out this explanation for the patterns we observe, we note that a recent study by Bartling et al. (2022) finds that increases in income also lead to increases in the purchases of socially responsible goods and thus a price sensitivity of moral behavior.

Furthermore, we find that diffused responsibility for the creation of the negative externality reduces the amount given to compensation in the main experiment and thus significantly raises net externalities. At the same time, contrary to our initial hypothesis, we do not find a robust significant effect of diffused harm relative to the baseline condition. With our data, we can address one possible explanation for lack of an effect in *DiffusedHarm* based on the perceived costs of the externality. In principle, the observation that diffused harm does not lead to an increase in externalities could be explained by the fact that subjects base their decisions on a linear social cost function for the externality. In this case, imposing the 120 sliders on two harmed parties would lead to exactly the same total damage as imposing 240 sliders on one harmed party. Yet, looking at the average willingness to pay to remove different levels of the externality (60, 120, 180 and 240 sliders), we find, in line with the common assumption in the literature as well as empirical findings for the slider task (Gill and Prowse, 2019), that the aggregate social cost function appears to be slightly convex (see also Figure A.4 in the Appendix). This suggests that the total perceived harm of 240 sliders is slightly lower when imposed on two subjects rather than when imposed on one subject only.

A reason why diffusion of harm did not robustly deteriorate social responsibility in our setting

[27] This conclusion is similar to the results of Kuhn and Uler (2019), who find that although there is positive demand for compensation, it is inefficiently low as experimental traders free ride extensively.

might be that the harm is not diffused enough to induce "compassion fade". We impose the externality on only 2 subjects, and not on a larger group which might dampen the effect of diffusion. Still, our findings are in line with the results of Bartling et al. (2019) who also find only weak evidence for diffusion of harm on the market share of a socially responsible good. In their experiment, they compare a condition with diffusion of harm across 6 subjects with a condition without diffusion.

The result from our new control treatment that beliefs about other's behavior correlate positively with own behavior is in line with other findings from the literature. For instance, the study by Andre et al. (2021) described above showed that social norm intervention might be effective to increase donations to mitigate climate change. Given that experimental buyers are on average willing to increase their compensation payment if others do so as well, interventions to shift beliefs in the direction of more pro-social/pro-environmental behavior might have positive effects in the presence of responsibility diffusion.

Finally, we acknowledge that our specific laboratory setting does not capture important elements that determine compensation decisions in practice. Our setting implements a very abstract decision task, and we focus on a negative externality that is directly imposed on other subjects and where the harm imposed by each subject is clearly quantifiable. Therefore, it might be possible that some buyers in our experiment interpret the decision as an allocation task between themselves and the third parties, and we cannot fully rule out this possibility. Under this interpretation, buyers have to decide between the assignment of the (equal) payments associated with not buying and the allocation of (potentially) unequal payments, and unequal effort costs induced by the externality. In this case, the payment for compensation would be determined based on the trade-off between self-interest and inequality in money and effort. Positive amounts paid by the buyer to reduce the task for the harmed party could then be motivated by a buyer's redistributive preferences, similar to the intuitions of models of inequality aversion (Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000), incorporating inequality both in the monetary and non-monetary domain.

Pro-environmental behavior in the field might, however, also be partially influenced by social preferences and the prosocial motivation to limit the damage imposed on others by one's own choices. In fact, as mentioned before, recent research has shown that related other-regarding motivations correlate with pro-environmental behavior (Lades et al., 2021, Enke et al., 2022).

Additional evidence shows that measures for altruism and social preferences are correlated with choices to avoid externalities (Riehm et al., 2022) and with donations to projects that fight climate change (Andre et al., 2021). Indeed, as some of our analyses show, buyers' compensation levels are positively correlated with a survey measure for altruism (Falk et al., 2018) in our setting.

Our specific abstract setting focuses on the prosocial motivation to avoid (and to reduce) harm imposed on others. At the same time, consumers in the real world might also offset their environmental externalities because they care about biodiversity and value the existence of certain species per se, rather than about their indirect effect on other people's well-being. In such cases, compensation might be used irrespective of the trade-off between personal benefit and how the harm created by a specific consumption decision affects other people. More generally, compensation decisions in the field can be expected to respond to the decision maker's disutility caused by the nature of a specific externality. Therefore, an interesting avenue for further research would be to shed more light on how the willingness to compensate relates to other characteristics and types of externalities.

Acknowledgments

Financial support by the Graduate School of Business and Economics of Maastricht University is gratefully acknowledged. Ethics approval was obtained from the Ethics Review Committee Inner City Faculties (ERCIC) at Maastricht University. We thank the Associate Editor, two anonymous reviewers, Björn Bartling, Nickolas Gagnon, Katrin Gödker, Lisa Spantig and seminar participants in Maastricht, Münster, Regensburg and at the Lyon-Maastricht Behavioral Workshop for valuable comments. The authors declare no conflict of interest.

References

Andre, P., Boneva, T., Chopra, F. and Falk, A. (2021). Fighting climate change: The role of norms, preferences, and moral values, *Working paper*.

Bartling, B., Fehr, E. and Özdemir, Y. (2023). Does market interaction erode moral values?, *Review of Economics and Statistics* **105**(1): 226—235.

Bartling, B. and Özdemir, Y. (2023). The limits to moral erosion in markets: Social norms and the replacement excuse, *Games and Economic Behavior* **138**: 143–160.

Bartling, B., Valero, V. and Weber, R. (2019). On the scope of externalities in experimental markets, *Experimental Economics* **22**(3): 610–624.

Bartling, B., Valero, V. and Weber, R. (2022). The causal effect of income growth on consumer social responsibility, *Working paper*.

Bartling, B., Weber, R. A. and Yao, L. (2015). Do markets erode social responsibility?, *The Quarterly Journal of Economics* **130**(1): 219–266.

Becker, G. M., DeGroot, M. H. and Marschak, J. (1964). Measuring utility by a single-response sequential method, *Behavioral Science* **9**(3): 226–232.

Berger, S., Kilchenmann, A., Lenz, O. and Schlöder, F. (2021). Willingness-to-pay for CO2 offsets: A large-scale field study on revealed preferences in the aviation industry, *Working paper*.

Blasch, J. and Farsi, M. (2014). Context effects and heterogeneity in voluntary carbon offsetting – a choice experiment in Switzerland, *Journal of Environmental Economics and Policy* **3**(1): 1–24.

Blasch, J. and Ohndorf, M. (2015). Altruism, moral norms and social approval: Joint determinants of individual offset behavior, *Ecological Economics* **116**: 251–260.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition, *American economic review* **91**(1): 166–193.

Branas-Garza, P., Duran, M. A. and Espinosa, M. P. (2009). The role of personal involvement and responsibility in unfair outcomes: A classroom investigation, *Rationality and Society* **21**(2): 225–248.

Butts, M. M., Lunt, D. C., Freling, T. L. and Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature, *Organizational Behavior and Human Decision Processes* **151**: 16–33.

Chen, D. L., Schonger, M. and Wickens, C. (2016). oTree – An open-source platform for laboratory, online, and field experiments, *Journal of Behavioral and Experimental Finance* **9**: 88–97.

Cryder, C. E. and Loewenstein, G. (2012). Responsibility: The tie that binds, *Journal of Experimental Social Psychology* **48**(1): 441–445.

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

Danz, D., Vesterlund, L. and Wilson, A. J. (2022). Belief elicitation and behavioral incentive compatibility, *American Economic Review* **112**(9): 2851–83.

Diederich, J. and Goeschl, T. (2014). Willingness to pay for voluntary climate action and its determinants: Field-experimental evidence, *Environmental and Resource Economics* **57**: 405–429.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences, *Journal of the European Economic Association* **9**(3): 522–550.

Engel, C. and Moffatt, P. G. (2014). dhreg, xtdhreg, and bootdhreg: Commands to implement double-hurdle regression, *The Stata Journal* **14**(4): 778–797.

Enke, B., Rodriguez-Padilla, R. and Zimmermann, F. (2022). Moral universalism and the structure of ideology, *The Review of Economic Studies* **forthcoming**.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018). Global evidence on economic preferences, *The Quarterly Journal of Economics* **133**(4): 1645–1692.

Falk, A., Neuber, T. and Szech, N. (2020). Diffusion of being pivotal and immoral outcomes, *The Review of Economic Studies* **87**(5): 2205–2229.

Falk, A. and Szech, N. (2013). Morals and markets, *Science* **340**(6133): 707–711.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation, *The quarterly journal of economics* **114**(3): 817–868.

Feldhaus, C., Gleue, M., Löschel, A. and Werner, P. (2022). Co-benefits motivate individual donations to mitigate climate change, *Working paper*.

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments, *American economic review* **100**(1): 541–556.

Fischbacher, U., Gächter, S. and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment, *Economics letters* **71**(3): 397–404.

Forest Trends' Ecosystem Marketplace (2021). Markets in motion - state of the voluntary carbon markets 2021, installment 1, `https://www.ecosystemmarketplace.com/publications/state-of-the-voluntary-carbon-markets-2021/` [Accessed: 24.09.2021].

Friedrichsen, J. (2017). Is socially responsible production a normal good?, *DIW Berlin Discussion Paper 1644 / 2017*.

Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition, *American Economic Review* **102**(1): 469–503.

Gill, D. and Prowse, V. (2019). Measuring costly effort using the slider task, *Journal of Behavioral and Experimental Finance* **21**: 1–9.

Gillingham, K., Rapson, D. and Wagner, G. (2016). The Rebound Effect and Energy Efficiency Policy, *Review of Environmental Economics and Policy* **10**(1): 68–88.

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

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE, *Journal of the Economic Science Association* **1**(1): 114–125.

Hamman, J. R., Loewenstein, G. and Weber, R. A. (2010). Self-interest through delegation: An additional rationale for the principal-agent relationship, *American Economic Review* **100**(4): 1826–46.

Harding, M. and Rapson, D. (2019). Does Absolution Promote Sin? A Conservationist's Dilemma, *Environmental and Resource Economics* **73**(3): 923–955.

Irlenbusch, B. and Saxler, D. J. (2019). The role of social information, market framing, and diffusion of responsibility as determinants of socially responsible behavior, *Journal of Behavioral and Experimental Economics* **80**: 141–161.

Jakob, M., Kübler, D., Steckel, J. C. and van Veldhuizen, R. (2017). Clean up your own mess: An experimental study of moral responsibility and efficiency, *Journal of Public Economics* **155**: 138 – 146.

Kesternich, M., Löschel, A. and Römer, D. (2016). The long-term impact of matching and rebate subsidies when public goods are impure: Field experimental evidence from the carbon offsetting market, *Journal of Public Economics* **137**: 70–78.

Kesternich, M., Römer, D. and Flues, F. (2019). The power of active choice: field experimental evidence on repeated contribution decisions to a carbon offsetting program, *European Economic Review* **114**: 76–91.

Kirchler, M., Huber, J., Stefan, M. and Sutter, M. (2016). Market design and moral behavior, *Management Science* **62**(9): 2615–2625.

Kotchen, M. J. (2009). Voluntary Provision of Public Goods for Bads: A Theory of Environmental Offsets, *The Economic Journal* **119**(537): 883–899.

Kuhn, K.-U. and Uler, N. (2019). Behavioral sources of the demand for carbon offsets: An experimental study, *Experimental Economics* **22**(3): 676–704.

Lades, L. K., Laffan, K. and Weber, T. O. (2021). Do economic preferences predict pro-environmental behaviour?, *Ecological Economics* **183**: 1–10.

Lange, A. and Ziegler, A. (2017). Offsetting Versus Mitigation Activities to Reduce CO2 Emissions: A Theoretical and Empirical Analysis for the US and Germany, *Environmental and Resource Economics* **66**(1): 113–133.

Löschel, A., Sturm, B. and Vogt, C. (2013). The demand for climate protection - empirical evidence from germany, *Economics Letters* **118**(3): 415–418.

Markowitz, E., Slovic, P., Västfjäll, D. and Hodges, S. (2013). Compassion fade and the challenge of environmental conservation, *Judgment and Decision Making* **8**(4): 397–406.

Ockenfels, A., Werner, P. and Edenhofer, O. (2020). Pricing externalities and moral behaviour, *Nature Sustainability* **3**: 872–877.

Pace, D. and van der Weele, J. (2020). Cutting carbon: An experiment on uncertainty and information about CO2 emissions, *Tinbergen Institute Discussion Paper TI 2020-059-I.*

Pigors, M. and Rockenbach, B. (2016a). The competitive advantage of honesty, *European Economic Review* **89**: 407–424.

Pigors, M. and Rockenbach, B. (2016b). Consumer social responsibility, *Management Science* **62**(11): 3123–3137.

Riehm, T., Fugger, N., Gillen, P., Gretschko, V. and Werner, P. (2022). Social norms, sanctions, and conditional entry in markets with externalities: Evidence from an artefactual field experiment, *Journal of Public Economics* **212**: 104701.

Schier, U. K., Ockenfels, A. and Hofmann, W. (2016). Moral values and increasing stakes in a dictator game, *Journal of Economic Psychology* **56**: 107–115.

Schumacher, H., Kesternich, I., Kosfeld, M. and Winter, J. (2017). One, two, many – insensitivity to group size in games with concentrated benefits and dispersed costs, *The Review of Economic Studies* **84**(3): 1346–1377.

Schwirplies, C. and Ziegler, A. (2016). Offset carbon emissions or pay a price premium for avoiding them? A cross-country analysis of motives for climate protection activities, *Applied Economics* **48**(9): 746–758.

Sorrell, S. and Dimitropoulos, J. (2008). The rebound effect: Microeconomic definitions, limitations and extensions, *Ecological Economics* **65**(3): 636 – 649.

Sutter, M., Huber, J., Kirchler, M., Stefan, M. and Walzl, M. (2020). Where to look for the morals in markets?, *Experimental Economics* **23**(1): 30–52.

Appendix

A Additional Analyses

A.1 Belief Dependence on Prices

Below we plot each buyer's probabilistic belief of their opponent's buying behavior. Most subjects believe that higher prices lead to lower buying probabilities, so they believe that other's will face a trade-off between own benefit and moral costs of creating the externality. Some subjects believe their opponents will buy independent of the price, and some (e.g., quite clearly so for subject 185) believe their opponent will always buy for any price < 70 ECU, and only afterwards be price sensitive. This observation could be explained by the parameters of our setting, where buying and fully compensating would surely be a Pareto improvement for prices below 70 ECU, whereas for prices above 70 ECU the welfare effect of buying and (partially) compensating depends on the perceived social costs of the externality.

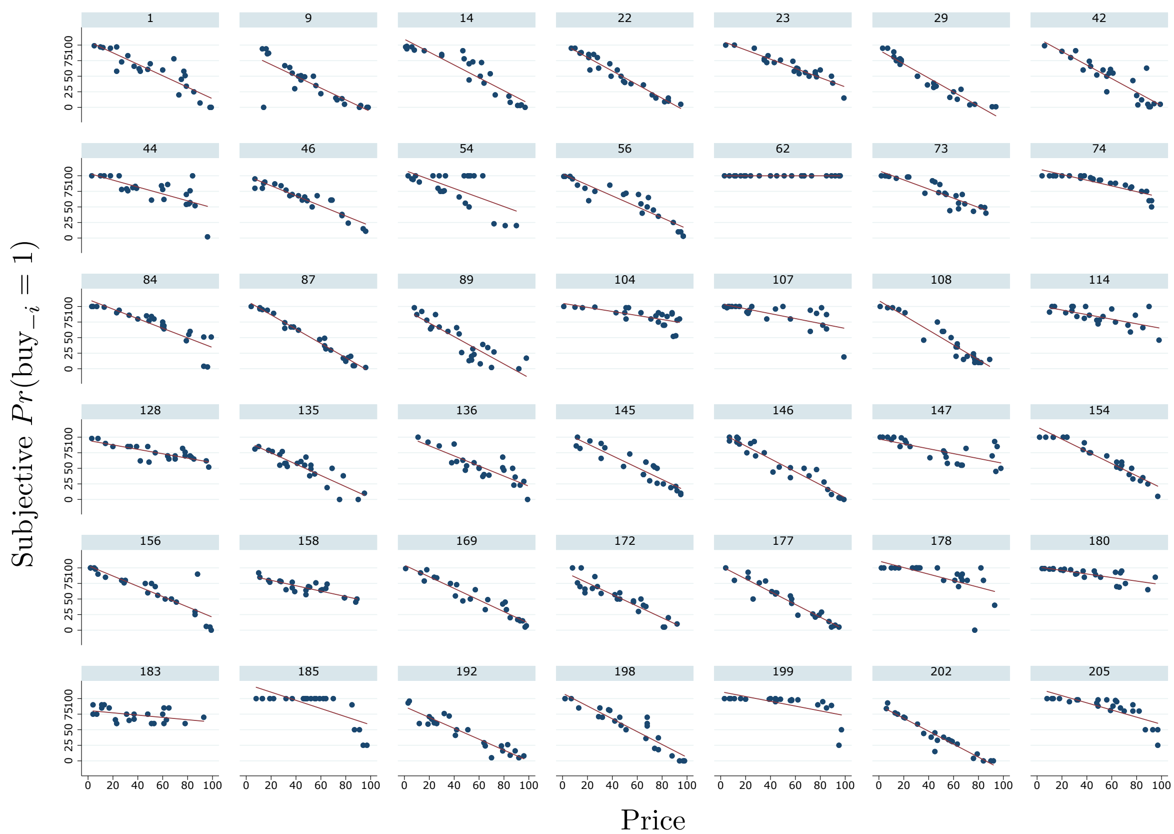

Figure A.1 – Correlation of price (x-axis) with subjective belief the other subject will buy (y-axis), for each subject.

Moving on to the correlation between prices and compensation beliefs, we find that the majority of subjects seems to believe that compensation behavior reacts to prices (and available surplus). As Figure A.2 below shows, higher prices tend to be associated with lower expected compensation amounts. This suggests that some subjects have a model in mind in which others trade-off the personal benefit from buying and creating the externality with some moral costs. Looking closer at the data, this is not the case for all subjects though. For example, some subjects believe that the other subject will never compensate, regardless of the price. Yet again, some other subjects have constant positive beliefs, indicating they think others will contribute the same positive amount independent of the price (one distinct example of this pattern is subject 192).

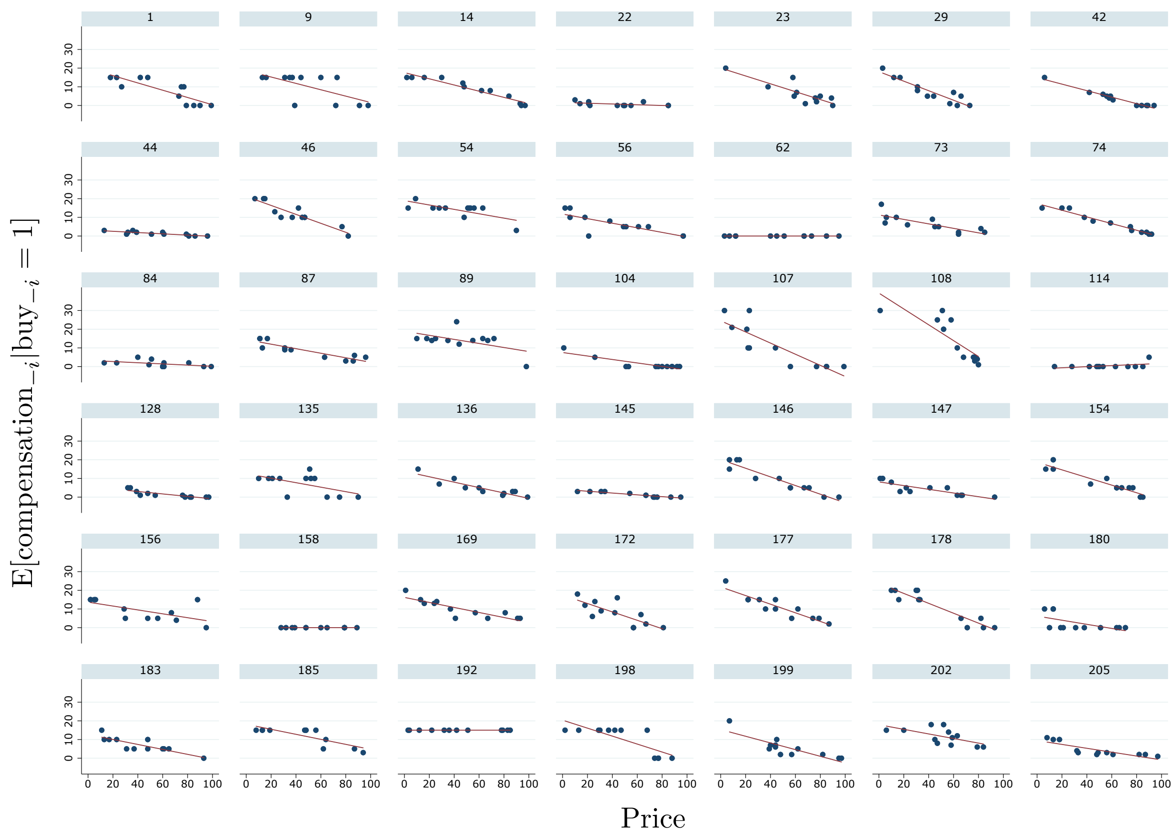

Figure A.2 – Correlation of price (x-axis) with expected compensation of the other buyer, conditional on having bought the good (y-axis), for each subject.

A.2 Belief Heterogeneity

Below we plot for every buyer in the *BeliefResponsibility* treatment, how their compensation amount correlates with their belief of the other's compensation amount (conditional on buying). We include the cases where we do not have an observed compensation amount, as the buyer decided not to buy in a given round at $y = -10$ in lighter grey. Variation in the expected compensation amount is mostly driven by variation in prices across rounds (see also Figure A.2, in the previous section). It becomes apparent that some subjects consistently do not compensate (amounts around 0), while having some variation in beliefs of their opponent's compensation behavior. Others display a clear positive correlation between compensation belief and own compensation. Finally, one group of subjects almost never compensate, and believe their opponents will do the same.

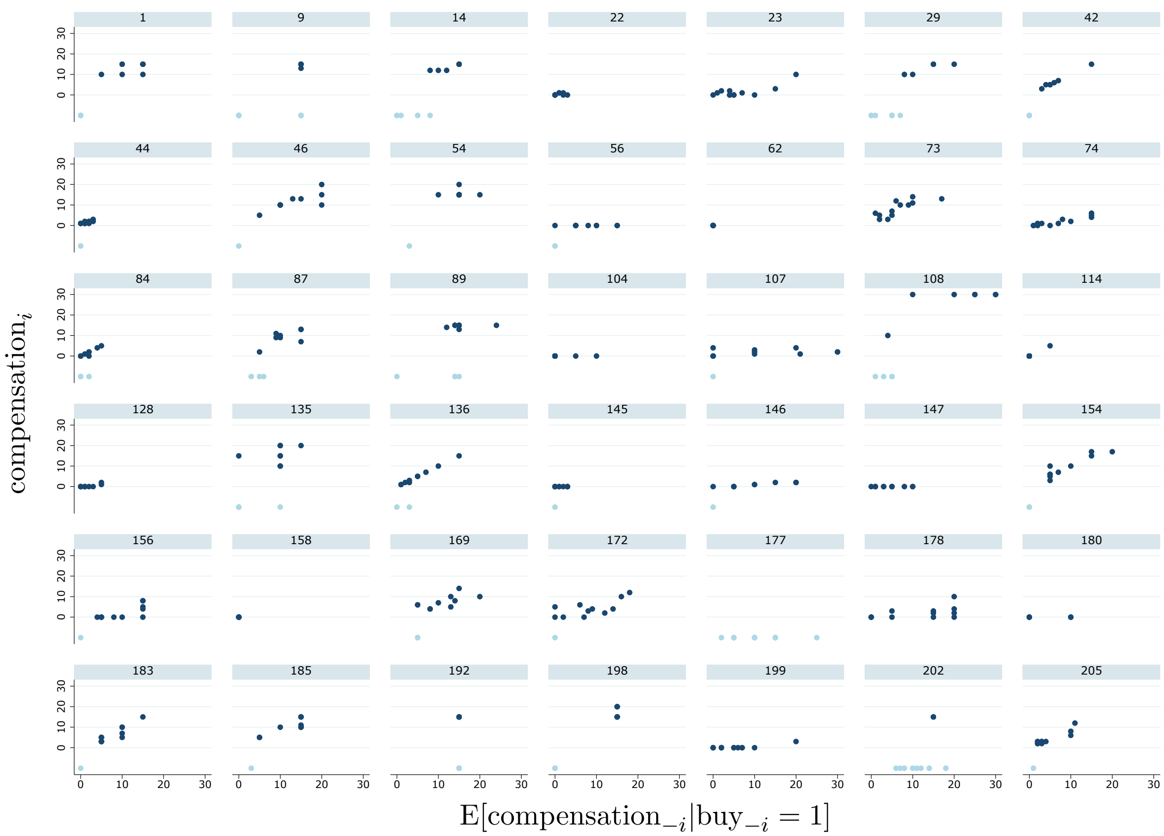

Figure A.3 – Correlation of belief of other's compensation amount conditional on buying (x-axis) with own compensation amount (y-axis) on subject level. Light blue dots are included at $y = -10$ to also display rounds where subjects did not enter the compensation stage, as they did not buy the good.

A.3 Regression Analysis of Compensation Behavior Including *ResponsibilityBelief* Treatment

Below we document the results of the regressions on compensation behavior using the full sample including the *ResponsibilityBelief* treatment. Specifications of the three models are the same as in the main text, only adding a treatment dummy for *ResponsibilityBelief*. *ResponsibilityBelief* is associated with lower compensation amounts similar to the *DiffusedResponsibility* treatment, according to Model (2). In Model (3), when including additional controls for economic preferences, while descriptively the effect persists, it is not significant anymore.

	(1)	(2)		(3)	
		Hurdle 1	Hurdle 2	Hurdle 1	Hurdle 2
DiffusedHarm	2.384	0.811*	-8.344***	1.151	-3.897
	(4.068)	(0.490)	(2.075)	(0.720)	(2.641)
DiffResponsibility	-3.645	0.398	-9.307***	0.538	-8.708**
	(3.348)	(0.406)	(2.335)	(0.702)	(3.556)
ResponsibilityBelief	-2.788	0.785*	-9.843***	7.685	-12.367
	(5.709)	(0.419)	(3.084)	(374.03)	(9.603)
Price	-0.117***		-0.124***		-0.123***
	(0.006)		(0.007)		(0.007)
WTP(240 sliders)	0.116**	0.005	0.062***	0.057*	0.014
	(0.052)	(0.007)	(0.022)	(0.034)	(0.034)
Altruism				-0.110	2.297***
				(0.214)	(0.352)
Risk aversion				-0.245*	1.936***
				(0.139)	(0.461)
Constant	6.943	-0.025	19.811***	1.156	-4.509
	(4.595)	(0.353)	(2.463)	(1.694)	(5.946)
Session dummies	Y	N	Y	N	Y
Observations	1,184	1,184		1,184	
# buyers	116	116		116	
Model	Tobit	Hurdle		Hurdle	
Clustered SE	No	No		No	

Table A.1 – RE regressions on compensation conditional on having bought the good. Model (1) describes the results of a RE-Tobit regression on compensation behavior. Models (2) and (3) give the results of a double hurdle model, where the first hurdle describes the probability of being the "compensator" type. The second hurdle in each case then uses again a panel specification on compensation, conditional on being the compensator type. *** p=0.01, ** p=0.05, * p=0.1

A.4 Social Cost of the Slider Task - WTP

The graph displays average willingness to pay in ECU to avoid having to work on the slider task as a function of different numbers of sliders, pooling across all player roles (i.e., including buyers and harmed subjects) in the main experiment. Willingness to pay was elicited using a budget of 100 ECU. Whiskers display standard errors. The average perceived social cost associated with the slider task is slightly convex.

As a rough indication for this convexity, we apply the idea of Jensen's inequality that for a convex function, a convex combination of its points are above the points of this function. Therefore, we check whether the convex combination of the average willingness to pay to avoid the lowest (60) and highest (240) number of sliders is larger than the average willingness to pay to avoid either 120 or 180 sliders. This is true for both 120 and 180 sliders, as 0.67*WTP(60) + 0.33*WTP(240) > WTP(0.67*60+ 0.33*240) and 0.33*WTP(60) + 0.67*WTP(240) > WTP(0.33*60+0.67*240), with WTP indicating subjects' average willingness to pay in order to avoid working on the respective number of sliders. Hence, it seems that the total perceived harm of 240 sliders is slightly lower when imposed on two subjects rather than when imposed on one subject only.

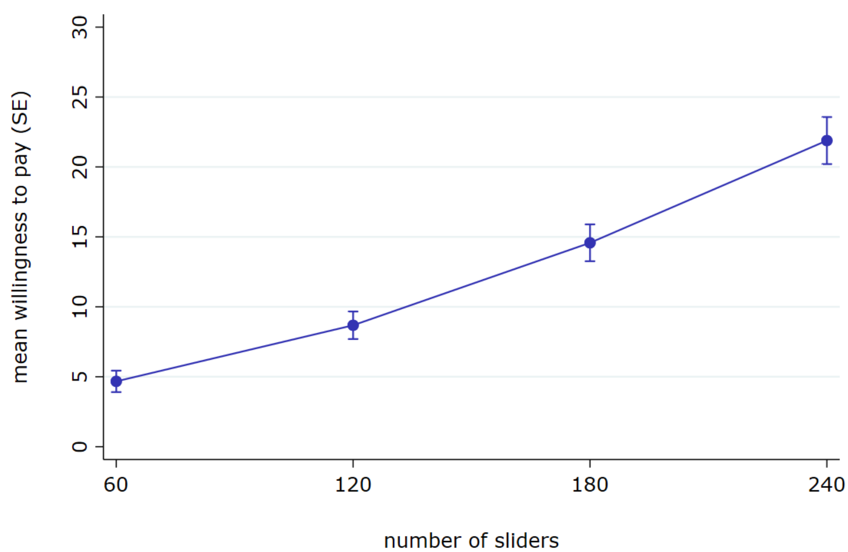

Figure A.4 – Estimated social cost of the slider task for various numbers of sliders. Points display average willingness to pay to avoid having to solve x sliders at the end of the experiment.). Whiskers display standard errors.

B Experimental Instructions

Below we include the full experimental instructions for the treatment condition *NoDiffusion*. Instructions for the other treatments were formulated in a similar way. We handed out the different parts of the instructions only once they were relevant.

l

General Instructions

Welcome to this experiment! In this experiment, you can earn money depending on the choices that are made by you and/or the other participants during the experiment. Therefore, please read the instructions carefully to make sure that you understand what decisions you can take and what the consequences of these decisions are. Your decisions during the experiment are entirely anonymous. If you have any questions about the instructions or during the experiment, please raise your hand and the experimenter will come to you to answer your question in private.

During the study, you are not allowed to talk to the other participants. Please also respect that no drinks or food are allowed inside the lab. Please switch off your mobile phone now.

Throughout the experiment, we will not talk of Euros but instead of points, you can earn. This means that your entire earnings will be, at first, calculated in points. At the end of the experiment, we will calculate your earnings in Euros, according to the following exchange rate

10 points = 2,50 Euro.

The experiment consists of three parts. The first part consists of one work task, which is followed by four rounds of a decision task. The second and third part then consist of twelve rounds each. In each of these rounds, you will face one decision task. At the end of the experiment, one of these 28 rounds (4 rounds from part 1 + 12 rounds from part 2 + 12 rounds from part 3) will be randomly selected, with each round being equally likely. The points resulting from the choices of the participants in this randomly selected round will be converted to Euro and paid out to you. No other participant will be informed about what you earn.

In the following, you find the instructions for the working task of the first part of the experiment. At the end of the work task, we will distribute the instructions for the decision task of part 1. The instructions for the second part of the experiment will be distributed after the end of the first part. At the end of the second part of the experiment, you will receive the instructions for the third part.

If you have understood this part of the instructions, please go on to the next page to read the instructions for the work task of part 1.

Instructions Part 1

The work task

The work task you have to solve is called slider task. In the slider task, a set of sliders is presented on your screen. You can adjust the slider to any number between 0 and 100 by clicking on the slider with your mouse and dragging it to the desired position. Your task is to drag the slider to the target position. The target position for each slider is the number 50. The red number on the right hand side of the slider tells you the current position of your slider. A slider is correctly adjusted when the current position is equal to 50.

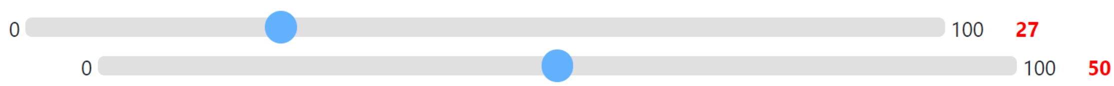

Figure 1 Two example sliders

For example, in the upper slider, the current position of the slider is 27. The slider is therefore not correctly placed yet. In the second example, you can see that the slider is correctly placed, as the position of the slider is directly at the target position. All sliders you have to solve will be displayed on one screen.

In the first part of the experiment, we ask you to solve 120 sliders to proceed to the next part of the experiment. You receive an initial payment of 100 points for working on the task.

Depending on the decisions you make in the four rounds of the decision task, your payment for part 1 of the experiment might differ from these 100 points, as will be explained after you have finished the slider task.

Do you still have any questions? If so, please raise your hand. We will then come to your seat to answer your question in private. If all the instructions are clear, please wait until we give you the password, which allows you to proceed.

l

Instructions Part 1

The decision task

Your initial payment for working on the slider task is 100 points.

In four rounds, we ask you now to state **the maximum number of points (between 0 points and 100 points) you are willing to pay** in order to avoid that you have to solve even more sliders at the end of the experiment.

In each round, a number X between 0 and 100 is drawn at random for each participant. Each number is equally likely to be drawn.

- If the number of points you stated is **greater than or equal to this randomly drawn number X, then you will pay X points**. In this case, you will not have to solve extra sliders at the end of the experiment. Instead, if the particular round is selected, we will give you your payoff for this round and you can leave the laboratory early. Your payoff for the round will then be:

  100 points – X points

- If the number of points you stated is **smaller than this randomly drawn number X, then you will keep the full 100 points from the working task**. However, you will then have to solve extra sliders at the end of the experiment, if the particular round is selected. Your payoff for the round will then be:

  100 points

Thus, the higher your answer in a given round, the lower the probability that you will have to solve additional sliders at the end of the experiment.

- In the first round, we ask for the maximum number of points you are willing to pay to avoid 60 more sliders.
- In the second round, we ask for the maximum number of points you are willing to pay to avoid 120 more sliders.
- In the third round, we ask for the maximum number of points you are willing to pay to avoid 180 more sliders.
- Finally, in the fourth round, we ask for the maximum number of points you are willing to pay to avoid 240 more sliders.

Note that your answer in a given round must be at least as high as or higher than the number, you stated in the round before.

Do you still have any questions? If so, please raise your hand. We will then come to your seat to answer your question in private. If all the instructions are clear, please wait until we give you the password, which allows you to proceed.

I

Instructions Part 2

There are two types of participants in part 2: Person A and Person B. The roles of Person A and Person B will be randomly assigned prior to the start of this part of the experiment. Each participant will keep her role throughout the second part.

Person A acts in the role of a buyer. Person B does not make a decision herself but can be affected by the buying decision of Person A.

Part 2 comprises 12 rounds. In every round, one Person A will be randomly matched with one Person B. It is ensured that the same Person As and Person Bs will never interact with each other in two consecutive rounds.

The decision task

Person A's decision

In each round, Person A and Person B are endowed with 75 points. Person A must then decide whether or not she wants to buy a fictitious product. Getting the product is worth 100 points to Person A. This means that when Person A decides to buy the product, she will receive an additional 100 points.

If Person A decides to buy the product, she has to pay the price for the product. The price is randomly determined in every round and can range between 1 and 100 points. Every price is equally likely.

If Person A decides to buy the product, this has consequences for the Person B she is matched with. If Person A buys the product, Person B will have to work on a task: At the very end of the experiment, after Person A has received her payoff for the experiment and left the laboratory, Person B will have to stay longer in the laboratory and solve altogether 240 sliders of the slider task in order to receive her payoff. The slider task will work exactly like in the first part of the experiment.

If, on the other hand, Person A decides not to buy the product, both Person A and Person B will keep their 75 points endowment, and Person B does not have to work on the slider task at the end of the experiment.

I

Payoffs

The payoffs in points in each round are determined as follows:

Person A's payoff:

If Person A decides to buy the product,

- Person A will get the 75 points she was endowed with. Additionally, she will get the 100 points the product is worth to her and pay the price of the product. Hence, Person A's payoff is

$$75 + 100 - price$$

If Person A decides not to buy the product,

- Person A will get the 75 points she was endowed with.

Person B's payoff:

If Person A decides to buy the product,

- Person B will get the 75 points endowment only if she solves 240 sliders in the slider task at the end of the experiment, after Person A has left the laboratory. If she does not solve 240 sliders in the slider task, she will not get any payoff.

If Person A decides not to buy the product,

- Person B will get the 75 points she was endowed with.

Once Person A has made her choice, a round ends and Person B is informed of the choice of Person A in the current round. After that, Person A is matched to a different Person B and the next round starts.

At the end of the experiment, we will start with the payment of those participants who do not have to solve any sliders. Every participant who has received her payment can leave the laboratory. Participants, who have to solve sliders, will work on the task and be paid afterwards.

<u>Illustrative Example 1</u>

Suppose the price of the product is 30 points in a given round. Person A decides to buy the product for the current price of 30 points. The final payoff in this round for Person A is therefore 145 points (= 75 points endowment + 100 points value of the good – 30 points price of the good). The final payoff in this round for Person B is 75 points. Additionally, Person B has to solve 240 sliders in the slider task at the end of the experiment to receive her payoff.

<u>Illustrative Example 2</u>

Assume now that the price of the product is 57 points in a given round. Person A decides not to buy the product. The final payoff in this round for Person A is therefore 75 points (= endowment). Person B's final payoff will be 75 points and she does not have to solve any sliders at the end of the experiment.

Do you still have any questions? If so, please raise your hand. We will then come to your seat to answer your question in private. If all the instructions are clear, please wait until we give you the password, which allows you to proceed with a few comprehension questions on the screen.

Instructions Part 3

Part 3 comprises 12 rounds. You keep the role you were assigned in part 2 (Person A or Person B) also in this part.

In every round, one Person A will be randomly matched with one Person B. It is ensured that the same Person As and Person Bs will never interact with each other in two consecutive rounds.

The decision task

In this part of the experiment, the decision task consists of two stages. The first stage is the same as the decision task in part 2. In this part of the experiment, however, there will be a second stage if Person A decides to buy the product in the first stage.

Person A's decision

Stage 1

As in part 2, Person A and Person B will be endowed with 75 points in each round. Person A must then decide whether or not she wants to buy a fictitious product. Getting the product is, again, worth 100 points to Person A. This means that when Person A decides to buy the product, she will receive an additional 100 points.

If Person A decides to buy the product, she has to pay the price for the product. The price is randomly determined in every round and can range between 1 and 100 points. Every price is equally likely.

As in part 2, if Person A decides to buy the product, this has consequences for the Person B she is matched with. If Person A buys the product, Person B will have to work on a task: At the very end of the experiment, after Person A has received her payoff for the experiment and left the laboratory, Person B will have to stay longer in the laboratory and solve altogether 240 sliders of the slider task in order to receive her payoff.

If, on the other hand, Person A decides not to buy the product, both Person A and Person B will get their 75 points endowment, and Person B does not have to work on the slider task at the end of the experiment.

Stage 2

If Person A decided to buy the good in stage 1, she enters stage 2. In stage 2, Person A can decide to use points of her payoff to reduce the number of sliders Person B has solve at the end of the experiment. Specifically, for each point Person A spends, Person B has to solve 8 fewer sliders.

I

Payoffs

The final payoff in one round depends on Person A's choice in stage 1 and stage 2. The payoffs in points in each round are determined as follows:

Person A's payoff:

If Person A decides to buy the product,

- Person A will get the 75 points she was endowed with. Additionally, she will get the 100 points the product is worth to her and pay the price of the product. Moreover, she has to pay the points she decided to spend in order to reduce the number of sliders for Person B. Hence, Person A's payoff is

$$75 + 100 - price\ - points\ spent\ to\ reduce\ number\ of\ sliders\ for\ Person\ B$$

If Person A decides not to buy the product,

- Person A will get the 75 points she was endowed with.

Person B's payoff:

If Person A decides to buy the product,

- Person B will get the 75 points endowment only if she solves

$$(240 - 8\ \text{x}\ points\ spent\ by\ Person\ A)$$

  sliders in the slider task at the end of the experiment, after Person A has left the laboratory. If she does not solve the sliders in the slider task, she will not get any payoff.

If Person A decides not to buy the product,

- Person B will get the 75 points she was endowed with.

Once Person A has made her choice, a round ends and Person B is informed of the choice of Person A in the current round. After that, Person A is matched to a different Person B and the next round starts.

At the end of the experiment, we will start with the payment of those participants who do not have to solve any sliders. Every participant who has received her payment can leave the laboratory. Participants, who have to solve sliders, will work on the task and be paid afterwards.

Illustrative Example 1

Suppose the price of the product is 10 points in a given round. Person A decides to buy the product, which gives her a payoff of 165 points (= 75 points endowment + 100 points value of the product – 10 points price of the product). Since Person A bought the product, Person B will have to solve 240 sliders in the slider task at the end of the experiment to receive her payoff.
Person A decides to use 10 points of her payoff to reduce the number of sliders by 80 sliders (= 8 sliders per point x 10 points). Instead of 240 sliders, Person B therefore has to solve 160 sliders (= 240 sliders – 80 sliders) at the end of the experiment in order to get a payoff of 75 points. Person A's final payoff in this round is thus 155 points (= 165 points from stage 1 – 10 points from stage 2).

I

<u>Illustrative Example 2</u>

Suppose now the price of the product is 50 points in a given round. Person A decides to buy the good, which gives her a payoff of 125 points (= 75 points endowment + 100 points value of the product – 50 points price of the product). Therefore, Person B will have to solve 240 sliders at the end of the experiment to receive her payoff.
Suppose now that in stage 2, Person A decides to use 30 points of her payoff to reduce the number of sliders by 240 sliders (= 8 sliders per point x 30 points). Instead of 240 sliders, Person B therefore has to solve no sliders (= 240 sliders – 240 sliders) at the end of the experiment to get a payoff of 75 points. Person A's final payoff in this round will be 95 points (= 125 points from stage 1 – 30 points from stage 2).

<u>Illustrative Example 3</u>

Suppose now the price of the product is 23 points in a given round. Person A decides to buy the good, which gives her a payoff of 152 points (= 75 points endowment + 100 points value of the product – 23 points price of the product). Therefore, Person B will have to solve 240 sliders at the end of the experiment to receive her payoff.
Suppose now that in stage 2, Person A decides not to use any points of her payoff to reduce the number of sliders for Person B. Therefore, Person B has to solve the full 240 sliders at the end of the experiment to get a payoff of 75 points. Person A's final payoff in this round will be 152 points (= 152 points from stage 1 – 0 points from stage 2).

Do you still have any questions? If so, please raise your hand. We will then come to your seat to answer your question in private. If all the instructions are clear, please wait until we give you the password, which allows you to proceed with a few comprehension questions on the screen.