MAX PLANCK INSTITUTE
FOR RESEARCH ON
COLLECTIVE GOODS

Florian Baumann
Tim Friehe
Pascal Langenbach

Discussion Paper
2020/8

FINES VERSUS DAMAGES: EXPERIMENTAL EVIDENCE ON CARE INVESTMENTS

Fines versus Damages: Experimental Evidence on Care Investments

Florian Baumann Tim Friehe Pascal Langenbach

July 17, 2023

Abstract

This paper studies the differential effects of fines and damages on people's investment in accident prevention. We report results from a series of experiments in which the level of monetary transfers after an accident is maintained across the two policy instruments, such that standard theory predicts the behavioral irrelevance of the instrument choice. However, we find that fines induce lower care investments than damages when fine revenue is used for prosocial ends. We discuss possible behavioral channels that may explain our findings.

1. INTRODUCTION

"The economic essence of tort law is its use of liability to internalize externalities created by high transaction costs" (Cooter and Ulen 2016, p. 190). Accordingly, threatening individuals with the obligation to cover damages is the most common instrument for controlling harmful behavior in the law-and-economics literature on tort law. However, it is commonly accepted that liability is only one instrument among many that may be used to control harm. Other formidable instruments are the corrective tax, which demands that the agent pays the state an amount equal to the expected harm *before* an accident, and a fine regime, which requires a payment equal to the harm level *after* an accident (e.g., Shavell 2007, 2011).[1] Theoretically, a fine regime induces the same level of care as a liability regime when the only difference between the two regimes is who receives the injurer's payment; assuming that injurers maximize expected utility from their payoffs, the regime choice is behaviorally irrelevant. In this paper, we test this suggested equivalence experimentally.

Using a laboratory experiment, we compare injurers' care investments when strictly liable injurers owe damages to their victim after an accident to care investments when injurers must pay a fine to society. Our treatment variation changes only who eventually receives the injurers' payment. The amount injurers pay after an accident is kept constant across treatments. In the laboratory, we can observe the injurer's incurred cost of precautions, which are impossible to track in the field (Guerra and Parisi 2022, van Velthoven 2009). Moreover, we can maintain the decision-making context across different regimes (e.g., regarding how the accident probability responds to variations in the level of care). In addition, we can collect information on potential injurers' pro-sociality, risk attitudes, and justice perceptions, as well as on their individual beliefs about what other participants expect them to do.

[1] In practice, some fines are already triggered by the violation of a regulation, that is, independent of the occurrence of an accident. Such fines are, for example, important in the context of road traffic (e.g., Goerke 2003). Fines that are due independent of the realization of an accident may be related to tort payments for creating expected harm (e.g., Miceli and Segerson 2005). In our experiment, we focus on fines due after an accident.

The theoretical literature that builds on the behavioral-irrelevance assumption simply states that fines are transfers from harm-doers to society (e.g., Becker 1968). Yet, when bringing this regime to the laboratory, different possibilities exist concerning how to represent "a transfer to society". We start from the notion that the fine is a transfer to the government, which provides public goods. Since charities also provide public goods (e.g., Karlan and List 2020, Kessler 2017), we use this relevant similarity between the state as the common fine-revenues recipient and charities by implementing the "transfer to society" as a transfer to a charity in our main experiment. An alternative interpretation of "a transfer to society" is that fine revenues allow for lower taxes from which all citizens benefit.[2] In a follow-up study, we implemented a fine regime by redistributing fine revenues to all participants.

Because monetary payoffs are constant across treatments, the standard model predicts that equilibrium care will be independent of whether injurers are subject to fines or damages after an accident. However, using insights from behavioral economics, we may think that individuals perceive regime differences in other ways: For example, while notions of fairness that require the injurer to pay for the harm caused are served in both scenarios, concerns about victim compensation are addressed only under liability (e.g., Kaplow and Shavell 2001). On a different note, a damages regime may be comparable to a fines regime that redistributes the fine to other participants because, in both regimes, peers will ultimately benefit from the injurer's payment. This comparability may be lost when the fine revenue is devoted to a charity. In this case, the use of the injurer's payment after an accident may be considered more valuable under a fine regime. Fines and damages may therefore induce asymmetric behavioral incentives.

Our results show that using damages for controlling harmful behavior can produce different levels of care than using fines – even though the injurer's expected payments are the same across

---

[2] This line of argument is similar to the double-dividend hypothesis in the context of pollution taxes (e.g., Fullerton and Metcalf 1998).

instruments. In our data, injurers who owe damages in the event of harm invest significantly more money in harm prevention than injurers who are required to pay a fine that will later be transferred to a charity. In contrast, when fine revenues are redistributed to the experimental participants, injurers' care investments in the fine regime are statistically indistinguishable from those in the damages regime.

Our paper complements the previous literature on the relative performance of different instruments to control harmful externalities. For instance, this literature has focused on potential informational or administrative differences between the instruments (e.g., Shavell 2013), assuming that instruments with the same financial incentives will induce the same behavior. Our findings highlight that the choice of who receives the injurers' transfers may critically influence injurers' care incentives even if financial obligations are constant across regimes.

By showing that the fine regime induces higher care incentives when fine revenues are redistributed to the participants instead of to a prosocial end in the form of a charity, we also contribute to the literature on the effects of earmarking policy proceeds. For example, in environmental economics, earmarking carbon tax revenues for environmental purposes tends to raise the public's acceptance of carbon taxes (e.g., Gevrek and Uyduranoglu 2015). Our results suggest that this greater acceptability of earmarked carbon taxes may come at the expense of a reduced behavioral effect. For the criminal context, Baumann et al. (forthcoming) show that different uses of fine revenue can change the deterrent effect of a fixed detection probability.

The structure of the paper is as follows. Next, we discuss the related literature. In Section 3, we present the experimental design and the procedures. In Section 4, we offer our results. We discuss the results and deliver the outcomes of follow-up studies we conducted to assess the robustness of the findings in Section 5. We conclude with a summary of our results and some policy implications of our research in Section 6.

2. LITERATURE

Our paper compares care levels of potential injurers obliged to pay either a fine or damages after an accident. So far, only a small experimental literature examining the behavioral effects of different liability rules exists (e.g., Angelova et al. 2014, Deffains et al. 2019, Guerra and Parisi 2022, Kornhauser and Schotter 1990, 1992, Wittman et al. 1997).

In the first paper in this area, Kornhauser and Schotter (1990) found that strict liability and negligence induce different care choices, although theory predicts behavioral equivalence. In their data, the negligence rule with the standard of due care set at the efficient level dominated strict liability. However, their experiment did not feature a real victim but only monetary consequences for the injurer. In contrast, Angelova et al. (2014) reported the predicted behavioral equivalence between strict liability and negligence. Yet, their setting only allowed for binary care levels, whereas Kornhauser and Schotter (1990) had a large set of alternative care levels.

Guerra and Parisi (2022) studied whether participants choose the same level of care when they act as potential injurers under strict liability or as potential victims under no liability. In analogy to our inquiry, they kept monetary payoffs constant across conditions and observed some behavioral differences. On a different note, Deffains et al. (2019) explored the interaction of obligations from tort law and social norms.[3]

In our design, we keep the injurer's financial obligation after an accident constant but vary who receives the payment, considering either a fine or damages.[4] To the best of our knowledge, we are the first to explore this distinction's implications for care choices in an incentivized laboratory

[3] Croson (2009) and Sullivan and Holt (2017) surveyed the use of experiments in law and economics and discussed some more contributions to the realm of tort law. Dopuch and King (1992), Dopuch et al. (1997), and King and Schwartz (1999) are examples from the accounting literature investigating liability rules in experimental settings.

[4] Attanasi et al. (2020) analyzed data from an experiment in which strictly liable firms may choose care and/or insurance to manage their expected liability when faced with uncertain harm. In this paper, we do not consider other risk management techniques than care investment.

experiment. Feldman and Teichman (2008) used a non-incentivized vignette study to compare how subjects' attitudes in a situation involving an externality depend on the timing of the payment (ex-ante or ex-post), the certainty of the payment (probabilistic or certain), and the stipulated receiver of the payment (the state or the harmed party). In their study, subjects were more likely to state their intent to perform the harmful activity (i.e., to produce a fertilizer according to the authors' scenario) if the harmed party received the eventual payment instead of the state. The authors argued that subjects might perceive the payment to the injured party more like a price for the harmful activity. Their result contrasts with the findings of our incentivized experiment in which subjects' actual precautionary behavior is measured.

Some other experimental studies also contrasted compensatory payments with non-compensatory payments, but these papers investigated frameworks very different from ours. Hoeppner et al. (2017) studied a principal-agent setup in which the principal decided to establish the principal-agent relationship. The agent then chose a level of effort that increased the project's success probability at a private cost to the agent. The authors contrasted possible arrangements for payments due after a project failure, including a control condition in which the principal does not receive the agent's payment. They did not find a significant difference between the control and damage conditions. Eisenberg and Engel (2014) contributed to the literature on public goods and punishment. They empowered one subject of any four-member group to reduce the earnings of one other group member when the public-good contribution of the other group members has fallen short of a reference level. They compared, among other things, behavior when the empowered player could transfer income for private gain to behavior when the player could choose to transfer income for personal gain or to destroy the other players' pay. Hence, in their study, whether losses enable transfers or not was selected by the subjects themselves. Desmet et al. (2022) compared the incentive effects of compensation and a fine payment when individuals could lie to another person for private gain. Whereas our design features incomplete compensation to ensure that subjects care about the bad event even in the damage treatment, Desmet et al. (2022) set the fine/compensation at the level of harm. In the first part of their experiment, subjects played a one-shot game. No treatment differences in the lying probability resulted. In the second part of their experiment, the game was repeated four times, and the data from this part suggest that a fine deters lying to a greater extent than compensation. However, this more substantial deterrent effect of a fine may be due to their specific experimental game. Lying to another participant is a clear norm violation, and a fine may have expressive value in this context. In contrast, we consider a tort setting in which precaution can be implemented at very different levels and where a simple social norm of the appropriate level of care-taking is less obvious.

3. EXPERIMENT DESIGN AND PROCEDURES

3.1 Design

Our experiment consists of four parts. In Part 1, participants earn their (uniform) endowment in a real-effort task. In Part 2, participants act either as injurers or victims in one of the treatments. In Part 3, participants perform in the same role as in Part 2 but play a different treatment. In Part 4, subjects participate in a social value orientation test, a risk elicitation task, a questionnaire on justice attitudes, and they assess the appropriateness of the policy instruments investigated in our study.

At the start of the experiment, participants receive information that either Part 2 or Part 3 will be randomly selected for payment. Part 4 is always payoff-relevant. Subjects first obtain a general introduction and, sequentially, receive instructions in hard copy for Part 1, Part 2, and Part 3. In other words, written instructions for each part are distributed only immediately before subjects begin the relevant part.[5] From the start, subjects know that the experiment consists of several parts. However, the specific content of later parts remains unknown during the previous parts. The tasks in Part 4 are explained on screen. For Parts 2 and 3, all participants have to solve control questions regarding the rules of the respective part before they can make their decisions. We now describe Parts 1 to 4, summarized in Table 1, in more detail.

---

[5] A translated version of our German instructions is included in Appendix A.

Table 1: Experimental Design & Treatments

Part 1		Real-effort task:
		Participants earn a fixed endowment of 750 points.
Instructions for Part 2: Participants learn their role (injurer or victim) and are matched.		
Part 2	Phase 1	Injurer chooses care investment.
	Phase 2	Nature determines whether an accident results.
		The victim loses 300 points in case of an accident.
	Phase 3	In case of an accident:
		• BASELINE: Injurer pays nothing.
		• DAMAGES: Injurer pays 270 points as damages to the victim.
		• FINE: Injurer pays 270 points as a fine (received by a charity).
	Incentivized belief elicitation regarding behavior in Phase 1 of Part 2.	
Instructions for Part 3: Participants keep their roles, are matched with a new partner, and play a different treatment.		
Part 3	Phase 1	Injurer chooses care investment.
	Phase 2	Nature determines whether an accident results.
		The victim loses 300 points in case of an accident.
	Phase 3	In case of an accident:
		• BASELINE: Injurer pays nothing.
		• DAMAGES: Injurer pays 270 points as damages to the victim.
		• FINE: Injurer pays 270 points as a fine (received by a charity).
	Incentivized belief elicitation regarding behavior in Phase 1 of Part 3.	
Part 4		Post-experimental tests
		Social Value Orientation
		Risk Attitude
		Justice Sensitivity
		Appropriateness of policy instrument used for the context at hand
		Demographic questionnaire

Notes: All injurers decide on the level of care investments in two treatments. We have participants in sequences BASELINE-FINE, BASELINE-DAMAGES, FINE-BASELINE, DAMAGES-BASELINE, FINE-DAMAGES, and DAMAGES-FINE.

3.1.1. Part 1: Real-Effort Task

Participants work to obtain an endowment of 750 points. We included the real-effort task to create a notion of entitlement (e.g., Oxoby and Spraggon 2008). For comparability, we selected a threshold task such that all subjects who make choices in Parts 2 to 4 have symmetric endowments (e.g., Duersch and Müller 2015). Participants must correctly count the number of zeros in tables of 150 randomly ordered zeros and ones. As Abeler et al. (2011, p. 473) emphasized, this boring task does not require prior knowledge, performance is easily measurable, comes at a positive cost of effort, and there is little learning possibility. Each participant has to count the number of zeros in three tables correctly.[6] Failing to solve three tables correctly within 10 minutes leads to exclusion from the experiment. Overall, 28 participants (out of 338) did not continue with the main experiment after Part 1, either because they had not completed the task within the time limit (23 participants) or because they had to be excluded to retain an even number of participants in the session in Parts 2 to 4 when other subjects failed to solve the task (5 participants).[7]

After Part 1, participants are informed about their role in Part 2. Roles are fixed throughout Parts 2 and 3. We distinguish *injurers* from *victims*. These roles are referred to as Person A and Person B in the neutrally framed instructions. Victims remain passive in Parts 2 and 3, except for stating their beliefs about the injurers' choices.

3.1.2. Parts 2 + 3: Care Choice

Parts 2 and 3 each consist of three phases. In Phase 1, injurers choose their level of care. Our design very closely follows the standard unilateral-care accident setup with expected social costs $x + \pi(x)h$, where x represents the linear cost of care, h the level of harm, and $\pi' < 0 < \pi''$ describes that the accident probability decreases at a diminishing rate with the care investment (see, e.g., Miceli 2017, Shavell 2007). In our framework, injurers choose from a discrete set of care investments, the level of harm is assumed to be fixed at 300 points, and the accident probability is approximately given by $\pi(x) = 1 - \sqrt{x/375}$. Table 2 and Figure 1 show the possible care levels and the implied levels of the accident probability. When selecting a care level of a given magnitude, injurers invest the specified cost of care using their 750 points endowments

[6] Participants had three trials per table. They received a new table after three unsuccessful trials.

[7] The possibility of being excluded from the rest of the experiment despite having completed the real-effort task in case an uneven number of participants remained after Part 1 was clearly communicated to participants up front (see the experimental instructions in the Appendix A for the exact wording). Naturally, subjects who were excluded to balance the number of participants were paid for their participation in the experiment and for solving the real-effort task.

from Part 1. Before injurers commit to a care investment, they can experiment with different levels of care using a visualization of the accident probability.[8] Based on the injurer's care choice, the computer randomly determines for each injurer-victim pair whether or not an accident occurs in Phase 2. In the event of an accident, the victim loses 300 points.

We distinguish three treatments: BASELINE, FINE, and DAMAGES. Payoff calculations in Phase 3 depend on the treatment. In treatment BASELINE, payoffs remain unchanged; that is, injurers keep their points after the care investment, and victims receive no compensation for the losses incurred. In treatment FINE, in the event of an accident, the injurer pays 270 points (90 percent of the victim's loss) as a fine, but the victim remains uncompensated. It is communicated that the payment by the injurer will be donated to a charity randomly chosen from a list of four charities presented to all subjects.[9] This design feature reflects the redistributive element of fines, justifying that fines are frequently treated as socially costless transfers in many theoretical contributions (e.g., Polinsky and Shavell 2007).[10] In treatment DAMAGES, in the event of an accident, the injurer transfers 270 points to the victim such that the victim's uncompensated harm amounts to only 30 points.[11]

After Part 2 (and again after Part 3), we elicit beliefs.[12] Victims state what care investment they expect from injurers; that is, we elicit victims' first-order beliefs. Injurers state their beliefs about

---

[8] Appendix B shows and example of the injurers' decision screen.

[9] The charities are: the German Red Cross, Médecins sans frontières, Welthungerhilfe, and SOS Children's Villages. A charity was selected randomly to show that individuals usually cannot determine what the fine revenue is used for. These charities are well-known and popular among subjects. For example, d'Albis et al. (2020) present results from an experiment in which subjects could select their charity and the Red Cross and the Médecins sans frontiers were the most and the third most popular in the ranking of charities, respectively.

[10] Moreover, in criminal proceedings in Germany, the legal system relevant for our participants, judges can under certain circumstances determine that monetary payments have to be made to a specified charity (see, for this practice, e.g., Weigend 2001).

[11] Incomplete compensation is commonly considered to be descriptive of tort liability and thus an important characteristic of comparisons of liability with other instruments (e.g., Shavell 2011).

[12] Appendix B shows the injurers' decision screen for the belief elicitation. Our design follows a choices-beliefs sequence in Parts 2 and 3, possibly raising the concern that the beliefs are influenced by the

which care level their matched victim expects from them. This is the injurer's second-order belief.[13]

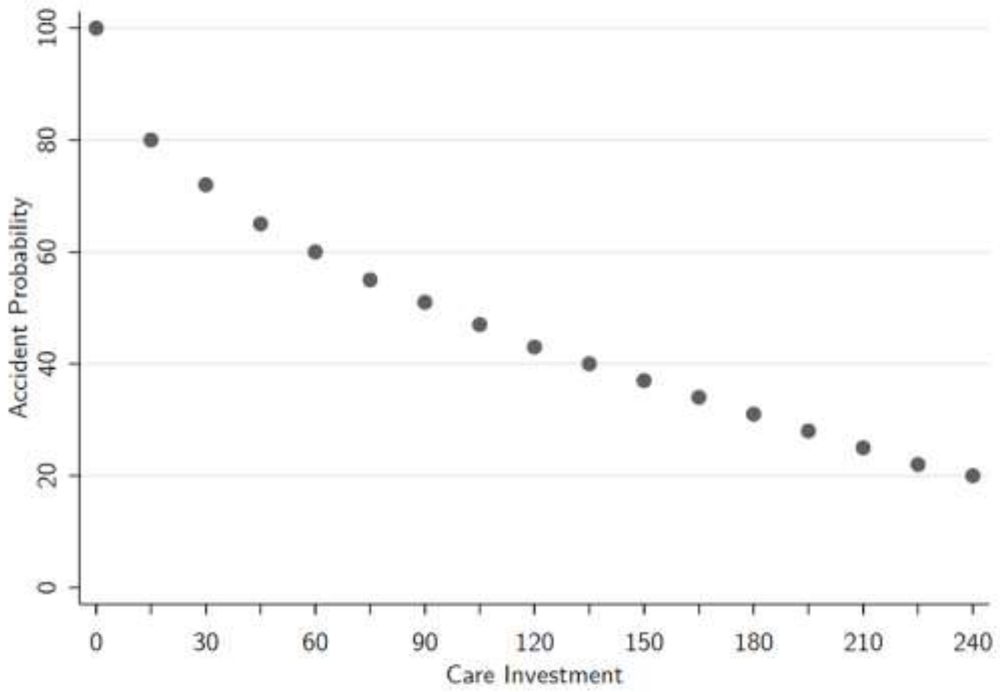
Table 2: Alternative Levels of Care and Implied Accident Probabilities

Care Investment	Accident Probability
0	100
15	80
30	72
45	65
60	60
75	55
90	51
105	47
120	43
135	40
150	37
165	34
180	31
195	28
210	25
225	22
240	20

choices assessed earlier. In the alternative sequence, one could have worried about the possibility that the elicitation of beliefs influences care choices – our main variable of interest. Attanasi and Nagel (2008), for example, discussed the two possible sequences and reported that the elicitation order often essentially did not affect results.

[13] Participants choose from a list of 17 intervals ranging from 0 to 240 points to state their beliefs. The belief elicitation is incentivized as follows (see, e.g., Cartwright 2019, Charness and Dufwenberg 2006): we use the average care investment by all injurers from the session except for the matched injurer to check the accuracy of a victim's first-order belief. The victim's first-order belief is considered accurate if the true average care investment lies inside the interval chosen by the victim. To test the precision of the injurer's second-order belief, we use the first-order belief of the matched victim. Subjects earn 25 points for each correct belief statement in the part of the experiment later randomly selected to be payoff relevant.

Figure 1: Alternative Levels of Care and Implied Accident Probabilities



Parts 2 and 3 differ only in the applicable treatment. The experimental procedures for the sequence of Parts 2 and 3 are as follows: After Part 2, participants receive no information about the care choices, the stated beliefs, or whether or not an accident occurred. When participants enter Part 3, they are informed that they will continue with the same role as in Part 2 but that they will be matched with a different individual and that there will be a change to the rules. This leads to six possible treatment sequences: BASELINE-FINE, BASELINE-DAMAGES, FINE-BASELINE, DAMAGES-BASELINE, FINE-DAMAGES, and DAMAGES-FINE.[14] The number of subjects per treatment sequence is shown in Table 3 in Section 3.2.

3.1.3 Part 4: Individual Characteristics

In Part 4, participants complete a battery of individual tasks. First, subjects complete a version of the incentivized social value orientation slider measure (Murphy et al. 2011), as programmed by Crosetto et al. (2019). We collect information about subjects' social value orientation because

[14] Letting participants play two treatments in Parts 2 and 3 allows us to obtain more data.

injurers face two trade-offs when selecting their care investment in our setting: injurers trade-off higher cost of care with both lower own expected (fine or liability) payments and lower expected (residual) harm for the victim. The subject's social value orientation could be an essential determinant of how the individual considers the latter element.

Next, we elicit participants' risk attitudes using the incentivized measure by Eckel and Grossman (Eckel and Grossman 2002, 2008, Dave et al. 2010), in which the participants choose between six binary lotteries, each with a 50/50 chance. In our experiment, the care decision by the injurer is risk-laden as no care investment level exists that can rule out the possibility of an accident. When trying to understand the regime's influence on the injurer's choice, we can control for any effects due to subjects' risk attitudes using this measure.

The SVO measure and the risk elicitation were payoff-relevant for all participants. For the SVO measure, payoffs for the involved subjects ranged from 100 to 170 points. The risk elicitation task specified (35,35) as the payoff vector in the risk-free lottery and (5,85) as the most uneven payoff vector. No payments are made for the remaining tasks to which we turn next.

Additionally, we ask participants about their justice sensitivity using the short version from Baumert et al. (2014) of the items introduced initially in Schmitt et al. (2010). Justice sensitivity might be relevant for our study as subjects' perceptions about different situations characterized by unfairness might theoretically explain their decision-making. For example, one question of the inventory asks whether the subject experiences feelings of guilt when she is enriched at the cost of another individual.

We also show participants a short vignette of a situation in which harm occurs to another person, and the injurer could influence the probability of harm. We ask participants how morally appropriate they find the three different forms of payment (no payment, damages, fine) on a six-item Likert-like scale (ranging from *0 = not appropriate at all* to *5 = fully appropriate*). The vignette read as follows (translated from German): "Imagine the following situation: Person A damaged

Person B's property in an accident. Person A would have been able to undertake measures to prevent the accident. To which degree do you consider the following rules morally appropriate? a) Person A pays nothing. b) Person A compensates Person B for the damage. c) Person A pays a fine which benefits the general public." We collected this information to understand better which regime subjects find more acceptable, as the acceptance of a regime can be important for the behavior within the limits of a regime. Finally, participants complete a demographic survey.

3.2 Procedures

We conducted the on-screen experiment in the *DecisionLab* at the Max Planck Institute for Research on Collective Goods in Bonn, Germany, in the fall of 2018. Participants were administered and recruited online via ORSEE (Greiner 2015) from the laboratory's subject pool. The experiment was implemented in z-Tree (Fischbacher 2007). We ran 12 sessions.[15] Only subjects who completed the real-effort task in Part 1 and correctly answered control questions for Parts 2 and 3 are included in the analysis. While 23 subjects failed to complete the real-effort task, five were excluded to maintain an even number of participants within the relevant sessions. Concerning the control questions in Parts 2 and 3, participants could try multiple times to answer them and could additionally request help from the experimenter. Nevertheless, two participants were not able to answer the control questions in Part 2 correctly, even after further explanations. Their data were excluded from the analysis. This leads to a number of 308 subjects for Parts 2 to 4, with 22 to 30 subjects per session. Table 3 presents the number of subjects per treatment sequence.

The subjects' mean age was 23 years. Sixty-six percent of our participants were female. The vast majority of subjects were university students (around 97 percent). Their fields of study included, amongst others, economics, law, linguistic science, agriculture and forestry, and medicine.

A typical session lasted around 90 minutes (including payment). Subjects could earn points

[15] The study was not preregistered.

during the experiment, which were converted to Euro at the end of the experiment at a conversion rate of 0.02 Euro per point. Subjects were paid in cash. The average earnings were around 14.20 Euro. The experimenter made the donation to the charities after all sessions had been concluded.

Table 3: Number of Subjects Overall per Treatment Sequence

Part 2/Part 3	BASELINE	FINE	DAMAGES
BASELINE	-	54	47
FINE	56	-	53
DAMAGES	48	50	-

Notes: There is an odd number of subjects in two treatment sequences because two observations were excluded from the data set. The excluded subjects did not master the control questions prior to Part 2.

3.3 Social Norms

In an additional study, using the paradigm developed by Krupka and Weber (2013), we elicited the social appropriateness of the different care investments in our treatments BASELINE, DAMAGES, and FINE. Third parties were invited for one treatment only and had to rate the social appropriateness of each care investment level on a four-item scale including "very socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", and "very socially appropriate". Subjects received a 7-Euro reward if their answer on the appropriateness scale for one randomly selected care investment level corresponded to the modal answer. We collected social norms data from 84 subjects.

4. RESULTS

In this section, we will first assess whether the order of treatments in Parts 2 and 3 was relevant for the chosen care levels (Section 4.1). Next, we will compare care investments across treatments at the group level (Section 4.2), followed by an analysis at the individual level (Section 4.3). Then, we will report the results on the subjects' beliefs about care investments and their relationship to actual care choices (Section 4.4.) as well as on the social norms data (Section 4.5).

4.1 No Order Effects in Care Investments

Each injurer chooses care investments in two treatments (in Parts 2 and 3). We do not find a statistically significant difference at the 10% level between care investments in Parts 2 and 3 for any of the three treatments (Mann-Whitney *U* tests (MWU), Part 2 vs. 3, BASELINE: p = 0.187, DAMAGES: p = 0.674, FINE: p = 0.550).[16] Based on this finding, we present our results at the group level using the pooled data from Parts 2 and 3. The implication of the absence of order effects is that subjects in our experiment seem not to condition their behavior in a given regime on whether they experienced a different regime before.[17]

4.2 Care at the Group Level

The average care investment in treatment BASELINE amounts to 49 points, is distinct from zero, and is considerably smaller than the average care investment in both, treatment FINE (99 points) and treatment DAMAGES (125 points), as shown in Figure 2.

The actual care levels may be compared with the care investment that minimizes the sum of care costs and the expected harm (i.e., the socially optimal level in textbook representations of the unilateral-care model)[18] and the care that minimizes the injurer's expected costs in FINE

[16] All reported tests are two-sided.

[17] This contrasts with observations in other experimental settings (e.g., Bruttel and Friehe 2014).

[18] Note that the care investment that minimizes the sum of care costs and expected harm is socially optimal only when agents have utility functions linear in wealth.

and DAMAGES, which is a care investment of 60 points and 45 points, respectively. The care investment of 45 points is also privately optimal for a risk-averse injurer with CRRA preferences and a coefficient of relative risk aversion of about one, for example, which accords with empirical estimates (e.g., Anderson et al. 2008).

Figure 2: Care Investments in Treatments BASELINE, DAMAGES, and FINE

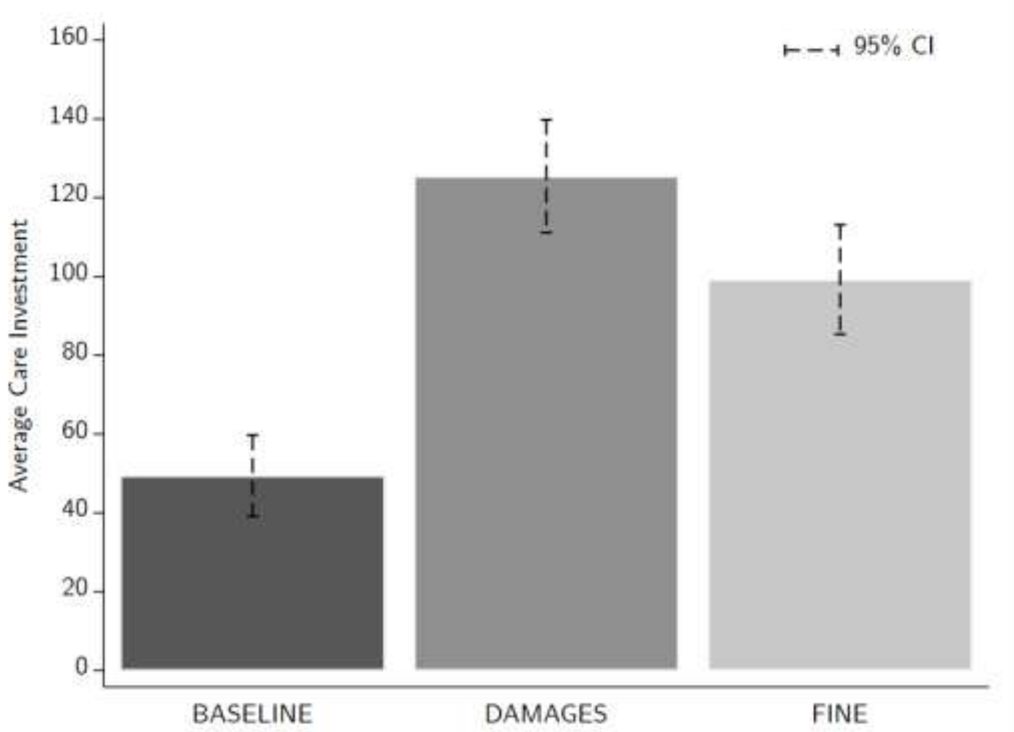
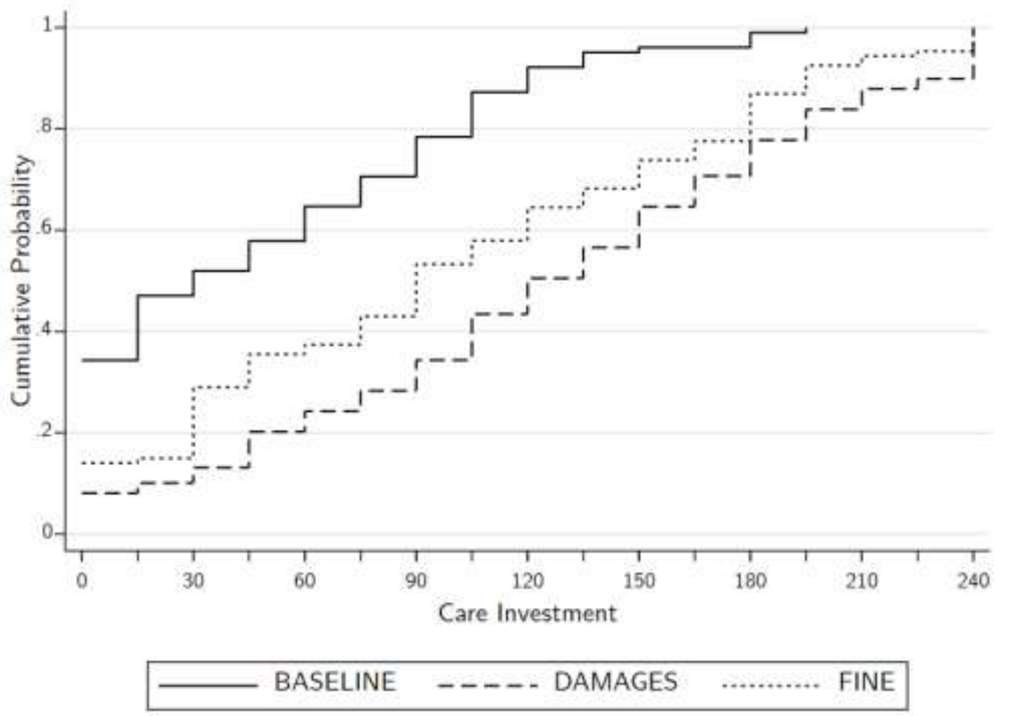


Figure 3 shows the cumulative distribution functions for the observed care investments in the three treatments. Overall, the care levels in treatment BASELINE are the lowest and the care levels chosen in treatment FINE are somewhat lower than those in treatment DAMAGES. About 64 (80) percent of subjects in FINE (DAMAGES) invest more than the 45 points which would maximize their own expected payoffs and only a small fraction decides not to invest at all (14 (8) percent). In contrast, 34 percent of subjects in treatment BASELINE choose a care investment of zero.

Figure 3: Cumulative Distribution Functions of Care Investments in Treatments



From these observations, we deduce that injurers invest more in treatment DAMAGES than in treatment FINE. This finding is supported statistically by a Wilcoxon signed-rank test (WSR) of the care levels in treatments DAMAGES and FINE for those injurers who participated in both treatments ($N = 52$, $p = 0.001$). In addition, care levels in treatments BASELINE and DAMAGES ($N = 47$, $p < 0.001$), and BASELINE and FINE ($N = 55$, $p < 0.001$) are significantly different according to WSR tests.

Note that the pooled sample consists of a mix of independent and dependent observations for all treatment comparisons. Above, we have reported the *within*-comparison of subjects who played both treatments. However, we can also compare care levels *between* subjects. To create a sample of independent observations, we start from the whole sample and, for subjects who played both treatments compared, drop the observations from Part 3. With this data, treatment differences are also significantly different according to MWU tests ($N = 154$, DAMAGES vs. FINE: $p = 0.028$; DAMAGES vs. BASELINE: $p < 0.001$; FINE vs. BASELINE: $p < 0.001$).

4.3 Care at the Individual Level

The results at the group level explained in Section 4.2 are confirmed by regression analyses which make use of the information elicited in Part 4. The fact that injurers chose their care levels in Parts 2 and 3 gives our data a panel structure. To exploit this panel structure, we use random-effects regressions for care choices in Parts 2 and 3. Regression results are presented in Table 4. The reference category is the behavior in the FINE treatment. In Model 1, the care investment is regressed on two treatment dummies and a "Part 3" dummy (equal to one when the choice was taken in Part 3). Model 2 includes subjects' social value orientation scores, risk preferences, and justice sensitivity (perpetrator scale) into the empirical model. Model 3 adds controls for subjects' demographic characteristics such as age, gender, and their field of study.

We confirm the treatment effects reported in Section 4.2 as indicated by the significance of the coefficients for treatment BASELINE and treatment DAMAGES in all three models. As the negative coefficient for the BASELINE dummy variable suggests, injurers invest significantly less in treatment BASELINE than in treatment FINE. In contrast, the positive coefficient for the DAMAGES dummy variable shows that injurers invest significantly more in treatment DAMAGES than in treatment FINE. The difference between the DAMAGES treatment and the BASELINE treatment is confirmed by Wald tests of the coefficients of the treatment dummies which display p-values smaller than 0.001 for all three models. These results are robust to the inclusion of subjects' social value orientation, risk preferences, and justice sensitivity as control variables in Model 2 and their demographics in Model 3.

4.4. Beliefs Main Experiment

We elicited the victims' first-order beliefs about the average care investment of injurers in their session. In addition, we elicited the injurers' second-order beliefs about the expectations of their matched victim. Figure 4 illustrates the respective average beliefs.[19] As the figure shows, mean

---

[19] We report results on the pooled data from Part 2 and 3 of the experiment. We find no order effect for subjects' beliefs in the DAMAGES and FINE treatment according to MWU tests (second-order beliefs,

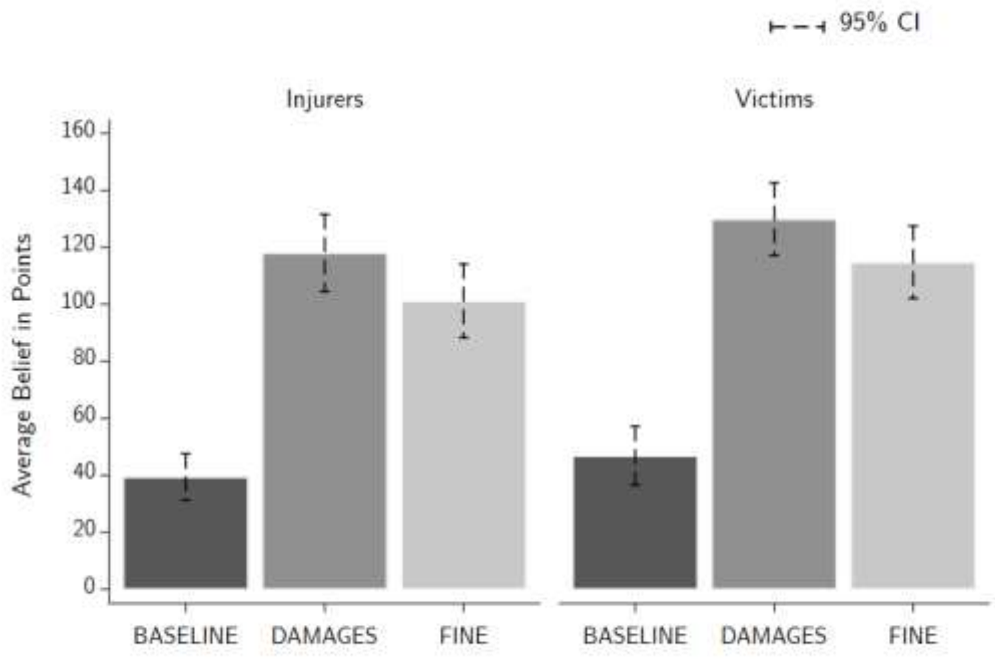
first-order beliefs and mean second-order beliefs are quite similar, and while the victims' first-order beliefs are descriptively higher than the injurers' second-order beliefs in each treatment, the beliefs of injurers and victims are not statistically different from each other (first-order beliefs vs. second-order beliefs, BASELINE: $p = 0.593$, DAMAGES: $p = 0.236$, FINE: $p = 0.129$, MWU).

We find that injurers' second-order beliefs are treatment-dependent and show the same ranking as the actual care investments.[20] Differences in second-order beliefs between treatment BASELINE and either the DAMAGES or the FINE treatment are statistically significant according to WSR tests ($p < 0.001$ for both comparisons) for the subjects who played both treatments. Differences in second-order beliefs between the DAMAGES and the FINE treatment are also significant (WSR, $p = 0.005$). Finally, the injurers' second-order beliefs and care investments are highly correlated. This is supported by a Spearman correlation coefficient of at least 0.672 for all three treatments ($p < 0.001$).

DAMAGES: $p = 0.312$, FINE: $p = 0.662$; first-order-beliefs, DAMAGES: $p = 0.122$, FINE: $p = 0.434$). However, in treatment BASELINE, second-order beliefs and first-order beliefs in Part 2 are significantly different from their counterparts in Part 3 $p \leq 0.034$ for both beliefs, MWU). Yet, this is not a concern since the differences in care investments between either the FINE or the DAMAGES and the BASELINE treatment can be established using only the data from Part 2 (FINE-BASELINE: $p = 0.001$, DAMAGE-BASELINE: $p < 0.001$, MWU).

[20] The victims' first-order beliefs are ranked similar to the injurers' second-order beliefs. WSR tests display p-values smaller than 0.001 for the comparisons of the BASELINE with the treatment DAMAGES or the treatment FINE, respectively, and a p-value of 0.003 for the comparison of the treatments DAMAGES and FINE (for those subjects who played both treatments).

Figure 4: Injurers' Second-Order Beliefs and Victims' First-Order Beliefs



In light of these results on injurers' beliefs and their potential influence on care investments, we consider the results from mediation analyses using structural equation modeling. First, these analyses allow us to study further the association between injurers' second-order beliefs and care investments. Second, they enable an understanding of whether treatment differences result only because second-order beliefs vary with the treatment (indirect effect) or whether an additional direct effect of treatment exists. Results of our analysis are given in Figures 5a and 5b whereas in the regressions FINE serves as reference treatment. The belief variable ranges from 1 to 17 and represents the intervals subjects could choose in the belief elicitation task.

Figure 5a Mediation Analysis Treatment DAMAGES

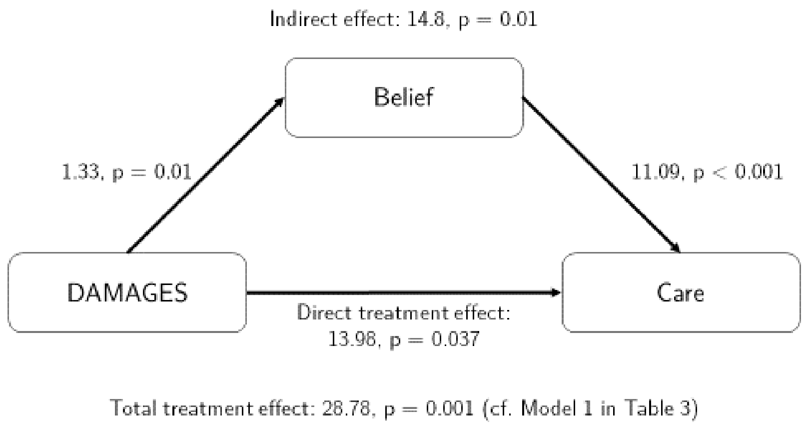
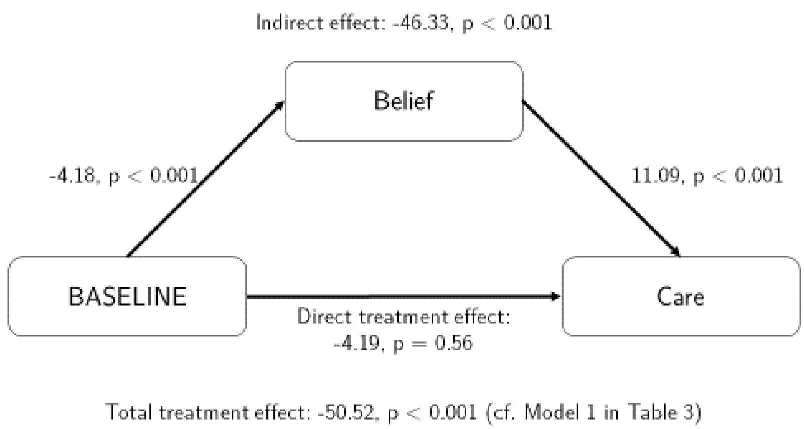


Figure 5b Mediation Analysis Treatment BASELINE



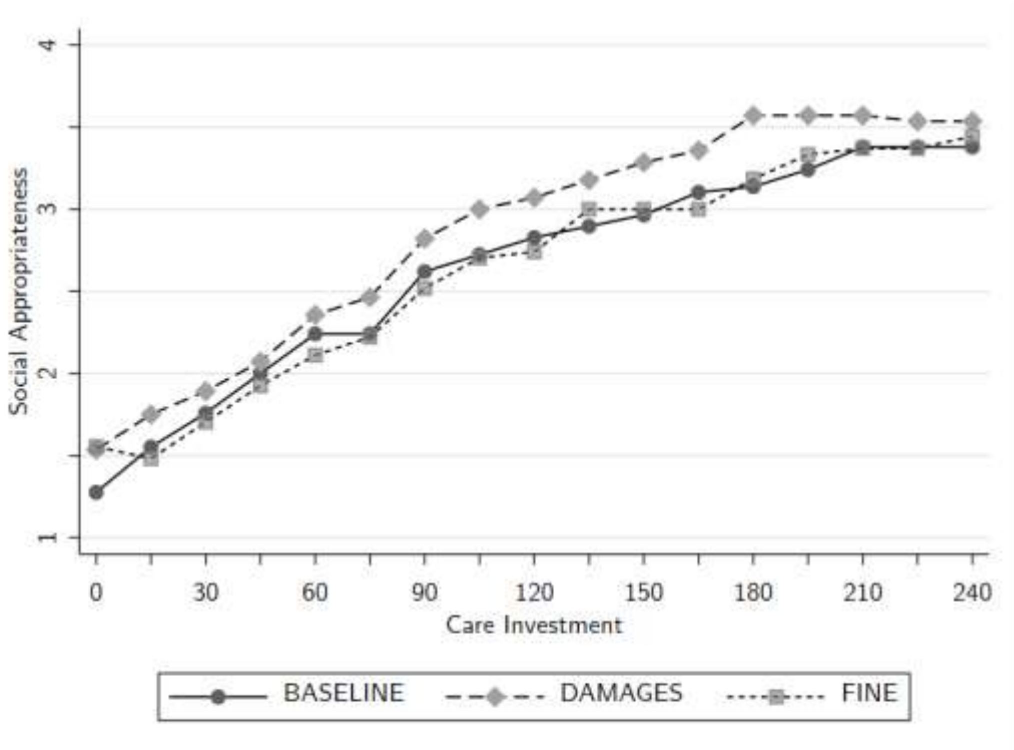
The differences between the treatment DAMAGES and FINE are only partly mediated by higher second-order beliefs of injurers in treatment DAMAGES (indirect effect). Yet, a direct effect of nearly equal size, which cannot be explained with beliefs, remains. In contrast, differences in care investments between treatments BASELINE and FINE are completely mediated by the injurer's beliefs.

Of course, the assumed causal structure of the relationship between injurers' beliefs and their care investments underlying these arguments has to be taken with a grain of salt. Both, beliefs and care investments, are endogenous (possibly explained by the same exogenous variables) and a causal influence from the earlier care choice on the later elicited beliefs also cannot be excluded.

4.5. Social Norms About Care Investments

Using the data on social norms about socially appropriate care investments from our additional study, we find that the average social appropriateness of any care investment level is descriptively the highest in the DAMAGES treatment for all positive care investments (Figure 6). The difference is most pronounced for intermediate to high care investment levels. In the DAMAGES treatment, care investment levels above 180 are not judged as more socially appropriate than investments of 180. In contrast, in treatments FINE and BASELINE, care investment levels above 180 seem more appropriate than an investment of 180. However, the differences between treatments are relatively small.

Figure 6: Social Norms Regarding Care Levels



This data shows that third parties seem to assign a relatively high value to reducing the expected harm compared to creating care costs. As stated above, the care investment level considered socially optimal in the standard account is 60.

Table 4: Treatment Effects on Care Investments at the Individual Level

	(1) Model 1	(2) Model 2	(3) Model 3
BASELINE	-50.04**	-49.85**	-50.96**
	(8.49)	(8.47)	(8.61)
DAMAGES	28.90**	28.36**	28.41**
	(8.57)	(8.55)	(8.69)
SVO		0.44	0.40
		(0.31)	(0.34)
Risk Attitude		3.35	3.27
		(2.49)	(2.76)
Justice Sensitivity		8.09*	5.91
		(3.92)	(4.40)
Part 3	-6.36	-6.36	-6.34
	(6.69)	(6.70)	(6.70)
Constant	101.63**	47.44*	-27.00
	(7.12)	(20.61)	(80.48)
Demographics	NO	NO	YES
N	308	308	308
No. of Groups	154	154	154

Notes: Results from random effects regressions. FINE treatment as reference category. Standard errors in parentheses. The dummy variables BASELINE and DAMAGES are equal to 1 for the BASELINE and the DAMAGES treatment, respectively. SVO controls for subjects' social value orientation. Risk Attitude controls for subjects' risk preferences (on a scale from 1 to 6). Justice Sensitivity (perpetrator) controls for subjects' justice sensitivity (average of two 6-item Likert-like scales). The Part 3-dummy equals 1 when the choice stems from Part 3. Demographic controls include participants' age, gender, experimental experience, number of siblings, a dummy whether the subject is a student, their field of study, their semester, and a dummy whether participants work for more than 10 hours per week. * and ** indicate significance at the 5% and 1% level.

5. DISCUSSION AND FOLLOW-UP EXPERIMENTS

Our results violate the predictions obtained from standard theory. Standard theory assumes subjects are only concerned about their material payoffs. According to this approach, care investments are only expected if injurers face a financial obligation in the accident state. Moreover, the fixed financial burden across the treatments FINE and DAMAGES should induce the same care level independent of how the transfers are ultimately used. In contrast, we observe a substantial care investment in the treatment BASELINE, in which an accident does not result in any monetary costs for the injurer. Furthermore, care investments are higher in the treatment DAMAGES than in the treatment FINE.

Concerning the first observation, we find that 66 percent of our subjects in treatment BASELINE invest some amount in lowering the probability of harm. Angelova et al. (2014) also provide evidence that subjects invest in accident avoidance even when the accident is not associated with a financial obligation for the injurer. In their very different dynamic setup, 73 percent of the subjects in the No-Liability treatment invested at least once in precautions. They refer to the fact that subjects might care for others as a possible explanation for this finding. Consistent with their argument, the subjects' social value orientation is a significant predictor of the care investment in treatment BASELINE.[21] Note that, at a very abstract level, the injurer's decision-making context in treatment BASELINE represents a dictator game in expected terms, where the injurer can use his resources to increase the victim's expected resources. When seen in this light, our results from treatment BASELINE are consistent with results from dictator games concerning, for example, how many dictators give (e.g., Engel 2011).

Next, we discuss the result concerning the relatively higher care investment in treatment DAMAGES than in treatment FINE. To explore potential explanations for the observed asymmetry of care investments, we test to what extent specific design features might be drivers of this result.

First, in experiments, subjects are often concerned with how well they perform relative to other participants. Suppose injurers fear that their performance will compare unfavorably to that of victims after an accident in treatment DAMAGES. In that case, they may be motivated to invest more in care than in treatment FINE, where the injurer's and the victim's payoffs are similar after an accident. To assess this line of argument, we conducted sessions in which injurers and

---

[21] Treatment specific regression results showing the influence of SVO on care choices are provided in Appendix C.

victims participated neither in the same room nor at the same time. Such design modifications tend to reduce the importance of payoff comparisons.

Second, in treatment FINE of our main experiment, the fine revenue was devoted to a prosocial end. One may speculate that subjects perceive paying money to a charity differently from paying money to a peer. Although there are real-world circumstances in which fine revenue can be devoted to prosocial ends (see Footnote 10), there are also circumstances where it seems more appropriate to assume that the fine revenue is perceived as redistribution to all society members, for example, in the form of lower taxes. To investigate this argument's relevance, we ran an additional experiment in which the fine revenue was redistributed to all participants.

In both follow-up studies, we only collected data for the treatments FINE and DAMAGES to focus on the key regime contrast. Given that we only elicit data from treatment sequences FINE-DAMAGES and DAMAGES-FINE, we compare data from the follow-up experiments (reported in subsections 5.1 and 5.2) only with the main-experiment data from subjects who also experienced the FINE-DAMAGES or DAMAGES-FINE treatment sequences. After all, although we did not detect any overall order effects on care choices in our main experiment, we cannot exclude the possibility that choosing the care investment in treatment FINE or DAMAGES after having made a choice in treatment BASELINE is somehow different from a choice in treatment DAMAGES (FINE) after having made a choice in treatment FINE (DAMAGES). We will explain the details of the two follow-up studies in the next subsections.

5.1 Payoff Comparisons as Driver in DAMAGES (Follow-up Study 1)

In our main experiment, injurers who care about how their own payoffs compare to their victims' payoffs have stronger care incentives in treatment DAMAGES than in treatment FINE. While the payoff comparison is independent of the treatment in the no-accident state, considerable payoff differences occur when an accident takes place: In the accident state, the injurer loses 270 points in both treatments whereas the victim loses 300 points in the FINE treatment, but only 30 points in the DAMAGES treatment. In contrast to these salient relative payoff implications for the usually small number of session participants in the roles of injurers and victims, in the real world, tort law often addresses circumstances in which injurers and victims are complete strangers before the accident. To that extent, having our results depend on strong inequity concerns would limit their external validity.

To check whether payoff comparisons with the victims explain the high level of care investments in treatment DAMAGES, we conducted additional sessions *online* (that means participants were not assembled in the laboratory) in which injurers and victims participated *at different points in time*. This fact that injurers and victims participated at different points in time was made clear to subjects. With these design changes, we significantly widen the social distance between injurer and victim while leaving our experimental protocol unchanged in other ways.

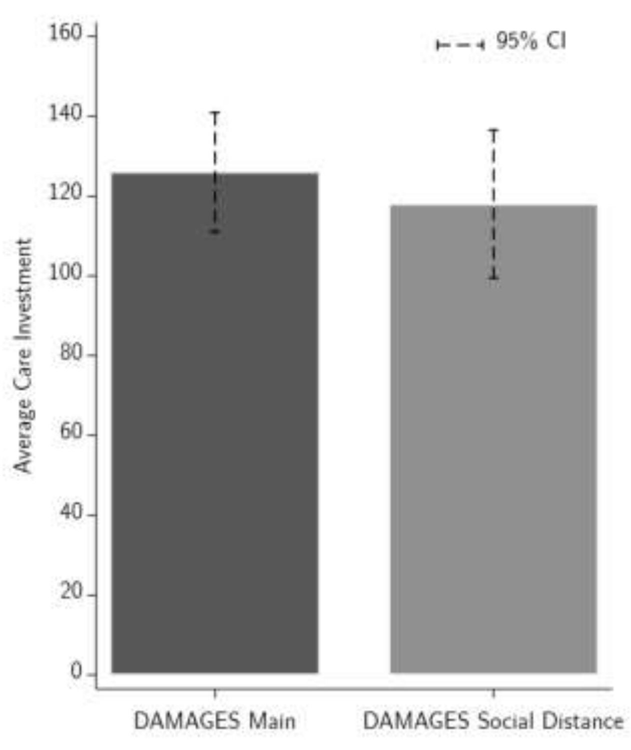
Overall, we collected data from 83 injurers in FINE-DAMAGES or DAMAGES-FINE sequences. The mean care investment level in treatment DAMAGES was 118 points and is not significantly different from the care investment in treatment DAMAGES of our main experiment (126 points, MWU, $p = 0.605$, Figure 7). In other words, we find that the care investments in the DAMAGES treatment with significant social distance are comparable to those from the main experiment. This contradicts the hypothesis that the concern about an unfavorable payoff comparison explains the asymmetry between care investments in treatment DAMAGES and treatment FINE in our main experiment.

In the online experiment, we also altered the design of treatment FINE. Instead of allocating fine revenue to a charity, we specified that any fine payment would be redirected to the experimenter's budget.[22] We find that the care investments in this variety of treatment FINE (mean 125 points) are not different from those in treatments DAMAGES of the online experiment (WSR, $p = 0.532$) or the main experiment (MWU, $p = 0.989$). This finding motivated a second

[22] This variation in the design of treatment FINE is irrelevant to the results from treatment DAMAGES reported above because we once again cannot find any order effects (MWU, $p > 0.287$).

robustness check in which we again modified treatment FINE relative to the design in our main experiment, but collected data from the laboratory to stay closer to our main experiment in this regard.

Figure 7: Care in DAMAGES in Main Experiment and Follow-Up Study 1

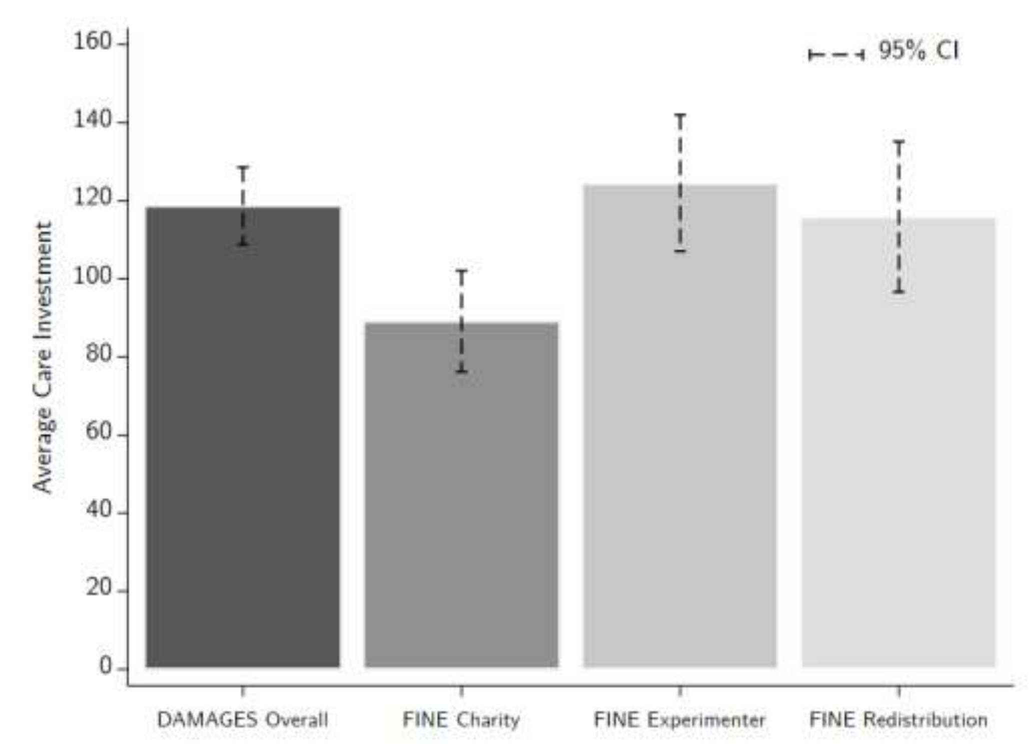


5.2 Prosocial Ends as Driver in FINE (Follow-Up Study 2)

In our main experiment, fine revenue was ultimately transferred to one of four charities. In experimental economics, charities are used for many different purposes. While also realistic for legal systems (see Footnote 10), allocating fine revenue to charities is a special case.

To check whether our results depend on how fine revenue is used, we analyze data from an experiment that closely followed our main experiment (e.g., it was conducted in the laboratory and included injurers and victims in common sessions,) except that the fine revenue was redistributed to all subjects after all sessions had been completed (each participant obtained the same share of the aggregate fine revenue). With 55 injurers in FINE-DAMAGES or DAMAGES-FINE sequences, we find that the mean care investment in the varied treatment FINE amounts to 116 points. The average care level in the DAMAGES treatment is 106 points. Care levels in the two treatments are statistically indistinguishable (WSR, p = 0.953). Comparing the care investments from this follow-up experiment to the care investments in the main experiment, we find no statistically significant difference in care investments between the two DAMAGES treatments (MWU, p = 0.166) and between the FINE treatment in the follow-up study and the DAMAGES treatment in the main experiment (MWU, p = 0.484). However, care investments in the FINE treatment with redistribution are significantly higher than investments in the FINE treatment of the main experiment (average care level FINE Main: 88 points, MWU, p = 0.047). Figure 8 displays the average care taken in the DAMAGES treatment (pooled over all studies) and in the different FINE treatments. Follow-up experiment 2, therefore, supports the conjecture that the asymmetry in care investments between treatments FINE and DAMAGES in our main analysis is mainly driven by the fact that the fine revenue is devoted to a charity.

Figure 8: Care in DAMAGES and FINE Treatments Across Studies



Such data is consistent with relatively greater unconditional altruism towards charities and their objectives. Fehr and Schmidt (2006), for example, report evidence consistent with unconditional altruism. If subjects perceive that the beneficiaries of the different charities are more deserving than the other participants in the experiment (either the victim in treatment DAMAGES or all other participants in treatment FINE of follow-up experiment 2), then care investments in treatment FINE of our main experiment should be the lowest. The intuition is that injurers consider the accident state in which they make a transfer of 270 points as less bad when the transfer goes to a prosocial end.

5.3 Further Discussion

In our main experiment, we find that injurers invest more in treatment DAMAGES than in treatment FINE where they invest more than in treatment BASELINE. Whereas material incentives are the same in treatments DAMAGES and FINE, it is possible that the regime itself communicates something to the subjects according to the literature on the expressive function of law (e.g., Sunstein 1996). Deffains et al. (2019), for example, argue that liability rules may crowd in concerns for the well-being of others because they create visible relationships between injurers and victims or generate a moral suasion effect. Our results in the main experiment seem aligned with this explanation, but the results from the follow-up experiments are not.

Different expressive-law meanings may manifest in different social norms. Treatment DAMAGES makes it salient that compensating those harmed by the act is important which could also mean that it is relatively more important to prevent accidents. In our social norms reported in Section 4.5, we see that the average social appropriateness of a high care investment is descriptively higher in the DAMAGES treatment, which provides some (weak) support of the expressive function of damages payments. However, the general pattern in terms of which care investment out of the alternative set is socially appropriate is similar across treatments.

From subjects' responses concerning which regime they considered to be morally appropriate, we find some support in our questionnaire data that subjects perceive the duty to compensate the victim of an accident as morally more appropriate than the duty to pay a fine to society at large. On a scale ranging from *0 = not appropriate* at all to *5 = fully appropriate*, participants perceive damages to be markedly more appropriate than a fine (mean rating for damages = 4.35, for a fine = 2.33, for no payment = 0.52, N = 262).[23] The psychology literature (e.g., Darley and Pittman 2003) explains that people require compensation from the harmed party to achieve justice when the injurer breached a standard of conduct, but that compensation may not be necessary in the absence of such a breach. Our results are consistent with this line of argument because the vignette can be read as if the injurer did not undertake any precaution.

6. CONCLUSION

Using experimental data, we compare the care levels induced by fines and damages in a design representing the standard unilateral-care model. In contrast to the prediction based on standard theory, we find that damages under a strict liability regime induce significantly higher care investments than a requirement to pay a fine that is as high as the damages payment but ultimately allocated to a charity. In contrast, the requirement to pay a fine induces care incentives comparable to those of the requirement to pay compensation if fine revenues are ultimately allocated to all experimental participants. Accordingly, care levels depend not only on the magnitude of the financial consequences but also on who receives eventual payments.

Our results are consistent with prosocial motivation on the injurers' part. An accident that produces a fixed social loss but, in the end, signifies that some resources are distributed to a prosocial end is not as bad as an accident that produces the same fixed social loss but, in the end, "only" induces some redistribution within the population of injurers and victims.

Our key finding is important for policy-making. Two policy instruments that seem equivalent can induce very different care investments. The policy-maker's choice with respect to the policy instrument in the harmful externality domain thus must incorporate that non-material incentives can drive a sizable wedge between behavioral outcomes of different (payoff-equivalent)

[23] These data are only available from the third session onwards, leading to a lower number of observations.

instruments. If policymakers intend to earmark proceeds from a policy intervention, they should be aware of the consequences assessed in our paper.

ACKNOWLEDGMENTS

We thank Michaela Pielmeier, Rovina Pinto, Alex Jude, and Rohil Rao for programming the experiments. We gratefully acknowledge the feedback and suggestions provided by Claudia Cerrone, Christoph Engel, Susann Fiedler, Elissa Gentry, Leonie Gerhards, Eric Helland, William H. J. Hubbard, Yoan Hermstrüwer, Oliver Himmler, Andreas Lange, Adi Leibovitch, César Martinelli, Marco Ottaviani, Hannah Schildberg-Hörisch, Alexander Stremitzer, Eyal Zamir, and two anonymous referees. Moreover, we have tremendously benefited from feedback received from participants of workshops in Bologna, Bonn, Nancy, and Marburg, the 2019 Annual Meeting of the American Law & Economics Association, the 2019 Annual Meeting of the European Association of Law & Economics, the 2019 Annual Conference of the Verein für Socialpolitik, and the 2019 Conference on Empirical Legal Studies.

REFERENCES

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. Reference Points and Effort Provision. *American Economic Review* 101:470-492.

Anderson, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. Eliciting Risk and Time Preferences. *Econometrica* 76:583-618

Angelova, Vera, Olivier Armantier, Giuseppe Attanasi, and Yolande Hiriart. 2014. Relative Performance of Liability Rules: Experimental Evidence. *Theory and Decision* 77:531-556.

Attanasi, Giuseppe, Laura Concina, Caroline Kamaté, and Valentina Rotondi. 2020. Firm's Protection Against Disaster: Are Investment and Insurance Substitutes or Complements? *Theory and Decision* 88:121-151.

Attanasi, Giuseppe, and Rosemarie Nagel, 2008. A Survey of Psychological Games: Theoretical Findings and Experimental Evidence. Pp. 204-232 in *Games, Rationality and Behavior. Essays on Behavioral Game Theory and Experiments*, edited by Alessandro Innocenti and Patrizia Sbriglia. London: Palgrave Macmillan.

Baumann, Florian, Sophie Bienenstock, Tim Friehe, and Maiva Ropaul. Forthcoming. Fines as Enforcers‘ Rewards or as a Transfer to Society at Large? Evidence on Deterrence and Enforcement Implications. *Public Choice.*

Baumert, Anna, Constanze Beierlein, Manfred Schmitt, Christoph J. Kemper, Anastassiya Kovaleva, Stefan Liebig, and Beatrice Rammstedt. 2014. Measuring Four Perspectives of Justice Sensitivity with Two Items Each. *Journal of Personality Assessment* 96:380-390.

Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169-217.

Blanco, Mariana, Dirk Engelmann, Alexander K. Koch, and Hans Theo Normann. 2010. Belief Elicitation in Experiments: Is There a Hedging Problem? *Experimental Economics* 13:412-438.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71:117-120.

Bruttel, Lisa, and Tim Friehe. 2014. On the Path Dependence of Tax Compliance. *European Economic Review* 65:90-107.

Cartwright, Edward. 2019. A Survey of Belief-Based Guilt Aversion in Trust and Dictator Games. *Journal of Economic Behavior and Organization* 167:430-444.

Charness, Gary, and Martin Dufwenberg. 2006. Promises and Partnership. *Econometrica* 74:1579-1601.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. oTree – An Open Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance* 9:88-97.

Cooter, Robert, and Thomas Ulen. 2016. *Law and Economics*. 6th edition. Boston: Addison-Wesley.

Crosetto, Paolo, Ori Weisel, and Fabian Winter. 2019. A Flexible z-Tree and o-Tree Implementation of the Social Value Orientation Slider Measure. *Journal of Behavioral and Experimental Finance* 23:46-53.

Croson, Rachel T. A. 2009. Experimental Law and Economics. *Annual Review of Law and Social Sciences* 5:25-44.

Darley, John M., and Thane S. Pittman. 2003. The Psychology of Compensatory and Retributive Justice. *Personality and Social Psychology Review* 7:324-336.

Dave, Chetan, Catherine C. Eckel, Cathleen A. Johnson, and Christian Rojas, 2010. Eliciting Risk Preferences: When Is Simple Better? *Journal of Risk and Uncertainty* 41:219-243.

d'Albis, Hippolyte, Giuseppe Attanasi, and Emmanuel Thibault. 2020. An Experimental Test of the Under-Annuitization Puzzle with Smooth Ambiguity and Charitable Giving. *Journal of Economic Behavior & Organization* 180:694-717.

Deffains, Bruno, Romain Espinosa, and Claude Fluet. 2019. Laws and Norms: Experimental Evidence with Liability Rules. *International Review of Law and Economics* 60:105858.

Desmet, Pieter T.M., Leonie Gerhards, and Franziska Weber. 2022. Is Compensation Fine? Sanction Schemes and Their Effects on Deterrence and Trust. Working Paper.

Dopuch, Nicholas, and Ronald R. King. 1992. Negligence Versus Strict Liability Regimes in Auditing: An Experimental Investigation. *Accounting Review* 67:97-120.

Dopuch, Nicholas, Daniel E. Ingberman, and Ronald R. King. 1997. An Experimental Investigation of Multi-Defendant Bargaining in 'Joint and Several' and Proportionate Liability Regimes. *Journal of Accounting and Economics* 23:189-221.

Duersch, Peter, and Julia Müller. 2015. Taking Punishment into Your Own Hands: An Experiment. *Journal of Economic Psychology* 46:1-11.

Eckel, Catherine C., and Philip J. Grossman. 2002. Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. *Evolution and Human Behavior* 23:281-295.

Eckel, Catherine C., and Philip J. Grossman. 2008. Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices. *Journal of Economic Behavior and Organization* 68:1-17.

Eisenberg, Theodore., and Christoph Engel. 2014. Assuring Civil Damages Adequately Deter: A Public Good Experiment. *Journal of Empirical Legal Studies* 11:301-349.

Engel, Christoph. 2011. Dictator Games: A Meta Study. *Experimental Economics* 14:583-610.

Fehr, Ernst, and Klaus Schmidt. 2006. The Economics of Fairness, Reciprocity, and Altruism – Experimental Evidence and New Theories. Pp. 615-691 in Vol. 1 of *Handbook on the Economics of Giving,* edited by Serge-Christophe Kolm and Jean Mercier Ythier. Amsterdam: Elsevier.

Feldman, Yuval, and Doron Teichman. 2008. Are All 'Legal Dollars' Created Equal? *Northwestern University Law Review* 102:223-262.

Fischbacher, Urs. 2007. z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics* 10:171-178.

Fullerton, Don, and Gilbert E. Metcalf. 1998. Environmental Taxes and the Double-Dividend Hypothesis: Did you Really Expect Something for Nothing? *Chicago-Kent Law Review* 73:221-256.

Gevrek, Z. Eylem, and Ayse Uyduranoglu. 2015. Public Preferences for Carbon Tax Attributes. *Ecological Economics* 118:186-197.

Greiner, Ben. 2015. Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association* 1:114-125.

Guerra, Alice, and Francesco Parisi. 2022. Injurers Versus Victims: (A)Symmetric Reactions to Symmetric Risks. *BE Journal of Theoretical Economics* 22:603-620.

Goerke, Laszlo. 2003. Road Traffic and Efficient Fines. *European Journal of Law and Economics* 15:65-84.

Hoeppner, Sven, Lars Freund, and Ben Depoorter. 2017. The Moral-Hazard Effect of Liquidated Damages: An Experiment on Contractual Remedies. *Journal of Institutional and Theoretical Economics* 173:1-22.

Kaplow, Louis, and Steven Shavell. 2001. Fairness Versus Welfare. *Harvard Law Review* 114:961-1388.

Karlan, Dean, and John A. List. 2020. How Can Bill and Melinda Gates Increase Other People's Donations to Fund Public Goods? *Journal of Public Economics* 191:104296.

Kessler, Judd B. 2017. Announcements of Support and Public Good Provision. *American Economic Review* 107:3760-3787.

King, Ronald R., and Rachel Schwartz. 1999. Legal Penalties and Audit Quality: An Experimental Investigation. *Contemporary Accounting Review* 16:685-710.

Kornhauser, Lewis A., and Andrew Schotter. 1990. An Experimental Study of Single-Actor Accidents. *The Journal of Legal Studies* 19:203-233.

Kornhauser, Lewis A., and Andrew Schotter. 1992. An Experimental Study of Two-Actor Accidents. Research Report No. 92–57, C. V. Starr Center for Applied Economics, New York.

Krupka, Erin L., and Roberto A. Weber. 2013. Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association* 11:495-524.

Miceli, Thomas J. 2017. *The Economic Approach to Law.* 3rd edition. Stanford: Stanford University Press.

Miceli, Thomas J., and Kathleen Segerson. 2005. Do Exposure Suits Produce a 'Race to File'? An Economic Analysis of a Tort for Risk. *Rand Journal of Economics* 36:613-627.

Murphy, Ryan O., Kurt A. Ackermann, and Michel J. J. Handgraaf. 2011. Measuring Social Value Orientation. *Judgment and Decision Making* 6:771-781.

Oxoby, Robert J., and John Spraggon. 2008. Mine and Yours: Property Rights in Dictator Games. *Journal of Economic Behavior and Organization* 65:703-713.

Polinsky, A. Mitchell, and Steven Shavell. 2007. The Theory of Public Enforcement of Law. Pp. 403-454 in Vol. 1 of *Handbook of Law and Economics*, edited by A. Mitchell Polinsky and Steven Shavell. Amsterdam: North Holland.

Schmitt, Manfred, Anna Baumert, Mario Gollwitzer, and Jürgen Maes. 2010. The Justice Sensitivity Inventory: Factorial Validity, Location in the Personality Facet Space, Demographic Pattern, and Normative Data. *Social Justice Research* 23:211-238.

Shavell, Steven. 2007. Liability for Accidents. Pp. 139-182 in vol. 1 of *Handbook of Law and Economics*, edited by A. Mitchell Polinsky and Steven Shavell. Amsterdam: North Holland.

Shavell, Steven. 2011. Corrective Taxation Versus Liability. *American Economic Review: Papers and Proceedings* 101:273-276.

Shavell, Steven. 2013. A Fundamental Enforcement Cost Advantage of the Negligence Rule Over Regulation. *The Journal of Legal Studies* 42:275-302.

Sullivan, Sean P., and Charles A. Holt. 2017. Experimental Economics and the Law. Pp. 78-103 in Vol. 1 of the *Oxford Handbook of Law and Economics*, edited by Francesco Parisi. New York: Oxford University Press.

Sunstein, Cass R. 1996. On the Expressive Function of Law. *University of Pennsylvania Law Review* 144: 2021-2053.

van Velthoven, Ben C. J. 2009. Empirics of Tort. Pp. 453-498 in *Tort Law and Economics*, edited by M. Faure. Cheltenham/Northampton: Edward Elgar.

Weigend, Thomas. 2001. Sentencing and Punishment in Germany. Pp. 188-221 in *Sentencing and Sanctions in Western Countries*, edited by Michael H. Tonry and Richard S. Frase. New York: Oxford University Press.

Wittman, Donald, Daniel Friedman, Stephanie Crevier, and Aaron Braskin. 1997. Learning Liability Rules. *The Journal of Legal Studies* 26:145-162.

APPENDICES

Appendix A: Translated instructions for the care game in the main experiment

General Information on the Experiment

Welcome to this experiment!

In this experiment, your decisions – and possibly the decisions of other participants – will have an influence on your payments. It is therefore very important that you read these instructions carefully. The experiment will be conducted in complete anonymity. In other words, you will not find out with whom you have interacted, just as those interacting with you will not find out anything about your identity. During the experiment, you must not speak to any of the other participants. Please raise your hand if you have any questions. We will come to you. Disobeying these rules will lead to exclusion from the experiment and all payments.

The experiment consists of 4 parts. These instructions will inform you about Part 1. We will distribute the instructions for Parts 2 and 3 to you just before those respective parts begin. The instructions for Part 4 will be shown to you on your screen later on.

You can earn money in all four parts of the experiment. Parts 1 and 4 are definitely relevant for your payment. Whether Part 2 or Part 3 will be paid out to you will depend on a random decision made by the computer at the end of the experiment. Whether Part 2 or Part 3 is chosen will be equally probable.

During the experiment, we will speak not of Euro, but of points. Your entire income will hence initially be calculated in points. The total number of points accumulated by you will be converted into Euro and paid out to you at the end of the experiment, at a conversion rate of:

1 Point = 0.02 Euro.

Information on the First Part of the Experiment

The first part of the experiment lasts a maximum of 10 minutes. During this time, you will be asked to solve the following task: You will be shown a table on your screen. A table consists of 10 lines, each of which has 15 numbers. These numbers are either 0 or 1. Your task is to guess correctly the number of zeros in the table, to write this figure in the appropriate box, and to click OK on your screen. If you have entered the correct number, a new table will automatically appear. If not, please double-check your solution for the current table, enter a new number, and confirm by clicking OK. If you enter the wrong value three times, this task will be considered unsolved and you will automatically be directed to a new table.

Your points account will be credited with 750 points if you correctly solve 3 tables within the time specified.

The first part of the experiment ends once either every participant has correctly solved three tables or once 10 minutes have elapsed. If one or several people are unable to solve three table tasks correctly within the specified time, the experiment is over for these participants. These people will only receive 250 points for showing up today. Since we require an even number of participants to continue, one person will be randomly selected in the case of an uneven number of remaining participants, and the experiment will end for this person as well. However, this person will receive the 750 points for solving the task. The participants concerned should remain in their booths for the further duration of the experiment.

[Treatment DAMAGES]

Information on the Second Part of the Experiment

In Part 2 of the experiment, there are two roles: A and B. One person who has Role A and another who has Role B are assigned to each other at random. Whether you are assigned Role A or Role B is determined by chance. You are told at the beginning of the second part which role you have been assigned.

Part 2 of the experiment consists of three phases:

Phase 1:

In Phase 1, Person A can use points. With the number of points used, Person A decides how high the probability is of 300 points being deducted from Person B in the second phase.
From Table 1, A chooses a number of points he or she wishes to invest. It is not possible to choose a number of points that is not in the table. If A does not invest any points, then the points deduction for B occurs with a probability of 100 percent, i.e., definitely. A higher investment of points by A reduces the probability of points being deducted from Person B. The lowest probability of a points deduction is 20 percent and occurs if 240 points are used. The relationship between the points investment by A and the probability of points being deducted from Person B is shown in Table 1. The points used by A are deducted from A's points account. B is not told how many points A has used.

Points Invested by A	Probability of Points Being Subtracted from B, in Percent
0	100
15	80
30	72
45	65
60	60
75	55
90	51
105	47
120	43
135	40
150	37
165	34
180	31
195	28
210	25
225	22
240	20

Table 1: Number of points used by A and the resulting probability level for a subtraction of points from B

Phase 2:

In Phase 2, the computer determines whether Person B is docked 300 points from his or her points account. For this the computer randomly chooses a number between 1 and 100. If this number is smaller than or the same as the probability determined by A for a points subtraction from B, then 300 points are deducted from B's account. If the random figure is higher, there is no deduction. Hence, the points invested by A in Phase 1 determine how probable a points deduction from B will be in Phase 2. A and B are told whether or not 300 points have been subtracted from B's account. The random number chosen by the computer is not communicated.

Phase 3:

Should B have suffered a points deduction in Phase 2, 270 points will be subtracted from A in the third phase. These 270 points will then be credited to B's account. If no points are subtracted from B in Phase 2, then nothing will happen in Phase 3.

Income after Part 2:

The income of A and B after the end of Part 2 therefore depends on whether points were subtracted from B's account.

If points were subtracted from B's account, then the income is calculated as follows:

Income A =

750 points from Part 1

– chosen points investment from Part 2 Phase 1

– 270 points to B from Part 2 Phase 3

Income B =

750 points from Part 1

– 300 points subtracted in Part 2 Phase 2

+ 270 points from A in Part 2 Phase 3

If no points were subtracted from B's account, then the income is calculated as follows:

Income A =

750 points from Part 1

– chosen points investment from Part 2 Phase 1

Income B =

750 points from Part 1

Please note that, at this point of the experiment, you do not yet receive the information from Phases 2 and 3, and hence have not yet received any information on your payments. You will be given this information at the end of the experiment.

Following Part 2, we ask both A and B to answer one question each on the points distribution in Phase 1. Answering this question can earn you further points. More information on this will be shown to you on your screen.

Please answer some control questions first. Then Part 2 of the experiment will begin.

[Treatment DAMAGES]

Information on the Third Part of the Experiment

In Part 3 of the experiment, there are also two roles: Role A and Role B. Each Person has the same role in Part 3 of the experiment which they had in Part 2. So, if in Part 2 you had Role A, then you will have Role A in Part 3 also. If you had Role B in Part 2, you will also have Role B in Part 3.

In Part 3, as before, one person with Role A and one Person with Role B are randomly assigned to each other. However, it is not possible for the person you are drawn with to be the same person who was already assigned to you in Part 2. You will therefore definitely interact with a different person in Part 3.

Part 3 of the experiment consists of three phases:

Phase 1:

Phase 1 in Part 3 is identical with Phase 1 in Part 2. Person A decides how many points, if any, to invest, thereby determining the probability with which 300 points will be deducted from the assigned Person B in the second phase. For this, A chooses a certain number of points from Table 1 (see the information on the second part of the experiment). The points invested are deducted from A's account.

Phase 2:

Phase 2 in Part 3 is identical with Phase 2 in Part 2. As described in the information on the second part, the computer will decide whether or not 300 points will be subtracted from Person B's account.

Phase 3:

Should B have suffered a points deduction in Phase 2, 270 points will be subtracted from A in the third phase. These 270 points will then be credited to B's account. If no points are subtracted from B in Phase 2, then nothing will happen in Phase 3.

Income after Part 3:

The income of A and B after Part 3 therefore depends on whether points were subtracted from B's account.

If points were subtracted from B's account, then the income is calculated as follows:

Income A =

750 points from Part 1

– chosen points investment from Part 3 Phase 1

– 270 points to B from Part 3 Phase 3

Income B =

750 points from Part 1

– 300 points subtracted in Part 3 Phase 2

+ 270 points from A in Part 3 Phase 3

If no points were subtracted from B's account, then the income is calculated as follows:

Income A =

750 points from Part 1

– chosen points investment from Part 3 Phase 1

Income B =

750 points from Part 1

Please note that, at this point of the experiment, you do not yet receive the information from Phases 2 and 3, and hence have not yet received any information on your payments. You will be given this information at the end of the experiment.

Following Part 3, we ask both A and B to answer one question each on the points distribution in Phase 1 of Part 3. Answering this question can earn you further points. More information on this will be shown to you on your screen.

Please answer some control questions first. Then Part 3 of the experiment will begin.

[Treatment FINE]

In Treatment FINE the text for Phase 3 read:

Phase 3:

Should B have suffered a points deduction in Phase 2, 270 points will be subtracted from A in the third phase. The value in Euro of these 270 points subtracted from Person A's account will be donated to one of the charities named in Table 2. You can find Table 2 at the end of these instructions. The charity will be chosen at random at the end of the experiment. If there is no points deduction for B in Phase 2, then nothing will happen in Phase 3.

Further the following table was attached to the instructions:

Table 2: Liste of Charities:

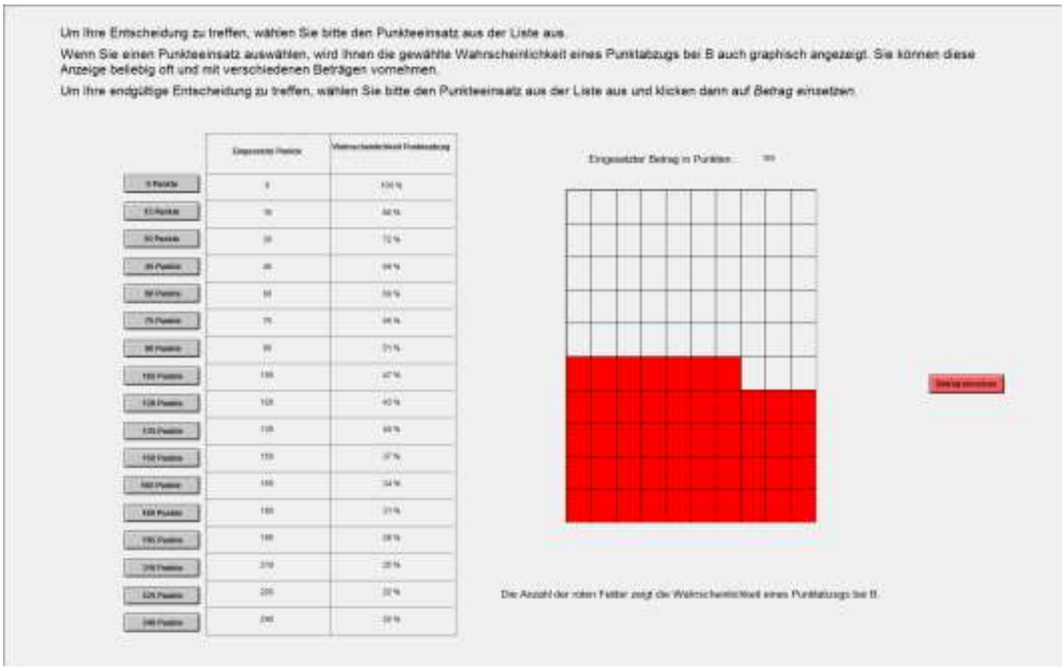
1. Deutsches Rotes Kreuz e. V.
2. Ärzte ohne Grenzen e. V.
3. Deutsche Welthungerhilfe e. V.
4. SOS-Kinderdörfer weltweit

[Treatment BASELINE]

In Treatment BASELINE the instructions only included descriptions for Phase 1 and Phase 2 as there was no Phase 3.

Appendix B: Screenshots of Care Decision & Belief Elicitation in Main Experiment

Figure 9: Injurers' decision screen



The table contains information about the cost of care and the accident probability. How many out of 100 squares are filled in red illustrates the accident probability based on the subject's preselection.

Figure 10: Injurers' second-order belief elicitation screen

We asked injurers to state their belief about the potential victim's first-order belief using the 17 categories listed beneath each other with explicit reference of the interval boundaries.

Appendix C: Treatment specific regression analyses main experiment

Table 5: Care Investments at the Individual Level

	(1) Baseline	(2) Damages	(3) Fine
SVO	1.44**	-0.24	0.40
	(0.41)	(0.64)	(0.64)
Risk Attitude	1.58	-2.77	6.07
	(3.26)	(4.75)	(5.07)
Justice Sensitivity	6.18	3.12	2.51
	(5.07)	(8.53)	(7.93)
Part 3	-15.84	2.29	-2.19
	(10.10)	(16.24)	(16.52)
Constant	71.46	5.10	40.56
	(80.60)	(89.93)	(130.83)
Demographics	YES	YES	YES
N	102	99	107

Notes: Results from OLS regressions. SVO controls for subjects' social value orientation. Risk Attitude controls for subjects' risk preferences (on a scale from 1 to 6). Justice Sensitivity (perpetrator) controls for subjects' justice sensitivity (average of two 6-item Likert-like scales). The Part 3-dummy equals 1 when the choice stems from Part 3. Demographic controls include participants' age, gender, experimental experience, number of siblings, a dummy whether the subject is a student, their field of study, their semester, and a dummy whether participants work for more than 10 hours per week. ** indicates significance at the 1% level.