Discussion Papers of the
Max Planck Institute for
Research on Collective Goods
2020/8

Fines versus Damages: Experimental Evidence on Care Investments

Florian Baumann
Tim Friehe
Pascal Langenbach

MAX PLANCK
SOCIETY

Discussion Papers of the
Max Planck Institute
for Research on Collective Goods

2020/8

Fines versus Damages:
Experimental Evidence on Care Investments

Florian Baumann / Tim Friehe / Pascal Langenbach

May 2020

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
https://www.coll.mpg.de

Fines versus Damages:

Experimental Evidence on Care Investments

Florian Baumann[1] Tim Friehe[2] Pascal Langenbach[3]

May 2020

Abstract

This paper studies the differential effects of fines and damages on people's investment in accident prevention. We report results from a laboratory experiment in which monetary payoffs are maintained across the two policy instruments. While standard theory predicts no difference in behavior, we find that potential injurers invest substantially more money in accident prevention when they are subject to damages instead of a fine. We discuss possible behavioral channels that may explain our findings.

JEL-Classification: H62, K42

Keywords: Externality; Care; Liability; Damages; Fine

[1] Center for Advanced Studies in Law and Economics (CASTLE), University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany. E-Mail: fbaumann@uni-bonn.de.

[2] Public Economics Group, University of Marburg, Am Plan 2, 35037 Marburg, Germany. CESifo, Poschingerstr. 5, 81679 Munich, Germany. E-Mail: tim.friehe@uni-marburg.de.

[3] Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, 53113 Bonn, Germany. E-Mail: langenbach@coll.mpg.de.

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 damages obligations 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 methods societies may use to control harm. Other formidable methods are the corrective tax, under which an agent pays the state an amount equal to the expected harm before an accident, and a fine regime, in which an agent pays the state an amount equal to the level of harm after an accident (e.g., Shavell 2007, 2011). Theoretically, a fine regime induces the exact same level of care as liability when the only difference between the two instruments is in terms of who eventually receives the payment. In this paper, we question this suggested equivalence using experimental data.

Our results show that different methods for controlling harmful behavior produce very different levels of care – even though the injurer's expected payments are the same across methods. In our experimental data, injurers who owe damages in the event of harm invest significantly more money in harm prevention than injurers who eventually owe a fine payment. Our paper complements the previous literature which compares the different instruments to control harmful externalities. While this literature, for example, has focused on potential informational or administrative differences between the instruments (e.g., Shavell 2013), it has always assumed that instruments with the exact same financial incentives will induce the exact same behavior.

Using a laboratory experiment, we compare the care investments of potential injurers when they owe damages after an accident to the care investments when they owe a fine. Our treatment variation thus only changes who eventually receives the payment while keeping the amount injurers have to pay constant. We compare strict liability, mandating injurers to pay damages to the victim after any accident, with a fine regime, forcing the injurer to pay a fine to society at large in the event of harm.[4]

Monetary payoffs are kept constant across treatments. This implies that the standard model predicts that equilibrium care will be independent of whether injurers are subject to fines or damages after an accident. However, individuals may perceive regime differences in other ways: For example, while notions of fairness that require the injurer to pay for harm caused are served in both scenarios, concerns about victim compensation are addressed only under liability (e.g., Kaplow and Shavell 2001). This already indicates that both conditions might provide different care motives for potential injurers. Below, we will elaborate on why care levels may depend on who eventually receives the payment even though monetary payoffs are the same. We will build on the recent literature in behavioral economics, particularly on theories of altruism, inequity aversion, and guilt aversion.

By means of a laboratory experiment, we are able to observe the injurer's incurred cost of precautions, whereas these are practically impossible to track in the field (e.g., 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 are able to 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.

In our experiment, we find that the average care investment in the damages treatment is about 25 percent higher than the one in the fine treatment. Both levels are significantly higher than the average care investment in a control condition without any financial obligation after an accident. In addition, we find a strong correlation between the injurers' actual care investments

[4] These payments may also be considered as liability to the state (Shavell 2019).

and beliefs about care-taking (both regarding potential victims' beliefs about the injurers' care levels and potential injurers' assessment of victims' beliefs).

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 describe behavioral hypotheses for our design. In Section 5, we present our results. We discuss our results and how they relate to our predictions in Section 6, before concluding with a brief summary of our findings and some policy implications of our research in Section 7.

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 (Angelova et al. 2014, Deffains et al. 2019, Kornhauser and Schotter 1990, 1992, Wittman et al. 1997). In the first paper in this area, Kornhauser and Schotter (1990) find 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) found 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 (2019) study whether participants choose the same level of care when they either act as a potential injurer under strict liability or as a potential victim under no liability. In analogy to our inquiry, they keep monetary payoffs constant across conditions and nevertheless observe behavioral differences. On a different note, Deffains et al. (2019) explore the interaction of obligations from tort law and social norms.[5]

[5] Croson (2009) and Sullivan and Holt (2017) survey the use of experiments in law and economics and discuss some more contributions to the realm of tort law. Dopuch and King (1992) and King and Schwartz (1999) are examples from the accounting literature investigating liability rules in experimental settings.

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. To the best of our knowledge, we are the first to explore this distinction's implications for care choices using laboratory data. There are other experimental papers contrasting compensatory payments with non-compensatory payments, but they investigate very different frameworks. Hoeppner et al. (2017) study a principal-agent setup in which the principal decides about establishing the principal-agent relationship. The agent then chooses a level of effort that increases the project's success probability at a private cost to the agent. The authors contrast different possible arrangements for payments due after a project failure, including a control condition in which the agent's payment is not received by the principal. In contrast to our study, they do not find a significant difference between the control condition and the damages condition. Eisenberg and Engel (2014) contribute to the literature on public goods and punishment. They empower one member of any group of four participants 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 compare, inter alia, behavior when the empowered player can transfer income for private gain to behavior when the player can destroy the other players' income. In Eisenberg and Engel (2014), in contrast to our study, whether losses enable transfers or not is chosen by the subjects themselves. Desmet et al. (2020) compare incentive effects of a compensation and a fine payment in a setup in which one participant may lie to another one for private gain. Whereas our design features incomplete compensation to ensure that subjects care about the bad event even in the damages treatment, Desmet et al. (2020) set the fine/compensation at the level of harm. In the first part of their experiment, subjects play the one-shot game and no statistically significant difference between treatments regarding the probability of lying results. In the second part of their experiment, the game is repeated four times and the data from this part suggest that the fine deters lying to a greater extent than a compensation requirement. The stronger deterrent effect of the 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 in terms 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 injurer or victim in one of the treatments. In Part 3, participants act 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 about the fact 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.[6] 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 are able to make their decisions. Next, we describe Parts 1 to 4, which are summarized in Table 1, in more detail.

[6] A translated version of our German instructions is included in the Supplementary Material.

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	**Stage 1**	Injurer chooses care investment
	Stage 2	Nature determines whether an accident results
		Victim loses 300 points in case of an accident
	Stage 3	In case of an accident • BASELINE: Injurer pays nothing • DAMAGES: Injurer pays 270 points as damages to victim • FINE: Injurer pays 270 points as a fine (received by a charity)
	Incentivized belief elicitation regarding behavior in Stage 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	**Stage 1**	Injurer chooses care investment
	Stage 2	Nature determines whether an accident results
		Victim loses 300 points in case of an accident
	Stage 3	In case of an accident • BASELINE: Injurer pays nothing • DAMAGES: Injurer pays 270 points as damages to victim • FINE: Injurer pays 270 points as a fine (received by a charity)
	Incentivized belief elicitation regarding behavior in Stage 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 about the level of care investments in two different treatments. We have participants in sequences BASELINE-FINE, BASELINE-DAMAGES, FINE-BASELINE, DAMAGES-BASELINE, FINE-DAMAGES, and DAMAGES-FINE.

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 (see, 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 (see, 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) emphasize, this boring task does not require any prior knowledge, performance is easily measurable, comes at a positive cost of effort, and there is little learning possibility. Each participant has to count correctly the number of zeros in three tables.[7] Failing to solve three tables correctly within 10 minutes leads to the exclusion from the experiment. Overall, 28 participants (out of 338) did not continue with the experiment after Part 1, either because they had not completed the task within the time limit or because they had to be excluded to retain an even number of participants in Parts 2 to 4 when other subjects failed to solve the task.[8]

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 players are referred to as Players A and B in the neutrally framed instructions. Victims remain passive in Parts 2 and 3, except for stating their beliefs about the injurers' choices.

Parts 2 + 3: Care Choice

Parts 2 and 3 each consist of three stages. In Stage 1, injurers choose their level of care. Our design very closely follows the unilateral-care accident setup (see, e.g., Shavell 2007). We distinguish three treatments: BASELINE, FINE, and DAMAGES. The possible care levels and their

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

[8] This 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 Supplementary Material 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.

respective effectiveness in terms of lowering the accident probability are represented in Table 2. The accident probability decreases at a diminishing rate with care to reflect the diminishing returns to care from the standard unilateral-care model. Before injurers commit to a care investment, they can experiment with different levels of care using a visualization of the accident probability.[9] Based on the injurer's care choice, the computer randomly determines for each injurer-victim pair whether or not an accident occurs in Stage 2. In the event of an accident, the victim loses 300 points. Payoff calculations in Stage 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 clearly 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.[10] This feature of our design reflects the redistributive element of fines justifying that fines are frequently treated as socially costless transfers in many circumstances (see, e.g., Polinsky and Shavell 2007).[11] 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.[12]

After Part 2 (and again after Part 3), we elicit beliefs. Victims state what care investment they expect from injurers, that is, we elicit victims' first-order beliefs. Injurers state their beliefs about what care level their matched victim expects from them. This is the injurer's second-order belief. Participants choose from a list of 17 intervals ranging from 0 to 240 points to state their belief.

[9] Figure A.1 in Appendix A shows the injurers' decision screen.

[10] 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.

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

[12] 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).

The belief elicitation is incentivized as follows (see, e.g., Cartwright 2018, 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.[13] Players earn 25 points for each correct belief statement.

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

[13] We guaranteed injurers that their matched victim will not learn their care investment. As victims were paid for correct beliefs about care investments, we asked them for their beliefs about care investment by *all other* injurers except their matched injurer in the session. No similar adjustment was necessary for the injurers' second-order beliefs.

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 actual 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. The number of subjects per treatment sequence is shown in Table 3 in Section 3.2.

Parts 4: Heterogeneity

In Part 4, participants complete a battery of individual tasks. First, subjects complete a version of the social value orientation slider measure (Murphy et al. 2011), as programmed by Crosetto et al. (2019). 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 lotteries. Additionally, we ask participants about their justice sensitivity using the short version from Baumert et al. (2014) of the items originally introduced in Schmitt et al. (2010). We also elicit participants' evaluation of the moral appropriateness of the different policy instruments that we consider in our treatments. 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. 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. 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.

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 we excluded two observations from the data set. The excluded subjects did not master the control questions prior to Part 2.

The subjects' mean age was 23 years. 66 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 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 donation to the charities was made by the experimenter after all sessions had been concluded.

4. Predictions

In this section, we describe behavioral predictions relying on different models from behavioral theory. Formal derivations are delegated to our Appendix B. We study standard theory, altruism, inequity aversion, and guilt aversion, focusing on the comparison of care investments in treatments FINE and DAMAGES.

Standard Theory. Standard theory assumes that subjects are only concerned about own material payoffs. Positive care investments result only if injurers face a financial obligation in the event of an accident (i.e., standard theory predicts no care in the BASELINE treatment). In treatments FINE and DAMAGES, subjects trade off higher costs of care and lower expected payments. Care investments only depend on the expected magnitude of the injurer's transfer, but not on the eventual recipient, and should therefore not differ between treatments FINE and DAMAGES. Denoting care in treatment $j, j \in \{FINE, DAMAGES\}$ by x^j we have

Hypothesis 1 (Standard Theory): *We expect* $x^{FINE} = x^{DAMAGES} > 0$.

Altruism. When subjects make their decision about care, they may also consider how it influences other subjects' payoffs. Fehr and Schmidt (2006), for example, report evidence that is consistent with this kind of unconditional altruism. The fact that victims incur harm in the event of an accident induces higher care in treatments FINE and DAMAGES (relative to the standard theory baseline) according to the weight of victim payoffs in the injurer's utility function. The fact that the injurer's payment increases the victim's or the society's payoffs in turn lowers care investments. Generally, altruism towards the victim may differ from altruism towards society at large. Thus, care will be lower (higher) in FINE than in DAMAGES when altruism towards society at large dominates (is smaller than) altruism towards the victim.

Hypothesis 2 (Altruism): *(i) We expect* $0 < x^{FINE} < x^{DAMAGES}$ *if altruism towards society at large is greater than altruism towards the victim. (ii) We expect* $x^{FINE} > x^{DAMAGES} > 0$ *if altruism towards society at large is smaller than altruism towards the victim.*

Inequity Aversion. Fehr and Schmidt (1999) proposed that much of behavioral data can be explained by considering that individuals dislike having payoffs that differ from the payoffs of others. The widespread use of inequity aversion theory has been attributed to its *empirical realism* combined with *analytical tractability* (Trautmann 2009). We assume that comparisons pertain to realized payoffs and note that − in treatment DAMAGES − the injurer experiences disadvantageous inequity in all circumstances. In the event of no accident, disadvantageous inequity is due to the fact that the injurer incurs the costs of care. In the event of an accident, the inequity is even magnified because the injurer additionally loses 270 points, whereas the victim only loses 30 points after receiving the damages payment. In treatment FINE, inequity is the same as in the DAMAGES treatment if no accident occurs. In the accident state, however, the inequity in the two treatments differs considerably, because in treatment FINE the injurer incurs costs of care and loses 270 points, while the victim loses 300 points. With a care investment larger than 30, the injurer experiences disadvantageous inequity in all circumstances in treatment FINE. When compared to the payoff difference in the accident state in the DAMAGES treatment, the payoff difference is much smaller in treatment FINE. Consequently, the injurer is less inclined to prevent the accident in treatment FINE than in treatment DAMAGES. We conclude:

Hypothesis 3 (Inequity aversion): *We expect* $0 < x^{FINE} < x^{DAMAGES}$.

Guilt Aversion. Assume that the injurer experiences guilt when she thinks that she disappointed the victim's payoff expectations (e.g., Battigalli and Dufwenberg 2007), supposing that the relevant comparison is between the victim's expected and actual payoffs. Actual victim payoffs fall short of expected payoffs when an accident occurs. Given some fixed care expectation, the difference between expected and actual payoffs in the accident state is larger in treatment FINE than in treatment DAMAGES. This ranking of payoff differences results because the victim is partly compensated in the DAMAGES treatment. The anticipated guilt from the accident state can thus be expected to be larger in the FINE treatment than in DAMAGES treatment, which should increase care in treatment FINE relative to care in treatment DAMAGES.[14]

Hypothesis 4 (Guilt Theory): *We expect* $x^{FINE} > x^{DAMAGES} > 0$.

[14] Endogenizing expected care using rational expectations, the above effect should be reinforced given the consistent beliefs by victims of higher care in treatment FINE than in treatment DAMAGES (the corresponding lower accident probability further compounds the feeling of guilt in the event of an accident in treatment FINE). See Appendix B.

5. 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 5.1). Next, we will compare care investments across treatments at the group level (Section 5.2), followed by an analysis at the individual level (Section 5.3). Finally, we will discuss the subjects' beliefs about care investments and the extent to which they match actual care choices (Section 5.4).

5.1 No Order Effects in Care Investments

Each injurer chooses care investments in two treatments (in Parts 2 and 3). When we test whether the order of treatments is relevant, 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).[15] Based on this finding, we present our results at the group level using the pooled data from Parts 2 and 3.

5.2 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 1. We thus find that injurers invest a greater amount 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 of the treatments (N = 52, p = 0.001). In addition, care levels in treatments BASELINE and DAMAGES (N = 47, p = 0.000), and BASELINE and FINE (N = 55, p = 0.000) are significantly different according to WSR tests.[16] In our pooled data, the mean care investment in treatment DAMAGES is about

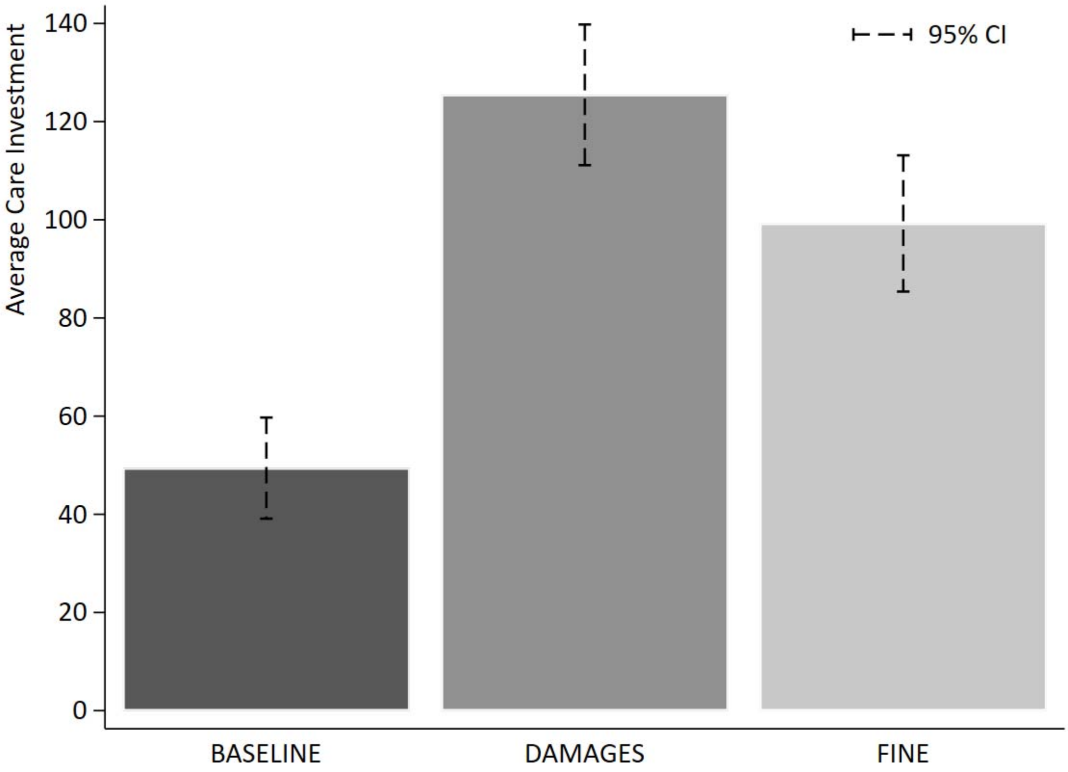
[15] All reported tests are two-sided.

[16] For all treatment comparisons, the pooled sample consists of a mix of independent and dependent observations. In the main text, we therefore report the within-comparison of subjects who played both

125% of the mean care investment in the FINE treatment, and the latter is about 200% of the mean care investment in treatment BASELINE.

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) [17] and the care that minimizes the injurer's expected costs in FINE and DAMAGES, which is a care investment of 60 points and 45 points, respectively.

Figure 1: Care Investments in Treatments BASELINE, DAMAGES, *and FINE*

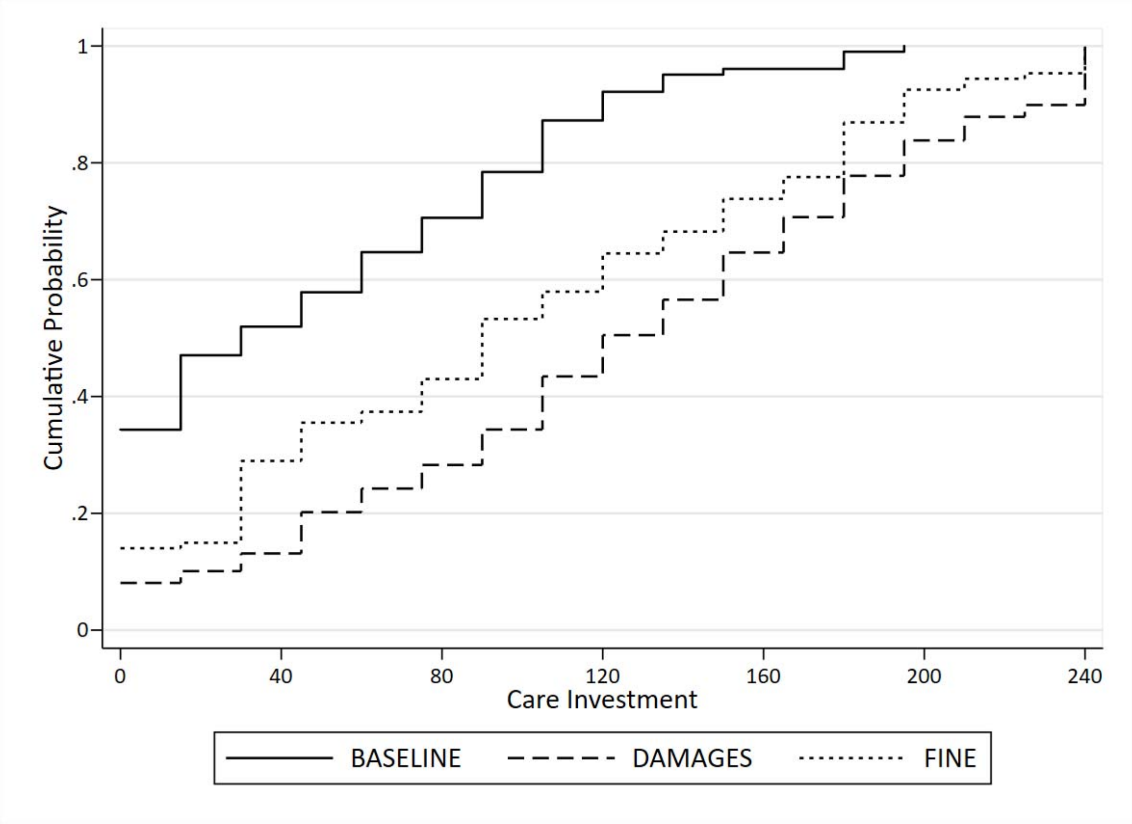


treatments. However, we could also compare care levels between subjects. To create a sample of independent observations, we start from the whole sample and, for subjects who played both treatments, drop the observations from Part 3. Treatment differences are significantly different according to MWU tests (N = 154, DAMAGES vs. FINES: p = 0.028; DAMAGES vs. BASELINE: p = 0.000; FINE vs. BASELINE: p = 0.000).

[17] 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 not featuring any of the behavioral characteristics mentioned above.

Figure 2 shows the cumulative distribution functions for the observed care investments. We find that the distribution for treatment BASELINE dominates the other two in the sense of first-order stochastic dominance, and that the distribution for treatment FINE dominates the distribution for treatment DAMAGES. About 64 (79) percent of subjects in FINE (DAMAGES) invest more than the 45 points, maximizing 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 do not invest in care at all.

Figure 2: Cumulative Distribution Functions of Care Investments in Treatments



5.3 Care at the Individual Level

The results at the group level explained in Section 5.2 are confirmed by regression analyses which take individual heterogeneity into account, using the information from Part 4. The fact that injurers chose their care levels in Parts 2 and 3 gives our data a panel structure. In order to exploit this panel structure, we use random-effects regressions for care choices in Parts 2 and 3. Regression results are presented in Table 3. 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 score, their risk preferences, and their 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.

Table 3: Treatment Effects on Care Investments at the Individual Level

	(1) Model 1	(2) Model 2	(3) Model 3
BASELINE	-50.04***	-49.85***	-50.95***
	(8.49)	(8.47)	(8.58)
DAMAGES	28.90***	28.36***	28.25***
	(8.57)	(8.55)	(8.68)
SVO		0.44	0.40
		(0.31)	(0.34)
Risk Attitude		3.35	3.04
		(2.49)	(2.67)
Justice Sensitivity		8.09**	5.72
		(3.92)	(4.34)
Part 3	-6.36	-6.36	-6.34
	(6.69)	(6.70)	(6.70)
Constant	101.63***	47.44**	65.92
	(7.12)	(20.61)	(55.49)
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. *, **, *** indicate significance at the 10%, 5%, and 1% level.

We confirm the treatment effects reported in Section 5.2 as indicated by the significance of the dummy variables 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, whereas the positive coefficient for the DAMAGES dummy variable shows that injurers invest significantly more money into accident avoidance 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 of 0.000 for all three models. These results are robust to the inclusion of the control variables in Model 2 and the participants' demographics in Model 3. We find that neither the subjects' social value orientation, nor their risk preferences, nor their justice sensitivity significantly explain care investments once the sociodemographic information is incorporated.

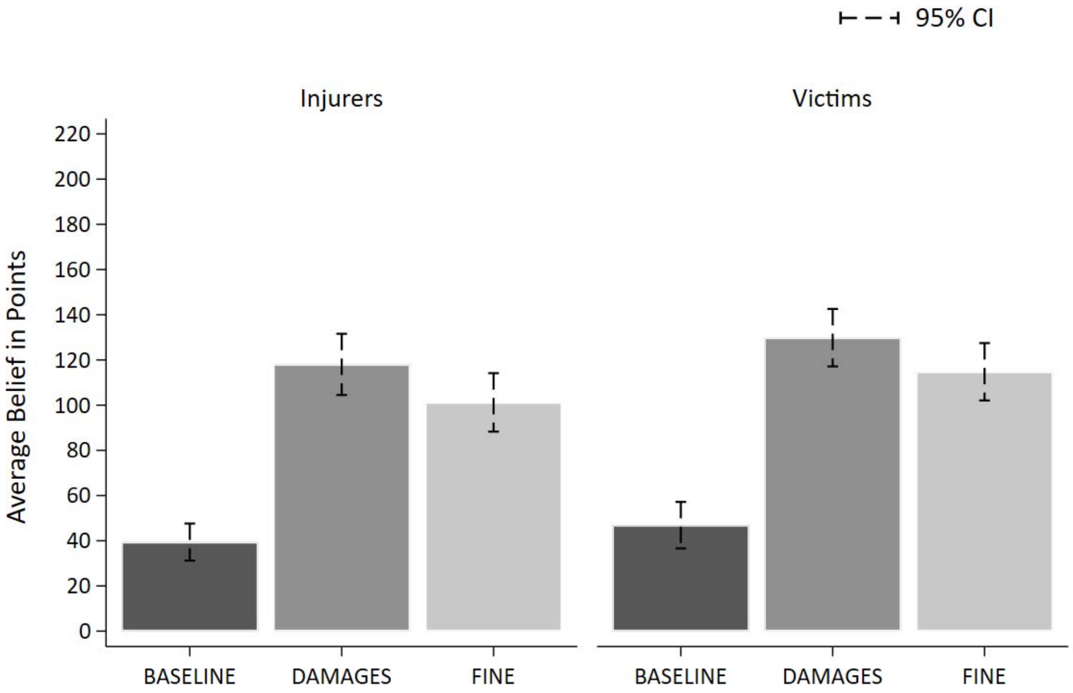
5.4 Beliefs

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 3 illustrates the respective average beliefs.[18] As the figures show, mean 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).

[18] 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, 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 <= 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.

We find that the injurers' second-order beliefs are treatment-dependent and show the same ranking as the actual care investments.[19] Differences in second-order beliefs between treatment BASELINE and either the DAMAGES or the FINE treatment are statistically significant according to WSR tests (p = 0.000 for both comparisons) for the subjects who played the 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 = 0.000).

Figure 3: Injurers' Second-Order Beliefs and Victims' First-Order Beliefs



[19] The victims' first-order beliefs are ranked similar to the injurers' second-order beliefs. Pairwise WSR tests display p-values of 0.000 for the comparison 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).

6. Discussion

Our results violate the standard-theory predictions, that is, we reject Hypothesis 1. Despite the fact that monetary payoffs were maintained across the treatments DAMAGES and FINE, the kind of payment proved to be important for care investments. Our findings are consistent with subjects being altruistic. According to Hypothesis 2 (i), treatment DAMAGES induces greater care investments than treatment FINE when altruism for society at large exceeds altruism regarding the victim. Our results are further consistent with Hypothesis 3 assuming that injurers are inequity-averse. Subjects invest more in accident avoidance when the accident state implies marked disadvantageous inequity (that is, in treatment DAMAGES). In contrast, the prediction derived assuming simple guilt aversion is inconsistent with our results (i.e., Hypothesis 4 is rejected by our data).

In addition to the approaches explained in Section 4, other factors may be relevant for our findings. For example, it may be that participants interpret the decision-making context they face in a given treatment in a treatment specific way that influences their choices. In the DAMAGES treatment, the institution mandates that the injurer pays a compensatory payment to the victim in the event of harm. In stark contrast, the institution in treatment FINE "tolerates" that the victim suffered harm in the event of an accident. The subjects' perception may thus be that, in the DAMAGES treatment, injurers are held responsible for causing harm *to their victims* (as injurers are responsible for their victims' integrity), while in treatment FINE, injurers are held responsible simply for having caused harm. By invoking a stronger feeling of responsibility in the DAMAGES treatment, this different perception of the respective institutions would lead to a ranking of care levels like that observed in our data. Along similar lines, Deffains et al. (2019) 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.

In the psychology literature, it is explained that people require compensation of 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 (e.g., Darley and Pittman 2003).[20] In data from our post-experimental questionnaire, we find some support 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. Specifically, in our questionnaire, we show participants a one-sentence vignette describing a situation in which an injurer could influence the accident probability and ultimately causes harm to another person. We then 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). We find that 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).[21] Against this backdrop, it may be hypothesized that our subjects act more in alignment with the perceived purpose of the policy when the policy itself is considered as more legitimate (see, e.g., Tyler 2003). The relatively more legitimate institution in the DAMAGES treatment should thus trigger relatively greater care investments than in treatment FINE, as is true for our data.

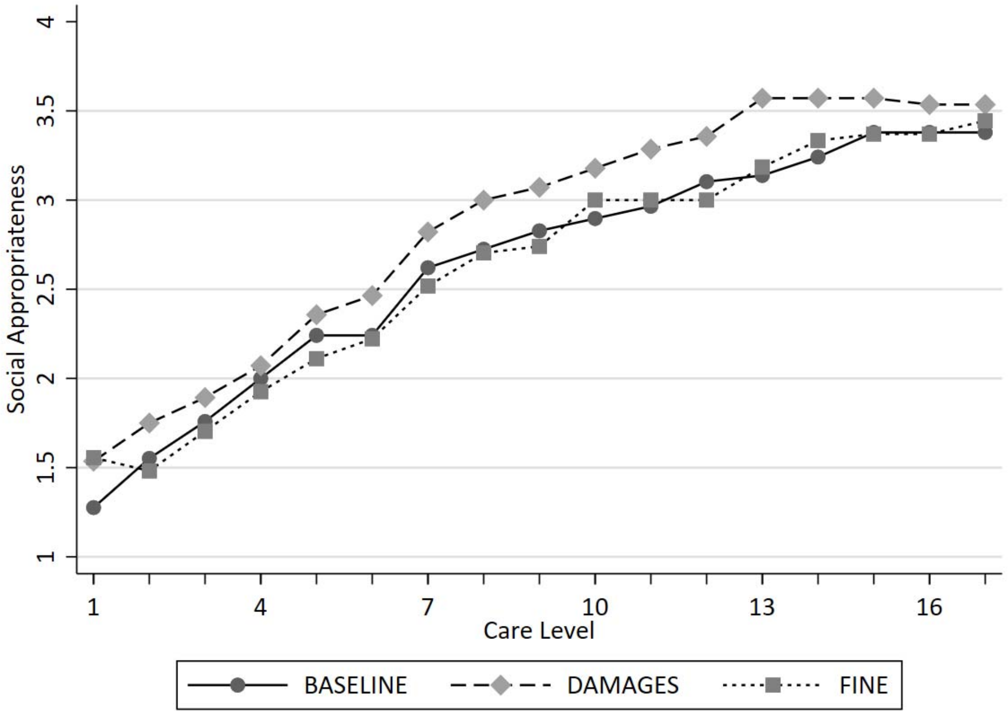
We followed up on the perceived appropriateness of the different institutions. In sessions conducted after the main experiment, we collected data about the social norms governing the different levels of care in our treatments (following Krupka and Weber 2013). Participants of that follow-up study were presented with the injurer's decision-making problem for one of the three treatments and, for every possible care choice, had to state whether they considered the care level either "very socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", or "very socially appropriate" (coded on a scale from 1 to 4). One of the 17 possible care choices was randomly chosen for payment, and subjects received 7 Euro if their rating equaled the modal response in the session. 84 subjects participated in four sessions in August 2019. We present the average ratings per care level in Figure 4. We find that the average

[20] In addition to the evaluation of the injurer's act in terms of intentions, it has been argued that whether or not compensation can repair damaged relationships depends on the level of compensation (e.g., De Cremer 2010, Desmet et al. 2011, Haesevoths et al. 2013).

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

social appropriateness of a given level of care is descriptively higher in the DAMAGES treatment, which may represent the overall greater approval of this institution (as explained above). However, other than that, we do not find any meaningful differences for the different regimes. Hence, the social appropriateness of the different care levels under the three institutions studied does not help to explain the reported treatment differences in care choices.

Figure 4: Social Norms Regarding Care Levels



7. 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. 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 different behavioral theories. If a liability regime creates a greater perceived responsibility for the integrity of the victim or is perceived as the relatively more legitimate institution for the context at hand, we would expect the ranking of care investments that we obtain in our data. Moreover, our findings are in line with predictions based on altruism – when altruism towards society at large (represented by charities in our design) exceeds altruism towards the victim – and inequity aversion.

Our key finding is important for policy-making. Two policy instruments that seem equivalent 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 instruments. We find that the fact that victims receive compensation – which at face value seems to lower the wrongfulness of the accident – raises observed care levels relative to a scenario in which injurers pay a fine to society at large. Our findings suggest that implementing the policy that serves the fairness ideals of many individuals also creates greater care investments. Our results are also important for recommendations regarding combining liability with fines, that is, payments to the state (e.g., Goerke 2002, 2003 or Shavell 2019), as the symmetry that is supposed in these contributions is questioned by our results.

Acknowledgements

We thank Michaela Pielmeier and Rovina Pinto for programming z-Tree and their research assistance. We gratefully acknowledge the feedback and suggestions provided by Claudia Cerrone, Christoph Engel, Susann Fiedler, Elissa Gentry, Leonie Gerhards, Eric Helland, Yoan Hermstrüwer, Oliver Himmler, Andreas Lange, Adi Leibovitch, César Martinelli, Marco Ottaviani, Hannah Schildberg-Hörisch, and Alexander Stremitzer. 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, J., Falk, A., Goette, L., and D. Huffman, 2011. Reference points and effort provision. *American Economic Review* 101, 470-492.

Angelova, V., Armantier, O., Attanasi, G., and Y. Hiriart, 2014. Relative performance of liability rules: Experimental evidence. *Theory and Decision* 77, 531-556.

Battigalli, P., and M. Dufwenberg, 2007. Guilt in games, *American Economic Review: Papers and Proceedings* 97, 170-176.

Baumert, A., Beierlein, C., Schmitt, M., Kovaleva, A., Liebig, S., and B. Rammstedt, 2014. Measuring four perspectives of justice sensitivity with two items each. *Journal of Personality Assessment* 96, 380-390.

Beranek, B., Cubitt, R., and S. Gächter, 2015. Stated and revealed inequality aversion in three subject pools. *Journal of the Economic Science Association* 1, 43-58.

Blanco, M., Engelmann, D., and H.T. Normann, 2011. A within-subject analysis of other-regarding preferences. *Games and Economic Behavior* 72, 321-338.

Bovenberg, A.L., and L.H. Goulder, 2002. Environmental taxation and regulation. In: A. J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, Vol. 3, North Holland: Elsevier.

Cartwright, E., forthcoming. A survey of belief-based guilt aversion in trust and dictator games. *Journal of Economic Behavior and Organization.*

Charness, G. and M. Dufwenberg, 2006. Promises and partnership. *Econometrica* 74, 1579-1601.

Charness, G., and M. Rabin, 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117, 817-869.

Charness, G. and M.O. Jackson, 2009. The role of responsibility in strategic risk-taking. *Journal of Economic Behavior and Organization* 69, 241-247.

Crosetto, P., Weisel, O., and F. 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, R., 2009. Experimental law and economics. *Annual Review of Law and Social Sciences* 5, 25-44.

Darley, J.M., and T.S. Pittman, 2003. The psychology of compensatory and retributive justice. *Personality and Social Psychology Review* 7, 324-336.

Dave, C., Eckel, C.C., Johnson, C.A., and C. Rojas, 2010. Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty* 41, 219-243.

De Cremer, D., 2010. To pay or to apologize? On the psychology of dealing with unfair offers in a dictator game. *Journal of Economic Psychology* 31, 843-848.

Deffains, B., Espinosa, R., and C. Fluet, 2019. Laws and norms: Experimental evidence with liability rules. *International Review of Law and Economics* 60, 105858.

Deffains, B., and C. Fluet, 2013. Legal liability when individuals have moral concerns. *Journal of Law, Economics, and Organization* 29, 930-955.

Desmet, P.T.M., De Cremer, D., and E. van Dijk, 2011. In money we trust? The use of financial compensations to repair trust in the aftermath of distributive harm. *Organizational Behavior and Human Decision Processes* 114, 75-86.

Desmet, P.T.M., Gerhards, L., and F. Weber, 2020. Is compensation fine? Sanction regimes and their effects on deterrence and trust. Mimeo.

Dopuch, N., and R.R. King, 1992. Negligence versus strict liability regimes in auditing: An experimental investigation. *Accounting Review* 67, 97-120.

Duersch, P., and J. Müller, 2015. Taking punishment into your own hands: An experiment. *Journal of Economic Psychology* 46, 1-11.

Eckel, C.C., and P.J. Grossman, 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior* 23, 281-295.

Eckel, C.C., and P.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, T. and C. Engel, 2014. Assuring civil damages adequately deter: A public good experiment. *Journal of Empirical Legal Studies* 11, 301-349.

Engelmann, D., and M. Strobel, 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94, 857-869.

Fehr, E., and K. Schmidt, 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817-868.

Fehr, E., and K. Schmidt, 2006. The economics of fairness, reciprocity, and altruism – Experimental evidence and new theories. In: Kolm, S.C., Ythier, J.M. (Eds.). *Handbook on the Economics of Giving.* Amsterdam: Elsevier.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171-178.

Füllbrunn, S., and T. Neugebauer, 2013. Limited liability, moral hazard, and risk taking: A safety net experiment. *Economic Inquiry* 51, 1389-1403.

Goerke, L., 2002. Accident law: Efficiency may require an inefficient standard. *German Economic Review* 3, 43-51.

Goerke, L., 2003. Road traffic and efficient fines. *European Journal of Law and Economics* 15, 65-84.

Greiner, B., 2015. Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association* 1, 114-125.

Guerra, A., and F. Parisi, 2019. Victims versus tortfeasors: An experimental test of the symmetric behavior hypothesis. Available at SSRN: https://ssrn.com/abstract=3133168 or http://dx.doi.org/10.2139/ssrn.3133168.

Haesevoets, T., Folmer, C.R., De Cremer, D., and A. van Hiel, 2013. Money isn't all that matters: The use of financial compensation and apologies to preserve relationships in the aftermath of distributive harm. *Journal of Economic Psychology* 35, 95-107.

Hoeppner, S., Freund, L., and B. Depoorter, 2017. The moral-hazard effect of liquidated damages: An experiment on contractual remedies. *Journal of Institutional and Theoretical Economics* 173, 1-22.

Kaplow, L., and S. Shavell, 2001. Fairness versus welfare. *Harvard Law Review* 114, 961-1388.

Kaplow, L. and S. Shavell, 2002. *Fairness versus welfare.* Harvard University Press: Cambridge, MA.

King, R.R., and R. Schwartz 1999. Legal penalties and audit quality: An experimental investigation. *Contemporary Accounting Review* 16, 685-710.

Koskela, E., and R. Schoeb, 1999. Alleviating unemployment: The case for green tax reform. *European Economic Review* 43, 1723-1746.

Murphy, R.O., Ackermann, K.A., and M.J.J. Handgraaf, 2011. Measuring social value orientation. *Judgment and Decision Making* 6, 771-781.

Landeo, C.M., Nikitin, M., and L. Babcock, 2007. Split-awards and disputes: An experimental study of a strategic model of litigation. *Journal of Economic Behavior and Organization* 63, 553-572.

Oxoby, R.J., and J. Spraggon, 2008. Mine and yours: Property rights in dictator games. *Journal of Economic Behavior and Organization* 65, 703-713.

Phaneuf, D.J., and T. Requate, 2017. *A course in environmental economics.* Cambridge, UK: Cambridge University Press.

Polinsky, A.M., and S. Shavell, 1994. A note on optimal cleanup and liability after environmentally harmful discharges. *Research in Law and Economics* 16, 17-24.

Schmitt, M., Baumert, A., Gollwitzer, M., and J. 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, S., 1993. The optimal structure of law enforcement. *Journal of Law and Economics* 36, 255-287.

Shavell, S., 2011. Corrective taxation versus liability. *American Economic Review: Papers and Proceedings* 101, 273-276.

Shavell, S., 2013. A fundamental enforcement cost advantage of the negligence rule over regulation. *Journal of Legal Studies* 42, 275-302.

Shavell, S., 2019. On the redesign of accident liability for the world of autonomous vehicles. *NBER Working Paper* 26220.

Sullivan, S.P., and C.A. Holt 2017. Experimental economics and the law. In: Parisi, F. (ed.), *Oxford Handbook of Law and Economics,* Oxford University Press, Chapter 5.

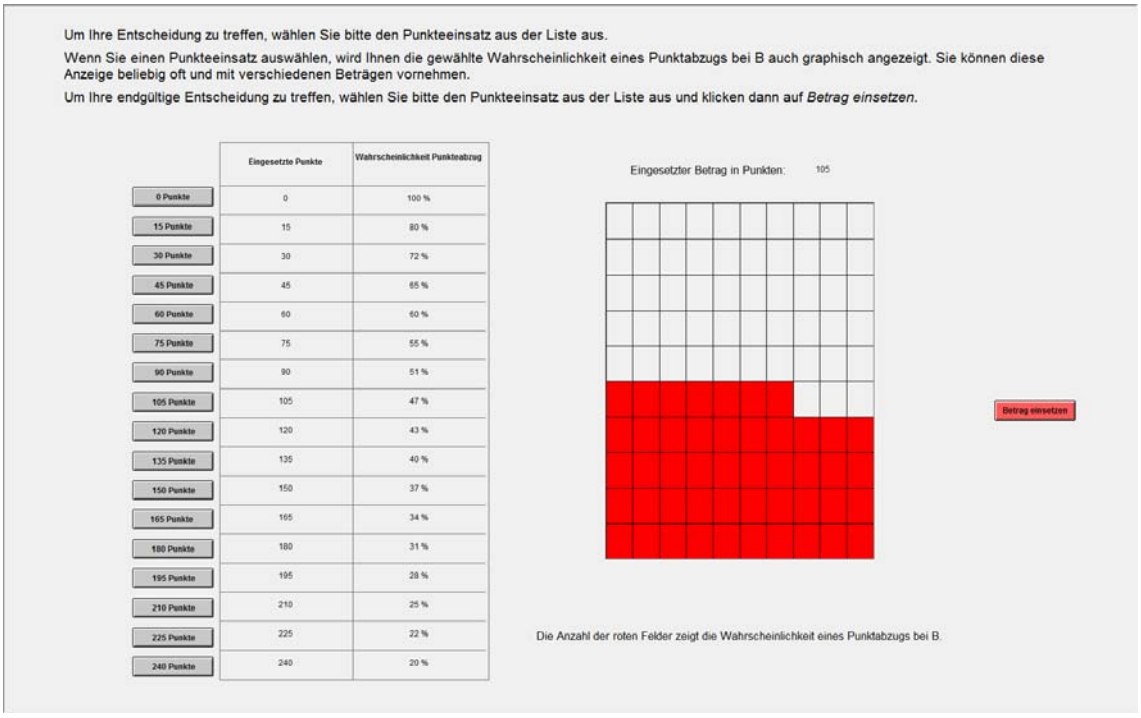
Trautmann, S.T., 2009. A tractable model of process fairness under risk. *Journal of Economic Psychology* 30, 803-813.

Tyler, T.R., 2003. Procedural justice, legitimacy, and the effective rule of law. *Crime and Justice* 30, 283-357.

Weigend T., 2001. Sentencing and punishment in Germany, in: Tonry, M. and R.S. Frase (eds.) *Sentencing and Sanctions in Western Countries,* Oxford University Press, Oxford.

Appendix A

Figure A.1: Injurers' decision screen



Um Ihre Entscheidung zu treffen, wählen Sie bitte den Punkteeinsatz aus der Liste aus.

Wenn Sie einen Punkteeinsatz auswählen, wird Ihnen die gewählte Wahrscheinlichkeit eines Punktabzugs bei B auch graphisch angezeigt. Sie können diese Anzeige beliebig oft und mit verschiedenen Beträgen vornehmen.

Um Ihre endgültige Entscheidung zu treffen, wählen Sie bitte den Punkteeinsatz aus der Liste aus und klicken dann auf *Betrag einsetzen*.

	Eingesetzte Punkte	Wahrscheinlichkeit Punkteabzug
0 Punkte	0	100 %
15 Punkte	15	80 %
30 Punkte	30	72 %
45 Punkte	45	65 %
60 Punkte	60	60 %
75 Punkte	75	55 %
90 Punkte	90	51 %
105 Punkte	105	47 %
120 Punkte	120	43 %
135 Punkte	135	40 %
150 Punkte	150	37 %
165 Punkte	165	34 %
180 Punkte	180	31 %
195 Punkte	195	28 %
210 Punkte	210	25 %
225 Punkte	225	22 %
240 Punkte	240	20 %

Eingesetzter Betrag in Punkten: 105

Die Anzahl der roten Felder zeigt die Wahrscheinlichkeit eines Punktabzugs bei B.

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 player's preselection.

Appendix B: Formal Derivation of Hypotheses

Along the lines of our design, we assume that, having earned an amount equal to y in the real-effort task, the injurer selects care x in Stage 1, implying investment costs equal to $cx^2/2$ and an accident probability amounting to $p(x) = 1 - x$ where $x \in [0, \overline{x} < 1]$. We assume quadratic cost of care and a linear accident probability for expositional convenience only. An accident in Stage 2 imposes harm amounting to h on the victim. In Stage 3, a payment D (F) is due after an accident occurred in the DAMAGES (FINE) treatment. The payment D is transferred to the victim (i.e., represents damages). The payment F is transferred to society at large (i.e., represents a fine). Our experimental setup assumes $y = c = 750$, $h = 300$, and $D = 270$ $(F = 270)$ in treatment DAMAGES (FINE).

Standard Theory. Standard theory assumes that the injurer's expected utility EU_I depends only on her own financial payoffs. Assuming risk neutrality for simplification, we have

$$EU_I = y - c\frac{x^2}{2} - (1 - x)(D + F)$$

from which we obtain equilibrium care as

$$x^S = \frac{D + F}{c}$$

which is the same for treatment FINE $(F = 270, D = 0)$ and treatment DAMAGES $(F = 0, D = 270)$.

Altruism. Suppose that the injurer's utility consists of not only of her own expected financial payoffs, but also of both the victim's payoffs with weight λ and society's payoffs with weight μ, where we assume $\lambda, \mu \ \epsilon$ [0,1). The parameters λ and μ describe the extent of altruism towards the victim and society at large, respectively. The injurer's objective function can be stated as

$$EU_I = y - c\frac{x^2}{2} - (1 - \ x)(D + F) + \lambda[y - (1 - \ x)(h - D)] + (1 - x)\mu F$$

The first-order condition

$$\frac{dEU_I}{dx} = -cx + (D + F) + \lambda(h - D) - \mu F = 0$$

is solved by the privately optimal level of care

$$x^A = \frac{(D+F) + \lambda(h-D) - \mu F}{c}.$$

Comparing the FINE and the DAMAGES treatment, care incentives are greater (lower) in treatment FINE than in DAMAGES if $\lambda > \mu$ ($\lambda < \mu$), that is, if the injurer cares more about the victim than society as a whole (more about society as a whole than the victim).

Inequity aversion. In our application of inequity aversion theory, the injurer's objective function can be stated as

$$\begin{aligned} EU_I &= (1-x)\pi_I^{accident} + x\ \pi_I^{no\ accident} \\ &-(1-x)[\alpha \max\{0, \pi_V^{accident} - \pi_I^{accident}\} + \beta \max\{0, \pi_I^{accident} - \pi_V^{accident}\}] \\ &\quad - x[\alpha \max\{0, \pi_V^{no\ accident} - \pi_I^{no\ accident}\} + \beta \max\{0, \pi_I^{no\ accident} - \pi_V^{no\ accident}\}] \end{aligned}$$

where π_I (π_V) indicates the realized injurer (victim) payoff with

$$\pi_I^{accident} = y - \frac{cx^2}{2} - F - D, \pi_V^{accident} = y - h + D$$

$$\pi_I^{no\ accident} = y - \frac{cx^2}{2}, \pi_V^{no\ accident} = y$$

The circumstance with *disadvantageous* inequity enters with a weight α that exceeds the weight β attached to the state with *advantageous* inequity, $\alpha > \beta \geq 0$ with $1 > \beta$.

In treatment DAMAGES, with $D = 0.9 * h$, the victim's payoffs always exceed payoffs for the injurer, creating disadvantageous inequity for the injurer in both states. We obtain the objective function

$$EU_I = y - \frac{cx^2}{2} - (1-x)D - \alpha\left[\frac{cx^2}{2} + (1-x)(2D-h)\right].$$

where the term in brackets displays the injurer's expected income disadvantage. Privately optimal care in the DAMAGES treatment results as

$$x^{DAMAGES} = \frac{D + \alpha(2D-h)}{c(1+\alpha)}.$$

Marginal costs of care are higher than standard theory predicts because higher care widens the gap between the victim's payoffs and the injurer's payoffs in both states of the world. However, the greater income difference in the accident state of the world implies an additional incentive to invest in accident avoidance.

In the FINE treatment, the victim's payoffs exceed those of the injurer in the event of no accident. After an accident, the victim's payoffs are higher than the injurer's payoffs only if harm falls short of the sum of the fine plus care costs, i.e., if $h < F + cx^2/2$. There is a level of care x^k defined by

$$c\frac{(x^k)^2}{2} = \ h - F$$

that divides the cases in which the injurer experiences advantageous inequity in the accident state of the world ($x < x^k$) from those cases in which disadvantageous inequity obtains ($x > x^k$). Assuming the parameter values used in our experiment, we obtain $x^k = 0.28$ corresponding to care costs of 30 and $x^k < x^{DAMAGES}$ for all $\alpha \geq 0$.[22] For care levels $x > x^k$ in treatment FINE, the injurer is experiencing disadvantageous inequity in all states of the world, and the objective function is given by

$$EU_I = y - \frac{cx^2}{2} - (1-x)F - \alpha\left[\frac{cx^2}{2} - (1-x)(h-F)\right].$$

If the privately optimal care level is above x^k, this care level is given by

$$x^{FINE} = \frac{F - \alpha(h-F)}{c(1+\alpha)}$$

In contrast to treatment DAMAGES, the income difference in the accident state is smaller than that in the no accident state. This reduces care incentives.

Guilt theory. Following our description in Section 4, the injurer maximizes

$$\begin{aligned} EU_I &= (1-x)\pi_I^{accident} + x\ \pi_I^{no\ accident} \\ &-\gamma(1-x)[(1-\bar{x})\pi_V^{accident} + \bar{x}\ \pi_V^{no\ accident} - \pi_V^{accident}] \\ &= y - c\frac{x^2}{2} - (1-x)(D+F) - \gamma\ (1-x)[\bar{x}(h-D)] \end{aligned}$$

[22] To put the care level x^k into perspective, note that, for example, we obtain $x^I_{DAMAGES} = 0.34$ for our parameter values and $\alpha = 1$. The mean level of α in the data set of Blanco et al. (2011), for example, is about 1.2. See also Beranek et al. (2015).

where $\bar{x}$ represents the reference point for own care and γ the importance of guilt aversion, $0 < \gamma < 1$. The reference point may stem from the injurer's beliefs about the victim's expectations, or from what the tortfeasor believes the average individual would do, or from what the injurer thinks is morally right to do (Cartwright 2018). Privately optimal care results as

$$x^G = \frac{D + F + \gamma\bar{x}(h - D)}{c}.$$

For a given reference point, care in treatment FINE exceeds that in DAMAGES. A higher reference point $\bar{x}$ increases care. In all likelihood, the reference point is treatment-dependent. Given that the care investment in treatment FINE exceeds that in treatment DAMAGES for given care expectations, it seems likely that the reference point in FINE surpasses that in DAMAGES. This reinforces the difference between scenarios in terms of what care level is expected from the injurer. This should hold if, for instance, the reference point is determined by beliefs about the behavior of the average individual or rational second-order beliefs. Should expectations be based on rational equilibrium beliefs, $x^G = \bar{x}$, we obtain

$$x^G = \frac{D + F}{c - \gamma(h - D)}$$

implying $x^{FINE} > x^{DAMAGES}$.

Supplementary Material: Translated instructions care game

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.