Pull yourself up by your bootstraps: Identifying procedural preferences against helping others in the presence of moral hazard

Rostislav Staněk[a,*], Ondřej Krčál[a], Katarína Čellárová[a]

[a] *Department of Economics, Masaryk University, Czech Republic*

Abstract

Governments and organizations often implement policies designed to help people affected by undesirable events. Such policies can make the society better off, but they may also create moral hazard. We use a laboratory experiment to examine two questions. First, can discretionary decisions to provide assistance overcome the problem of moral hazard and lead to higher efficiency? Second, if so, will people prefer this discretionary procedure to the strict liability policy in which no assistance is provided? We find that assistance is more efficient than a strict liability procedure. However, people still prefer the strict liability regime over assistance. We conduct additional treatments that show that this effect is not driven by the presence of human discretion, nor by aversion to risk, ambiguity, loss or inequality. This suggests that when moral hazard is a concern people have procedural preferences in favor of strict liability.

Keywords: Strict liability, Assistance, Procedural preferences, Experiment, Moral hazard
JEL: C91, D90

*Corresponding author
Email address: rostislav.stanek@econ.muni.cz (Rostislav Staněk)

1. Introduction

This paper studies whether people prefer public provision of costly assistance to people affected by undesirable events or whether they prefer such people to cover their losses themselves. We are interested in situations where such assistance is efficient (its benefits outweigh its costs) and where the effort made to avoid an undesirable event is not observable. These situations are also characterized by moral hazard that may undermine incentives to take the unobservable precautionary measures (Ehrlich and Becker, 1972). Such situations are common and economically relevant. They include decisions to help people suffering from drug addiction (Doleac and Mukherjee, 2018) and HIV positive patients (Lakdawalla et al., 2006). They also include decisions to reallocate work from one team member to another (Chan, 2016), to provide assistance to victims of natural disasters (Browne and Hoyt, 2000), or to bail out banks or companies (Farhi and Tirole, 2012).

We will use the following stylized example to guide us through this paper. Imagine a passenger who arrives late at the airport. The airline faces a choice between speeding up his/her security procedures at a cost to its other passengers or having the passenger miss his/her flight. The passenger's late arrival might be the result of bad luck (e.g. unexpected traffic problems) or could be due to negligence. Therefore, the airline's discretionary decision cannot be based directly on the level of precautions the passenger took. The discretionary decision to speed up the passenger's security procedures is likely to be efficient: the benefits of the passenger not missing the flight outweigh the costs. However, such choices also create moral hazard: In this case, the result could be an inefficient situation in which a large number of passengers arrive late.

This paper proposes an experimental design in which helping people in need (e.g. by speeding up procedures for passengers who arrive late to the airport) may or may not lead to higher payoffs, depending on the severity of the moral hazard problem. It introduces an effort-provision game in which agents try to avoid a bad outcome by expending costly effort. What matters, however, is the *observable effort*, which is equal to the sum of the actual effort chosen by the agents and a non-positive random variable called *bad luck*. If the observable effort falls below a certain threshold, the bad outcome is not avoided. The game has two regimes: *strict liability* and *assistance*. In strict liability, agents themselves bear the losses created by insufficient effort or bad luck; in assistance, an official might transfer those losses to all the other participants. The officials are given a (vaguely stated) standard, which advises them to provide assistance only to participants whose actual level of effort is above an optimally set threshold. However, it is difficult for the officials to apply this standard because they do not see the participants' actual effort, only their observable effort. In our guiding example, in the strict liability regime all agents (passengers) arriving late would miss their flights. In the assistance regime, the official (airline) would apply the standard to speed up procedures for the passengers who had made a sufficient effort to arrive on time, at a cost to the others.

We address two questions: First, we investigate whether the participants overcome the moral hazard problem so that assistance is more efficient in terms of monetary payoffs to participants than strict liability. Second, we test whether the participants' preferences for assistance or strict liability correspond to the respective monetary outcomes.[1] To answer these questions, each subject plays the effort-provision game in both regimes. We then let the subjects vote on which regime they preferred and implement their preferred regime in the last part of the experiment. In assistance, monetary efficiency, measured by the sum of monetary payoffs to participants, depends on the official's ability to avoid the moral hazard problem. This can be done by providing assistance in a way that motivates agents to expend sufficient effort levels. We find that the monetary payoffs in the assistance regime are higher than in the strict liability regime, which means that the officials are able to enforce the standard and the moral hazard problem can be avoided. Despite this, the majority of participants vote for the strict liability regime and this preference for strict liability is confirmed by a discrete choice model which controls for payoff differences between the regimes.

In addition to the version of the effort-provision game explained above, in which the official is played by a participant in the experiment (which we refer to as the HUMAN treatment), we also introduce two additional between-subject treatments to rule out some possible explanations for the strict liability preference. This preference may be related to some more

[1] This latter question is motivated by cases of assistance programs that generate opposition, which materializes in citizen protests or elections and often leads to such programs being limited or terminated. Such protests are often driven by people feeling that these programs are arbitrary and unfair; such resentment intensifies in times of economic hardship (see e.g. the description of opposition to the Moving to Opportunity program in Goering et al., 2003).

general behavioral mechanisms: People have a well-documented tendency to avoid situations in which another person determines their outcome. In a trust game, people require higher expected payoffs if there is any chance that they could be betrayed by a human opponent, compared to the equivalent game against a computer (Bohnet and Zeckhauser, 2004; Bolton and Ockenfels, 2010). In principal-agent experiments when people are asked to delegate decisions directly to someone else, they sacrifice monetary gain in order to make the decision themselves (Owens et al., 2014; Fehr et al., 2013; Bartling et al., 2014). In line with these results, the preference for the strict liability regime that our experiment reveals might be related to the presence of the human official in the assistance regime. In order to investigate this explanation, we conduct a NATURE treatment in which the human official is replaced by a known probability distribution taken from the officials' choices made in the sessions with human officials. We find that people still prefer strict liability over assistance, which means the unpopularity of the assistance is not driven by the presence of a human official but by the liability assignment aspect of the procedure.

The preference for the assistance regime might also reflect differences in the distribution of round payoffs between these regimes: the payoffs in assistance have lower variance and fewer of them are negative. This might lead to a preference for assistance if subjects are averse to inequality, risk, ambiguity, or loss. The CONTROL treatment generates the same payoff distribution for both regimes as in HUMAN and NATURE, but the subjects do not make any choices. The preference for *strict liability* is no longer present in this treatment, which shows that this result is not driven by aversion to risk, ambiguity, loss or inequality.

These findings about the preference for strict liability are in line with recent literature on *procedural preferences* showing that individuals value institutions and procedures for their intrinsic value. Most of this literature makes use of pie-splitting games to show that people prefer procedures that provide fair randomization over unequal outcomes (Bolton et al., 2005; Karni et al., 2008), or guarantee a kind distribution of outcomes (Sebald, 2010). Our finding, however, seems to be driven by what are known as *purely procedural preferences*, i.e. preferences for procedural properties that are unrelated to payoffs or outcomes (Chlaß et al., 2019). Chlaß et al. (2019) document that people consider procedural simplicity, efficiency, the distribution of the decision and information rights in their choices. In line with the notion of purely procedural preferences, Sausgruber and Tyran (2014) showed in a laboratory experiment that people prefer uniform taxes over discriminatory taxes that are equally efficient and produce the same expected outcomes. We contribute to this literature by showing that the contradiction between efficiency and preferences also applies in the case of (public) assistance in situations fraught with moral hazard. Moreover, we find that the preference for the strict liability regime is correlated with deontological attitudes[2]. We interpret that correlation as suggestive evidence that the voting decision in our experiment reflects the discrepancy between efficiency and a norm that people should be fully liable for their own losses.

Our paper is also related to the literature on moral hazard problems in loss-sharing situations.[3] Besides one empirical study that documents the existence of moral hazard problems in many real-world contexts, a few laboratory experiments have examined the effects of loss-sharing on loss-reducing investment (Füllbrunn and Neugebauer, 2013; Mol et al., 2020), risky lottery choices (Bixter and Luhmann, 2014) and overtreatment at a credence good market (Huck et al., 2016). The experiment most closely related to the current study is that presented by Füllbrunn and Neugebauer (2013), who consider a situation in which participants can make an investment in order to avoid losses. The experiment includes a full liability treatment, in which participants are fully liable for their loss, and a limited liability treatment in which the losses are shared equally within a group. They find that limited liability leads to lower loss-avoidance investment, however any efficiency comparison in their experiment would be trivial because individual preferences are fully aligned with social welfare in the full liability treatment. Unlike Füllbrunn and Neugebauer (2013), we study a situation in which loss-sharing is not given by a rule known in advance, but rather it depends on human discretion guided by a vaguely stated standard. Moreover, we focus on situations in which providing loss coverage has direct benefits in terms of social welfare. We contribute to the literature by investigating whether losses caused by a decrease of loss-avoidance effort outweigh the direct benefits of loss coverage.

[2]People with deontological attitudes judge an action based on whether that action is right or wrong under a set of rules, rather than based on the consequences of the action.

[3]Note that our problem differs from the substantial literature on moral hazard in teams (Holmstrom, 1982). In teams, shirking behavior is pervasive since individual effort levels are substitutes and team members are paid according to their aggregate effort. In our case, individual effort levels are not substitutes and cannot be aggregated.

Our experiment also resembles the determination of liability in tort law models with unilateral accident (Shavell, 2007). These models assume that the overall value of the loss does not depend on the assignment of liability. Our experimental design, instead, follows the assumption that the loss is larger when the agent has to cover it by himself. Given the lack of real-world data, several experiments have been conducted to examine the relative advantages of strict liability rules and negligence standards. These papers have focused primarily on the effects of strict liability and negligence on effort levels (Kornhauser and Schotter, 1990; Angelova et al., 2014; Deffains et al., 2019). The experiment we present in this paper is different from these contributions both in terms of its aim and the experimental design. Our experimental design contains a human decision-maker in the role of an official. This is motivated by i) the need to interpret generally-defined standards; ii) our focus on moral hazard, where we investigate whether human officials are able to credibly enforce the given standard. We contribute to the literature by providing evidence that ranking different liability rules based on revealed preferences does not coincide with ranking these rules based on monetary payoffs.

The rest of this paper is structured as follows: Section 2 formulates the theoretical framework. Section 3 introduces the experimental design and discusses the predictions. Section 4 presents the results, which are then briefly discussed in section 5.

2. Theoretical framework

The structure of our theoretical framework follows our two research questions. First, we compare the efficiency of the assistance and strict liability regimes in the effort-provision game. Second, we discuss procedural preferences for strict liability.

2.1. Efficiency of regimes

In the effort-provision game, the agent chooses the effort level e. The effort is costly and the monetary costs are given by the function $c(e)$ which is increasing and non-concave, i.e. $c'(e) > 0$ and $c''(e) \geq 0$. Bad luck b is a random variable with support $[-\underline{b}, 0]$ and probability distribution function $f(b)$ which is increasing, $f'(b) > 0$. The effort level and bad luck determine whether the outcome is bad or good. A bad outcome occurs if the sum of the effort and bad luck falls below the threshold T, i.e. $b + e < T$. In our guiding example, e is the effort expended to arrive at the airport in time, such as getting up early, using more secure means of transport, etc. Bad luck can take many forms, such as the passenger facing unexpected traffic problems. The natural situation in which the passenger misses their flight unless assistance is provided constitutes the threshold.

Only the sum $b + e$ is observable to the third party. The effort itself e is not observable. If the bad outcome happens, there is a loss that needs to be covered. The loss can be covered by the agent or by the society (other agents in the group). In the former case the value of the loss is L_A, in the latter case it is L_S. We assume that $L_A > L_S$, which represents that the society as a whole is better off if the loss is covered by others.[4] In our example, the airport personnel observes late arrival, but cannot usually verify to what extent the situation was caused by negligence. L_A is the loss resulting from a missed airplane, while L_S would be the extra waiting time or other inconveniences or risks caused by speeding up the security checks.

The regimes differ in the way the liability for the loss is assigned. Under the *strict liability* regime, the responsible agent always pays the loss. The *assistance* regime is complemented by a standard which states that the loss will be paid by the agent if he did not exert sufficient effort to prevent the loss. A benevolent official decides whether this requirement was met. The official can observe $e + b$, which we call observable effort. In our guiding example, the observable effort could be the arrival time of the passenger coupled with the general traffic situation known to the official.

The welfare function in this modelling framework is given by the negative value of total monetary costs

$$-Pr(b + e < T)l - c(e).$$

It comprises of three elements: i) the probability that the total effort falls below the threshold $Pr(b + e < T)$; ii) the loss l which is paid by the agent or society $l \in \{L_A, L_S\}$; iii) the costs of exerting effort $c(e)$.

We provide the solutions of the model under four different conditions: strict liability regime (SL), assistance without commitment (A), assistance with commitment (AC) and first-best solution (FB); and we compare their welfare consequences. In the strict liability regime, the loss is always paid

[4] When the losses are equal $L_S = L_A$, the first-best solution can be achieved by a simple strict liability regime.

by the agent. The agent chooses the effort level e^{SL} in order to minimize the expected loss and effort costs. In the assistance regime with commitment, the official first chooses the standard D. If $D < e+b < T$, the official lets the society to cover the loss. In the second stage, the agent chooses the optimal effort level e^{AC}.

Note that a benevolent but myopic official may not necessarily be willing to enforce the standard D. The problem is that enforcing a standard is dynamically inconsistent in a one-shot game. Once the effort decision is made and effort costs are sunk, the official is tempted to deviate from the standard and let the society pay for the loss even if the observable effort falls below D. When the agent realizes this dynamic inconsistency, he/she puts in zero effort. This situation is labeled as the assistance regime without commitment. The first-best solution provides a welfare benchmark by choosing both variables l and e to maximize the welfare function.

The following proposition compares of the effort and welfare levels in these four situations. The proof and details of calculations are to be found in the Supplementary material.

Proposition 2.1. *The effort levels under the different regimes rank as follows* $e^A < e^{AC} \leq e^{FB} < e^{SL}$*. The welfare under the different regimes ranks as follows* $W^A < W^{SL} < W^{AC} < W^{FB}$.

We will use this proposition to formulate hypotheses about the outcomes of the experiment.

2.2. Preference for regimes

The literature shows that people have preferences regarding procedures that generate monetary outcomes (Bolton et al., 2005; Karni et al., 2008; Sebald, 2010). In our framework, people might prefer the strict liability regime for purely procedural reasons (Chlaß et al., 2019). The following reasons, as discussed by Chlaß et al. (2019), are relevant in our setting:

- *Transparency:* In strict liability, each agent's payoff is a function of his/her own choice and the realization of the random variable bad luck; his/her payoff function does not depend on choices which he/she cannot observe. In assistance, however, the agent's payoff function depends on fellow players' choices (the application of a vague standard, effort) which the agent does not observe. Assistance is therefore an nontransparent procedure.

- *Inequality in information:* In both regimes, information is unequally distributed amongst players. In the assistance regime, agents are at a disadvantage compared with officials. The officials learn about all the agents' observable effort and have private information about whether or not they will apply the vague standard at the time of their decision. In the strict liability regime, agents have an information advantage over officials since the latter have no information at all. Hence if agents prefer to have the information advantage they will prefer strict liability.

- *Simplicity*: The strict liability procedure is simpler because agents do not need to take the official's choice into account and hence do not need to reason about so many strategies.

In addition to the purely procedural concerns covered by Chlaß et al. (2019), agents might dislike their payoffs being reduced because of other agents' possible negligence.

3. Experimental design and predictions

Our experiment consists of the effort provision game and a voting procedure. The effort provision game has two regimes: strict liability and assistance. These two regimes are played in the first two stages of the experiment. Voting in the third stage determines which regime will be played in the final stage.

Subjects are randomly matched into groups of five. Four subjects are given the role of *agents*, and one subject has the role of the *official*. The matching remains fixed during the whole experiment in order to strengthen the learning effect. The experiment consists of four stages: the strict liability regime, the assistance regime, the voting stage, and the final stage.

The strict liability regime has 15 periods. The officials are inactive: they do not make any decisions and they do not receive any feedback about the other agents' behavior. At the beginning of each period, agents are endowed with 140 CZK[5] and 6 tokens. The subjects know that they will lose between zero and six tokens according to a predetermined probability distribution.

[5] At the time of the experiment, 1 USD was equivalent to 22 Czech Crowns (CZK) and 1 EUR was equivalent to 25 CZK. A standard wage for an hour of unqualified student labour was approx. 100 CZK.

The probability distribution is presented in Table 1. Agents are given the opportunity to buy zero to six additional tokens. Each token costs 10 CZK. After each agent has made their purchase decision, a random draw determines how many tokens are lost. Any agent left with fewer than six tokens must cover this loss by paying 100 CZK.

Table 1: Probability distribution

Number of tokens lost	0	1	2	3	4	5	6
Probability	0.26	0.20	0.16	0.12	0.12	0.08	0.06

The assistance regime also consists of 15 periods. The only difference between the assistance regime and the strict liability regime is the official's active role. Officials do not observe the extent of bad luck, i.e. how many tokens are lost: they are only informed about the number of tokens remaining. If that number falls below six, the official chooses whether the loss is to be paid by the agent with the insufficient number of tokens or by the other three agents in the group. In the former case, the affected agent pays 100 CZK; in the latter case, each of the other three group members pays 25 CZK. Officials are instructed to let the agent with the insufficient number of tokens pay for the loss if they think that he/she bought fewer than two tokens. This instruction represents the standard and the agents are also aware of it. After each round, agents receive feedback about how many tokens they have and their own payoff. For each of the other agents, they are told whether the remaining number of tokens was below or above the threshold and how the official decided. This information about the other agents is displayed in random order, so neither the agents nor the official are able to track the identity of the other agents during the subsequent periods.

The purpose of the first two stages is twofold. First, we can test whether the assistance regime is more efficient, i.e. whether it leads to agents obtaining higher monetary payoffs. Second, agents become familiar with both the assistance regime and the strict liability regime. They learn what monetary payoffs can be gained in both regimes, enabling them to make competent voting decisions.[6] To control for possible order effects, half of the sessions

[6]This is why the order of the voting stage is fixed. We believe that any experimenter demand effects from this order are unlikely because it is not clear from our neutral instructions what our research question is, nor which of the regimes should be preferred.

were conducted with the assistance regime first and the strict liability regime second, while the other half were conducted with the regimes in the reverse order.

In the voting stage, the agents in each group vote on which of the regimes should be played in the final stage. They can vote for either strict liability or assistance. For each group, the regime that receives the majority of votes is chosen; if both regimes receive two votes, the regime for the final stage is chosen randomly (with a 50% probability of each). In the final stage, the participants play according to the rules of the regime chosen in the voting stage. The number of periods in the final stage is random. After each period, there is a 0.3 probability of the game ending. The random number of periods ensures that the final stage does not take too much time and the officials, if active, still face a trade-off between enforcing a sufficient level of effort and capturing gains by letting the group members pay for the loss.

The experiment has three treatments, which are used to rule out possible reasons for the observed preference for the strict liability regime. The baseline HUMAN treatment corresponds to the description above. The NATURE treatment replaces the official with a random device. If we find a preference for strict liability in NATURE, we may rule out possible ambiguity aversion or betrayal aversion related to the official's discretionary choices. The CONTROL treatment eliminates all choices by both officials and agents, so that the agents only passively observe the outcomes in both regimes. If the preference for strict liability is absent in CONTROL, this means that the preference for strict liability is not explained by aversion to inequality, risk, ambiguity or loss related to differences in the outcome distribution between the regimes.

The experiment follows standard procedures. The experimental instructions are read aloud at the beginning of each stage and subjects follow with their own copy. The instructions use neutral language and the subjects receive the instructions for each stage separately. At no stage are the subjects informed about what will happen in subsequent stages. At the end of the experiment, one randomly-selected period from each of the first two stages (strict liability regime, assistance regime), and the last period from the final stage are selected for payoff. The official obtains a payoff equal to the average payoff among the four agents from the selected periods, calculated separately for the assistance regime, the strict liability regime and the final stage. Note that this payment scheme provides incentives for the official to maximize the group's overall monetary payoff.

After the experiment, the subjects fill in a questionnaire. Most importantly, the questionnaire includes a shorter version of the consequentialist scale by Robinson et al. (2015), which we use to check for any ethical component in the preference for strict liability. This short version consists of four questions that assess endorsement of utilitarian or deontological beliefs.[7] Participants indicate how much they agree with each statement on a 5-point Likert scale. The total score ranges between 4 and 20, with higher scores showing a tendency towards a more utilitarian attitude and lower scores pointing towards a more deontological attitude. Additionally, the questionnaire contains the standard socio-demographic variables, measures of personality traits (the Big Five personality traits by Rammstedt and John (2007), each trait with 2 questions on a 5-point Likert scale and with the total score ranging between 2 and 10), self-reported risk attitude on a 10-point Likert scale (Dohmen et al., 2011) and tolerance to ambiguity (Budner, 1962) with a total score ranging between 4 and 20 (4 questions, 5-point Likert scale).

Table 2 presents the equilibrium predictions, based on the parameters and functional forms used in the experimental design, for all the theoretical regimes discussed in subsection 2.1: the strict liability regime (SL), assistance without commitment (A), assistance with commitment (AC), and first-best solution (FB).

Table 2: Equilibrium predictions

	Solution			
	SL	AC	A	FB
Effort levels (purchased tokens)	4	2	0	2
Bad outcome probability	0.14	0.38	0.74	0.38
Welfare (expected payoff)	86	88	84.5	91.5

Our experiment contains only two of these regimes: assistance and strict liability. While the strict liability regime corresponds closely to the theoretical SL regime, the choices the officials make and the agents' reactions to those choices determine whether the outcomes of the assistance regime more

[7]These questions are: "Rules and laws should only be followed when they maximize happiness."; "When deciding what action to take, the only relevant factor to consider would be the outcome of the action."; "Some rules should never be broken."; "It is never morally justified to cause someone harm."

closely resemble those of the theoretical AC or A regimes. The officials are instructed to follow the standard and assist those who are reasonably close to the threshold derived from the AC regime. We expect the agents to follow the standard and thereby avoid the moral hazard problem and achieve outcomes close to those predicted in the AC regime. This leads to our first hypothesis:

Hypothesis 1: Effort, measured by purchased tokens, is lower in assistance than in strict liability.

The model predicts that if agents reduce their level of effort and the officials follow the standard, AC will lead to higher monetary payoffs than SL. This generates our second hypothesis.

Hypothesis 2: The average monetary payoff is higher in assistance than in strict liability.

Preferences between the experimental regimes of strict liability and assistance are elicited in the voting stage, in which agents (but not officials) vote on which of the regimes should be played in the final stage. Based on the discussion of procedural preferences in subsection 2.2, we formulate the following conjecture about the outcome of the voting decision:

Conjecture 1: Contrary to their monetary incentives, agents prefer strict liability over assistance.

There are other possible explanations for such a preference for strict liability, besides procedural preferences. A preference for strict liability might be driven by the presence of a human official. Ambiguity-averse subjects might want to avoid the ambiguity related to the official's discretionary power to determine the outcome or they might feel betrayed if they comply with the standard and the official makes them liable for the loss. To rule these concerns out, we compare the results of the HUMAN treatment, in which the official is a human subject, with those of the NATURE treatment, which is identical to HUMAN in all aspects except that the official is played by a computer program. The computer decides according to a function that defines the probability that the loss is paid by other members of the group conditional on the observable effort. The value of probabilities was established as the fraction of choices in which the group members had to pay for the loss in the HUMAN treatment sessions. The decisions made by nature therefore mimicked the decisions made by the human officials. The probabilities are presented in Table 5, and the subjects were informed about the values of these probabilities in the instructions.

Preferences for strict liability might also result from different distributions of payment in the first stages of the experiment, which consist of 15 rounds each. More specifically, subjects are more likely to have negative payoffs in the assistance regime and the variance of those payoffs is higher than in the strict liability regime (see Figure 1 for the distribution of payoffs in our experiment). This might lead to a preference for strict liability if subjects are averse to loss, risk, ambiguity or inequality.[8] To address these concerns, we introduce an additional treatment called CONTROL, which is similar to the NATURE treatment but with one difference: subjects do not choose the effort. Instead, the number of purchased tokens is generated by the computer from the empirical distribution of purchased tokens in the NATURE and HUMAN treatments. This treatment exogenously generates the same distribution of payoffs.

Suppose the NATURE and CONTROL treatments reveal that ambiguity aversion or betrayal aversion created by the official's discretionary choices, and aversion to loss, risk, ambiguity or inequality due to the different distribution of payoffs among agents cannot explain a preference for strict liability. As the CONTROL treatment also rules out any reasons related to differences in expected payoffs and, ex-ante, agents expect the same equilibrium payoff in both treatments, such a preference cannot be the result of a preference for equal expected payoffs (Bolton et al., 2005; Karni et al., 2008). A preference for strict liability may then be due to several other aspects of the procedure, such as information transparency or inequality.

Since Chlaß et al. (2019) categorize information transparency and inequality as ethical concerns, we check the ethical component of the preference for strict liability (analogously to Chlaß et al. (2019)) by correlating it with a consequentialist scale (Robinson et al., 2015).

[8] As we discuss in the subsequent section, agents receive anonymized information about other agents' payoffs, so they are not able to calculate the total payoff to other subjects (in all 15 rounds). Inequality aversion might not be an issue here if they do not expect the other subjects' payoffs to differ substantially from theirs.

4. Results

This section presents the data and the results of the experiment. Our description of the results focuses on two separate questions. First, we test whether subjects exerted sufficient effort and whether monetary payoffs were higher in the assistance regime. Second, we analyse the subjects' preferences, which were elicited via voting.

4.1. Data

The experiment was conducted in October 2018 at the Masaryk University Experimental Economics Laboratory (MUEEL) in Brno, Czech Republic. In total, we recruited 328 student subjects using hroot (Bock et al., 2014). The experimental environment was programmed in zTree (Fischbacher, 2007). The experiment took about two hours and participants received 254 CZK (10 EUR) on average. There were 16 experimental sessions in total: 6 sessions of the HUMAN treatment with 120 subjects, 6 sessions of the NATURE treatment (116 subjects) and 4 sessions of the CONTROL treatment (92 subjects). Since the HUMAN subjects were divided into groups of 5 (4 agents and 1 official) and the subjects in the NATURE and CONTROL sessions were divided into groups of 4, this resulted in 24 groups (independent observations) in HUMAN, 29 in NATURE, and 23 in CONTROL.[9]

Table 3 shows the means of selected variables in the three between-subject treatments: HUMAN, NATURE and CONTROL. The table includes information about the number of agents and sessions (also split by the order of the first two stages), socio-demographic variables, psychological scales, and choice variables differentiated by the regime. Table 4 uses bootstrapped confidence intervals to test the effects of the two between-subject manipulations of interest. First, it shows that the order does not affect the number of tokens purchased or the payoffs in either regime. It also shows that there

[9]To test hypotheses 1 and 2, we planned to collect data from a total of 50 groups in the HUMAN and NATURE treatments. Our power analysis, which was based on a simulation with the distribution of purchased tokens observed in the pilot session as an input, indicated that this would be sufficient. With $N = 50$, we found a significant difference in 98 % of cases. The significance level was 5 % in a paired t-test based on group averages, and N refers to the number of groups (independent observations). When using the equilibrium values of purchased tokens (see Table 2) instead of pilot data, we obtain power 0.78 for $N = 50$. We collected data from 53 groups in HUMAN and NATURE.

are no significant differences between the outcomes in the HUMAN and NATURE treatments. In fact, all three treatments generate the same payoff distributions. We conclude that our computer algorithm in the NATURE treatment successfully simulated the choices made by human officials, and the setup of the CONTROL treatment successfully mimicked all the choices in the HUMAN and NATURE treatments.

Table 3: Descriptive statistics

	HUMAN	NATURE	CONTROL
Agents (sessions)	96 (6)	116 (6)	92 (4)
- Order 1: Assistance first	48 (3)	56 (3)	44 (2)
- Order 2: Strict liability first	48 (3)	60 (3)	48 (2)
Female	0.51	0.56	0.46
Age	21.3	22.0	22.4
Students of economics or business	0.63	0.69	0.62
Risk (10-point Likert scale)	5.28	5.43	5.54
Ambiguity scale (total score: 4–20)	11.33	11.17	11.58
Consequentialist scale (total score: 4–20)	8.31	8.89	8.33
Big five personality traits (BF)			
- BF extraversion (total score: 2–10)	5.46	5.30	5.15
- BF agreeableness (total score: 2–10)	5.16	5.11	5.37
- BF conscientiousness (total score: 2–10)	5.69	5.69	5.62
- BF neuroticims (total score: 2–10)	5.50	5.25	5.20
- BF openness (total score: 2–10)	6.40	6.27	6.24
Purchased tokens			
- Assistance	2.86	2.73	2.79
- Strict liability	3.61	3.84	3.79
Monetary payoffs			
- Assistance	86.6	86.8	86.9
- Strict liability	83.5	83.9	83.8
Frequency of loss			
- Assistance	0.29	0.30	0.30
- Strict liability	0.20	0.18	0.19

4.2. Effort and efficiency

This subsection provides the tests of Hypotheses 1 and 2. Hypothesis 1 states that fewer tokens will be purchased in the assistance regime than in the strict liability regime. Table 4 presents the differences in group averages between the two regimes (assistance − strict liability) and bootstrapped confidence intervals. In the assistance regime, agents in both the HUMAN and NATURE treatments purchase significantly fewer tokens (these decisions are not made in CONTROL). According to Hypothesis 2, the agents' average monetary payoff is higher in the assistance regime. As can be seen in Table 4, the shift from the strict liability regime to the assistance regime increases the monetary payoff in both HUMAN and NATURE treatments. Both hypotheses are thus confirmed by our data.

Table 4: Mean differences in tokens and payoffs based on group averages

Differences in averages between...	Regime/ treatment	Variable: Purchased tokens	Variable: Monetary payoffs
...order 1 and order 2	Assistance	−0.22 (−0.47, 0.05)	0.64 (−1.30, 2.54)
	Strict liability	0.11 (−0.13, 0.34)	1.04 (−1.23, 3.36)
...treatments HUMAN and NATURE	Assistance	0.13 (−0.21, 0.52)	−0.19 (−2.41, 1.87)
	Strict liability	−0.23 (−0.58, 0.09)	−0.4 (−2.93, 2.37)
...regimes assistance and strict liability	NATURE	−0.74* (−0.97, −0.52)	3.08* (0.75, 5.58)
	HUMAN	−1.11* (−1.32, −0.89)	2.85* (0.01, 5.76)

Note: The brackets report bootstrapped 95% confidence intervals. Individual observations (groups, each consisting of 120 observations = 4 agents × 30 periods) were resampled. * denotes significance at the 5% level.

These results suggest that the officials were able to keep the effort close to optimum in the assistance regime. Table 5 shows how likely officials were to decide that other members of the group should cover a loss per amount of remaining tokens. Table 5 also indicates whether two neighboring probabilities are significantly different from each other. This was tested by logit models, one for each neighboring pair of remaining tokens, in which the official's choice to cover loss is explained by a constant and a dummy variable for the number of tokens. The same empirical probabilities were used by the computer in the NATURE and CONTROL treatments. The best-response of the agents to this behavior is to purchase two tokens, which is the optimal amount under the AC solution. This shows that the officials were able to enforce the standard.[10]

[10]We also test whether the group averages are different from the theoretical prediction, which stated that agents should purchase 4 tokens in the SL regime and 2 tokens in the AC regime. The results show that the agents have a tendency to over-invest in the

Table 5: Official's behavior

Remaining tokens	5		4		3		2		1		0
Covered loss probability	0.88	$>^{**}$	0.66	$>^{***}$	0.38	$>$	0.25	$>$	0.24	$>^{**}$	0.0

Note: $^{**}p < 0.05$; $^{***}p < 0.01$. For each pair of remaining tokens we estimated a logit model in which a dummy (1 for a lower number of tokens) is regressed against a constant term, standard errors were clustered at individual level.

Our hypotheses are also supported by the regression models in Table 6, in which we control for individual (agent) fixed effects. The estimates present the within-individual effect of the assistance regime. They also show how the effect of assistance interacts with order and treatment. The estimates are based on data from all 30 periods. The standard errors are clustered at the group level (a group consists of 120 observations: 4 agents times 30 periods). The Poisson regressions in models 1 and 2 explain the number of purchased tokens, which ranges from zero to six. The OLS regressions in models 3 and 4 explain the agents' monetary payoffs. In line with our hypotheses, the table shows that the assistance regime leads to lower effort and higher payoffs. The interaction *Assistance* $\times$ *Order 1* is negative in model 2, which strengthens the effect of assistance on the number of purchased tokens. Conversely, the effect on monetary payoff is weakened by the interaction with order (opposite signs), but still the effect is statistically significant in Order 1 overall. When we split the results further by treatment, the significance is at the 5% level in HUMAN ($p = 0.028$) and at the 10% level in NATURE ($p = 0.062$). The *lost tokens* variable measures the number of tokens lost due to bad luck. It is a random number ranging from zero to six drawn from the distribution presented in Table 1. The number of lost tokens strongly predicts the monetary payoff, as losing more tokens increases the probability of falling below the threshold and experiencing a loss.

4.3. Voting

Next, we test Conjecture 1, stating that participants vote for the less efficient strict liability regime with higher frequency. In the HUMAN treatment,

assistance regime (Wilcoxon S-R test $p < 0.001$) and slightly under-invest in the strict liability regime (Wilcoxon S-R test $p < 0.001$). This confirms that the officials were able to overcome the moral hazard problem and the participants exerted sufficient effort in the assistance regime.

Table 6: Efficiency of the assistance regime

	Dependent variable:			
	Purchased tokens		Monetary payoff	
	Poisson marginal effects		*OLS*	
	(1)	(2)	(3)	(4)
Assistance	−0.894***	−0.490***	3.008***	3.490**
	(0.043)	(0.076)	(0.68)	(1.31)
Assistance × Order 1		−0.439***		−0.82
		(0.080)		(1.39)
Assistance × NATURE		−0.322***		−0.146
		(0.082)		(1.40)
Lost tokens			−10.65***	−10.65***
			(0.38)	(0.38)
Individual fixed effect	Yes	Yes	Yes	Yes
Observations	6,360	6,360	6,360	6,360
R^2			0.32	0.33

Note: Standard errors clustered at the group level including 120 observations (4 agents × 30 periods). $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

60 % of participants obtain higher payoffs in the assistance regime but only 38 % of them vote for the assistance regime. In the NATURE treatment, 60 % of participants obtain higher payoffs in the assistance regime but only 45 % of them vote for it.

Table 7 uses the following identification strategy to identify preferences for the strict liability regime. Assume that agents have a utility function

$$U_i(m_i, S) = \alpha_0 S + \alpha_1 m_i + \epsilon_i,$$

where m is the agent's monetary payoff, S is a dummy variable which takes the value one in the assistance regime and ϵ_i is the unobserved portion of utility. Based on the voting decision and actual payoffs in the first two stages of the experiment, we can use discrete choice techniques to identify the parameters α_0 and α_1. The parameter α_0 is interpreted as an alternative-specific constant indicating the utility of the assistance regime not related to monetary payoff. Negative values of this parameter suggest preference in favor of strict liability. The dependent variable *Voting* takes a value of one if the agent voted for the assistance regime and zero if he/she voted for the strict liability regime. The variable *Payoff difference* is the difference in the participants' average payoff between the assistance regime and the strict liability regime. We estimate the intrinsic utility of the assistance regime (alternative-specific variable α_0) for all three treatments separately. We label these variables *HUMAN*, *NATURE* and *CONTROL* in Table 7. We test whether the alternative specific constants are different from zero. A significant and negative result would mean that people prefer strict liability over assistance even after controlling for the payoff differences. We also control for the order effect and the interactions of order with payoff difference and treatments.

Table 7: Logit model explaining voting decision: average marginal effects

	Dependent variable: Voting		
	(1)	(2)	(3)
HUMAN	-0.187^{***}	-0.168^{**}	-0.165^{**}
	(0.052)	(0.339)	(0.077)
NATURE	-0.117^{**}	-0.146^{***}	-0.142^{***}
	(0.061)	(0.234)	(0.051)
CONTROL	0.015	0.061	
	(0.074)	(0.094)	
Payoff difference	0.017^{***}	0.014^{***}	0.013^{***}
	(0.003)	(0.004)	(0.004)
Order 1	0.032	−0.054	0.079
	(0.061)	(0.528)	(0.081)
Payoff difference × Order 1		0.004	0.003
		(0.006)	(0.006)
HUMAN × Order 1		0.047	−0.063
		(0.165)	(0.128)
NATURE × Order 1		0.142	
		(0.150)	
Consequentialist			0.059^{**}
			(0.030)
Observations	304	304	212

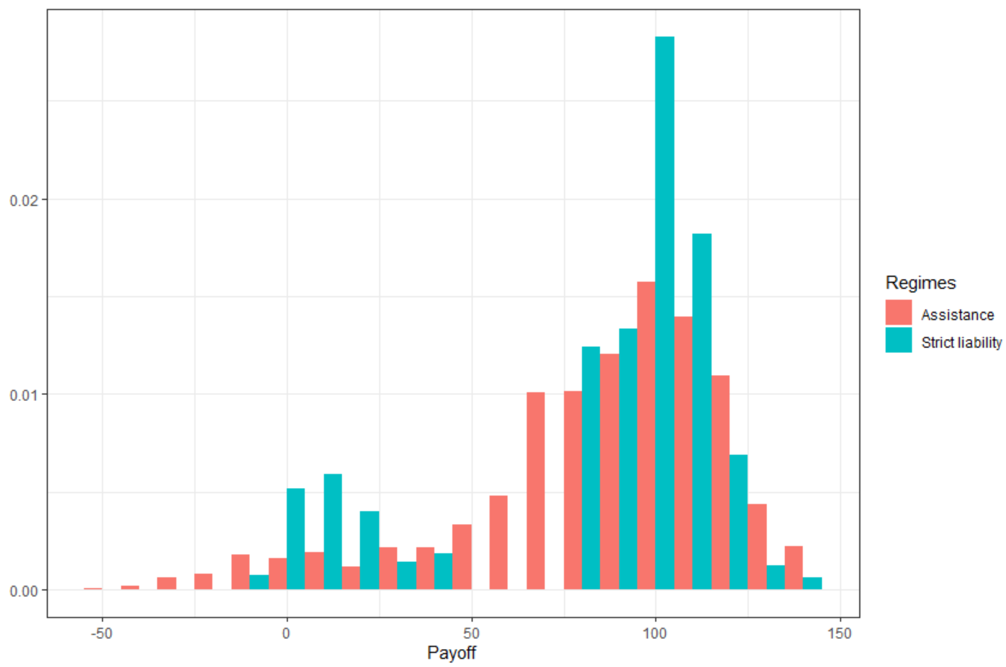
Note: The dependent variable *Voting* takes a value of one if the agent voted for the assistance regime and zero otherwise. Standard errors clustered at the group level (group includes 4 agents). $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Two main results stand out from model 1 in Table 7. First, there is a significant and substantial preference for the strict liability regime. The average willingness to pay to avoid the assistance regime the in HUMAN treatment is around 12 CZK (calculated as a ratio of the parameters *HUMAN*/*Payoff difference* from column 1 of Table 7), which is approximately four times the average payment difference between the regimes. Second, the relative unfavorability of the assistance regime is not driven solely by the presence of a human official. The preference for the strict liability regime is present even in the NATURE treatment. Although the presence of a human official makes the strict liability preference stronger, there is no statistical difference between the HUMAN and NATURE treatments. The average marginal effect of the human official is 0.061 ($p = 0.332$).

It is conceivable that our participants voted against the assistance regime because their monetary preferences are not fully captured by the difference in payoff between the strict liability and assistance regimes. Their voting decisions might be driven by the fact that they are averse to risk, ambiguity, loss or inequality. Indeed, the distribution of monetary payoffs in the assistance regime has not only a higher mean but also larger support. The minimum possible value of monetary payoff in the strict liability regime is -10 (the participant purchases 5 additional tokens and loses 6 tokens), while in the assistance regime it is -85 (the participant purchases 5 additional tokens, loses 6 tokens and pays an extra 75 in external costs). Figure 1 shows the actual distribution of monetary payments in the assistance and strict liability regimes, and confirms that the payoff distributions are different with the assistance regime having larger support. Although we control the payment difference between the regimes, this may not be sufficient since risk-averse or loss-averse agents might take the whole monetary payoff distribution into account when making their voting decisions.

In order to address this concern, we look at voting decisions in the CONTROL treatment. Recall that the CONTROL treatment exogenously generates the same payoff distribution as the NATURE or HUMAN treatments. Subjects in the CONTROL treatment then simply reveal their preference for the payoff distribution generated by the assistance regime or by the strict liability regime. If these participants only cared about monetary payoffs and the probabilities of securing them, the voting results in the CONTROL treatment should be the same as those found in the NATURE and HUMAN treatments. However, procedural preferences may create a wedge between the voting decisions. The coefficient for the CONTROL treatment shows

Figure 1: Histogram of payoffs in the assistance and strict liability regimes.



that preference for the strict liability regime completely disappears in this treatment. Participants in the NATURE and HUMAN treatments voted for the strict liability regime more often than participants in the CONTROL treatment. The average marginal effect is 0.20 for the HUMAN treatment ($p = 0.009$) and 0.13 for the NATURE treatment ($p = 0.074$). We do not observe any order effects. This result shows that participants' preference for the strict liability regime is not driven by risk aversion, loss aversion or any other preferences related to the payoff distribution. Overall, the results suggest that it is the strict liability aspect common to both the NATURE and HUMAN treatments that makes the strict liability regime preferable.

The model in column 3 uses data from the HUMAN and NATURE treatments, where we identify the preference for strict liability. It explores the correlation between the preference for strict liability and the consequentialist scale standardized to have mean 0 and standard deviation 1. We can see that the consequentialist scale is related to the preference for strict liability. Participants who agreed more with the position that some rules should be honoured in all circumstances were more likely to vote for the strict liability regime. Participants who expressed extreme utilitarian views (i.e. two standard deviations from the mean) did not manifest this preference. Using the specification of model 3, we estimated an additional seven models with other survey measures (risk, ambiguity scale, and Big Five scales) instead of the *consequentialist* variable, but we found that none of these were significantly related to voting.

5. Discussion

Societies and organizations implement many policies that prevent individuals from suffering losses or falling into hardship. Such policies face the challenge of recognizing whether a particular individual has suffered a loss due to bad luck or through his/her own negligence. Although some signals about a given agent's negligence are usually available, we focus on a situation in which it is not possible to design any rule to describe the complex nature of those signals and specify how the policy decision should depend on them. However, human officials are able to observe those signals and decide whether to provide assistance. Our experiment investigates whether the discretion of such officials, guided by a negligence standard, makes the society better off compared to a strict liability regime. We consider two dimensions of what it means to be better off: higher monetary payoffs and revealed preferences elicited through a vote.

We find that assistance provision guided by a general negligence standard leads to higher monetary payoffs than strict liability. This result shows that non-verifiable information can be valuable in moral hazard situations. Standard contract theory argues that an optimal contract in moral hazard situations should condition payoffs on verifiable outcomes correlated with effort provision. Non-verifiable information can be valuable only if it can be truthfully reported to a third party (Hart and Moore, 1999; Maskin and Moore, 1999). In our experiment, we assume that the information is observable by a third party, but it cannot enter the contract or legal norm. The solution in this case might be to set a vaguely stated standard and grant decision-making power to an official who is able to observe the non-verifiable information.

More interestingly, our experiment provides evidence, based on revealed preferences, that people prefer the strict liability regime over the assistance regime. Our results further demonstrate that this preference for strict liability cannot be fully explained by the presence of a human official and that the preference for strict liability vanishes when the agent cannot influence the level of effort and therefore the probability of loss occurring. This suggests that the preference for strict liability is procedural and seems to be related to people's aversion to having their payoff influenced by other, possibly negligent, agents and a (random) mechanism allocating the loss to everyone else but the negligent agent. This conclusion is supported by the fact that this preference is positively correlated with deontological attitudes. People with utilitarian attitudes do not exhibit any preference for the strict liability regime.

These results suggest that people might be reluctant to vote for policies that offer costly assistance to people who are at least partly responsible for their misfortunes. This finding is consistent with experimental evidence (Lefgren et al., 2016) and with the positive cross-country correlation between social spending and the belief that luck is the main factor determining income (Alesina et al., 2001). Our results complement these findings by showing that, within the specific setup we study, people do not favor policies that provide costly assistance even when those policies lead to higher monetary wealth for them and for society as a whole.

One obvious limitation of this research is the composition of our sample, which consisted exclusively of Czech students; this raises the question of how generalizable our results are. Since we find that preferences for strict liability are correlated with consequentialist attitudes, we believe that the results can be generalized at least for any population that has a similar composition of utilitarian vs. deontological attitudes. On the other hand, the aversion to the assistance regime that we identified might disappear in populations with more utilitarian attitudes. Cross-culture comparisons of consequentialist attitudes are scarce. To obtain at least an indicative idea of how utilitarian the Czech population is compared to other countries, we can look at the data from a world-wide moral machine project (Awad et al., 2018) that gathers preferences regarding the moral decisions made by self-driving cars. In particular, we focus on the extent to which people are willing to spare those who cross the road legally compared to those who cross the road on a red light. This decision is close to the deontological vs. utilitarian distinction, since it considers punishing people heavily for minor offenses. It shows that Czechs have views close to the average for the sample of 130 countries in terms of the probability of sparing rule-followers.[11] This suggests that the observed preference for strict liability could be generalized outside our sample. Still, more research into cross-country differences in procedural preferences would be of great interest.

Acknowledgments

We would like to thank the audiences at the 2019 NOeG Conference, the SABE-IAREP Conference in 2019, and the European Meeting of ESA in 2019. We are also grateful to two anonymous referees' comments and suggestions which substantially improved the paper. We gratefully acknowledge financial support from the Czech Science Foundation through Grant 17-00496S.

References

Alesina, A., Glaeser, E., Sacerdote, B., Durlauf, S.N., Levy, F., 2001. Why doesn't the united states have a european-style welfare state?/comments and discussion. Brookings Papers on Economic Activity , 187.

Angelova, V., Armantier, O., Attanasi, G., Hiriart, Y., 2014. Relative performance of liability rules: experimental evidence. Theory and decision 77, 531–556.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., Rahwan, I., 2018. The moral machine experiment. Nature 563, 59–64.

Bartling, B., Fehr, E., Herz, H., 2014. The intrinsic value of decision rights. Econometrica 82, 2005–2039.

Bixter, M.T., Luhmann, C.C., 2014. Shared losses reduce sensitivity to risk: A laboratory study of moral hazard. Journal of economic psychology 42, 63–73.

Bock, O., Nicklisch, A., Baetge, I., 2014. hroot:hamburg registration and organization online tool. European Economic Review 71, 117–120.

[11] The country-level average causal effect of lawful behavior on being spared is 0.352. Standard deviation is 0.044. The causal effect for the Czech Republic is 0.374, which is half a standard deviation from the mean.

Bohnet, I., Zeckhauser, R., 2004. Trust, risk and betrayal. Journal of Economic Behavior & Organization 55, 467–484.

Bolton, G.E., Brandts, J., Ockenfels, A., 2005. Fair procedures: Evidence from games involving lotteries. The Economic Journal 115, 1054–1076.

Bolton, G.E., Ockenfels, A., 2010. Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states: Comment. American Economic Review 100, 628–33.

Browne, M.J., Hoyt, R.E., 2000. The demand for flood insurance: empirical evidence. Journal of risk and uncertainty 20, 291–306.

Budner, S., 1962. Intolerance of ambiguity as a personality variable 1. Journal of personality 30, 29–50.

Chan, D.C., 2016. Teamwork and moral hazard: evidence from the emergency department. Journal of Political Economy 124, 734–770.

Chlaß, N., Güth, W., Miettinen, T., 2019. Purely procedural preferences-beyond procedural equity and reciprocity. European Journal of Political Economy 59, 108–128.

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

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

Doleac, J.L., Mukherjee, A., 2018. The moral hazard of lifesaving innovations: naloxone access, opioid abuse, and crime. Opioid Abuse, and Crime (September 30, 2018) .

Ehrlich, I., Becker, G.S., 1972. Market insurance, self-insurance, and self-protection. Journal of political Economy 80, 623–648.

Farhi, E., Tirole, J., 2012. Collective moral hazard, maturity mismatch, and systemic bailouts. American Economic Review 102, 60–93.

Fehr, E., Herz, H., Wilkening, T., 2013. The lure of authority: Motivation and incentive effects of power. American Economic Review 103, 1325–59.

Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Experimental economics 10, 171–178.

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

Goering, J.M., Feins, J.D., et al., 2003. Choosing a better life?: Evaluating the moving to opportunity social experiment. The Urban Insitute.

Hart, O., Moore, J., 1999. Foundations of incomplete contracts. The Review of Economic Studies 66, 115–138.

Holmstrom, B., 1982. Moral hazard in teams. The Bell Journal of Economics , 324–340.

Huck, S., Lünser, G., Spitzer, F., Tyran, J.R., 2016. Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. Journal of Economic Behavior & Organization 131, 78–105.

Karni, E., Salmon, T., Sopher, B., 2008. Individual sense of fairness: an experimental study. Experimental Economics 11, 174–189.

Kornhauser, L., Schotter, A., 1990. An experimental study of single-actor accidents. The Journal of Legal Studies 19, 203–233.

Lakdawalla, D., Sood, N., Goldman, D., 2006. Hiv breakthroughs and risky sexual behavior. The Quarterly Journal of Economics 121, 1063–1102.

Lefgren, L.J., Sims, D.P., Stoddard, O.B., 2016. Effort, luck, and voting for redistribution. Journal of Public Economics 143, 89–97.

Maskin, E., Moore, J., 1999. Implementation and renegotiation. Review of Economic studies , 39–56.

Mol, J.M., Botzen, W.W., Blasch, J.E., 2020. Risk reduction in compulsory disaster insurance: Experimental evidence on moral hazard and financial incentives. Journal of Behavioral and Experimental Economics 84, 101500.

Owens, D., Grossman, Z., Fackler, R., 2014. The control premium: A preference for payoff autonomy. American Economic Journal: Microeconomics 6, 138–61.

Rammstedt, B., John, O.P., 2007. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. Journal of research in Personality 41, 203–212.

Robinson, J.S., Joel, S., Plaks, J.E., 2015. Empathy for the group versus indifference toward the victim: Effects of anxious and avoidant attachment on moral judgment. Journal of Experimental Social Psychology 56, 139–152.

Sausgruber, R., Tyran, J.R., 2014. Discriminatory taxes are unpopular—even when they are efficient and distributionally fair. Journal of Economic Behavior & Organization 108, 463–476.

Sebald, A., 2010. Attribution and reciprocity. Games and Economic Behavior 68, 339–352.

Shavell, S., 2007. Liability for accidents. Handbook of law and economics 1, 139–182.

Supplementary material

First-best solution

The first-best solution is given as a solution of the following problem, where a benevolent dictator maximizes the welfare function by choosing the effort level e and variable $l \in \{L_S, L_A\}$ which determines who is responsible for the loss

$$\max_{e,l} -Pr(b + e < T)l - c(e).$$

In the first-best solution, the loss is always paid by the society, i.e. $l = L_S$ and the first-best effort level is given by the following first order condition

$$f(T - e^{FB})L_S = c'(e^{FB}). \tag{1}$$

Strict liability regime

In the strict liability regime, the agent always pays the costs whenever the sum of his effort and bad luck falls below the threshold T. Hence, the agent chooses an effort level that maximizes his own payoff

$$\max_{e} -Pr(b + e < T)L_A - c(e).$$

The solution of the problem is given by the following first order condition that implicitly defines the optimal effort under the strict liability regime e^{SL}.

$$f(T - e^{SL})L_A = c'(e^{SL}) \tag{2}$$

This paper only considers situations in which strict liability is a better outcome than a situation in which the agent does not exert any effort and the loss is always paid by the society. This is assured by the following assumption:

$$-Pr(b < T)L_S < -Pr(b + e^{SL} < T) - c(e^{SL}) \tag{3}$$

If this assumption does not hold, there would be no need to consider the strict liability regime at all. Instead, the paper would be reduced to a discussion of whether the identified optimal threshold level of $b + e$ for assistance should be applied.

Assistance regime without commitment

In the assistance regime without commitment, the official's reaction function is to let the society always pay the loss $l = L_S$ and the agent's optimal effort level is equal to zero, $e^A = 0$.

Assistance regime with commitment

In the assistance regime with commitment, the agent chooses the effort level that maximizes his own payoff

$$\max_{e} -Pr(b + e < D)L_A - c(e).$$

The solution of the problem is given by the following condition

$$f(D - e^*)L_A = c'(e^*). \tag{4}$$

This condition implicitly defines the agent's best-response function $e^*(D)$. By applying the implicit function theorem we can derive the slope of this best-response function

$$e^{*\prime}(D) = \frac{f'(D - e)L_A}{f'(D - e)L_A + c''}.$$

Since the cost function is non-concave, i.e. $c'' \geq 0$, the slope is positive but less or equal to one, $e^{*\prime}(D) \in (0, 1]$. The official chooses the threshold D in order to maximize the welfare function given the agent's best-response function[12].

$$\max_{D} -Pr(b + e^* < D)L_A - Pr(D < b + e^* < T)L_S - c(e^*) \tag{5}$$

The solution of this problem is given by the following first order condition

$$f(T - e)\, e^{*\prime}\, L_S - f(D - e)(1 - e^{*\prime})(L_A - L_S) = c'(e)\, e^{*\prime}. \tag{6}$$

Proof of effort ranking

It follows from the first-order conditions that $e^S = 0$ and the condition (3) ensures that $e^{AC} > 0$. This proves that $e^{AC} > e^A$. By comparing the first order conditions (1) and 2, we can see that the last inequality $e^{FB} < e^{SL}$ holds. The optimal effort level e^{AC} satisfies the condition (6) which can be rewritten as $-\frac{1-e'}{e'} f(D - e^{AC})(L_A - L_S) + f(T - e^{AC})L_s = c'$. Now, suppose by contradiction that $e^{AC} > e^{FB}$. The condition (1) together with the assumptions that marginal cost are non-decreasing $c' \geq 0$ and the probability

[12]The officials problem with N agents would be the same since all agents have the same best-response function

function is increasing $f' > 0$ imply that $f(T - e^{AC})H_s < c'$. For the condition (6) to be satisfied, it has to be the case that $\frac{1-e'}{e'} f(D - e^{AC})(L_A - L_S)$ is negative. This cannot be true since the slope of the best-response function e' is positive. Hence, we have a contradiction which proves the second inequality $e^{AC} \leq e^{FB}$.

Proof of welfare ranking

The first inequality $W^A < W^{SL}$ holds by assumption (3). The prove the second inequality $W^{SL} < W^{AC}$ consider the welfare in a assistance regime with commitment as a function of the assistance threshold $W(D) = -Pr(b + e^* < D)L_A - Pr(D < b + e^* < T)L_S - c(e^*)$ where e^* is given by the condition (6). The welfare W^{AC} is the maximum value of the welfare function $W(D)$ given by problem (5). The welfare in strict liability regime is a equal to this welfare evaluated at T, i.e. $W^{SL} = W(T)$. Therefore, we only need to show that the inequality is strict. When we calculate the first derivative of the welfare and plug-in for c' from condition (6) we get $f(T - e)\, e^{*\prime}\, L_S - f(D - e)(1 - e^{*\prime})(L_A - L_S) = f(D - e)L_A\, e^{*\prime}$. By evaluating the first derivative as point T we have $-f(T - e)\,(L_A - L_S) < 0$. Since the welfare function is decreasing at T, it holds that $W^{AC} > W^{SL}$. The third inequality $W^{AC} < W^{FB}$ also holds as strict because it cannot be simultaneously the case that $D = 0$ and $e^{AC} = e^{FB}$.