

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

4 Rostislav Staněk^{a,*}, Ondřej Krčál^a, Katarína Čellárová^a

5 ^a*Department of Economics, Masaryk University, Czech Republic*

6 **Abstract**

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

20 *Keywords:* Strict liability, Assistance, Procedural preferences, Experiment,
21 Moral hazard

22 *JEL:* C91, D90

*Corresponding author

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

23 1. Introduction

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

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

49 This paper proposes an experimental design in which helping people in
50 need (e.g. by speeding up procedures for passengers who arrive late to the
51 airport) may or may not lead to higher payoffs, depending on the severity of
52 the moral hazard problem. It introduces an effort-provision game in which
53 agents try to avoid a bad outcome by expending costly effort. What matters,
54 however, is the *observable effort*, which is equal to the sum of the actual
55 effort chosen by the agents and a non-positive random variable called *bad*
56 *luck*. If the observable effort falls below a certain threshold, the bad outcome
57 is not avoided. The game has two regimes: *strict liability* and *assistance*.
58 In strict liability, agents themselves bear the losses created by insufficient
59 effort or bad luck; in assistance, an official might transfer those losses to all

60 the other participants. The officials are given a (vaguely stated) standard,
61 which advises them to provide assistance only to participants whose actual
62 level of effort is above an optimally set threshold. However, it is difficult for
63 the officials to apply this standard because they do not see the participants'
64 actual effort, only their observable effort. In our guiding example, in the strict
65 liability regime all agents (passengers) arriving late would miss their flights.
66 In the assistance regime, the official (airline) would apply the standard to
67 speed up procedures for the passengers who had made a sufficient effort to
68 arrive on time, at a cost to the others.

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

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

¹This 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).

91 general behavioral mechanisms: People have a well-documented tendency to
92 avoid situations in which another person determines their outcome. In a trust
93 game, people require higher expected payoffs if there is any chance that they
94 could be betrayed by a human opponent, compared to the equivalent game
95 against a computer (Bohnet and Zeckhauser, 2004; Bolton and Ockenfels,
96 2010). In principal-agent experiments when people are asked to delegate de-
97 cisions directly to someone else, they sacrifice monetary gain in order to make
98 the decision themselves (Owens et al., 2014; Fehr et al., 2013; Bartling et al.,
99 2014). In line with these results, the preference for the strict liability regime
100 that our experiment reveals might be related to the presence of the human
101 official in the assistance regime. In order to investigate this explanation, we
102 conduct a NATURE treatment in which the human official is replaced by a
103 known probability distribution taken from the officials' choices made in the
104 sessions with human officials. We find that people still prefer strict liability
105 over assistance, which means the unpopularity of the assistance is not driven
106 by the presence of a human official but by the liability assignment aspect of
107 the procedure.

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

117 These findings about the preference for strict liability are in line with
118 recent literature on *procedural preferences* showing that individuals value in-
119 stitutions and procedures for their intrinsic value. Most of this literature
120 makes use of pie-splitting games to show that people prefer procedures that
121 provide fair randomization over unequal outcomes (Bolton et al., 2005; Karni
122 et al., 2008), or guarantee a kind distribution of outcomes (Sebald, 2010).
123 Our finding, however, seems to be driven by what are known as *purely proce-*
124 *dural preferences*, i.e. preferences for procedural properties that are unrelated
125 to payoffs or outcomes (Chlaß et al., 2019). Chlaß et al. (2019) document
126 that people consider procedural simplicity, efficiency, the distribution of the
127 decision and information rights in their choices. In line with the notion of
128 purely procedural preferences, Sausgruber and Tyran (2014) showed in a

129 laboratory experiment that people prefer uniform taxes over discriminatory
130 taxes that are equally efficient and produce the same expected outcomes.
131 We contribute to this literature by showing that the contradiction between
132 efficiency and preferences also applies in the case of (public) assistance in
133 situations fraught with moral hazard. Moreover, we find that the preference
134 for the strict liability regime is correlated with deontological attitudes². We
135 interpret that correlation as suggestive evidence that the voting decision in
136 our experiment reflects the discrepancy between efficiency and a norm that
137 people should be fully liable for their own losses.

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

²People 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.

³Note 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.

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

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

182 **2. Theoretical framework**

183 The structure of our theoretical framework follows our two research ques-
184 tions. First, we compare the efficiency of the assistance and strict liability
185 regimes in the effort-provision game. Second, we discuss procedural prefer-
186 ences for strict liability.

187 *2.1. Efficiency of regimes*

188 In the effort-provision game, the agent chooses the effort level e . The
189 effort is costly and the monetary costs are given by the function $c(e)$ which
190 is increasing and non-concave, i.e. $c'(e) > 0$ and $c''(e) \geq 0$. Bad luck b is
191 a random variable with support $[-\underline{b}, 0]$ and probability distribution function
192 $f(b)$ which is increasing, $f'(b) > 0$. The effort level and bad luck determine
193 whether the outcome is bad or good. A bad outcome occurs if the sum of
194 the effort and bad luck falls below the threshold T , i.e. $b + e < T$. In our

195 guiding example, e is the effort expended to arrive at the airport in time,
 196 such as getting up early, using more secure means of transport, etc. Bad
 197 luck can take many forms, such as the passenger facing unexpected traffic
 198 problems. The natural situation in which the passenger misses their flight
 199 unless assistance is provided constitutes the threshold.

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

210 The regimes differ in the way the liability for the loss is assigned. Under
 211 the *strict liability* regime, the responsible agent always pays the loss. The
 212 *assistance* regime is complemented by a standard which states that the loss
 213 will be paid by the agent if he did not exert sufficient effort to prevent the
 214 loss. A benevolent official decides whether this requirement was met. The
 215 official can observe $e + b$, which we call observable effort. In our guiding
 216 example, the observable effort could be the arrival time of the passenger
 217 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

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

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

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

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

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

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

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

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

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

247 2.2. Preference for regimes

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

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

- 260 • *Inequality in information:* In both regimes, information is unequally
261 distributed amongst players. In the assistance regime, agents are at a
262 disadvantage compared with officials. The officials learn about all the
263 agents' observable effort and have private information about whether or
264 not they will apply the vague standard at the time of their decision. In
265 the strict liability regime, agents have an information advantage over
266 officials since the latter have no information at all. Hence if agents
267 prefer to have the information advantage they will prefer strict liability.
- 268 • *Simplicity:* The strict liability procedure is simpler because agents do
269 not need to take the official's choice into account and hence do not need
270 to reason about so many strategies.

271 In addition to the purely procedural concerns covered by Chlaß et al.
272 (2019), agents might dislike their payoffs being reduced because of other
273 agents' possible negligence.

274 3. Experimental design and predictions

275 Our experiment consists of the effort provision game and a voting proce-
276 dure. The effort provision game has two regimes: strict liability and assis-
277 tance. These two regimes are played in the first two stages of the experiment.
278 Voting in the third stage determines which regime will be played in the final
279 stage.

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

285 The strict liability regime has 15 periods. The officials are inactive: they
286 do not make any decisions and they do not receive any feedback about the
287 other agents' behavior. At the beginning of each period, agents are endowed
288 with 140 CZK⁵ and 6 tokens. The subjects know that they will lose between
289 zero and six tokens according to a predetermined probability distribution.

⁵At 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.

290 The probability distribution is presented in Table 1. Agents are given the
 291 opportunity to buy zero to six additional tokens. Each token costs 10 CZK.
 292 After each agent has made their purchase decision, a random draw determines
 293 how many tokens are lost. Any agent left with fewer than six tokens must
 294 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

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

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

⁶This 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.

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

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

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

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

356 After the experiment, the subjects fill in a questionnaire. Most impor-
 357 tantly, the questionnaire includes a shorter version of the consequentialist
 358 scale by Robinson et al. (2015), which we use to check for any ethical com-
 359 ponent in the preference for strict liability. This short version consists of
 360 four questions that assess endorsement of utilitarian or deontological beliefs.⁷
 361 Participants indicate how much they agree with each statement on a 5-point
 362 Likert scale. The total score ranges between 4 and 20, with higher scores
 363 showing a tendency towards a more utilitarian attitude and lower scores
 364 pointing towards a more deontological attitude. Additionally, the question-
 365 naire contains the standard socio-demographic variables, measures of person-
 366 ality traits (the Big Five personality traits by Rammstedt and John (2007),
 367 each trait with 2 questions on a 5-point Likert scale and with the total score
 368 ranging between 2 and 10), self-reported risk attitude on a 10-point Likert
 369 scale (Dohmen et al., 2011) and tolerance to ambiguity (Budner, 1962) with
 370 a total score ranging between 4 and 20 (4 questions, 5-point Likert scale).

371 Table 2 presents the equilibrium predictions, based on the parameters
 372 and functional forms used in the experimental design, for all the theoretical
 373 regimes discussed in subsection 2.1: the strict liability regime (SL), assistance
 374 without commitment (A), assistance with commitment (AC), and first-best
 375 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

376 Our experiment contains only two of these regimes: assistance and strict
 377 liability. While the strict liability regime corresponds closely to the theoret-
 378 ical SL regime, the choices the officials make and the agents’ reactions to
 379 those choices determine whether the outcomes of the assistance regime more

⁷These 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.”

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

386
387 **Hypothesis 1:** Effort, measured by purchased tokens, is lower in assistance
388 than in strict liability.

389
390 The model predicts that if agents reduce their level of effort and the offi-
391 cials follow the standard, AC will lead to higher monetary payoffs than SL.
392 This generates our second hypothesis.

393
394 **Hypothesis 2:** The average monetary payoff is higher in assistance than in
395 strict liability.

396
397 Preferences between the experimental regimes of strict liability and as-
398 sistance are elicited in the voting stage, in which agents (but not officials)
399 vote on which of the regimes should be played in the final stage. Based on
400 the discussion of procedural preferences in subsection 2.2, we formulate the
401 following conjecture about the outcome of the voting decision:

402
403 **Conjecture 1:** Contrary to their monetary incentives, agents prefer strict
404 liability over assistance.

405
406 There are other possible explanations for such a preference for strict liabil-
407 ity, besides procedural preferences. A preference for strict liability might be
408 driven by the presence of a human official. Ambiguity-averse subjects might
409 want to avoid the ambiguity related to the official's discretionary power to
410 determine the outcome or they might feel betrayed if they comply with the
411 standard and the official makes them liable for the loss. To rule these con-
412 cerns out, we compare the results of the HUMAN treatment, in which the
413 official is a human subject, with those of the NATURE treatment, which is
414 identical to HUMAN in all aspects except that the official is played by a com-
415 puter program. The computer decides according to a function that defines
416 the probability that the loss is paid by other members of the group condi-
417 tional on the observable effort. The value of probabilities was established as

418 the fraction of choices in which the group members had to pay for the loss
419 in the HUMAN treatment sessions. The decisions made by nature therefore
420 mimicked the decisions made by the human officials. The probabilities are
421 presented in Table 5, and the subjects were informed about the values of
422 these probabilities in the instructions.

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

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

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

⁸As 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.

450 **4. Results**

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

456 *4.1. Data*

457 The experiment was conducted in October 2018 at the Masaryk Univer-
458 sity Experimental Economics Laboratory (MUEEL) in Brno, Czech Repub-
459 lic. In total, we recruited 328 student subjects using hroot (Bock et al., 2014).
460 The experimental environment was programmed in zTree (Fischbacher, 2007).
461 The experiment took about two hours and participants received 254 CZK (10
462 EUR) on average. There were 16 experimental sessions in total: 6 sessions of
463 the HUMAN treatment with 120 subjects, 6 sessions of the NATURE treat-
464 ment (116 subjects) and 4 sessions of the CONTROL treatment (92 subjects).
465 Since the HUMAN subjects were divided into groups of 5 (4 agents and 1
466 official) and the subjects in the NATURE and CONTROL sessions were di-
467 vided into groups of 4, this resulted in 24 groups (independent observations)
468 in HUMAN, 29 in NATURE, and 23 in CONTROL.⁹

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

⁹To 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.

477 are no significant differences between the outcomes in the HUMAN and NA-
478 TURE treatments. In fact, all three treatments generate the same payoff
479 distributions. We conclude that our computer algorithm in the NATURE
480 treatment successfully simulated the choices made by human officials, and
481 the setup of the CONTROL treatment successfully mimicked all the choices
482 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 neuroticisms (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

483 4.2. *Effort and efficiency*

484 This subsection provides the tests of Hypotheses 1 and 2. Hypothesis 1
485 states that fewer tokens will be purchased in the assistance regime than
486 in the strict liability regime. Table 4 presents the differences in group aver-
487 ages between the two regimes (assistance – strict liability) and bootstrapped

488 confidence intervals. In the assistance regime, agents in both the HUMAN
489 and NATURE treatments purchase significantly fewer tokens (these decisions
490 are not made in CONTROL). According to Hypothesis 2, the agents' aver-
491 age monetary payoff is higher in the assistance regime. As can be seen in
492 Table 4, the shift from the strict liability regime to the assistance regime
493 increases the monetary payoff in both HUMAN and NATURE treatments.
494 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	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 \times 30 periods) were resampled. * denotes significance at the 5% level.

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

¹⁰We 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.

507 Our hypotheses are also supported by the regression models in Table 6, in
508 which we control for individual (agent) fixed effects. The estimates present
509 the within-individual effect of the assistance regime. They also show how
510 the effect of assistance interacts with order and treatment. The estimates
511 are based on data from all 30 periods. The standard errors are clustered
512 at the group level (a group consists of 120 observations: 4 agents times 30
513 periods). The Poisson regressions in models 1 and 2 explain the number of
514 purchased tokens, which ranges from zero to six. The OLS regressions in
515 models 3 and 4 explain the agents' monetary payoffs. In line with our hy-
516 potheses, the table shows that the assistance regime leads to lower effort and
517 higher payoffs. The interaction *Assistance* \times *Order 1* is negative in model 2,
518 which strengthens the effect of assistance on the number of purchased tokens.
519 Conversely, the effect on monetary payoff is weakened by the interaction with
520 order (opposite signs), but still the effect is statistically significant in Order
521 1 overall. When we split the results further by treatment, the significance is
522 at the 5% level in HUMAN ($p = 0.028$) and at the 10% level in NATURE
523 ($p = 0.062$). The *lost tokens* variable measures the number of tokens lost
524 due to bad luck. It is a random number ranging from zero to six drawn from
525 the distribution presented in Table 1. The number of lost tokens strongly
526 predicts the monetary payoff, as losing more tokens increases the probability
527 of falling below the threshold and experiencing a loss.

528 4.3. Voting

529 Next, we test Conjecture 1, stating that participants vote for the less effi-
530 cient 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

	<i>Dependent variable:</i>			
	Purchased tokens		Monetary payoff	
	<i>Poisson marginal effects</i>		<i>OLS</i>	
	(1)	(2)	(3)	(4)
Assistance	-0.894*** (0.043)	-0.490*** (0.076)	3.008*** (0.68)	3.490** (1.31)
Assistance × Order 1		-0.439*** (0.080)		-0.82 (1.39)
Assistance × NATURE		-0.322*** (0.082)		-0.146 (1.40)
Lost tokens			-10.65*** (0.38)	-10.65*** (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). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

531 60 % of participants obtain higher payoffs in the assistance regime but only
532 38 % of them vote for the assistance regime. In the NATURE treatment, 60
533 % of participants obtain higher payoffs in the assistance regime but only 45
534 % 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

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

535 where m is the agent's monetary payoff, S is a dummy variable which takes
536 the value one in the assistance regime and ϵ_i is the unobserved portion of
537 utility. Based on the voting decision and actual payoffs in the first two
538 stages of the experiment, we can use discrete choice techniques to identify
539 the parameters α_0 and α_1 . The parameter α_0 is interpreted as an alternative-
540 specific constant indicating the utility of the assistance regime not related
541 to monetary payoff. Negative values of this parameter suggest preference
542 in favor of strict liability. The dependent variable *Voting* takes a value of
543 one if the agent voted for the assistance regime and zero if he/she voted for

544 the strict liability regime. The variable *Payoff difference* is the difference
545 in the participants' average payoff between the assistance regime and the
546 strict liability regime. We estimate the intrinsic utility of the assistance
547 regime (alternative-specific variable α_0) for all three treatments separately.
548 We label these variables *HUMAN*, *NATURE* and *CONTROL* in Table 7.
549 We test whether the alternative specific constants are different from zero. A
550 significant and negative result would mean that people prefer strict liability
551 over assistance even after controlling for the payoff differences. We also
552 control for the order effect and the interactions of order with payoff difference
553 and treatments.

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

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

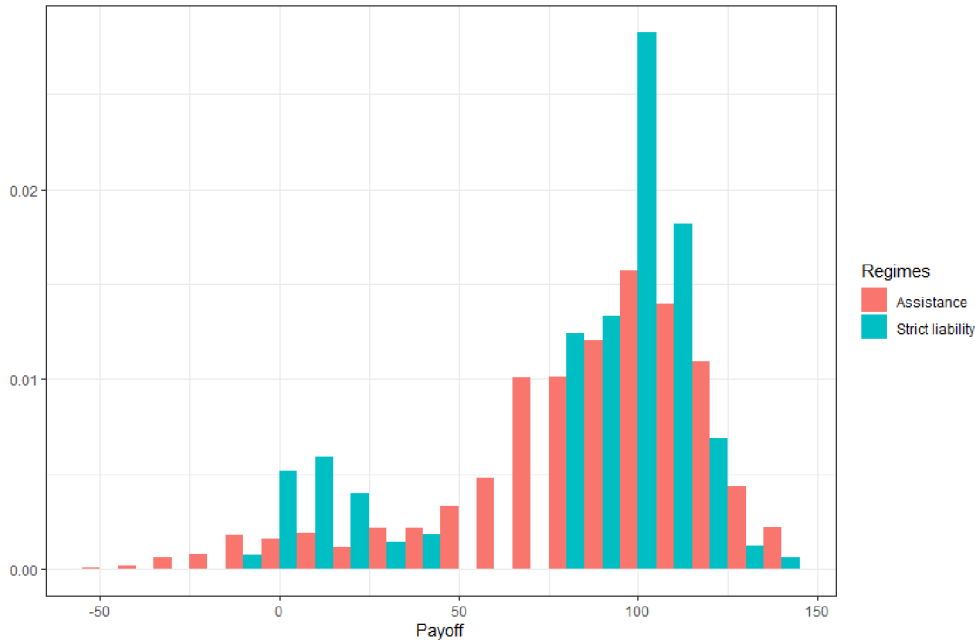
Note: The dependent variable *Voting* takes a value of one if the agent voted for the assistance regime and zero otherwise. Standard errors clustered at the group level (group includes 4 agents). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

554 Two main results stand out from model 1 in Table 7. First, there is a sig-
555 nificant and substantial preference for the strict liability regime. The average
556 willingness to pay to avoid the assistance regime the in HUMAN treatment
557 is around 12 CZK (calculated as a ratio of the parameters *HUMAN/Payoff*
558 *difference* from column 1 of Table 7), which is approximately four times the
559 average payment difference between the regimes. Second, the relative unfa-
560 vorability of the assistance regime is not driven solely by the presence of a
561 human official. The preference for the strict liability regime is present even in
562 the NATURE treatment. Although the presence of a human official makes
563 the strict liability preference stronger, there is no statistical difference be-
564 tween the HUMAN and NATURE treatments. The average marginal effect
565 of the human official is 0.061 ($p = 0.332$).

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

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

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



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

602 The model in column 3 uses data from the HUMAN and NATURE treat-
 603 ments, where we identify the preference for strict liability. It explores the
 604 correlation between the preference for strict liability and the consequential-
 605 ist scale standardized to have mean 0 and standard deviation 1. We can see
 606 that the consequentialist scale is related to the preference for strict liability.
 607 Participants who agreed more with the position that some rules should be
 608 honoured in all circumstances were more likely to vote for the strict liability

609 regime. Participants who expressed extreme utilitarian views (i.e. two stan-
610 dard deviations from the mean) did not manifest this preference. Using the
611 specification of model 3, we estimated an additional seven models with other
612 survey measures (risk, ambiguity scale, and Big Five scales) instead of the
613 *consequentialist* variable, but we found that none of these were significantly
614 related to voting.

615 5. Discussion

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

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

641 More interestingly, our experiment provides evidence, based on revealed
642 preferences, that people prefer the strict liability regime over the assistance
643 regime. Our results further demonstrate that this preference for strict li-
644 ability cannot be fully explained by the presence of a human official and

645 that the preference for strict liability vanishes when the agent cannot influ-
646 ence the level of effort and therefore the probability of loss occurring. This
647 suggests that the preference for strict liability is procedural and seems to
648 be related to people’s aversion to having their payoff influenced by other,
649 possibly negligent, agents and a (random) mechanism allocating the loss to
650 everyone else but the negligent agent. This conclusion is supported by the
651 fact that this preference is positively correlated with deontological attitudes.
652 People with utilitarian attitudes do not exhibit any preference for the strict
653 liability regime.

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

663 One obvious limitation of this research is the composition of our sample,
664 which consisted exclusively of Czech students; this raises the question of
665 how generalizable our results are. Since we find that preferences for strict
666 liability are correlated with consequentialist attitudes, we believe that the
667 results can be generalized at least for any population that has a similar
668 composition of utilitarian vs. deontological attitudes. On the other hand,
669 the aversion to the assistance regime that we identified might disappear in
670 populations with more utilitarian attitudes. Cross-culture comparisons of
671 consequentialist attitudes are scarce. To obtain at least an indicative idea
672 of how utilitarian the Czech population is compared to other countries, we
673 can look at the data from a world-wide moral machine project (Awad et al.,
674 2018) that gathers preferences regarding the moral decisions made by self-
675 driving cars. In particular, we focus on the extent to which people are willing
676 to spare those who cross the road legally compared to those who cross the
677 road on a red light. This decision is close to the deontological vs. utilitarian
678 distinction, since it considers punishing people heavily for minor offenses.
679 It shows that Czechs have views close to the average for the sample of 130

680 countries in terms of the probability of sparing rule-followers.¹¹ This suggests
681 that the observed preference for strict liability could be generalized outside
682 our sample. Still, more research into cross-country differences in procedural
683 preferences would be of great interest.

684 **Acknowledgments**

685 We would like to thank the audiences at the 2019 NOeG Conference,
686 the SABE-IAREP Conference in 2019, and the European Meeting of ESA
687 in 2019. We are also grateful to two anonymous referees' comments and
688 suggestions which substantially improved the paper. We gratefully acknowl-
689 edge financial support from the Czech Science Foundation through Grant
690 17-00496S.

691 **References**

- 692 Alesina, A., Glaeser, E., Sacerdote, B., Durlauf, S.N., Levy, F., 2001. Why
693 doesn't the united states have a european-style welfare state?/comments
694 and discussion. *Brookings Papers on Economic Activity* , 187.
- 695 Angelova, V., Armantier, O., Attanasi, G., Hiriart, Y., 2014. Relative per-
696 formance of liability rules: experimental evidence. *Theory and decision* 77,
697 531–556.
- 698 Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon,
699 J.F., Rahwan, I., 2018. The moral machine experiment. *Nature* 563, 59–64.
- 700 Bartling, B., Fehr, E., Herz, H., 2014. The intrinsic value of decision rights.
701 *Econometrica* 82, 2005–2039.
- 702 Bixter, M.T., Luhmann, C.C., 2014. Shared losses reduce sensitivity to risk:
703 A laboratory study of moral hazard. *Journal of economic psychology* 42,
704 63–73.
- 705 Bock, O., Nicklisch, A., Baetge, I., 2014. hroot:hamburg registration and
706 organization online tool. *European Economic Review* 71, 117–120.

¹¹The 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.

- 707 Bohnet, I., Zeckhauser, R., 2004. Trust, risk and betrayal. *Journal of Eco-*
708 *nomic Behavior & Organization* 55, 467–484.
- 709 Bolton, G.E., Brandts, J., Ockenfels, A., 2005. Fair procedures: Evidence
710 from games involving lotteries. *The Economic Journal* 115, 1054–1076.
- 711 Bolton, G.E., Ockenfels, A., 2010. Betrayal aversion: Evidence from brazil,
712 china, oman, switzerland, turkey, and the united states: Comment. *Amer-*
713 *ican Economic Review* 100, 628–33.
- 714 Browne, M.J., Hoyt, R.E., 2000. The demand for flood insurance: empirical
715 evidence. *Journal of risk and uncertainty* 20, 291–306.
- 716 Budner, S., 1962. Intolerance of ambiguity as a personality variable 1. *Journal*
717 *of personality* 30, 29–50.
- 718 Chan, D.C., 2016. Teamwork and moral hazard: evidence from the emer-
719 gency department. *Journal of Political Economy* 124, 734–770.
- 720 Chlaß, N., Güth, W., Miettinen, T., 2019. Purely procedural preferences-
721 beyond procedural equity and reciprocity. *European Journal of Political*
722 *Economy* 59, 108–128.
- 723 Deffains, B., Espinosa, R., Fluet, C., 2019. Laws and norms: Experimental
724 evidence with liability rules. *International Review of Law and Economics*
725 60, 105858.
- 726 Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G.,
727 2011. Individual risk attitudes: Measurement, determinants, and behav-
728 ioral consequences. *Journal of the European Economic Association* 9, 522–
729 550.
- 730 Doleac, J.L., Mukherjee, A., 2018. The moral hazard of lifesaving innova-
731 tions: naloxone access, opioid abuse, and crime. *Opioid Abuse, and Crime*
732 (September 30, 2018) .
- 733 Ehrlich, I., Becker, G.S., 1972. Market insurance, self-insurance, and self-
734 protection. *Journal of political Economy* 80, 623–648.
- 735 Farhi, E., Tirole, J., 2012. Collective moral hazard, maturity mismatch, and
736 systemic bailouts. *American Economic Review* 102, 60–93.

- 737 Fehr, E., Herz, H., Wilkening, T., 2013. The lure of authority: Motivation
738 and incentive effects of power. *American Economic Review* 103, 1325–59.
- 739 Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic ex-
740 periments. *Experimental economics* 10, 171–178.
- 741 Füllbrunn, S., Neugebauer, T., 2013. Limited liability, moral hazard, and risk
742 taking: A safety net game experiment. *Economic Inquiry* 51, 1389–1403.
- 743 Goering, J.M., Feins, J.D., et al., 2003. Choosing a better life?: Evaluating
744 the moving to opportunity social experiment. The Urban Insitute.
- 745 Hart, O., Moore, J., 1999. Foundations of incomplete contracts. *The Review*
746 *of Economic Studies* 66, 115–138.
- 747 Holmstrom, B., 1982. Moral hazard in teams. *The Bell Journal of Economics*
748 , 324–340.
- 749 Huck, S., Lünser, G., Spitzer, F., Tyran, J.R., 2016. Medical insurance
750 and free choice of physician shape patient overtreatment: A laboratory
751 experiment. *Journal of Economic Behavior & Organization* 131, 78–105.
- 752 Karni, E., Salmon, T., Sopher, B., 2008. Individual sense of fairness: an
753 experimental study. *Experimental Economics* 11, 174–189.
- 754 Kornhauser, L., Schotter, A., 1990. An experimental study of single-actor
755 accidents. *The Journal of Legal Studies* 19, 203–233.
- 756 Lakdawalla, D., Sood, N., Goldman, D., 2006. Hiv breakthroughs and risky
757 sexual behavior. *The Quarterly Journal of Economics* 121, 1063–1102.
- 758 Lefgren, L.J., Sims, D.P., Stoddard, O.B., 2016. Effort, luck, and voting for
759 redistribution. *Journal of Public Economics* 143, 89–97.
- 760 Maskin, E., Moore, J., 1999. Implementation and renegotiation. *Review of*
761 *Economic studies* , 39–56.
- 762 Mol, J.M., Botzen, W.W., Blasch, J.E., 2020. Risk reduction in compulsory
763 disaster insurance: Experimental evidence on moral hazard and financial
764 incentives. *Journal of Behavioral and Experimental Economics* 84, 101500.

- 765 Owens, D., Grossman, Z., Fackler, R., 2014. The control premium: A prefer-
766 ence for payoff autonomy. *American Economic Journal: Microeconomics*
767 6, 138–61.
- 768 Rammstedt, B., John, O.P., 2007. Measuring personality in one minute
769 or less: A 10-item short version of the big five inventory in english and
770 german. *Journal of research in Personality* 41, 203–212.
- 771 Robinson, J.S., Joel, S., Plaks, J.E., 2015. Empathy for the group versus
772 indifference toward the victim: Effects of anxious and avoidant attachment
773 on moral judgment. *Journal of Experimental Social Psychology* 56, 139–
774 152.
- 775 Sausgruber, R., Tyran, J.R., 2014. Discriminatory taxes are unpopular—even
776 when they are efficient and distributionally fair. *Journal of Economic*
777 *Behavior & Organization* 108, 463–476.
- 778 Sebald, A., 2010. Attribution and reciprocity. *Games and Economic Behavior*
779 68, 339–352.
- 780 Shavell, S., 2007. Liability for accidents. *Handbook of law and economics* 1,
781 139–182.

782 **Supplementary material**

783 *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

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

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

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

786 *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

$$\max_e -Pr(b + e < T)L_A - c(e).$$

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

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

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

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

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

796 *Assistance regime without commitment*

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

800 *Assistance regime with commitment*

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

$$\max_e -Pr(b + e < D)L_A - c(e).$$

801 The solution of the problem is given by the following condition

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

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

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

802 Since the cost function is non-concave, i.e. $c'' \geq 0$, the slope is positive but
 803 less or equal to one, $e^{*'}(D) \in (0, 1]$. The official chooses the threshold D
 804 in order to maximize the welfare function given the agent's best-response
 805 function¹².

$$\max_D -Pr(b + e^* < D)L_A - Pr(D < b + e^* < T)L_S - c(e^*) \quad (5)$$

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

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

807 *Proof of effort ranking*

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

¹²The official's problem with N agents would be the same since all agents have the same best-response function

815 function is increasing $f' > 0$ imply that $f(T - e^{AC})H_s < c'$. For the condition
 816 (6) to be satisfied, it has to be the case that $\frac{1-e'}{e'}f(D - e^{AC})(L_A - L_S)$ is neg-
 817 ative. This cannot be true since the slope of the best-response function e' is
 818 positive. Hence, we have a contradiction which proves the second inequality
 819 $e^{AC} \leq e^{FB}$.

820 *Proof of welfare ranking*

821 The first inequality $W^A < W^{SL}$ holds by assumption (3). To prove the
 822 second inequality $W^{SL} < W^{AC}$ consider the welfare in an assistance regime
 823 with commitment as a function of the assistance threshold $W(D) = -Pr(b +$
 824 $e^* < D)L_A - Pr(D < b + e^* < T)L_S - c(e^*)$ where e^* is given by the condition
 825 (6). The welfare W^{AC} is the maximum value of the welfare function $W(D)$
 826 given by problem (5). The welfare in strict liability regime is equal to this
 827 welfare evaluated at T , i.e. $W^{SL} = W(T)$. Therefore, we only need to show
 828 that the inequality is strict. When we calculate the first derivative of the
 829 welfare and plug-in for c' from condition (6) we get $f(T - e)e^{*'}L_S - f(D -$
 830 $e)(1 - e^{*'})(L_A - L_S) = f(D - e)L_A e^{*'}$. By evaluating the first derivative at
 831 point T we have $-f(T - e)(L_A - L_S) < 0$. Since the welfare function is
 832 decreasing at T , it holds that $W^{AC} > W^{SL}$. The third inequality $W^{AC} <$
 833 W^{FB} also holds as strict because it cannot be simultaneously the case that
 834 $D = 0$ and $e^{AC} = e^{FB}$.