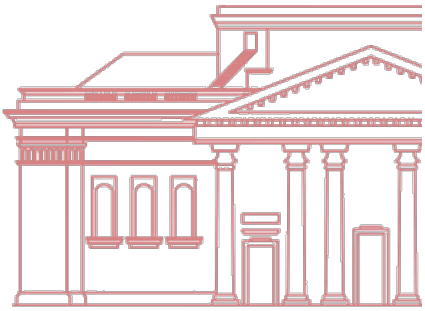
CENTRE DE RECHERCHES EN ÉCONOMIE ET DROIT
Paris Center for Law and Economics



CRED WORKING PAPER N^o 2020-6

Laws and Norms: Experimental Evidence with Liability Rules

February 2020

Bruno Deffains* Romain Espinosa† Claude Fluet‡

*Department of Economics and CRED, University Paris II Panthéon Assas, email: bruno.deffains@paris2.fr
†CNRS, CREM - Université Rennes 1. Email: romain.espinosa@univ-rennes1.fr
‡Université Laval, CRÉFiR, CRED. Email: claude.fluet@fsa.ulaval.ca

1 Introduction

Actions causing harm to third parties are often governed by liability rules which allow the harmed parties to claim compensation from the tortfeasor. A main purpose of such rules is to deter socially inefficient behavior generating negative externalities. Under the so-called strict liability rule, tortfeasors are required to compensate the harm they caused irrespective of precautions or circumstances. Under the negligence rule, they are required to compensate if their behavior fell short of some legal standard of conduct given the circumstances. In either case, and provided the negligence rule is appropriately designed, the standard prediction is that perfectly enforced liability rules will yield socially efficient decisions in the sense of maximizing overall welfare. Conversely, when legal liability is nonexistent or is imperfectly enforced, for instance when tortfeasors are seldom detected, the standard prediction is that socially inefficient behavior will be under-deterred.

There are reasons to question these predictions. A voluminous literature in behavioral economics shows that individuals are not purely motivated by their material self-interest, e.g., they may also be concerned by others' wellbeing or may be motivated by reciprocity or self-image considerations. On the one hand, the existence of other-oriented motives is likely to decrease opportunistic and socially inefficient behavior, and might therefore usefully complement poorly enforced liability rules. On the other hand, such motives may also generate inefficiencies by driving out self-interested but socially efficient behavior, i.e., they could lead to *over-deterrence* of actions causing harm to others. In addition, the literature has showed that material sanctions may crowd in or crowd out informal motivations. Should crowding out occur, the efficiency gains of formal liability rules would be smaller when enforcement costs are taken into account.

We investigate how individuals react to liability rules under different settings. Our objective is threefold. First, we aim at evaluating the effect of introducing liability rules: Do theoretically equivalent canonical liability rules achieve similar levels of efficiency? Are they able to perform better than the benchmark scenario without legal rules, even when the rules are poorly enforced and are in principle nondeterrent? Second, we seek to understand to which extent liability rules are able to influence individual behavior beyond the changes in material incentives they generate: Do deviations in behavior from the no-law scenario result solely from payoff relevant deterrence effects? Do liability rules crowd in or crowd out concerns for the group? Third, we investigate how behavior regulated by liability rules is perceived by individuals. Do individuals express disapproval of individuals who cause harm to others? Does it matter if causing harm was socially efficient? Or do individuals mostly express disapproval of those who violated standards established by the liability rules? Does the type of rule then matter?

To answer these questions, we develop a *Liability Game* that shares with Public Good experiments the property that each participant's final payoff depends on what other participants do. Individuals repeatedly interact in a Stranger Matching protocol and are both potential tortfeasors and victims. This replicates many real-life situations where individuals can both cause harm and suffer harm caused by others, e.g., one may be both a driver and a pedestrian, each tenant in an apartment building may impose nuisances on others, each worker in a plant may put others at risk. As in many everyday situation, negative externalities can result from either socially inefficient or socially efficient actions, depending on the private circumstances of each individual. Ex ante, our participants face the same probability distribution of circumstances. Ex post, however, when actions must be chosen, participants will be in different circumstances and will not know the circumstances of others. Typically, in our experiment, socially efficient behavior leads to unequal payoffs. Participants also have the possibility to express disapproval of other group members at the end of each period after the realization of payoffs. Sanctions can be either symbolic (non-monetary punishment) or involve small costs to both parties (monetary punishment).

We introduce treatment variations with respect to the legal rules and the enforcement policy. We compare the cases where behavior is regulated by standard liability rules (i.e., strict liability or the negligence rule) to a benchmark without legal rules (*No Law*). In addition, we vary the degree of enforcement of the liability rules and consider the cases where they are either perfectly enforced (*Severe Law*) or enforced only with probability one half (*Mild Law*).

Our results first show that both the strict and negligence liability regimes achieve a substantially higher level of efficiency than the no-law benchmark scenario. This holds both when liability rules are perfectly or imperfectly enforced. When rules are perfectly enforced, we confirm the theoretical prediction that both the negligence rule and strict liability achieve the same level of efficiency. However, when the legal rules are imperfectly enforced, one of the two liability rules appears to do better in deterring socially inefficient harm-causing actions. Strict liability outperforms the negligence rule in this respect. On the other hand, strict liability also appears to over-deter socially efficient harm-causing actions, a feature that does not arise under the negligence rule. Second, we show that liability rules affect individual behavior beyond the associated material deterrence effects. In our setting, social preferences in the form of inequality aversion cannot explain socially efficient behavior in the No Law or Mild Law treatments, but motivations such as altruism and efficiency concerns can play a role. When we decompose individual utilities into self-interested and efficiency-oriented components, we observe larger weights on efficiency concerns under either liability rules than in the No Law benchmark. This suggests that liability rules crowd in concerns for others' well-being. Third, we find that participants mainly disapprove of or sanction players who caused harm, irrespective of the prevailing legal rule or of whether the harmful action is socially efficient.

In the standard approach of Law and Economics, legal liability for harm amounts to the price put on harmful behavior. However, there are clearly other dimensions, in addition to the threat of legal sanctions. Legal liability increases the salience of harmful actions by categorizing them as actions that trigger a right to compensation, perhaps only in some circumstances. When harmful actions would otherwise arise anonymously, legal liability also make these actions attributable, which allow informal sanctions to come at play. Although liability rules do not prohibit any particular actions, they may also convey the message that some actions are not right, at least in some circumstances, which may influence behavior by itself, a moral suasion or normative effect. Legal liability may also affect expectations of how others will behave, which in turn influences behavior when preferences are other-regarding. Our experiment seeks to characterize the overall effects of legal liability.

The paper proceeds as follows. Section 2 briefly reviews the related experimental literature. Section 3 presents the experimental design. Section 4 draws on the theoretical literature on social preferences to derive predictions. Section 5 presents the results. Section 6 discusses the interpretation and Section 7 concludes.

2 Related Literature

There are relatively few experimental studies of liability rules in the law and economics literature as such; see Sullivan and Holt (2017). An early paper is Kornhauser and Schotter (1990). They compare strict liability and the negligence rule in the so-called single-actor (or unilateral) accident framework where agents can invest in precautions to reduce the probability of causing harm to others. More recently Angelova et al. (2014) considered No Law, strict liability and the negligence rule in the same set-up with only two precaution levels ("care" versus "no-care"). They find that either liability rule provides socially efficient incentives but that roughly half of the subjects also invest in care under No Law. In both these papers, consistent with the traditional approach in law and economics (Becker, 1968), legal obligations are backed by deterrent incentives. In our

experiment, by contrast, the focus is on nondeterrent law. In addition, the occurrence of harm is non-stochastic, i.e., it follows deterministically from one's actions. Socially efficient "care" then depends on the choice between actions given one's private circumstances.[1] Non-stochastic harm minimizes the potentially confounding effects of risk aversion and of confusion due to the subjects' computations under uncertainty.

Closely related to this paper is the experimental literature on the effect of "legal obligations" in linear Public Good games (henceforth PG games). Such games, also known as Voluntary Contribution Mechanisms, have been widely studied because they epitomize the conflict between self-interest and group interest in social dilemma situations; see Chaudhuri (2011) and Villeval (2012) for recent surveys. A smaller strand of this literature has studied the role of weakly incentivized obligations to contribute to the public good (e.g., Tyran and Feld, 2006; Galbiati and Vertova, 2008, 2014; Kube and Traxler, 2011; Riedel and Schildberg-Hörisch, 2013). The aim is to test the notion of "expressive law", i.e., law not backed by binding incentives as defined in Cooter (1998) or McAdams and Nadler (2005) among others.[2]

A rationale often invoked to explain the role of expressive law is that obligations impact behavior because they affect the individuals' beliefs about the behavior of others. If individuals have other-regarding preferences (e.g., inequality aversion or a predisposition for conditional cooperation), social dilemma situations in terms of material payoffs become coordination games in terms of utilities (Fehr and Schmidt, 1999, 2006). Weakly incentivized obligations can then help coordinating on more cooperative equilibria. Another rationale is that obligations may play a role by themselves because of the emotional cost of disobeying obligations. Galbiati and Vertova (2014) show that expressive law works through both channels. Non-binding obligations influence beliefs, which triggers conditional contributions. They also affect preferences, which is captured by a rightward shift in the individuals' conditional contribution schedule, i.e., individuals contribute more even after controlling for their expectations of others' behavior.

Recently, Eisenberg and Engel (2014, 2016) addressed some of the most important questions related to the effect of legal liability and in particular the fact that individuals may attach some moral value to legal obligations. In their 2014 paper, they use a linear PG game experiment to explore the deterrent effects of legal damages imposed by "victims" on individuals who do not efficiently contribute to the public good. They consider treatments where the probability of liability is in principle not powerful enough to enforce the first best. They find that imperfectly imposed liability nevertheless has some deterrence effect and that deterrence is sensitive to the certainty of liability and to its severity, e.g., whether legal demages are compensatory or punitive. In their 2016 contribution, they seek to unpack the motivations to refrain from causing harm in the unilateral accident framework. The design induces a trade-off between individual profit and a reluctance to impose harm. The authors find that the negligence rule works through three independent additive effects: the rule activates norms of conduct, it exposes the plaintiff to disapproval (a formally expressed blame), and to the risk of having to pay damages (a pecuniary deterrent). Surprisingly, the informal effects operate even if there is no (human) victim, and therefore no reason to activate morality.

Our experimental design considers a game, the *Liability Game*, which shares some features with canonical PG games but also differs on several dimensions. On the one hand, we study situations

[1] Our definition of strict liability and the negligence rule is the same as in the economic model of the public enforcement of law discussed in Polinsky and Shavell (2007). The "accident model" more commonly used in the theoretical economic analysis of tort law is reviewed in Shavell (2007).

[2] See Fluet and Galbiati (2016) for a survey. Some of the experimental literature has been concerned with the "democratic dividend" when non-binding obligations are introduced endogeneously by the participants (Tyran and Feld, 2006; Markussen et al., 2014).

where individuals (potentially) face a trade-off between private and social interests, where they are in symmetrical situations, and where the composition of groups changes over time. These characteristics are shared with many of the canonical PG games. On the other hand, the situations we consider differ on the following dimensions. First, canonical PG games imply linear returns of individual contributions such that social efficiency always requires full contribution. While liability rules also aim at deterring selfish behavior when it runs counter to social welfare, they must be careful not to deter welfare-increasing selfish strategies. We therefore consider liability rules in an environment where selfish behavior can either increase or decrease social welfare, depending on circumstances. Second, the very setting of canonical PG games implies that players know exactly what decisions other group members should have taken to maximize social welfare (i.e., full contribution to the public good). In many situations involving liability rules, one cannot directly infer whether the tortfeasor's action was socially undesirable. In the Liability Game, imposing a loss on others might be either welfare increasing or welfare decreasing, according to the (unobserved) situation the tortfeasor faces. Third, canonical PG games assume perfect information about individual contributions to the public good. Most of the literature on liability rules assumes on the contrary that the tortfeasor's action or conditions associated with the action are not directly observable by others in general, but can be partially or fully revealed by enforcement authorities. The Liability Game proposes a setting where tortfeasors' actions are not directly observable. Fourth, canonical PG games start with a situation where players own their endowment and can partially or fully invest it in the public good. Situations where liability rules are involved usually consider the opposite: endowments or equivalently property rights are initially defined and those who suffer a loss caused by tortfeasors are referred to as victims. Finally, when legal obligations are introduced, there is a subtle difference in the message conveyed by the obligation. In the canonical PG game, the message is that one may be fined for not contributing to the public good, which presumably conveys the obligation to contribute. In our case, as is true of actual liability rules, the explicit message is not that one should not cause harm but rather that one will need to compensate harm caused, possibly only in some circumstances or with some probability.

We also borrow from the Public Good games literature the possibility that individuals impose symbolic or small monetary punishments on others. Informal costly sanctions have been shown to reduce free-riding, both in the Stranger and Partner matching protocols (Fehr and Gächter, 2000, 2002, among others). We emphasize that our aim here is not to study the extent to which weakly enforced formal legal sanctions are supplemented by effective informal punishment for bad behavior. Rather we seek to capture the expression of social disapproval as a reflection of the underlying social norms. We impose a one-to-one ratio between the small cost to the punished and the small cost to the punisher. From the literature, we would therefore expect relatively little effect on cooperation in our experiment.[3] We also allow for purely symbolic punishment as an expression of disapproval. The mere fear of disapproval has also been shown to affect behavior (Masclet et al., 2003; Rege and Telle, 2004; Noussair and Tucker, 2005; Dugar, 2013). In our Liability Game, however, one's actions or circumstances are only imperfectly observed by others. They can be inferred, if at all, only from the overall frequency of harmful actions or from findings of liability under the prevailing rule and enforcement policy. As with real life liability rules, subjects are allowed to express diffuse disapproval of certain actions or more targeted disapproval of individuals who have been found liable but will otherwise remain strangers. Other participants' net final payoff always remains private information.[4]

[3] Social punishment has been shown to be effective in sustaining cooperation in public good games run only when the ratio is substantially above unity, i.e., a high impact-low cost ratio; see Nikiforakis and Normann (2008), Egas and Riedl (2008). With the linear punishment technology, a ratio of at least three-to-one is required

[4] Costly social sanctions in Public Good games under noisy monitoring have been studied by Grechenig et al.

3 The Experiment

The experiment consists of two phases. The first phase is the core of the experiment. Subjects play the *Liability Game*, a game in which legal rules and the enforcement policy change across treatments. Each round of play involves two stages. In the first stage participants choose their actions and may or may not be subject to legal sanctions. In the second stage they have the opportunity to informally sanction other participants. The game is repeated 10 times with nonmonetary punishments and is followed by a modified version of the same game in which nonmonetary punishments are replaced by small monetary punishments, also repeated 10 times. Although participants knew that the experiment was made of several parts, they did not know during the first ten periods that ten subsequent periods with monetary punishments would follow. Finally, in the second phase, we run a questionnaire for demographics and additional control variables. Instructions are displayed in the Appendix.

3.1 Liability Game

Common Set-up. After several control questions, participants are introduced to the Liability Game with a new set of instructions. They are told that they are going to play a game that will be repeated 10 times. At each round, participants are randomly and anonymously matched into groups of 4 participants.

Participants start each round with an initial endowment of 20 ECU. At each round they need to choose between two actions, Y and X. Action Y yields an income of 6 ECU and does not affect the other participants' earnings. Action X yields a state-dependent income and reduces by 4 ECU the earnings of each of the three other participants in the group. While participants are told about actions X and Y, we shall refer to them as actions *Harmful* and *Harmless* respectively for the remainder of the paper. At the beginning of each round, a random state is drawn for each participant among four possible states, A, B, C and D, with equal probability (i.e., 25%). States are independently drawn and are private information throughout the game. The states define a participant's circumstances with respect to the private benefit of action *Harmful*: this action yields an income of 14 ECU in state A, 16 ECU in state B, 20 ECU in state C, and 22 ECU in state D. Note that an individual's circumstances in each period is private information[5] known before choosing which action to undertake and that, at no moment during the experiment, do participants have a feedback about the income of other players.

Treatments. Participants were subjected to five different treatments defined in terms of the prevailing liability rule and the enforcement policy. The between-subject treatment variations on the three liability rules and the two enforcement policies are summarized in Table 1.[6]

- **No Law (NL).** In the first treatment, there is no liability rule. Each participant has to bear the losses caused by the actions *Harmful* of other participants in the group. A participant's net payoff per period, in addition to the endowment at the start of the round, is therefore (i) the private benefit from one's own action *Harmless* or *Harmful*, (ii) minus the losses caused by the actions *Harmful* of other participants.

(2010) and Ambrus and Greiner (2012).

[5] In the instructions: "The other participants of your group won't be able to observe your situation (and you won't be able to know theirs)". See Instructions 2 in the Appendix.

[6] We thank a referee for suggesting this table.

- **Severe Strict Liability (SSL).** In the second treatment, participants are told they will be required to compensate the other group members for the losses caused by their decision to engage in action *Harmful*. Therefore no one suffers from the other participants' decision to engage in *Harmful*. The net period income equals (i) the private benefit from one's own action *Harmless* or *Harmful*, (ii) minus the damages (12 ECU) for compensation if action *Harmful* was chosen.

- **Severe Negligence Rule (SNR).** In the third treatment, participants are told they will be required to compensate for the losses caused by their action *Harmful* if they are in state A or B. A participant therefore suffers from the other participants' actions *Harmful* only if these actions were undertaken in the circumstances C or D. The net period income equals (i) the private benefits from one's own action *Harmless* or *Harmful*, (ii) minus the damages (12 ECU) for compensation if they choose *Harmful* in state A or B, (iii) minus the losses caused by the other participants' decision to engage in *Harmful* in the circumstances C or D.

- **Mild Strict Liability (MSL).** The fourth treatment is similar to the second treatment, except that participants who engage in *Harmful* are made to compensate only with a probability equal to 0.5, henceforth the detection or enforcement probability.[7] The net period income equals (i) the private benefit from one's own action *Harmless* or *Harmful*, (ii) minus the eventual damages (12 ECU) for compensation if they choose *Harmful* and are detected, (iii) minus the losses caused by the other participants' decisions to engage in *Harmful* that were not detected.

- **Mild Negligence Rule (MNR).** The fifth treatment is similar to the third treatment, except that participants who engage in *Harmful* in state A or B are made to compensate only with a probability of one half. The net period income equals (i) the private benefit from one's own action *Harmless* or *Harmful*, (ii) minus the eventual damages (12 ECU) for compensation if the participant chooses *Harmful* in state A or B and is detected, (iii) minus the losses caused by the other participants' decision to engage in *Harmful* when either they were in state C or D or they were in the states A or B but were not detected.

Framing. Liability rules in the instructions were described in a neutral way to avoid any framing effects beyond those naturally involved in liability rules.[8] Specifically, regarding the description of the actions' consequences on others, the instructions were worded as follows:

> *Action Y affects only your own income. On the contrary, action X negatively affects the income of the other participants of your group: it generates a loss of 4 ECU to each of them. Thus, choosing action X will lead to a total loss of 12 ECU for the group (i.e. 4 ECU for each of the three other players).*

Regarding the compensation mechanisms, we describe them as a mechanism aiming at compensating damages caused by action *Harmful*. For instance, for the Severe Negligence Rule, we explain the compensation rules as follows[9]:

[7] Detection was randomly decided by the computer at the individual level.

[8] Liability rules are aimed at *compensating* victims of negative externalities. We therefore talked about "compensations". However, we did not influence participants as to whether action *Harmful* is more or less desirable than action *Harmless*. See the discussion in Section 6.

[9] All instructions can be found in the Appendix.

Table 1: Overview of the treatment variations

		Liability Rules		
		Strict Liability	**Negligence**	**No Law**
Enforcement Policy	**Mild Law** Imperfect enforcement	Perfect compensation for any loss caused to others, if any. Probability of detection: 50%	Perfect compensation for any loss caused to others, if any, only under specific circumstances. Probability of detection: 50%	No compensation for any loss caused to others, if any.
	Severe Law Perfect enforcement	Perfect compensation for any loss caused to others, if any. Probability of detection: 50%	Perfect compensation for any loss caused to others, if any, only under specific circumstances. Probability of detection: 100%	

Once all the participants of your group, you included, have decided between action X and Y, the computer will compute all individual payoffs. Each individual who has caused damages to others in situations A or B will have to compensate them. Participants who undertook action X in situations C or D will not compensate for the damages caused.

Nonmonetary punishments and payoffs. At the end of each period, participants learn the total number of other group members who chose action *Harmful*. In all treatments but *No Law*, participants also know whether each of the other three group members (anonymously identified as player 1, 2 or 3) had to compensate other participants, i.e., was held "legally liable". Individual actions and states of nature are therefore private information except in so far as actions and states can be inferred from the assignment of liability or from the total number of actions *Harmful*. This is the only information obtained about other participants.

After receiving this information and learning their period payoff, participants have the opportunity to assign disapproval points (between 0 and 6) to each other participant in their group. The disapproval points are individual. After the assignment of disapproval points, a final screen displays to each participant the amount of disapproval points the participant received.

More formally, the status of individual i is defined as follows:

$$\text{liable}_i = \begin{cases} 0 & \text{in No Law} \\ HF_i & \text{in Severe Strict Liability} \\ \mathbb{1}_{A,B} HF_i & \text{in Severe Negligence Rule} \\ \mathbb{1}_{\text{det}} HF_i & \text{in Mild Strict Liability} \\ \mathbb{1}_{\text{det}} \mathbb{1}_{A,B} HF_i & \text{in Mild Negligence Rule} \end{cases}$$

where HF is a dummy variable equal to 1 if action *Harmful* is undertaken, $\mathbb{1}_{A,B}$ is a dummy variable equal to 1 if individual i is in state A or B and equal to 0 otherwise; $\mathbb{1}_{\text{det}}$ is a dummy variable equal to 1 if individual i is detected after engaging in action *Harmful*, and is 0 otherwise.

Monetary punishments and payoffs. After the ten rounds of the Liability Game with disapproval points, participants are given a new set of instructions and learn that they will play

another 10 rounds of the previous game. The only change compared to the 10 previous rounds is that disapproval points are replaced by costly sanction points.[10] Participants have the opportunity to impose sanction points on other group members. Each point decreases both the participant's and the target's payoff by 0.5 ECU.[11] Each participant can assign up to 6 sanction points to each group member.

3.2 Control Questions.

Before playing the game, participants were asked a series of questions to ensure that the game was well understood. We generated a mock stage game in which we displayed the actions, situations and liability status of a virtual group of participants. Control questions were designed to address all the mechanisms that affect the period payoffs. Participants had to fill in, step by step, a table which required to compute the losses each participant imposed and bore, the compensation each of them gave and received, and their final payoff.

3.3 Questionnaire

After completion of the Liability Game, participants are asked to fill out a questionnaire on demographics (age, gender) and on preferences and self-perception. These include (i) self-declared political orientation, (ii) attitude with respect to state intervention in the economy, (iii) self-assessed risk aversion, (iv) the extent to which they see themselves as selfish, (v) how much they think others see them as selfish, (vi) the extent to which they feel concerned about the well-being of others, (vii) how much they think others see them as being concerned by the well-being of others. For cross-study comparison purposes, the last four questions were adapted from Angelova et al. (2014).[12]

3.4 Sessions.

The experiment was computerized using z-Tree (Fischbacher 2007). We ran 10 sessions (two per treatment) in May, July and September 2016 in Québec (Canada) and Strasbourg (France).[13] Each session included 20 participants, amounting to five groups of four subjects at each round, except for one No Law session that included only 16 participants. Overall, 196 participants took part in the experiment. An ECU was convertible to Canadian dollars at 30 ECU = 1 dollar or to Euros at 40 ECU = 1 Euro. The same co-author supervised all sessions. All participants in Strasbourg were recruited by ORSEE (Greiner (2015)) and by a similar software developed by Laval University in Québec. Participants earned on average 19.15 Euros and the sessions lasted between 75 and 90 minutes.

[10] The within-subject treatment variation regarding the monetary cost of sanctions is kept constant for all liability rules and enforcement policies. Our goal is not to explore the impact of sanctions *per se*, but to capture social norms by measuring how subjects respond to the same type of sanction under different liability rules.

[11] Given the one-to-one ratio between the cost to the punished and the punisher, informal sanctions do not change relative payoffs. This eliminates the possibility that informal sanctions are driven by aversion to disadvantageous inequality. As noted earlier, a high impact-low cost ratio is needed to sustain cooperation in public good games; see footnote 3.

[12] In this experiment, we use non-incentivized elicitations of preferences regarding risk-aversion and concerns for others, as in Angelova et al. (2014). While non-incentivized elicitations can have some benefits (Lönnqvist et al., 2015), other experimental works take the advantage of the laboratory to use incentivized elicitation methods (e.g., Eisenberg and Engel, 2016).

[13] We ran 3 sessions in Québec (1 SSL, 1 SNR, 1 NL), and all remaining sessions in Strasbourg. All regression results presented in the paper hold if we include a dummy variable that accounts for the location of the session.

4 Models and Predictions

Socially efficient behavior, in the sense of maximizing total group wealth, is for action *Harmless* to be chosen in the circumstances A or B and action *Harmful* in the circumstances C or D. However, with purely self-interested agents, action *Harmful* is strictly dominant under No Law in all circumstances.[14] Under Mild Law, the same is true for a risk neutral under both legal regimes, but in circumstances A and B a sufficiently risk averse might choose action *Harmless* to avoid bearing risk. Nevertheless, given the initial endowments and the probability of detection, only an extremely risk averse would do so. For example, with the Constant Relative Risk Aversion utility function, an individual in the circumstance A will choose *Harmful* if his coefficient of relative risk aversion is below 3.5. A coefficient above this threshold is outside the normal range in laboratory experiments, where relative risk aversion coefficients are often estimated to be between 0.5 and 1.5 (see for instance Holt and Laury 2002). We conclude that our Mild Law is non-deterrent except possibly for unreasonably risk averse individuals. In what follows, for the case of imperfect enforcement, expected values are taken as an acceptable approximation of certainty equivalents.[15]

Prediction 1 Under No Law or under Mild Law, purely self-interested agents choose the action *Harmful* in all circumstances. Under Severe Law, in either legal regimes, they choose the action *Harmless* in the circumstances A or B and the action *Harmful* in the circumstances C or D.

The experimental literature suggests some departure from pure self-interest. This has been analyzed mostly in terms of interpersonal comparisons or efficiency concerns. We discuss the implications of some well-known formulations.

Inequality aversion. In Fehr and Schmidt's canonical model of social preferences, individuals have a utility function of the form:

$$u_i = \pi_i - \frac{\alpha_i}{n-1}\sum_{j\neq i}\max\{\pi_j - \pi_i, 0\} - \frac{\beta_i}{n-1}\sum_{j\neq i}\max\{\pi_i - \pi_j, 0\}$$

where π_i is individual i's total payoff[16], n is the number of individuals in the group, and α_i and β_i are the so-called envy and empathy (or guilt) parameters respectively, with $0 \leq \beta_i \leq \alpha_i$ and $\beta_i < 1$.

In many PG games, the above utility function has been shown to be consistent with an efficient outcome (among multiple Nash equilibria), provided all individuals have sufficiently large empathy.[17] This arises in particular because the efficient outcome is characterized by equal payoffs. This is not so in our set-up: efficient behavior is most likely to yield unequal payoffs, irrespective of the legal regime. Because envy is a stronger driver than empathy, inequality aversion is not conducive to efficient behavior. We only give the informal argument. Consider the No Law regime. Suppose individual i expects others to behave efficiently and finds himself in the circumstance A. The probability that at least one of the three other players is in the circumstances C or D equals $1 - .5^3 = 0.875$. This is also the probability that at least one of the other three players, behaving efficiently, will choose the action *Harmful*, thereby inflicting harm on i and providing for himself a large payoff. In other words, individual i will most surely be *behind*. Therefore, in the circumstance

[14] Proceeding by backward induction, the equilibrium prediction is that of a one-shot game at each period of play.

[15] In the subsequent analysis, we rejected the hypothesis that the certainty equivalent significantly differs from the expected value.

[16] This is the total payoff as defined in Section 3.1. It therefore depends on the individual's action and circumstances, on the actions of others, and on the legal regime.

[17] See the elaborate discussion in Markussen et al. (2014), in particular their online appendix.

A, he will not be motivated by empathy (but rather by envy) and he will choose the socially inefficient action *Harmful*. In turn, if an individual expects others to always choose *Harmful*, he will himself do the same in all circumstances.[18]

The same argument applies under Mild Law, given the enforcement probability of one half and our calibration of payoffs. By contrast, and as expected, efficient behavior is a Nash equilibrium under Strong Law. Thus, in our set-up, inequality aversion yields the same predictions about the possibility of efficient behavior as would standard self-interested preferences. A similar argument applies with other models of inequality aversion, such as Bolton and Ockenfels (2000).

Altruism and efficiency concerns. The main concurrent approach views departures from pure self-interest as motivated by altruism or a concern for efficiency. Borrowing from Charness and Rabin (2002) and others[19], let individuals have a utility function of the form:

$$u_i = (1 - \lambda_i)\pi_i + \lambda_i \sum_j \pi_j \tag{1}$$

where $\lambda_i \in [0, 1]$ is the weight the individual puts on social efficiency as measured by the total group payoff, including that of individual i. Note that (1) is equivalent to

$$u_i = \pi_i + \lambda_i \sum_{j \neq i} \pi_j.$$

Individual i's utility is always increasing in his own payoff and is never decreasing in the total payoff of other participants. If all individuals have λ_i equal to zero, preferences reduce to pure self-interest. An individual with a positive λ_i is willing to sacrifice some private payoff provided this increases sufficiently the total group payoff, including his own.[20]

Due to the linearity of (1), the equilibrium is in dominant strategies: one's best response is independent of the (expected) actions of others and the losses these may cause. Denote by g_p the private benefit from individual i's action, net of the legal damages he may be required to pay. Denote by g_s the effect on the total group payoff, including individual i. The private gain from action *Harmless* is y, that of action *Harmful* is x, which depends on circumstances, and the external harm this action causes is h. Therefore, for action *Harmless*, $g_p = g_s = y$. For action *Harmful*, in expected value, $g_p = x - ph$ and $g_s = x - h$ where p is the probability of having to compensate the harm. Depending on circumstances, the legal regime and the enforcement policy, p is either zero, one half or unity.

Let us denote with capital letters the net private and social consequences of action *Harmful* compared to action *Harmless*:

$$G_p = x - y - ph, \quad G_s = x - y - h.$$

Then, in the Nash equilibrium, individual i chooses *Harmful* if

$$\Delta u_i = (1 - \lambda_i)G_p + \lambda_i G_s \tag{2}$$

[18] There may be another equilibrium. If *all* individuals have $\beta_i > 2/3$, then there is an equilibrium where all individuals always choose *Harmless*. However, such a degree of aversion to advantageous inequality is inconsistent with results found elsewhere; see for instance Bellemare et al. (2008).

[19] See Andreoni and Vesterlund (2001), Engelman and Strobel (2004), Charness et al. (2016)

[20] Such an individual is altruistic in the sense that he values others' payoffs, but not to the point of redistributing his own wealth as this would not change the total group payoff.

is positive. The utility differential can be rewritten as

$$\Delta u_i = G_p - \lambda_i(G_p - G_s)$$

where $G_p - G_s = (1-p)h$ is the expected *uncompensated* harm imposed on others. Thus, the social concern parameter λ_i can be interpreted as the rate at which the individual trades off his own private benefit net of legal sanctions against the uncompensated loss caused to third parties. Note that $G_p \geq G_s$.

If action *Harmful* is socially efficient, i.e., $G_s > 0$, then (2) is positive and action *Harmful* is chosen irrespective of the value of λ_i, the legal regime and the enforcement policy. If action *Harmful* is socially inefficient, i.e., $G_s < 0$, and the law perfectly internalizes the harm, i.e., $G_p = G_s$, then (2) is negative and action *Harmless* is chosen. It follows that the social concern parameter λ_i matters only in circumstances where self and group interest conflicts (i.e., $G_p > 0$ and $G_s < 0$) and liability rules are imperfectly enforced or non existent.

Whether social concern then makes a difference depends on the size of λ_i and the choices facing the individual. In the circumstances A or B, an individual with a sufficiently large λ_i will choose action *Harmless* even under No Law. In the same circumstances, an individual with a smaller λ_i may choose *Harmful* under No Law but *Harmless* under nondeterrent Mild Law, because the private gain G_p net of legal liability is now smaller while the social gain G_s is unaffected. Also, under No Law or under Mild Law, an individual who would have chosen *Harmless* in the circumstance A may well choose *Harmful* in the circumstance B because both G_p and G_s are increasing in the gross benefit x from action *Harmful*.

Prediction 2 When agents are efficiency concerned, (i) in the circumstances A or B, the proportion of agents choosing *Harmful* may decrease from No Law to Mild Law; (ii) under No Law or Mild Law the proportion of agents choosing *Harmful* may increase from circumstance A to circumstance B; (iii) in the circumstances C or D, all agents choose *Harmful* irrespective of the legal regime.

Reciprocity, normative effects of obligations, and other considerations. Other things equal, some individuals may strictly prefer not to cause harm. In particular, when both actions are equally efficient, i.e., $G_s = 0$, and the legal system perfectly internalizes the harm caused to others, i.e., $G_p = G_s$, some individuals may strictly prefer action *Harmless*. This modifies the expression in (2) to

$$\Delta u_i = (1-\lambda_i)G_p + \lambda_i G_s + \delta_i$$

where $\delta_i \leq 0$ captures a willingness to avoid action *Harmful* and reflects a variety of motivations other than related to private and social gains, e.g., a pure reluctance to cause harm or the disutility from social disapproval when one is found (or is believed) to have caused harm.[21] In Section 5.2, δ_i is replaced by βX_i where β is a vector of coefficients and X_i is a vector of variables, including socio-demographic variables and occurrences during the game such as disapproval points received. An individual with a sufficiently negative $\delta_i = \beta X_i$ could refrain from the action *Harmful* even when it is the socially efficient action.

Reciprocity refers to the possibility that the weight an agent puts on others' payoffs depends on the perceived social concern of others. For instance, $\lambda_i = 0$ when others misbehave while $\lambda_i = \lambda^* > 0$

[21] The latter relates to the incentivizing role of social or self-esteem concerns, as in Bénabou and Tirole (2006, 2011); see also Deffains and Fluet (2013) for an application to liability rules in the accident model.

when others behave well.[22] Under No Law, reciprocity raises the possibility of multiple equilibria ranked in terms of the extent of socially efficient behavior.[23]

As discussed in Section 2, the literature on public good games with nondeterrent legal obligations has emphasized that obligations may affect expectations and help coordinate on the better equilibria. Another possibility is that obligations directly modify preferences through moral suasion, irrespective of sanctions and expectations.[24] We capture such effects by regime dependent coefficients. The expression in (2) is modified to

$$\Delta u_i = (1 - \lambda_{ir})G_p + \lambda_{ir}G_s + \delta_{ir} \tag{3}$$

where the legal regime is denoted by r and is either No Law (NL), strict liability (SL) or the negligence rule (NR).

If legal liability crowds in social concerns, the λ_{ir}'s will be at least as large under SL or NR as under NL. In the circumstances A and B, a change from No Law to Mild Law then reduces the proportion of individuals choosing *Harmful* in two ways: first, the private material benefit of action *Harmful* is reduced because of the expected liability cost; secondly, the law may reinforce social concerns.

Prediction 3 Under strict liability or the negligence rule, individuals are at least as much concerned about the net uncompensated losses caused to others as under No Law.

Disapproval behavior. In PG games or prisoners' dilemma experiments, abstracting from revenge punishments, informal sanctions target those who behave badly, i.e., do not contribute or defect (e.g., Falk et al., 2005). This is consistent with the role of indignation in social sanctioning behavior (e.g., Hopfensitz and Reuben, 2009; Carpenter and Matthews, 2012). The analogue in our set-up, if individuals value efficiency concerns, would be for informal sanctions to target those who are thought to behave inefficiently.

If others are expected to be motivated by self-interest and possibly also by some degree of efficiency concerns, it should be inferred that those who did not cause harm behaved efficiently. A participant observes the number of actions *Harmful* among the three other participants and he observes who has been held liable, an *identification effect*. From these two pieces of information, a participant derives the probability that any of the other participant engaged in action *Harmful*. Under No Law or Mild Law, one who causes harm is more likely to have behaved inefficiently, hence disapproval (or sanction) points assigned to other participants should be increasing in the probability that they caused harm. In addition, under the negligence rule, liability reveals for sure that behavior was inefficient. More generally, informal sanctions may depend on the perceived prevalence of inefficient behavior.

Prediction 4 Disapproval is targeted at those who are perceived to inefficiently cause harm.

[22] This is in line with the simple model of reciprocity in Section II of Charness and Rabin (2002).

[23] Segal and Sobel (2007) assume that individuals have preferences over the strategies of others rather than only over outcomes. An individual's preferences can then be represented as a linear combination of all the agents' utility functions over outcomes, where the weights depend on the equilibrium strategies. This is consistent with the decision function in (2) where the individuals' λ_i are obtained at equilibrium.

[24] See Dal Bó and Dal Bó (2014). Galbiati and Vertova (2014) find that both channels are at play.

5 Results

To explore the results of our experiment, we proceed as follows. First, we discuss the overall efficiency of each liability system. Second, we test the explanatory power of a simplified version of the theoretical model proposed in section 4. The goal is to assess how much the individuals' decisions can be explained in a framework combining only self-interest and social efficiency concerns together with a pure reluctance term. Next we take into account the socio-demographic variables and additional factors that may impact the decision to undertake the action *Harmful*. This may be interpreted as disaggregating the reluctance term. Finally, we explore the decisions to impose social sanctions on other participants.

5.1 Efficiency of liability rules

Figure 1 displays the proportion of actions *Harmful* undertaken in each of the four possible circumstances under every legal regime. Clearly the presence of a legal system and the extent to which it is enforced greatly impact decisions when self and group interests conflict, i.e., in the circumstances A and B. In these circumstances, the proportion of actions *Harmful* is greatest under No Law (88.2% and 96.9% under circumstances A and B respectively) and reaches the smallest level under Severe Law, both under strict liability (9.5% and 18.3% under circumstances A and B respectively) and under the negligence rule (15.0% and 15.0% under circumstances A and B respectively). Proportion tests and multivariate analyses[25] confirm that individuals are more likely to undertake the *Harmful* action in circumstances A and B in No Law, than in Severe Strict Liability ($p < 0.001$ for circumstances A and B) or in Severe Negligence Rule ($p < 0.001$ for circumstances A and B). Thus, it seems that a perfectly enforced legal system successfully achieves its main objective.

Result 1 The introduction of a priori deterrent *perfectly* enforced liability rule significantly reduces the proportion of harmful welfare-decreasing actions.

However, the data also partially contradict the theoretical predictions based on purely self-interested preferences. First, in the circumstances A or B, there is a significantly smaller proportion of actions *Harmful* under Mild Law than in No Law, for both the strict liability (41.8% and 63.9% under circumstances A and B respectively) and the negligence rule (54.3% and 75.2% under circumstances A and B respectively). In these cases, it is privately inefficient to undertake *Harmless* but almost half of the participants choose to do so. Proportion tests and multivariate analyses reject the equality of behaviors under circumstances A and B for No Law compared to Mild Strict Liability ($p < 0.001$ for circumstances A and B) and Mild Negligence Rule ($p < 0.001$ for circumstances A and B).

Result 2 The introduction of a priori non-deterrent *imperfectly* enforced liability rule significantly reduces the proportion of welfare-decreasing actions.

Moreover, the proportion of actions *Harmful* increases with the circumstances: more participants tend to undertake *Harmful* in situation B compared with A or in situation D compared with C, although standard preferences would predict similar choices.

[25] We estimated a random effect probit model that include all control variables presented in section 3.3. Standard errors were clustered at the session level. We report the p-values of the coefficient associated with the treatment dummy variable. Results are available on demand.

So far, the results are consistent with Prediction 2. In addition, we also observe that, under Mild Law, strict liability yields a smaller proportion of actions *Harmful* than the negligence rule in the circumstances A and B. Indeed, 41.8% (resp. 63.9%) of the participants choose the action *Harmful* in circumstance A (resp. B) in Mild Strict Liability, compared to 54.3% (resp. 75.2%) for Mild Negligence Rule. These differences are statistically significant in a proportion test ($p = 0.015$ and $p = 0.014$ for circumstances A and B respectively) and in multivariate analyses ($p < 0.001$ and $p = 0.012$).

Result 3 When legal rules are imperfectly enforced, strict liability is more likely to deter harmful welfare-decreasing actions than the negligence rule.

Finally, we further observe that both under Mild and Severe Law, strict liability also yields a smaller proportion of actions *Harmful* in the circumstances C and D. In these circumstances, it is both privately and socially optimal to choose *Harmful*. In Severe Law, participants under strict liability undertake the action *Harmful* 83.2% of the time in circumstance C (resp. 91.3% in circumstance D), against 94.0% for participants who play under the negligence rule (resp. 97.0% in circumstance D). These differences are statistically significant in proportion tests ($p < 0.001$ and $p = 0.0158$ for circumstances C and D respectively) and in multivariate analyses ($p = 0.010$ and $p < 0.001$). In Mild law, we observe a similar pattern: 83.7% participants facing strict liability in circumstance C choose the *Harmful* action (resp. 96.8% in circumstance D), against 99.1% for participants facing the negligence rule (resp. 98.9% in circumstance D). The difference is statistically significant for circumstance C in proportion tests ($p < 0.001$ and $p = 0.169$ for circumstances C and D respectively) and for both circumstances in multivariate analyses ($p < 0.001$ and $p < 0.001$).

Result 4 When legal rules are imperfectly or perfectly enforced, strict liability is more likely to deter harmful welfare-increasing actions than the negligence rule.

5.2 Social concerns and private interests

We next turn to a simple approach for understanding the data on the basis of the model developed in Section 4. The next section will build on this preliminary analysis, integrating additional econometric considerations. For simplicity, we assume that all participants share the same fixed set of preferences except for an additive noise term in maximizing utility. This term crudely captures preference heterogeneity among participants. We fit the logistic regression

$$P = \frac{e^{\gamma \Delta u}}{1 + e^{\gamma \Delta u}} \tag{4}$$

where P is the probability of undertaking action *Harmful*, Δu is the difference in utility between the actions *Harmful* and *Harmless* as defined in (3), and γ is the parameter capturing the sensitivity of behavior to differences in utility as defined in the model. The size of γ reflects the explanatory power of the model. Thus, we estimate a binary-response logit with the propensity score

$$\gamma \left[(1 - \lambda_r) G_p + \lambda_r G_s + \delta \right], \, r \in \{NL, SL, NR\}$$

Table 5 shows the regression results for a variety of restrictions on the parameter values. In Model 1, all parameters are constrained to equal zero except γ. This specification corresponds to purely self-interested preferences, i.e., the propensity score in the logit is γG_p. In Model 2, we allow

for social concerns with a parameter that does not depend on the legal regime. The parameter is highly significant and the log-likelihood improves markedly. Individuals put non-negligible weight on the group payoff in addition to their own private benefit. In Model 3, we allow the constant term to differ from zero. The term has the expected sign indicating a reluctance to choose action *Harmful* everything else equal, but the gain in explanatory power is slight and the average willingness to pay to avoid action *Harmful* is very small (0.27 ECU).

In the last three models, we allow the social concern parameter to differ between legal regimes. The major gain here is with respect to the value of the parameter under strict liability. In Model 4a, λ_{SL} may differ from $\lambda_{NL} = \lambda_{NR}$. The difference is highly significant. Individuals put greater weight on others' payoff under strict liability compared with No Law and the negligence rule. Model 4b does the same with respect to λ_{NR} but this is much less successful. Finally, in Model 5, all restrictions are removed. The social concern parameters are quantitatively important in all legal regimes and they differ significantly between regimes. In this more flexible model, the reluctance parameter δ essentially vanishes. Figure 2 illustrates the predictions of the model.[26] To summarize, we have the following results, which are consistent with Prediction 3.

Result 5 Individuals care about others' payoff. In the absence of legal obligations, there is a trade-off rate of 28% between private benefits and the losses imposed on others.

Result 6 Legal rules crowd in social concerns. The trade-off rate between the net private benefits and the net uncompensated losses imposed on others increases to 37% under the negligence rule and to 50% under the strict liability rule.

5.3 Determinants of action *Harmful*

The model of Section 4 explains a great deal of the data. About 43% of the individual decisions[27] are explained by expected private benefits net of legal liability, the effect on group payoff, and the impact of legal rules on the social concerns coefficient. We now consider further refinements. We exploit the repeated choices of each individual by adding individual effects. In addition, we include the control variables of the questionnaire and also consider the impact of social disapproval and informal costly sanctions on the decision to undertake *Harmful*.

In the Tables 6[28] and 7, the decisions to choose *Harmful* over *Harmless* are analyzed by logistic regression with individual random effects and clustered at the session level.[29] The variable G is defined as $G = G_s - G_p$, i.e., it is the expected uncompensated loss imposed on others expressed as a negative value. In the regression of Table 6, the social concern variable is constrained to be the same in all legal regimes. For simplicity, the estimated coefficients are presented in their raw form. For comparison with Table 5, the coefficient of G_p should be read as $\beta_1 = \gamma$; the coefficient of G should be read as $\beta_2 = \gamma\lambda$. Hence the social concern parameter is computed as $\lambda = \beta_2/\beta_1$. In Table 7, the social concern parameter is allowed to differ between legal regimes, so the coefficient of G is interacted with a regime specific dummy variable.[30]

[26] Allowing δ to differ between legal rules was not significant. We also tested for the role of risk aversion under imperfect enforcement with dummy variables subtracted from the expected private gain. This was also not significant.

[27] On the basis of the Pseudo R^2

[28] Note that the results hold if we exclude the 10th and the 20th periods.

[29] We cluster data at the session level to take into account the potential dynamic session effects. See Fréchette (2012).

[30] Ai and Norton (2003) and Norton et al. (2004) discuss the difficulty of directly interpreting interaction terms in non-linear models. This difficulty originates from the fact that two variables changes at the same time. Here, our specification does not include treatment effects, which allows us to interpret the coefficients directly. Given that, we

As in the previous section, the participants' behavior is partly explained by a concern for the net uncompensated losses they can impose on others. In Table 6, the estimated coefficient of the net loss variable G is significant in all specifications. Depending on the specification, the value of the implied social concern parameter λ varies between 0.37 and 0.39 (in the columns 2a to 5a) which is very close to the value estimated in Model 3 of Table 5. When we allow the coefficient to differ between legal regimes as in Table 7, we again obtain that the social concern parameter in the presence of liability rules is substantially larger than under No Law. The parameter is around 0.25 under No Law, 0.53 under the negligence rule and 0.57 under strict liability. The difference between strict liability and the negligence rule is no longer significant.

Our experiment allows participants to express disapproval of their fellow participants at the end of each round. Social sanctions have been showed to influence behavior in public good experiments. Our results suggest that individuals are indeed less likely to undertake *Harmful* when they have received nonmonetary punishment points at the previous round. This effect is more pronounced with monetary punishment points. The behavioral impact of social sanctions occurs mainly in situations where individuals were disapproved of after undertaking *Harmful*, whereas disapproval following action *Harmless* does not affect subsequent decisions.[31] Last, the proportion of *Harmful* actions by other participants in the past history of the game does not seem to affect one's decision to undertake action *Harmful*. Note also that there is clearly much heterogeneity between individuals. Everything else equal, those who have chosen action *Harmful* in the past are also significantly more likely to do so in the future.

Result 7 Individuals are less likely to reiterate a harmful action when they have been disapproved of or informally sanctioned at the previous period in situations where they have engaged in the harmful action.

5.4 Determinants of social sanctions

Our main objective for introducing social sanctions, which we now deal with, is to identify the determinants of disapproval as a reflection of the underlying social norms. Because our groups are rematched after each round, disapproval can only be expressed about essentially anonymous persons, e.g., those who have been found liable or who are believed to have caused harm.

Participants largely took advantage of the possibility to express disapproval. Individuals received on average 4 to 8 nonmonetary punishment points per round depending on the treatment. Costly sanctions are assigned less often, i.e., they are sensitive to the cost of punishment. When a legal regime is in force, nonmonetary punishment appears to be mainly concentrated on participants who were held liable, except under the negligence rule under Mild Law; see Figure 3. By contrast, with costly sanctions, there is no difference between liable and not liable under strict liability, whether Mild or Severe, while the difference is significant under the Severe negligence rule and somewhat so under the Mild negligence rule; see Figure 4. The latter result (but not that Figure 3) is roughly consistent with Prediction 4. Also, there is much disapproval of individuals who have not been held liable. Under No Law, no one is ever held liable but participants received on average 8 disapproval points per round, which is also consistent with Prediction 4.

computed the marginal effects associated with the regressions of table 7 and found similar results to what is presented below. Note that, if we exclude individual effects, we still observe statistically different effects of G between SL and NL, but the significance decreases for NR vs. NL.

[31] Note that including past behaviors in the regression can create endogeneity issues. Correcting endogeneity using Arellano-bond like correction would require GMM estimations, which do not allow for logistic regressions with random individual effects and clustering at the session level (Stata package *melogit*). Following the literature on Public Good Contribution, we prefer to correct for individual heterogeneity and clustering.

We ran a series of regressions on the decision of individual i to punish the other participants in his group at round t. Recall that, at the end of each round, participant i observes whether participant j was held liable and he also observes the total number of actions *Harmful* for which individuals were not held liable. From the latter, participant i can infer the probability that a non liable participant j undertook the harmful action. This inference does not depend on the legal regime and in some cases the up-dated probability may very well equal unity. This allows us to disentangle two effects: (i) how one's belief that j engaged in *Harmful* affects one's disapproval of j; (ii) the effect of liability *per se*, in addition to its role in revealing that j engaged in *Harmful*.

The results are displayed in Table 8. The coefficient of the explanatory variable liable_j is the number of punishment points assigned on average to a participant who has been found liable. The variable $(1 - liable_j)pHF$ is the probability that a non liable participant j engaged in action *Harmful*. Its estimated coefficient can therefore be interpreted as the number of punishment points assigned to a non liable participant who is believed to have engaged in *Harmful* for sure, i.e., when $(1 - liable_j)pHF$ equals unity. Overall the difference between the coefficients of the two variables is small or not significant. There is no significant effect of liability *per se* in addition to its informational role.

Engaging in action *Harmful* does not necessarily imply socially inefficient behavior. The same observation holds with respect to legal liability, except under the negligence rule. We further inquired whether there is additional disapproval of a participant who may be believed to have *inefficiently* engaged in *Harmful*, whether the participant was found liable or not. The variable pHF_AB captures participant i's belief that participant j engaged in *Harmful* in the circumstances A or B. The aim is now to disentangle between disapproval of individuals who are believed to have caused harm and of those who are believed to have caused harm inefficiently.[32] The coefficient of pHF_AB has the predicted sign but is not significant. We conclude that there is no evidence that efficiency concerns constitute a strong determinant of disapproval.

Finally, a tendency for "blind revenge " with respect to nonmonetary punishment points is also observed. Participants are more likely to sanction other participants the more they have been sanctioned themselves at the previous round. The effect is significant but quantitatively small and is non existent when social sanctions are costly. Although "blind revenge " is more likely to occur when an individual was disapproved of after undertaking *Harmless* as opposed to *Harmful*, the difference is not statistically significant.

Result 8 The main driver of social disapproval is the belief that a participant has engaged in the harmful action, whether or not the participant was held legally liable and irrespective of whether behavior was believed to be inefficient.

6 Discussion

Framing effects in Public Good games have been extensively discussed. For instance, subjects contribute more when the positive externality of contributing to the public good is stressed rather

[32]The variable pHF_AB is the probability that a participant undertook action *Harmful* in situation A or B. It is the probability of the joint events *Harmful* and AB. We compute it as the product of the probability that an individual undertook *Harmful* and the conditional probability that an individual undertook action *Harmful* in situation A or B given that he undertook action *Harmful*. For treatments SSL, MSL and NL, the conditional probability equals the ratio of the number of actions *Harmful* undertaken in A or B over the total number of actions *Harmful* undertaken in the treatment, i.e. the empirical frequency. In SNR, it is equal to 1 if the individual is liable, and to 0 if he is not liable. Under MNR, the variable equals 1 if the individual is liable and equals the empirical frequency if he is not liable.

than the negative externality of not contributing (e.g., Andreoni, 1995; Park, 2000). In a review paper, Cartwright (2016) remarks that positive-negative framing should be distinguished from give-take framing, i.e., whether subjects are asked to contribute to a public good or take from a public resource. He also suggests a third dimension related to how the participants' initial allocation is presented.

By their very nature, liability rules involve some degree of framing. This is because they posit a reference point equal to the individual's endowment (an *entitlement*) and with respect to which compensable losses are defined. Under No Law, the instructions to the participants specified endowments, the private benefits from possible actions, and the fact that some actions caused losses to others and that one could suffer losses due to the actions of others. When legal liability was introduced, it was described as the obligation to compensate the losses caused to others, possibly only in some circumstances. In Cartwright's typology, our instructions can therefore be described as a negative-take-endowment framing. Obviously the consequences of one's actions could have been presented differently. However, because our goal is to assess the behavioral impact of liability rules, our framing of consequences is part to the very effect we aim at capturing. It should be stressed that at no point did the instructions suggest to participants that they should refrain from the action *Harmful*. Consistent with the legal notion of liability rules, the message only concerned the obligation to compensate and in what circumstances.

We also analyzed the determinants of social disapproval as an expression of underlying social norms.[33] Participants unambiguously express disapproval of individuals who cause losses, whether in No Law or when an obligation to compensate is conveyed. In the treatments with Law, those actually found liable face somewhat greater disapproval but the difference is barely significant. Also, whether the action causing harm is socially efficient does not seem to matter in the expression of disapproval. This can be read directly from the Figures 5 and 6. When liability rules are perfectly enforced, most individuals behave efficiently in terms of the overall group income. Under strict liability, individuals found liable therefore most likely behaved efficiently, i.e., they must have chosen the harm causing action only in the circumstances C or D. These individuals should face no disapproval if participants primarily disapproved of actions that inefficiently cause harm.[34] However, individuals found liable under strict liability receive on average the same number of disapproval points as those found liable under the negligence rule, who for sure behaved inefficiently. Moreover, whether harm was compensated or not has no effect either on the expression of disapproval. Individuals found liable under the perfectly enforced strict liability rule face roughly the same disapproval level as the average individual under No Law, who most likely has caused harm but without compensating. Finally, individuals react (at least slightly) to disapproval received but only after choosing the action causing harm. Presumably this is when they feel disapproval was 'justified'. In our experiment, bad behavior seems to be equated to causing losses to others. The finding that subjects punish others who cause damage, whether this is social welfare enhancing or decreasing, has also been found in other experiments; see for instance Wilkinson-Ryan and Baron (2009) for the case of breach of contract. Causing damage albeit for higher personal gain is not something people seem to endorse.[35]

We posited a utility function with social welfare concerns augmented with additional consid-

[33] Our definition of social norms is the usual one "as behavioral standards based on socially shared beliefs about how individuals ought to behave in a given situation " (Harris et al, 2015, p. 292; see also Bernhard et al. 2006a,b).

[34] This is also the prediction in Deffains and Fluet (2013) in a model where heterogeneous individuals earn social esteem when they are thought to trade-off appropriately private benefits and losses to others. Under perfectly enforced strict liability, everyone behaves the same, hence being found liable is not stigmatizing.

[35] It may also be remarked that, in our experiment, one may choose to cause harm in order to increase one's private benefit, not to avoid a private loss. Wilkinson-Ryan and Baron (2009) remark that individuals are more prone to sanction tortfeasors when harmful actions are motivated by cupidity rather than loss avoidance.

erations. Our set-up, however, is one-sided: individuals can benefit the group only by sacrificing private benefits in order to avoid causing losses. The social concern parameter is therefore best interpreted as the rate at which individuals trade-off private benefits against losses imposed on others, i.e., individuals would perhaps not be driven by the same parameter if sacrificing private benefits increased the wealth of others. We allowed for the possibility that 'observed' coefficients be equilibrium-dependent and in particular that they differ between legal regimes. We also allowed for a reluctance to undertake the action causing harm, unrelated to the trade-off between private benefits and losses to others. Reluctance as defined here acts as a fixed cost of engaging in the harmful action. In the complete analysis, it also captures the threat of social disapproval. Both social efficiency concerns and reluctance seem to be at work and may explain why under strict liability, with either Mild or Severe Law, some individuals refrain from the harmful action even when it is desirable from a collective point of view. The risk of legal sanctions combined with other motivations then yields some overdeterrence.

7 Conclusion

An extensive theoretical literature has discussed the merits of strict liability and fault-based legal systems in deterring behavior that generates negative externalities. This body of research concludes that both types of rules, when perfectly enforced, achieve efficiency by aligning self-interest on collective interest.

We designed an experiment to investigate how agents behave with or without liability rules and when rules are weakly enforced. Our set-up is related to PG games in which participants are randomly matched with strangers and where private and group interests potentially conflict. In our setting, participants must decide between two actions, one of which generate negative externalities. In principle the legal rules considered completely align private and group interests but they are insufficiently deterrent when poorly enforced.

Our paper compares theoretically equivalent liability rules (strict liability and negligence) when enforcement is not perfect and when engaging in a harm causing action may either increase or decrease welfare. In a broader sense, this relates to the literature on welfare-enhancing and welfare-reducing corruption. For example, Cameron et al. (2009) investigate whether individuals' tendency to engage in and punish corrupt behaviour is different when the payoff loss to society is greater than the private benefits to the bribers. Our findings can contribute to improve the economic models of tort law by showing that apparently equivalent liability rules do not in fact achieve the same levels of efficiency.

The participants' behavior contradicts the predictions under standard self-interested preferences. First, the evidence suggests that individuals are willing to trade-off private gains against the losses imposed on others. Second, our experiment reveals that the weight given to social concerns relative to private benefits is increased by the introduction of liability rules. This effect is net of or in addition to the expected formal sanction introduced by legal liability, when nondeterrent, or of the fact that legal liability facilitates informal sanctions by making harmful actions attributable. We do not investigate the precise channel through which the effect operates, e.g., whether it is through changed expectations triggering conditional cooperation or through pure norm activation. We find weak evidence that the effect is stronger under strict liability than under the negligence rule. Compared with the negligence rule, the obligation under strict liability perhaps appears unequivocal and more equitable: in principle at least, one is always "responsible" for the harm caused to others.

By comparison with the extant literature, our paper clarifies the conditions under which law could induce compliance by norm activation (and not by deterrence). Among others, Tyran and Feld

(2006) use a PG game to explain that compliance results from a "democratic dividend" phenomenon. On the contrary, our paper shows that the norm activation process may arise even when the law is exogenously imposed. In our experiment, norm activation appears to be related to the design of the liability rules. We think this result is important to analyze the role of expressive law, i.e., the normative effects of legal obligations (Galbiati and Vertova, 2008, 2014 ; Dal Bó and Dal Bó, 2014). These effects presumably depend on the *meaning* of liability rules.

A further contribution of our paper consists in analyzing the determinants of social disapproval in this setting. As expected, participants are indeed somewhat less likely to engage in harmful conduct when they have caused harm and have been informally sanctioned by other group members. In particular, we find that individuals tend to disapprove of other group members who have or may be believed to have engaged in actions generating negative externalities, irrespective of legal liability or of socially inefficient behavior. Causing harm rather than harm combined with socially inefficient behavior is the main driver of social disapproval. A possible interpretation of our findings is that strict liability, by contrast with the negligence rule, is more in tune with the underlying social norm.

Our results are potentially useful for policy makers. To consider an example, under Mild Law, we observe that strict liability yields more deterrence than the negligence rule in circumstances where causing harm is socially inefficient (i.e., the circumstances A and B in our setting). Suppose for instance that such circumstances occur 85% of the time. The possibility that strict liability over-deters in the remaining 15% of circumstances (where causing harm is socially efficient) may then be largely offset by greater deterrence in the most frequent circumstances. This observation can therefore help in identifying situations where strict liability should be preferred to simple negligence.

A variant of our analysis would be to consider the above *Liability Game* without informal sanctions, as in Eisenberg and Engel's (2014) main analysis. These authors obtain results similar to that in our paper, but in a setting closer to the canonical linear PG Game. By contrast, our framework is closer to the usual setting where real life liability rules operate, i.e., a setting with well defined initial endowments and with losses defined accordingly. Discarding informal sanctions may be viewed as allowing an evaluation of the "pure" effect of (imperfectly enforced) legal liability, i.e., independently of the effect of informal sanctions. However, one may also consider that social disapproval of misconduct is always potentially at work in the background even when formal liability is nonexistent, albeit only in a diffuse form because wrongdoers cannot be identified. An essential feature of legal liability, even when legal sanctions have little bite, is then to allow some wrongdoers to be identified. As a result, social disapproval may be more finely targeted. Admittedly, the trade-off in our experiment is then that several effects operate simultaneously. A natural extension of the present paper would therefore be to attempt to disentangle the various effects, as done in Eisenberg and Engel (2016) in the case of the unilateral accident model, but now in a setting where individuals interact as both potential victims and potential injurers and where social disapproval is endogenous.

References

1. Ambrus A., Greiner B. (2012). Imperfect Public Monitoring with Costly Punishment: An Experimental Study. American Economic Review, 102, 3317-332.

2. Ai C., Norton E. C. (2003). Interaction terms in logit and probit models. Economics letters, 80(1), 123-129.

3. Angelova V., Armentier O., Attanasi G., Hiriart Y. (2014). Relative Performance of Liability Rules: Experimental Evidence. Theory and Decision, 77, 531-556.

4. Andreoni J. (1995). Warm-Glow versus Cold Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments. Quarterly Journal of Economics 110, 1-21.

5. Becker G. (1968). Crime and Punishment: An Economic Approach. Journal of Political Economy, 76, 169-217.

6. Bellemare C., Kroeger S., Van Soest A. (2008). Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. Econometrica, 76, 815-839.

7. Bénabou R., Tirole J. (2006). Incentives and Prosocial Behavior. American Economic Review, 96, 1652-1678.

8. Bénabou R., Tirole J. (2011). Laws and Norms. NBER wp 17579.

9. Bernhard H., Fehr E., Fischbacher U. (2006a). Group Affiliation and Altruistic Norm Enforcement. American Economic Review, 96, 217–221.

10. Bernhard H., Fischbacher U., Fehr E. (2006b). Parochial Altruism in Humans. Nature, 442, 912–915.

11. Carpenter J.P., Matthews P.H. (2012). Norm enforcement: Anger, Indignation or Reciprocity. Journal of the European Economic Association, 10, 555–572.

12. Cartwright E. (2016), A Comment on Framing Effects in Linear Public Good Games. Journal of the Economic Science Association, 1-12.

13. Charness G., Rabin M. (2002) Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics, 117, 817–869.

14. Charness G., Rigotti L., Rustichini A. (2016). Social Surplus Determines Cooperation Rates in the One-Shot Prisoner's Dilemma. Games and Economic Behavior, 100, 113-124.

15. Chaudhuri A. (2011). Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature. Experimental Economics, 14, 47-83.

16. Cooper, D. J., Kagel, J. H. (2003). The impact of meaningful context on strategic play in signaling games. Journal of Economic Behavior & Organization, 50(3), 311-337.

17. Cooter R. (1998). Expressive Law and Economics. Journal of Legal Studies, 27, 585-608.ion on Cooperation. Journal of Public Economics, 117, 28-38.

18. Cooter R., Porat, A. (2000). Does risk to oneself increase the care owed to others? Law and economics in conflict. The Journal of Legal Studies, 29(1), 19-34.

19. Dal Bó E., Dal Bó P. (2014). Do the Right Thing: The Effect of Moral Suasion on Cooperation. Journal of Public Economics, 117, 28-38.

20. Deffains B., Fluet C. (2013). Legal Liability when Individuals Have Moral Concerns. Journal of Law, Economics, and Organization, 29, 930-955.

21. Dugar S. (2013). Non-monetary Incentives and Opportunistic Behavior: Evidence from a Laboratory Public Good Game. Economic Inquiry, 51, 1374-1388.

22. Egas M., Riedl A. (2008). The Economics of Altruistic Punishment and the Maintenance of Cooperation. Proceedings of the Royal Soceity B: Biological Sciences, 275(1637), 871-878.

23. Eisenberg, T., Engel, C. (2014). Assuring civil damages adequately deter: A public good experiment. Journal of Empirical Legal Studies, 11(2), 301-349.

24. Eisenberg, T., Engel, C. (2016). Unpacking Negligence Liability: Experimentally Testing the Governance Effect. Journal of Empirical Legal Studies, 13(1), 116-152.

25. Falk A., Fehr E., Fischbacher F. (2005). Driving Forces behind Informal Sanctions. Econometrica, 73, 2017-2030.

26. Falk A., Fehr E., Zehnder C. (2006). Fairness perceptions and reservation wages – The behavioral effects of minimum wage laws. Quarterly Journal of Economics, 121, 1347–1381.

27. Fehr E., Gächter S. (2000). Cooperation and Punishment in Public Goods Experiments. American Economic Review, 90, 980-994.

28. Fehr E., Gächter S. (2002). Altruistic Punishment in Humans. Nature, 415, 137-140.

29. Fehr E., Schmidt K. (1999). A Theory of Fairness, Competition and Cooperation. Quarterly Journal of Economics, 114, 817-68.

30. Fehr E., Schmidt K. (2006).The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence qnd New Theories. In Kolm S.-C. and J. Mercier Ythier (eds.), Handbook of the Economics of Giving, Elsevier, pp. 615-691.

31. Fluet C., Galbiati R. (2016). Lois et normes: les enseignements de l'économie comportementale. Actualité économique: revue d'analyse économique, 92, 191-215.

32. Fréchette, G. (2012). Session-effects in the Laboratory. Experimental Economics, 15, 485-498.

33. Funk P. (2007). Is there an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines. American Law and Economics Review, 9, 135-159.

34. Galbiati R., Vertova P. (2008). Obligations and Cooperative Behavior in Public Good Games. Games and Economics Behavior 64, 146-70.

35. Galbiati R., Vertova P. (2014). How Laws Affect Behavior: Obligations, Incentives and Cooperative Behavior. International Review of Law and Economics 38, 48-57.

36. Grechenig K., Nicklish A., Thöni C. (2010). Punishment Despite Reasonable Doubt – A Public Good Experiment with Sanctions and Uncertainty. Journal of Empirical Legal Studies 7, 847-867.

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

38. Harris D., Herrmann B., Kontoleon A., Newton J. (2015). Is it a Norm to Favour your Own Group? Experimental Economics, 18, 491-521.

39. Holt C. A., Laury S.K. (2002). Risk Aversion and Incentive Effects. American Economic Review, 92, 1644–1655.

40. Hopfensitz A., Reuben E. (2009). The Importance of Emotions for the Effectiveness of Social Punishment. Economic Journal, 119, 1534-1559.

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

42. Kube S., Traxler C. (2011). The Interaction of Legal and Social Norm Enforcement. Journal of Public Economic Theory, 13, 639-660.

43. Lönnqvist, J. E., Verkasalo, M., Walkowitz, G., Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. Journal of Economic Behavior & Organization, 119, 254-266.

44. McAdams R., Nadler J. (2005). Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/dove Game. Journal of Empirical Legal Studies, 2, 87-123.

45. McAdams R.H., Rasmusen E.B. (2007). Norms in Law and Economics. In Polinsky, A. M. and S. Shavell (eds.), Handbook of Law and Economics, Vol. 1, North-Holland.

46. Markussen T., Putterman L., Tyran J.-R. (2014). Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes. Review of Economic Studies, 81, 301-324.

47. Masclet D., Noussair C., Tucker S., Villeval M.C. (2003). Monetary and Non-Monetary Punishment in the Voluntary Contribution Mechanism. American Economic Review, 93, 366-380.

48. Mueller, P. A., Solan, L. M., Darley, J. M. (2012). When does knowledge become intent? Perceiving the minds of wrongdoers. Journal of Empirical Legal Studies, 9(4), 859-892.

49. Norton E. C., Wang H., Ai C. (2004). Computing interaction effects and standard errors in logit and probit models. Stata Journal, 4, 154-167.

50. Noussair C., Tucker S. (2005). Combining Monetary and Social Sanctions to Promote Cooperation. Economic Inquiry, 43, 649-660.

51. Nikiforakis N., Normann H.T. (2008). A Comparative Analysis of Punishment in Public Good Experiments. Experimental Economics, 11, 358-369.

52. Polinsky M.A., Shavell S. (2007). The theory of public enforcement of law. In Polinsky A. M. and S. Shavell (eds.), Handbook of Law and Economics, Vol. 1, New York: North-Holland.

53. Rege M., Telle K. (2004). The Impact of Social Approval and Framing on Cooperation in Public Good Situations. Journal of Public Economics, 88, 1625-1644.

54. Riedel N., Schildberg-Hörisch H. (2013). Asymmetric Obligations. Journal of Economic Psychology, 35, 67-80.

55. Segal U., Sobel J. (2007). Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings. Journal of Economic Theory, 136, 197-216.

56. Shavell S. (2007), Liability for Accidents. In Polinsky A. M. and S. Shavell (eds.), Handbook of Law and Economics, Vol. 1, New York: North-Holland.

57. Sullivan S.P., Holt C.A. (2017). Experimental Economics and the Law. In Parisi F. (ed.), The Oxford Handbook of Law and Economics, Oxford University Press, pp. 78-103.

58. Tyran J., Feld L. (2006). Achieving Compliance When Legal Sanctions are Non-Deterrent. Scandinavian Journal of Economics, 108, 135-156.

59. Villeval M.-C. (2013). Contributions aux biens publics et préférences sociales. Apports récents de l'économie comportementale. Revue Économique, 63, 389-420.

60. Wilkinson-Ryan, T., Baron, J. (2009). Moral judgment and moral heuristics in breach of contract. Journal of Empirical Legal Studies, 6(2), 405-423.

A Tables

Table 2: Description of the variables.

Variable Name	Description
expGains_{it}	Expected private gains of individual i at round t of undertaking *Harmful* rather than *Harmless* given the situation $k(it)$.
socialGains_{it}	Social contribution to the group welfare of individual i at round t of undertaking *Harmful* rather than *Harmless* given the situation $k(it)$.
recDisap_{it}	Number of non-monetary punishment points received by individual i at round t.
recSanctions_{it}	Number of monetary punishment points received by individual i at round t.
HF_{it}	Dummy variable equal to 1 if the individual i undertakes action *Harmful* at round t, 0 if she/he chooses *Harmless*.
givDisap_{ijt}	Number of non-monetary punishment points given by individual i to participant j at round t.
$\text{givSanctions}_{ijt}$	Number of monetary punishment points points given by individual i to participant j at round t.
historyHF_{it}	Proportion of actions *Harmful* that other group members of individual i undertook until round t.
liable_{jt}	Dummy variable equal to 1 if individual j was held liable at round t, 0 otherwise.
uncompens_{it}	Number of uncompensated accidents (other than those of individual i) for group $g(i)$ at round t.

Table 3: Descriptive Statistics

Variable	Mean	St. Dev.	Min	Max
expGains	6.505	6.759	-4	16
socialGain	.0133	3.151	-4	4
recDisap	6.262	4.615	0	18
recSanctions	.789	1.823	0	12
Harmful (HF)	.7059	.4557	0	1
givDisap	2.087	2.441	0	6
givSanctions	.263	1.032	0	6
historyHF	.696	.1839	0	1
liable	.2616	.4396	0	1
uncompens	.2629	1.032	0	6

Table 4: Descriptive statistics by treatment

Variable	SSL	SNR	MSL	MNR	NL
expGains	-0.055	5.983	6.205	8.958	11.981
	(3.130)	(9.078)	(3.153)	(6.017)	(3.199)
socialGain	-0.055	-0.0175	0.205	-0.05	-0.0194
	(3.130)	(3.181)	(3.153)	(3.097)	(3.199)
recDisap	5.63	0.075	6.993	6.628	8.028
	(4.970)	(0.264)	(4.708)	(4.232)	(4.406)
recSanctions	1.445	0.405	1.118	0.59	0.342
	(4.970)	(1.391)	(1.934)	(1.492)	(1.160)
Harmful (HF)	0.495	0.5525	0.729	0.82	0.958
	(0.500)	(0.498)	(0.445)	(0.384)	(0.200)
givDisap	1.877	1.403	2.331	2.209	2.676
	(2.429)	(2.096)	(2.520)	(2.469)	(2.488)
givSanctions	0.482	0.135	0.373	0.197	0.114
	(1.454)	(0.730)	(1.136)	(0.877)	(0.677)
historyHF	0.5229	0.566	0.679	0.805	0.929
	(0.127)	(0.116)	(0.119)	(0.09)	(1.160)
liable	0.495	0.075	0.326	0.15	0
	(.500)	(0.264)	(0.469)	(0.367)	.

Means. Standard deviations in parentheses.

Table 5: Structural Econometrics. (N=3,920)

Model	Restrictions	γ	λ_{NL}	λ_{SL}	λ_{NR}	δ	LL
(1)	$\lambda_{NL} = \lambda_{SL} = \lambda_{NR} = \delta = 0$	.299*** (.01)					-1439.1
(2)	$\lambda_{NL} = \lambda_{SL} = \lambda_{NR}; \delta = 0$	.471*** (.019)	.391*** (.022)	.391*** (.022)	.391*** (.022)		-1366.9
(3)	$\lambda_{NL} = \lambda_{SL} = \lambda_{NR}$	.46*** (.019)	.349*** (.033)	.349*** (.035)	.349*** (.035)	-.273* (.158)	-1365.4
(4-a)	$\lambda_{NL} = \lambda_{NR}$	.464*** (.019)	.319*** (.035)	.488*** (.045)	.319*** (.035)	-.141 (.045)	-1357
(4-b)	$\lambda_{NL} = \lambda_{SL}$	.46*** (.02)	.351*** (.035)	.351*** (.035)	.344*** (.044)	-.276* (.159)	-1365.4
(5)	none	.463*** (.019)	.278*** (.044)	.498*** (.046)	.374*** (.044)	-.078 (.162)	-1355

Significance level: *** significant at 1% level; ** significant at 5% level; * significant at 10% level.
LL is the log-likelihood.
For Model 4-a: $H_0 : \lambda_{NL} = \lambda_{SL}$; p<1%.
For Model 4-b: $H_0 : \lambda_{NL} = \lambda_{NR}$; p=84.7%.
For Model 5: $H_0 : \lambda_{NL} = \lambda_{SL}$; p<1%.
$H_0 : \lambda_{NL} = \lambda_{NR}$; p<1%.
$H_0 : \lambda_{SL} = \lambda_{NR}$; p<1%.

Table 6: Regression of the decision to undertake action *Harmful*.

Method	RE Logit					
Model	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
Gp	0.588***	0.552***	0.563***	0.769***	0.765***	0.599***
	(0.0897)	(0.0764)	(0.0733)	(0.169)	(0.167)	(0.0943)
G	0.237**	0.203**	0.208**	0.285**	0.296**	0.274**
	(0.105)	(0.0936)	(0.0937)	(0.129)	(0.130)	(0.116)
recDisap_{t-1}		-0.0503***				
		(0.0141)				
$\text{recDisap}_{t-1} \times HF_{t-1}$			-0.0770***			
			(0.0137)			
$\text{recDisap}_{t-1} \times HL_{t-1}$			0.0429			
			(0.0499)			
$\text{recSanctions}_{t-1}$				-0.139***		
				(0.0537)		
$\text{recSanctions}_{t-1} \times HF_{t-1}$					-0.202***	
					(0.0324)	
$\text{recSanctions}_{t-1} \times HL_{t-1}$					-0.0310	
					(0.0596)	
HF_{t-1}			0.431**		0.440***	
			(0.204)		(0.140)	
historyHF						1.557
						(0.959)
Period	0.0386***	0.0685***	0.0699***	0.143***	0.137***	0.0455***
	(0.0145)	(0.0254)	(0.0267)	(0.0445)	(0.0424)	(0.0146)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual RE	Yes	Yes	Yes	Yes	Yes	Yes
N	3920	1960	1960	1,764	1,764	3724

Significance level: *** significant at 1% level; ** significant at 5% level; * significant at 10% level.
Robust standard errors clustered at the session level.
Control variables include all variables presented in subsection 3.3.

Table 7: Regression of the decision to undertake action *Harmful* (cont'd).

Method	RE Logit					
Model	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
Gp	0.608***	0.575***	0.589***	0.783***	0.779***	0.615***
	(0.0849)	(0.0725)	(0.0688)	(0.150)	(0.149)	(0.0906)
G $\times \mathbb{1}_{NL}$	0.167**	0.145*	0.146*	0.201	0.216*	0.190*
	(0.0842)	(0.0834)	(0.0844)	(0.126)	(0.127)	(0.104)
G $\times \mathbb{1}_{NR}$	0.333***	0.307***	0.321***	0.390*	0.392**	0.349***
	(0.127)	(0.0992)	(0.0992)	(0.201)	(0.193)	(0.134)
G $\times \mathbb{1}_{SL}$	0.367***	0.326***	0.336***	0.442***	0.443***	0.379***
	(0.106)	(0.0838)	(0.0821)	(0.145)	(0.142)	(0.111)
recDisap$_{t-1}$		-0.0527***				
		(0.0124)				
recDisap$_{t-1}$ $\times HF_{t-1}$			-0.0819***			
			(0.0506)			
recDisap$_{t-1}$ $\times HL_{t-1}$			0.0467			
			(0.0527)			
recSanctions$_{t-1}$				-0.128**		
				(0.0510)		
recSanctions$_{t-1}$ $\times HF_{t-1}$					-0.184***	
					(0.0346)	
recSanctions$_{t-1}$ $\times HL_{t-1}$					-0.0322	
					(0.0634)	
HF_{t-1}			0.458**		0.411***	
			(0.203)		(0.125)	
historyHF						1.042
						(0.903)
Period	0.0389***	0.0712***	0.0728***	0.144***	0.138***	0.0464***
	(0.0145)	(0.0257)	(0.0271)	(0.0443)	(0.0423)	(0.0151)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual RE	Yes	Yes	Yes	Yes	Yes	Yes
N	3920	1960	1960	1,764	1,764	3724
$H_0 : \beta_{NL} = \beta_{NR}$ (p-val)	0.051	0.013	0.011	0.317	0.326	0.045
$H_0 : \beta_{NL} = \beta_{SL}$ (p-val)	< 0.001	<0.001	<0.001	0.031	0.029	0.002
$H_0 : \beta_{NR} = \beta_{SL}$ (p-val)	0.592	0.6555	0.750	0.723	0.707	0.600

Significance level: *** significant at 1% level; ** significant at 5% level; * significant at 10% level.
Robust standard errors clustered at the session level.
Control variables include all variables presented in subsection 3.3.

Table 8: Decision of individual i to give non-monetary or monetary punishment points to individual j at period t.

Method	GLS with RE									
	Non-monetary punishment					Monetary punishment				
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
liable_j	1.191***	1.809***	1.735***	1.733***	1.726***	0.211***	0.261***	0.238***	0.211***	0.210***
	(0.184)	(0.204)	(0.264)	(0.249)	(0.250)	(0.0483)	(0.0650)	(0.0432)	(0.0502)	(0.0508)
(1-liable_j) pHF		1.493***	1.470***	1.437***	1.452***		0.115	0.107	0.118	0.120
		(0.211)	(0.195)	(0.211)	(0.202)		(0.148)	(0.134)	(0.134)	(0.140)
pHF_AB			0.209	0.0579	0.0870			0.0614	0.123	0.127
			(0.450)	(0.394)	(0.387)			(0.202)	(0.210)	(0.223)
recDisap_{t-1}				0.0595***	0.0844***					
				(0.0124)	(0.0116)					
$\text{recDisap}_{t-1}\times$ HF_{t-1}					-0.0217					
					(0.0172)					
$\text{recSanctions}_{t-1}$									0.0322***	0.0182
									(0.00874)	(0.0119)
$\text{recSanctions}_{t-1}\times$ HF_{t-1}										0.0203
										(0.0204)
HF_{t-1}					-0.128					-0.0301
					(0.173)					(0.110)
Period	0.0536***	0.0432**	0.0429**	0.0124	0.0116	-0.00938	-0.0102	-0.0103	-0.00763	-0.00740
	(0.0193)	(0.0210)	(0.0211)	(0.0201)	(0.0196)	(0.00871)	(0.00975)	(0.00990)	(0.00991)	(0.00975)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Individual × Period) RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$H_0 : \beta_{liable_j} = \beta_{(1-liable_j)pHF}$ (p-value)		0.0793	0.271	0.150	0.171		0.167	0.286	0.455	0.497
N	5,880	5,880	5,880	5,292	5,292	5,880	5,880	5,880	5,292	5,292

Significance level: *** significant at 1% level; ** significant at 5% level; * significant at 10% level.
Robust standard errors clustered at the session level.
Control variables include all variables presented in subsection 3.3.

B Figures

Figure 1: Proportion of actions *Harmful* across treatments and situations. (Spikes representing 95% confidence intervals.)

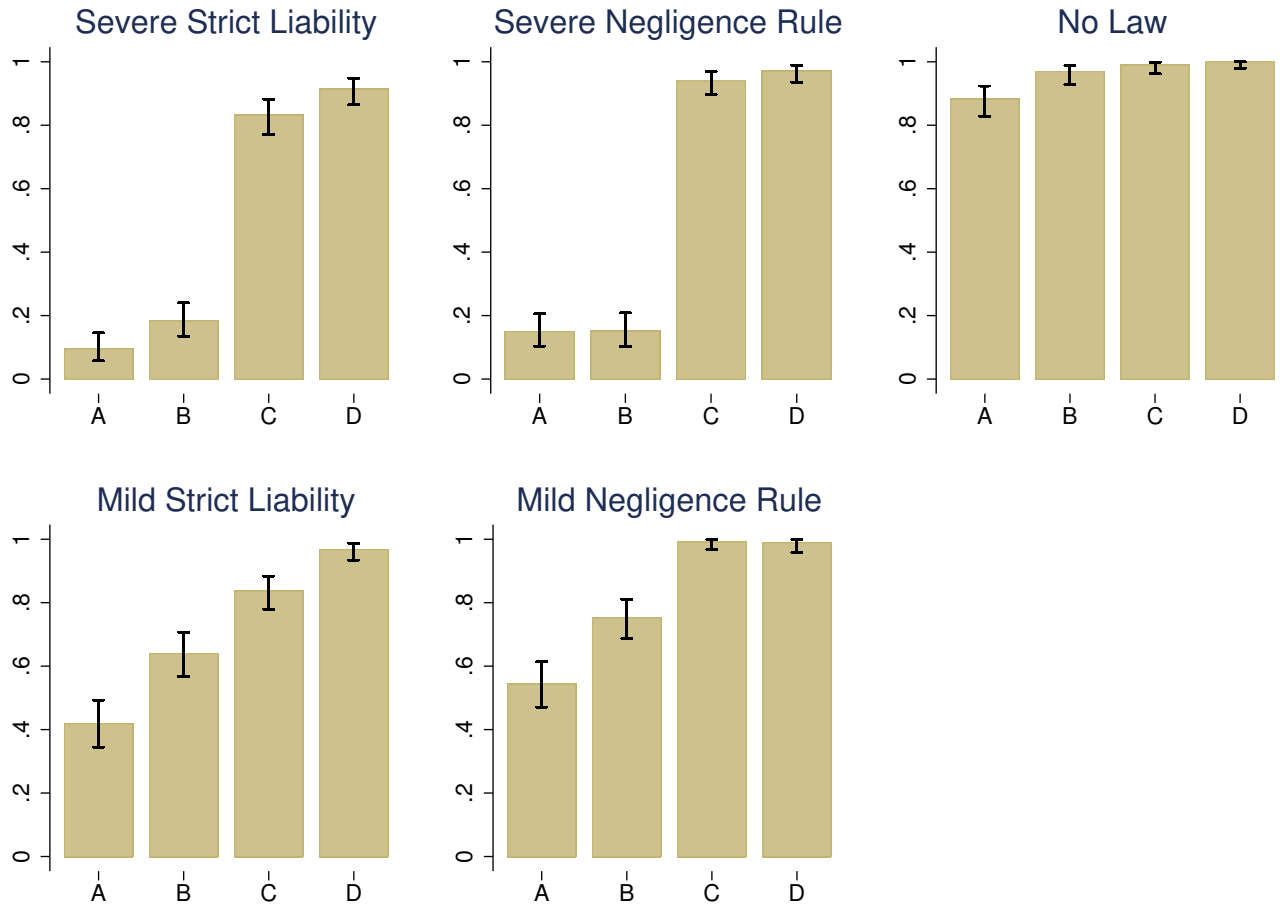


Figure 2: Prediction of behaviors after logit regression (Model 5, table 5).

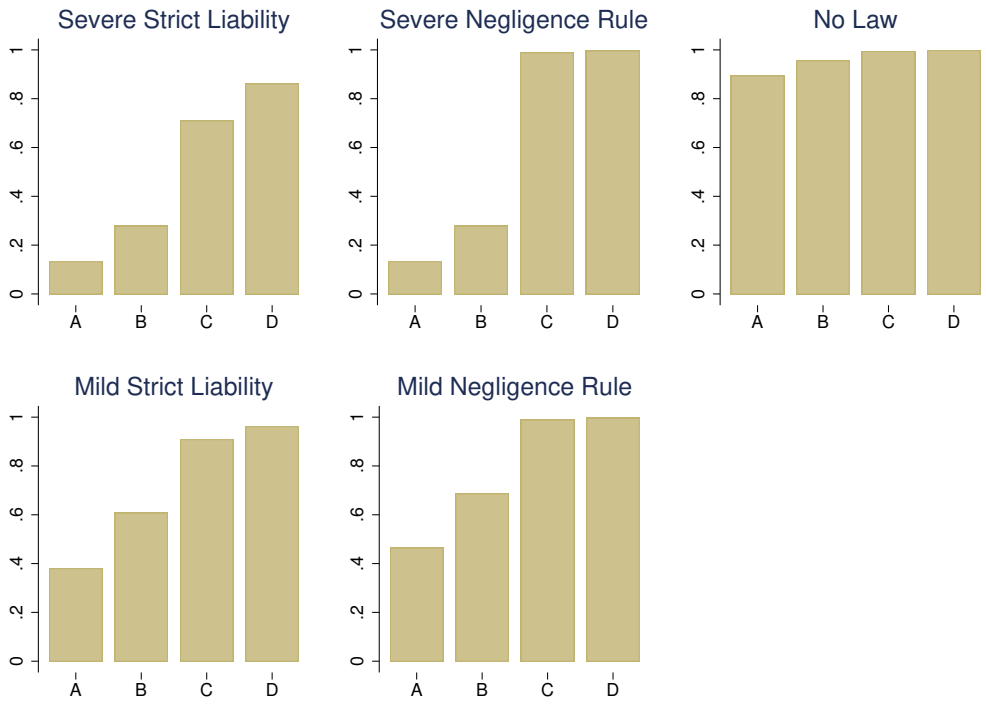


Figure 3: Average non-monetary punishment points per condemnation status and treatment.

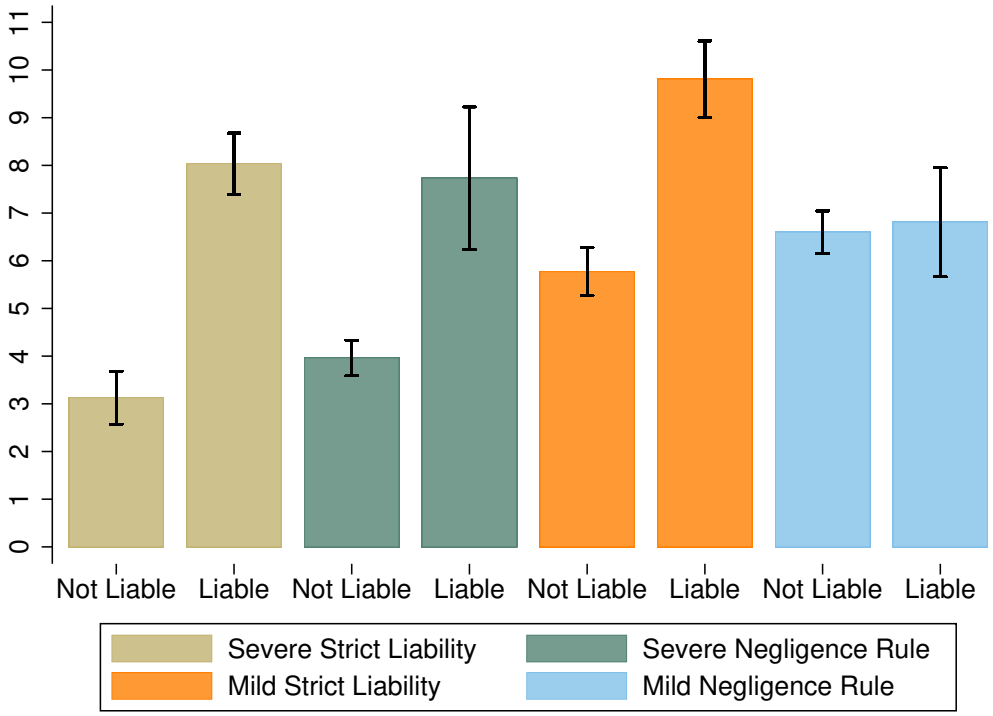
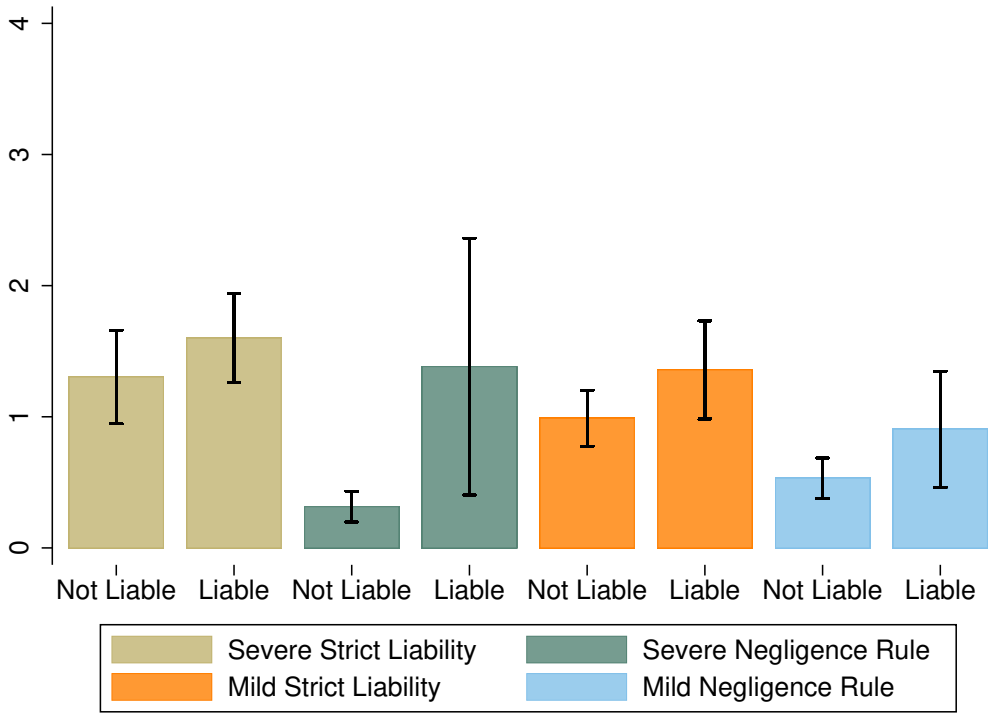


Figure 4: Average monetary punishment points per condemnation status and treatment.



Online Appendix
Translated Instructions

– Instructions 1 –

Translation: Original version in French

Welcome everyone!

You are about to take part in an experiment, and we thank you for this. This experiment consists of three parts. The present instructions deal with the first part of the experiment only. The instructions related to the second and third parts will be handed to you at the beginning of each part. If the instructions are not clear, or if a question remains unanswered, please raise your hand and wait for an instructor to come to you.

The experiment includes moments of individual decisions as well as interactions with other persons in the room. At some steps in the experiment, your decision may impact the payoff of other participants, and reciprocally. For this reason, it is strictly forbidden to communicate with other participants during the entire session. In case of violation, we will be forced to exclude you from the room.

During the experiment, you will obtain ECUs (*Experimental Currency Unit*). At the end of the experiment, the obtained ECUs will be converted into Euros. The exchange rate is 40 ECUs for 1 Euro.

The first part of the experiment is made of five small games. Note that the outcomes of each of these games will be revealed at the end of the experiment only.

– Instructions 2 –
Severe Strict Liability

Translation: Original version in French

Thank you for achieving the first part of the experiment. The second part is about to start.

In this second part, you will play a game where your decisions will affect your own payoff as well as other participants' revenue. This game will be played 10 times. For each repetition, you will be grouped with three other participants. Groups will be generated randomly at the beginning of each repetition.

At the beginning of each round, you will receive 20 ECU that will be your base revenue. Then, you will need to choose between two actions that will affect your revenue: actions Y and X. Action Y yields 6 ECU per round, whatever the situation. The revenue generated by action X depends on random conditions. At the beginning of each round, you will learn whether you are in situation A, B, C or D. For each of these situations, action X will give you 14, 16, 20 or 22 ECU respectively. The other participants of your group won't be able to observe your situation (and you won't be able to know theirs). Situations A, B, C and D have the same probability to occur (25% each).

Action Y affects only your own revenue. On the contrary, action X negatively affects the revenue of the other participants of your group: it generates a loss of 4 ECU to each of them. Thus, choosing action X will lead to a total loss of 12 ECU for the group (i.e. 4 ECU for each of the three other players). The table below summarizes the effects of the two actions:

		Situation A (25 %)	Situation B (25 %)	Situation C (25 %)	Situation D (25 %)
Your revenue	Action Y	+6	+6	+6	+6
	Action X	+14	+16	+20	+22
The revenue of each of the other players of your group	Action Y	+0	+0	+0	+0
	Action X	-4	-4	-4	-4

Once all the participants of your group, you included, have decided between action X and Y, the computer will compute all individual payoffs. Each individual who has caused damages to others will have to compensate them. Thus, if you choose action X, you will have to compensate each other player of the group by 4 ECU, whatever the situation you are in. Your final revenue will be equal to:

- your individual profit resulting from action X or Y given the situation you are in;

- minus the potential compensations you will have to pay to the other participants if you chose action X.

After this, you will have the possibility to give disapproval points to each of the other participants of your group. You will be able to give between 0 (no disapproval) to 6 (strong disapproval) points to each other player in your group. The participants of your group will be displayed in an anonymous manner, with a number-id that is randomly generated and renewed at each round. Participants who had to compensate will be showed as such to the rest of the group. After each participant assigned his/her disapproval points to the other group members, you will learn the total number of disapproval points other participants gave you. Disapproval points don't affect your revenue.

This game will be played 10 times in a row. After each iteration, participants within the session will be randomly rematched.

The steps of the game that will be repeated 10 times can be summarized as follows:

1) You are randomly matched in a group of 4 participants.
2) You decide to undertake action X or Y.
3) Gains and losses associated to actions X and Y as described above take place.
4) You learn your revenue as well as how it has been determined. You also learn which individuals compensated you.
5) You assign disapproval points to each member of your group.
6) You learn the total number of disapproval points others have given you.

– Instructions 2 –
Severe Negligence Rule

Translation: Original version in French

Thank you for achieving the first part of the experiment. The second part is about to start.

In this second part, you will play a game where your decisions will affect your own payoff as well as other participants' revenue. This game will be played 10 times. For each repetition, you will be grouped with three other participants. Groups will be generated randomly at the beginning of each repetition.

At the beginning of each round, you will receive 20 ECU that will be your base revenue. Then, you will need to choose between two actions that will affect your revenue: actions Y and X. Action Y yields 6 ECU per round, whatever the situation. The revenue generated by action X depends on random conditions. At the beginning of each round, you will learn whether you are in situation A, B, C or D. For each of these situations, action X will give you 14, 16, 20 or 22 ECU respectively. The other participants of your group won't be able to observe your situation (and you won't be able to know theirs). Situations A, B, C and D have the same probability to occur (25% each).

Action Y affects only your own revenue. On the contrary, action X negatively affects the revenue of the other participants of your group: it generates a loss of 4 ECU to each of them. Thus, choosing action X will lead to a total loss of 12 ECU for the group (i.e. 4 ECU for each of the three other players). The table below summarizes the effects of the two actions:

		Situation A (25 %)	Situation B (25 %)	Situation C (25 %)	Situation D (25 %)
Your revenue	Action Y	+6	+6	+6	+6
	Action X	+14	+16	+20	+22
The revenue of each of the other players of your group	Action Y	+0	+0	+0	+0
	Action X	-4	-4	-4	-4

Once all the participants of your group, you included, have decided between action X and Y, the computer will compute all individual payoffs. Each individual who has caused damages to others in situations A or B will have to compensate them. Participants who undertook action X in situations C or D won't compensate for the caused damages. Thus, if you choose action X in situations A or B, you will have to compensate each other player of the group by 4 ECU. Your final revenue will be equal to:

- your individual profit resulting from action X or Y given the situation you are in;

• minus the potential compensations you will have to pay to the other participants if you chose action X (if you were in situations A or B);

• minus the potential damages caused by other participants who chose action X when they were in situations C or D.

After this, you will have the possibility to give disapproval points to each of the other participants of your group. You will be able to give between 0 (no disapproval) to 6 (strong disapproval) points to each other player in your group. The participants of your group will be displayed in an anonymous manner, with a number-id that is randomly generated and renewed at each round. Participants who had to compensate will be showed as such to the rest of the group. After each participant assigned his/her disapproval points to the other group members, you will learn the total number of disapproval points other participants gave you. Disapproval points don't affect your revenue.

This game will be played 10 times in a row. After each iteration, participants within the session will be randomly rematched.

The steps of the game that will be repeated 10 times can be summarized as follows:

1) You are randomly matched in a group of 4 participants.
2) You decide to undertake action X or Y.
3) Gains and losses associated to actions X and Y as described above take place.
4) You learn your revenue as well as how it has been determined. You also learn which individuals compensated you.
5) You assign disapproval points to each member of your group.
6) You learn the total number of disapproval points others have given you.

– Instructions 2 –
Mild Strict Liability

Translation: Original version in French

Thank you for achieving the first part of the experiment. The second part is about to start.

In this second part, you will play a game where your decisions will affect your own payoff as well as other participants' revenue. This game will be played 10 times. For each repetition, you will be grouped with three other participants. Groups will be generated randomly at the beginning of each repetition.

At the beginning of each round, you will receive 20 ECU that will be your base revenue. Then, you will need to choose between two actions that will affect your revenue: actions Y and X. Action Y yields 6 ECU per round, whatever the situation. The revenue generated by action X depends on random conditions. At the beginning of each round, you will learn whether you are in situation A, B, C or D. For each of these situations, action X will give you 14, 16, 20 or 22 ECU respectively. The other participants of your group won't be able to observe your situation (and you won't be able to know theirs). Situations A, B, C and D have the same probability to occur (25% each).

Action Y affects only your own revenue. On the contrary, action X negatively affects the revenue of the other participants of your group: it generates a loss of 4 ECU to each of them. Thus, choosing action X will lead to a total loss of 12 ECU for the group (i.e. 4 ECU for each of the three other players). The table below summarizes the effects of the two actions:

		Situation A (25 %)	Situation B (25 %)	Situation C (25 %)	Situation D (25 %)
Your revenue	Action Y	+6	+6	+6	+6
	Action X	+14	+16	+20	+22
The revenue of each of the other players of your group	Action Y	+0	+0	+0	+0
	Action X	-4	-4	-4	-4

Once all the participants of your group, you included, have decided between action X and Y, the computer will compute all individual payoffs. Each individual who has caused damages to others will have to compensate them if he/she is detected by the computer (which occurs with a probability of 50%), whatever the situation you were in. Thus, if you choose action X, you will have to compensate each other player of the group by 4 ECU, whatever the situation you are in. Your final revenue will be equal to:

- your individual profit resulting from action X or Y given the situation you are in;

• minus the potential compensations you will have to pay to the other participants if you chose action X (50% probability);

• minus the potential uncompensated damages made by other participants (50% probability if they chose action X).

After this, you will have the possibility to give disapproval points to each of the other participants of your group. You will be able to give between 0 (no disapproval) to 6 (strong disapproval) points to each other player in your group. The participants of your group will be displayed in an anonymous manner, with a number-id that is randomly generated and renewed at each round. Participants who had to compensate will be showed as such to the rest of the group. After each participant assigned his/her disapproval points to the other group members, you will learn the total number of disapproval points other participants gave you. Disapproval points don't affect your revenue.

This game will be played 10 times in a row. After each iteration, participants within the session will be randomly rematched.

The steps of the game that will be repeated 10 times can be summarized as follows:

1) You are randomly matched in a group of 4 participants.
2) You decide to undertake action X or Y.
3) Gains and losses associated to actions X and Y as described above take place.
4) You learn your revenue as well as how it has been determined. You also learn which individuals have compensated you.
5) You assign disapproval points to each member of your group.
6) You learn the total number of disapproval points others have given you.

– Instructions 2 –
Mild Negligence Rule

Translation: Original version in French

Thank you for achieving the first part of the experiment. The second part is about to start.

In this second part, you will play a game where your decisions will affect your own payoff as well as other participants' revenue. This game will be played 10 times. For each repetition, you will be grouped with three other participants. Groups will be generated randomly at the beginning of each repetition.

At the beginning of each round, you will receive 20 ECU that will be your base revenue. Then, you will need to choose between two actions that will affect your revenue: actions Y and X. Action Y yields 6 ECU per round, whatever the situation. The revenue generated by action X depends on random conditions. At the beginning of each round, you will learn whether you are in situation A, B, C or D. For each of these situations, action X will give you 14, 16, 20 or 22 ECU respectively. The other participants of your group won't be able to observe your situation (and you won't be able to know theirs). Situations A, B, C and D have the same probability to occur (25% each).

Action Y affects only your own revenue. On the contrary, action X negatively affects the revenue of the other participants of your group: it generates a loss of 4 ECU to each of them. Thus, choosing action X will lead to a total loss of 12 ECU for the group (i.e. 4 ECU for each of the three other players). The table below summarizes the effects of the two actions:

		Situation A (25 %)	Situation B (25 %)	Situation C (25 %)	Situation D (25 %)
Your revenue	Action Y	+6	+6	+6	+6
	Action X	+14	+16	+20	+22
The revenue of each of the other players of your group	Action Y	+0	+0	+0	+0
	Action X	-4	-4	-4	-4

Once all the participants of your group, you included, have decided between action X and Y, the computer will compute all individual payoffs. Each individual who has caused damages to others in situations A or B will have to compensate them, if he/she is detected by the computer (which occurs with a probability of 50%). Participants who undertook action X in situations C or D won't compensate for the caused damages. Thus, if you choose action X in situations A or B, you will have to compensate each other player of the group by 4 ECU with a probability of 50%. Your final revenue will be equal to:

- your individual profit resulting from action X or Y given the situation you are in;

• minus the potential compensations you will have to pay to the other participants if you chose action X (if you were in situations A or B) and you were detected (probability of 50%);

• minus the potential damages caused by other participants who chose action X when they were in situations C or D, or when they were in situation A or B but were not detected.

After this, you will have the possibility to give disapproval points to each of the other participants of your group. You will be able to give between 0 (no disapproval) to 6 (strong disapproval) points to each other player in your group. The participants of your group will be displayed in an anonymous manner, with a number-id that is randomly generated and renewed at each round. Participants who had to compensate will be showed as such to the rest of the group. After each participant assigned his/her disapproval points to the other group members, you will learn the total number of disapproval points other participants gave you. Disapproval points don't affect your revenue.

This game will be played 10 times in a row. After each iteration, participants within the session will be randomly rematched.

The steps of the game that will be repeated 10 times can be summarized as follows:

1) You are randomly matched in a group of 4 participants.
2) You decide to undertake action X or Y.
3) Gains and losses associated to actions X and Y as described above take place.
4) You learn your revenue as well as how it has been determined. You also learn which individuals compensated you.
5) You assign disapproval points to each member of your group.
6) You learn the total number of disapproval points others have given you.

– Instructions 2 –
No Law

Translation: Original version in French

Thank you for achieving the first part of the experiment. The second part is about to start.

In this second part, you will play a game where your decisions will affect your own payoff as well as other participants' revenue. This game will be played 10 times. For each repetition, you will be grouped with three other participants. Groups will be generated randomly at the beginning of each repetition.

At the beginning of each round, you will receive 20 ECU that will be your base revenue. Then, you will need to choose between two actions that will affect your revenue: actions Y and X. Action Y yields 6 ECU per round, whatever the situation. The revenue generated by action X depends on random conditions. At the beginning of each round, you will learn whether you are in situation A, B, C or D. For each of these situations, action X will give you 14, 16, 20 or 22 ECU respectively. The other participants of your group won't be able to observe your situation (and you won't be able to know theirs). Situations A, B, C and D have the same probability to occur (25% each).

Action Y affects only your own revenue. On the contrary, action X negatively affects the revenue of the other participants of your group: it generates a loss of 4 ECU to each of them. Thus, choosing action X will lead to a total loss of 12 ECU for the group (i.e. 4 ECU for each of the three other players). The table below summarizes the effects of the two actions:

		Situation A (25 %)	Situation B (25 %)	Situation C (25 %)	Situation D (25 %)
Your revenue	Action Y	+6	+6	+6	+6
	Action X	+14	+16	+20	+22
The revenue of each of the other players of your group	Action Y	+0	+0	+0	+0
	Action X	-4	-4	-4	-4

Once all the participants of your group, you included, have decided between action X and Y, the computer will compute all individual payoffs. Your revue will be equal to your individual payoff resulting from action X and Y given the situation you were in, and the potential damages caused by other participants.

After this, you will have the possibility to give disapproval points to each of the other participants of your group. You will be able to give between 0 (no disapproval) to 6 (strong disapproval) points to each other player in your group. The participants of your group will be displayed in an anonymous manner, with a number-id that is randomly generated and renewed at each round. After each participant assigned his/her

disapproval points to the other group members, you will learn the total number of disapproval points other participants gave you. Disapproval points don't affect your revenue.

This game will be played 10 times in a row. After each iteration, participants within the session will be randomly rematched.

The steps of the game that will be repeated 10 times can be summarized as follows:

1) You are randomly matched in a group of 4 participants.
2) You decide to undertake action X or Y.
3) Gains and losses associated to actions X and Y as described above take place.
4) You learn your revenue as well as how it has been determined.
5) You assign disapproval points to each member of your group.
6) You learn the total number of disapproval points others have given you.

- Instructions 3 –

Translation: Original version in French

Thank you for having completed the second part of the experiment. The third part is about to start.

In this third part of the experiment, you will be asked to play the same game as previously with a modification of the rules. Indeed, the disapproval points described previously will be replaced by sanction points. A sanction point costs you 0.5 ECU and decreases the targeted person's revenue by 0.5 ECU.

The game takes place in the same way as before, with the only exception of sanction points. This new game will be repeated 10 times with random rematching at each repetition.