Is Compensation Fine?
Sanction Schemes and their Effects on Deterrence and Trust

Pieter Desmet* Leonie Gerhards† Franziska Weber‡

November, 2022

Abstract

Both fine and compensation payments are commonly used to sanction misbehaviour. Interestingly, they are typically not consistently applied across different jurisdictions and their comparative strengths and weaknesses are empirically not yet well established and understood. Our experiment allows us to, on the one hand, contrast the compliance-inducing effects of fines and compensation on potential infringers. On the other hand, it enables us to examine their respective capabilities to maintain or restore potential victims' trust. We find that fines induce more compliance than compensation. Yet, there is a mismatch with the counterpart's trust: compared to a situation without sanctions, potential victims' trust is higher if there is a fine regime in place. Nonetheless, trust is no longer shaped by the sanctions scheme per se when victims have first-hand experienced misbehaviour.

Keywords: Deterrence, Trust, Compensation, Fine, Experiment
JEL Codes: C91, D02, K42.

*Rotterdam Institute of Law and Economics/Erasmus School of Law, Erasmus University Rotterdam; Burg. Oudlaan 50, 3062 PA Rotterdam. Email: desmet@law.eur.nl

†King's Business School, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, UK. Email: leonie.gerhards@kcl.ac.uk.

‡Rotterdam Institute of Law and Economics/Erasmus School of Law, Erasmus University Rotterdam; Burg. Oudlaan 50, 3062 PA Rotterdam. Email: weber@law.eur.nl.

1 Introduction

As an effective legal system is the cornerstone of a viable society, it is imperative to identify and put in place the sanction scheme that best induces compliant behaviour and ensures that people can trust others not to betray them. Both fines and compensation are widely used sanction schemes, but even for the same infringements, they are applied inconsistently across jurisdictions. Unfair commercial practices law is a case in point. It regulates the way in which traders can market their products towards consumers[1] and serves to offset the information asymmetry that exists between trader and consumer with a view to the quality of the products offered. Different countries traditionally used different sanctions against the very same unfair practices. In Spain, for instance, since 2010 consumers can make individual claims for compensation.[2] In Italy on the other hand, the provisions are traditionally enforced by way of public fines.[3]

Given the fundamental role of sanctions for society, it is striking that there are so little empirical insights on the relative effects of fines versus compensation as they are rarely assessed against each other. Also prior research on sanctions has primarily focused on infringers. Yet, from a social perspective it is equally important to understand to what extent sanction schemes make potential victims willing to be vulnerable to the actions of others who may potentially betray them (i.e. trust Rousseau et al., 1998).

To fill these gaps, we design a novel laboratory experiment that allows us to simultaneously evaluate the relative effects of fine and compensation schemes on potential infringers' misbehaviour as well as on their counterparts' trust. We opted for the controlled environment of an experiment which allows us to record in particular infringers' misbehaviour to an extent that would be infeasible with field data.

We set out to compare fine and compensation schemes on a potential victim's and a potential infringer's behaviour in a simple two-player set-up with perfect stranger matching. We operationalise misbehaviour as lying about a state of the world to make a profit at the expense of the other player (cf. Agranov and Buyalskaya, 2022). Analogously, we interpret the potential victims' propensity to act upon potential lies as the degree of trust in their counterpart. Lying is detected and sanctioned with a pre-defined probability that is common knowledge to both parties. For a clean comparison of their respective effects, across treatments Fine and Comp,

[1] It is, for instance, forbidden to share untruthful information or employ aggressive marketing techniques. Consumers may end up buying a product due to unfair marketing practices which they would otherwise not have bought.

[2] See Art. 33 (5) and 33 (1) of the Spanish Act on Unfair competition as introduced by Law 29/2009.

[3] See Part II Art. 27 of Codice del Consumo as introduced by legislative decree No. 146 of 2007.

we vary the type of sanction (fines vs. compensation), but keep the size of the sanction payment the same. We also include a control treatment without any sanction regime (No S).

The experiment consists of three parts. In part 1 we test for ex-ante compliance and trust effects by observing participants in a one-shot game. Part 2 enables us to observe learning effects after having experienced the treatment-specific sanction scheme. On the one hand, we study infringers' future compliance after having experienced the law by ways of having been checked for or actually been caught lying. On the other hand, we analyse which sanction scheme is more successful in maintaining and/or restoring trust after a victim's counterpart was checked for or caught lying. In part 3, sanctions are lifted in order to test for lasting effects of the previous sanction schemes.

In the first two parts of the experiment, we observe least lying in Fine, comparably more lying in Comp and most lying in No S. These treatment differences are significant in particular in the repeated setting of part 2. Interestingly, even when sanctions are lifted, we observe less lying in Fine and Comp than in No S in part 3 of the experiment, pointing towards a sustained compliance effect of both sanction schemes.

Somewhat in line with these results regarding infringer behaviour, trust is significantly higher in Fine than in No S in part 2 of the experiment. Apart from that we do not observe any other significant treatment difference in trust in any of the three parts. This suggests that trust on the potential victims' side is not primarily determined by the type of sanction scheme. Rather, previous (bad) experiences with an infringer seem to shape victims' future trust.

The paper proceeds as follows. We first review existing research on the effects of different sanction regimes on deterrence by way in which they induce compliance. After that, we review existing research on the effects of sanctions on trust and introduce our experiment and finally, discuss its results and implications.

2 Related literature

2.1 Sanctions and infringers – Deterrence

The literature on sanctions so far has been rather infringer-centred and the focus has primarily been on the ex ante perspective, that is, the extent to which sanctions deter potential infringers from becoming infringers in the first place (see, for instance, Stigler, 1970, Garoupa, 2001, Cooter, 1988 or Andreoni, 1991). The classic economic model of deterrence by Becker (1968) considers both the size of the sanction and the probability of detection and conviction,

as paramount to incentivise deterrence. According to deterrence theorists, the interplay between substantive laws and their enforcement forms the incentives and, therefore, the deterrent backbone of compliance (Veljanovski, 1984). Deterrence theory assumes that if the expected benefits of violating the law are outweighed by the expected costs (mainly determined by the size of the sanction and the probability of detection) the individual will comply with rather than violate the law.

The effects of deterrence theory could often be corroborated, see for instance Engel's (2018) review of empirical and experimental research in criminal law or Slemrod's (2016) survey of the literature on tax compliance. However, the matter can be more complex. In a theoretical model, Dari-Mattiacci and Raskolnikov (2021) extend the basic deterrence model by relaxing some of the original assumptions and discuss contexts in which compliance is not necessarily increasing in expected sanctions as well as situations in which equally sized rewards and punishments do not produce the same incentive effects. Gneezy and Rustichini (2000) show in their well-known field experiment how a newly introduced sanction can even lead to an increase in violations, when the fine payment is perceived as a price. Schildberg-Hörisch and Strassmair (2012) test deterrence theory in a laboratory experiment. Only in the case of very high incentives, they do find support for the conjecture that crime (weakly) decreases in deterrent incentives. The authors argue that deterrence incentives can crowd out intrinsic motivation to act pro-socially. In a comparable experimental setup, Khadjavi (2015) corroborates Schildberg-Hörisch and Strassmair's (2012) explanation. Furthermore, he is able to link this type of crowding out to potential infringers' emotional state when they take decisions. Agranov and Buyalskaya's (2022) laboratory experiment reveals that sanction schemes that communicate only partial information (the minimum fine in particular) are more effective at increasing compliance than full information schemes.

Arguably, due to its simplicity and its associated methodological advantages, fine schemes have been subject of many studies. Contrarily, compensation payments as a sanction have received comparably less attention (for an exception with hypothetical vignettes, see Cardi et al., 2012).

Even fewer studies compare fines and sanctions: Only recently, scholars have begun to argue that sanctions of equal monetary value can be more or less effective depending on how they are framed (Mulder, 2018). Kurz et al. (2014), for example, study the effects of identical sanctions, framed as either retributive or compensatory, on the occurrence of late-coming to

a lab experiment. They observe that participants are more punctual when the sanction is framed retributively rather than compensatory, suggesting that compensation schemes may be less deterring than fine schemes. However, in their study both retributive and compensatory sanctions have the same beneficiary (the experimenter), which is irreconcilable with the crucial difference between compensation and fines.

Other researchers go beyond mere framing effects and in fact study the effects of whether or not victims are the beneficiary of sanctions, which is the core characteristic of compensation. Eisenberg and Engel (2014) compare the effects of three different types of damages and also include a treatment where the player who enacts the sanction can decide to forfeit some or all of the infringer's period income, with no benefit to themselves – in essence a fine. However, in their experiment this forfeiture is merely introduced as an option for the punisher, meant to signal their intentions. Adding the forfeit option does not make a difference in terms of deterrence and moreover, the option is also only rarely chosen.In a recent vignette study, Desmet and Weber (2022) observe that infringers' willingness to pay is higher under compensation than under a fine scheme and that infringers are similarly willing to take precautionary measures under both sanction schemes. However, their study operationalises infringements as unintentional acts and does not focus on deterrence in particular. Baumann et al. (2020) conduct an experiment to compare the effects of fines and compensation on people's investment in accident prevention. They observe that potential injurers invest substantially more money in accident prevention when they are subject to compensation instead of a fine. However, in their study too, harm is not intentionally inflicted by the infringers themselves.

We close by noting that, as highlighted by Baumann et al. (2020), an important aspect to consider in the comparison of fines and compensation schemes is their interplay with potential infringers' guilt. Guilt aversion considers in how far people care about others' expectations and anticipate feelings of guilt if they fall short of such expectations (cf. Charness and Dufwenberg, 2006 and Battigalli and Dufwenberg, 2007). A defining difference between compensation and fines in this respect is that infringers have the possibility to ex-post "morally cleanse" themselves in the first, but not in the latter type of sanction scheme. In compensation schemes, infringers' guilt can be reduced if they are convicted to compensate the victim. Putting it the other way round, potential infringers have an additional guilt aversion incentive to comply in a fine scheme which is absent under compensation.

2.2 Sanctions and victims – Maintaining trust and trust repair

While it is essential to know to what extent different sanctions have the potential to induce compliance among potential infringers, from a societal point of view an equally important question is under what sanction regime people are more willing to make themselves vulnerable to the actions of others who may potentially betray them (that is, display trust, Rousseau et al., 1998).

Two conditions must exist for trust to arise: interdependence and risk. Interdependence refers to the reliance on another to achieve one's interests; Risk entails the probability of loss (Rousseau et al., 1998). The perceived risks involved in trust decisions are therefore interpersonal risks and depend on the assessment of the intentions or behaviour of the person to trust.

The presence of a sanction system can reduce risk by decreasing the probability of loss. Yet not all sanctions may achieve this in the same way. The introduction of a compensation regime directly affects the probability of loss for victims by creating a possibility to recoup some of the losses and increasing the expected payoffs in case of betrayal. Compensation in this sense functions as an insurance mechanism that (potentially) safeguards payoffs. Fines on the other hand are a punitive response aimed primarily at infringers. Because of that, the presence of fines can only indirectly signal to potential victims that a loss is less likely to occur. How the decision to trust someone is affected by the presence of fines therefore only depends on a subjective appraisal of the other's reaction to the presence of fines, whereas under compensation regimes, potential victims also have a more certain reduction in the probability of loss, irrespective of the perceived deterrent capacity of compensation.

Trust that mainly depends on the appraisal of the deterrents that a potential infringer faces, is referred to in the literature as deterrence-based trust (Lewicki et al., 1996). According to this view, the introduction of a sanction regime will only increase potential victims' willingness to be vulnerable to the actions of others to the extent that the introduced sanctions are seen as deterring. If potential victims were to assume that infringers behave consistent with the classic deterrence framework, they will view infringers as being mainly deterred by the size and probability of the sanction. Victims would as a result prima facie be just as willing to trust their interaction partner under a compensation vs. a fine regime.

If we look at the existing literature on sanctions and trust, some critical gaps become clear. First of all, many studies that looked at the effects of sanctions on trust have taken an ex-post

perspective, focusing on actual victims' reactions to receiving compensation or to seeing an infringer being punished, rather than on the ex ante tendency to make themselves vulnerable under different sanction systems (see e.g. Bottom et al., 2002; Desmet et al., 2010; Desmet et al., 2011). Moreover, those studies only looked at the repair of trust and cooperation within the same relation. That is, they studied the decline and restoration of trust between the same interaction partners, ignoring the one-shot nature of many interactions where betrayal occurs and ignoring the spill-over effects that betrayal and sanctions may have on trust in subsequent interactions with new interaction partners.

Also, similar as with the literature on sanctions and infringers, studies that did look into ex-ante trust have not directly compared the relative effects of compensation and fines on potential victims' trust. Vollan (2011) observe the effects of (potential) third party punishment on ex ante trust using one-shot trust games for people interacting with strangers and find that they increase trust significantly. Malhotra and Murnighan (2002) look at how contracts that guarantee a certain payoff for trustors affect initial trust and trust building between two players in a trust game. They observe that the certainty of receiving a guaranteed pay-off increases potential victims' trust, supporting the idea that reducing the risk of receiving lower payoffs (by e.g. a compensation regime) will increase trust. However, these authors do not exactly study the presence of a compensation system but rather look at the effects of automatic contract enforcement where the probability of enforcement is 100%. Also focusing on trust in contractual relations, Bohnet et al. (2001) studied the behaviour of first movers who have to decide whether they want to enter a contract without knowing whether the second mover will perform. The authors observe that the contractual stipulation of damages in case of breach can stimulate trust. Using a one-shot trust game, Bohnet and Baytelman (2007) observe that the option to punish untrustworthy behaviour induces potential victims to trust more. All of the above studies, however, do not directly compare compensation with fines and do not consider a setting with repeated decision making, which allows to study how the experience of wrongdoing affects future trust.

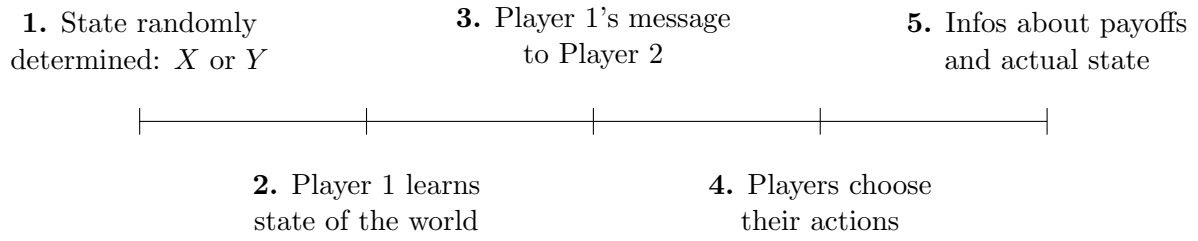
3 The experiment

3.1 Design

We opted for the simplest design that allows us to simultaneously study infringers' and victims' behaviour. Always two players are matched to take decisions in their roles of the potential

infringer and the potential victim. In the neutrally-framed instructions, we refer to them as Player 1 and Player 2, respectively. The experimental game evolves in five stages, summarised in Figure 1.

Figure 1: The experimental game



In the first stage, the state of the world is randomly determined to be either X or Y with equal probability and both players know this. Next, only Player 1 learns the prevailing state of the world. In the third stage, Player 1 chooses to send one of two standardised messages to Player 2. They can either send "State X prevails" or "State Y prevails". Empty messages are ruled out by design. It is common knowledge that Player 1's knows the prevailing state of the world and that their message does not have to be truthful. In the fourth stage, both players choose their actions. Player 1 chooses between actions A and B, while Player 2 decides between actions C and D.

Table 1 summarises the monetary payoffs (in points[4]) players obtain given the actions chosen and the prevailing state of the world. This table is similarly presented and carefully explained in the instructions so that all the following is common knowledge: In state X, Player 1 always prefers A, irrespective of Player 2's choice, while Player 2 always prefers C, resulting in action profile (A,C). Similarly, in state Y, Player 1 always prefers B, while Player 2 always prefers D, resulting in (B,D). However importantly, in state Y, action profile (B,C) would yield a higher payoff to Player 1. Player 1 thus has an incentive to lie about the state of the world in state Y to make Player 2 choose C instead of D. In the instructions, reproduced in Appendix B, we did not openly encourage participants to lie. However, in the control questions, we asked participants (independent of role) to calculate both players' payoffs in action profiles (B,D) and (B,C) in state Y, thereby making explicit Player 1's incentive to lie. Player 1 has no incentive to lie in state X.[5] When studying treatment differences in lying, we therefore focus on state Y.

[4]In the experiment, one point corresponds to 0.40 Euro.

[5]Lying in state X would drive Player 2s to choose action D and hence decrease Player 1's payoffs relative to a situation in which Player 1 had told the truth. Compare Player 1's payoffs in action profiles (A,C) and (A,D) in state X, 20 vs. 10.

Analogously, we analyse treatment differences in Player 2s' trust by considering their propensity to choose action C when their matched Player 1 reported that state X prevails.

Table 1: Player 1's and Player 2's payoffs

Monetary payoffs in state X:

		Player 2	
		C	D
Player 1	A	20,20	10,10
	B	10,10	0,0

Monetary payoffs in state Y:

		Player 2	
		C	D
Player 1	A	10,0	0,10
	B	20,10	10,20

Notes: Payoffs denoted in points.

In the final stage, payoffs are realised and both players are informed about the actual state of the world. Hence, irrespective of treatment, every Player 2 finally learns if they have been lied to. Depending on treatment, both players are moreover informed whether Player 1 is sanctioned for lying.

In treatment No S (short for "No Sanction") Player 1 is never sanctioned for lying. In treatments Fine and Comp (short for "Compensation"), conversely, a third of Player 1s' messages are randomly checked for correctness. If caught lying, Player 1 is sanctioned. In Fine, 10 points are deducted from the infringer's earnings. Effectively, the money goes back to the experimenter. In Comp, similarly, 10 points are deducted from the infringer's earnings. However, in this case the amount is transferred to Player 2. That is, across Fine and Comp, we only vary the type of sanction, not its size.

In both treatments, the 10 point sanction payment correspond to the victim's loss that results from reaching (B,C) instead of (B,D) in state Y.[6] Given the experimental parameters, the sanction is non-deterrent in expectations (similar to what we see in reality like not buying a parking or bus ticket or speeding). A lie costs the infringer 3.33 points in Fine and Comp. We chose this size of sanction payment to, on the one hand, increase the number of lies that we could study in the experiment, while, on the other hand, taking into account the potential victims' desire to receive at least some meaningful expected compensation in Comp when having been lied to.

[6]Lying Player 1s are sanctioned irrespective of whether they actually harm the matched Player 2. That means, in case Player 2 chooses D, not C in state Y, they do not incur a loss of 10 points, but a lying Player 1 is sanctioned nevertheless.

3.2 Procedures

The experiment comprises three parts. The treatment specific instructions for part 1 are distributed at the beginning of the experiment. In this part we implement a one-shot version of the above described experimental game. All participants have to answer a series of control questions to ensure that everyone understands the rules of the game. Only thereafter, the computer randomly assigns participants the role of Player 1 or Player 2. Participants stay in their randomly allocated role for the entire duration of the experiment.

The instructions for part 2 and 3 are distributed only at the beginning of the respective parts. In part 2, participants play the same treatment-specific experimental game as in part 1 for another four times. Player 1s and Player 2s are randomly re-matched in every round in a perfect stranger matching fashion. Thus, while part 1 allows us to observe participants' decisions in a true one-shot game, part 2 enables us to study potential learning effects, while reputational effects are ruled out by design. By analysing behaviour of Player 1s who have been caught lying in previous rounds (in part 1 or 2), part 2 permits to test for any additional effects of having experienced the law on future compliance. Similarly, we can study which type of sanction is more successful in restoring trust, after a victim's previously matched counterpart was caught lying. Lastly, in part 3 participants play one round of treatment No S with a randomly selected new matching partner (again, a "perfect stranger"). This final round allows us to test for lasting effects of the previous sanction schemes of Comp and Fine.

The experiment was conducted at the University of Hamburg between Winter 2018 and Spring 2019. The sessions lasted about 60 minutes which included reading the instructions, taking decisions in the three parts of the experiment, a brief computerised survey on socio-economic information and personal characteristics and the payment of participants at the end. In every session, 24 participants took part, half of them in either role (Player 1 or Player 2). We recorded behaviour of 96 participants in No S and further 144 participants each in Fine and Comp. Due to the implemented stranger matching design, these individual participant observations are organised in 32 matching groups of 12 (that is, each matching group contains 6 participants in either role). On average, participants earned 11.47 Euro. In No S 52% of participants were female, in Fine 51% and in Comp 52%.

The vast majority of participants managed to answer the control questions at the beginning of the experiment correctly without further assistance. To further check for participants' understanding of the experimental game, we consider some key decisions they took. Firstly, 99%

of Player 1s choose their payoff maximising, strictly dominant action A in state X; 98% do so (that is, they choose B) in state Y. Secondly, in state X, 99% of Player 1s report the state truthfully. Thirdly, 93% of Player 2 choose their payoff maximising action D if their matched Player 1 reports that state Y prevails.[7] We hence conclude that the overwhelming part of participants understands the rules of the game.[8]

3.3 Predictions

In the following, we focus our attention to the treatment differences in Player 1s' compliance and Player 2s' trust that emerge from the treatment-specific sanction schemes.

Given the chosen monetary incentives and the one-shot character of the game, rational Player 1s should always and independent of treatment choose their state-specific payoff-maximising, dominant action (that is, A in state X and B in state Y). Moreover, in all treatments, they have an incentive to lie about the state of the world when state Y prevails, in order to increase the chances of Player 2 choosing C. The implemented sanctions are non-deterrent in expectations: a lie can yield an additional 10 points if successful – and costs an infringer at most 3.33 points in expectations (3.33 points in FINE and COMP and nothing at all in NO S). Player 1s have no incentive to lie in state X.

Therefore, rational Player 2s always act upon Player 1s' messages if the latter report state Y and choose their dominant action D. If, on the other hand, Player 1s report state X, Player 2s cannot rely on the message and remain uncertain about the state of the world.

If Player 2s always play C, they can secure an expected payoff $15 (= \frac{1}{2} \times 20 + \frac{1}{2} \times 10$ in NO S and FINE, $16\frac{2}{3} = \frac{1}{2} \times 20 + \frac{1}{2} \times 13\frac{1}{3}$ in COMP). The same is true if Player 2s always play D or if they mix, i.e. play C with probability $p_C \in [0, 1]$.[9]

Ample evidence shows that many individuals do not act in a purely selfish manner, but

[7] Note that Player 2s who choose C instead of D in this situation do not necessarily behave irrationally. They might also intentionally reward their matched Player 1s for telling the truth.

[8] The findings remain essentially unchanged if we restrict our sample to participants who choose the payoff maximising action and report state X truthfully. These results are available from the authors upon request.

[9] If Player 2s always play D, in NO S and FINE: $15 = \frac{1}{2} \times 10 + \frac{1}{2} \times 20$, in COMP: $16\frac{2}{3} = \frac{1}{2} \times 10 + \frac{1}{2} \times 23\frac{1}{3}$. If Player 2s mix, i.e. play C with probability $p_C \in [0, 1]$, in NO S and FINE: $15 = \frac{1}{2}(20p_C + 10(1 - p_C)) + \frac{1}{2}(10p_C + 20(1 - p_C))$, in COMP: $16\frac{2}{3} = \frac{1}{2}(20p_C + 10(1 - p_C)) + \frac{1}{2}(13\frac{1}{3}p_C + 23\frac{1}{3}(1 - p_C))$. As is always the case with these type of games, there exists many Perfect Bayesian Equilibria. However, we can rule out the existence of "truthtelling equilibria", in which Player 1s always report the true state as they have an incentive to lie in state Y. It is similarly straightforward to prove that there exist (i) equilibria in which Player 1s always (i.e. irrespective of state) send message "State X prevails", play their state-specific dominant action and Player 2s always (i.e. irrespective of message) play C, (ii) equilibria in which Player 1s always send message "State X prevails", play their state-specific dominant action and Player 2s always play D, as well as (iii) equilibria in which Player 1s always send message "State X prevails", play their state-specific dominant action and Player 2s play C with probability p, where $p \in [0, 1]$ in NO S and $p \in [\frac{1}{3}, 1]$ in FINE and COMP.

reveal social preferences and exhibit norm-abiding behaviour.[10] For the context of the present experiment, one can plausibly argue that the sanction regimes in FINE and COMP convey social norms that condemn and deter lying, even if the implemented sanctions are non-deterrent in expectations (cf. section 3.1). This arguably increases average compliance on the part of Player 1s and consequently promotes average trust on the part of Player 2s in FINE and COMP compared to NO S (cf. deterrence-based trust, Lewicki et al. (1996)). This leads to our first behavioural prediction:

Hypothesis 1: *Due to compliance effects in* FINE *and* COMP*, Player 1s lie less often and Player 2s trust more often in* FINE *and* COMP *than in* NO S *in part 1 and 2 of the experiment.*

Apart from that, postulating that Player 1s are to some degree guilt averse (cf. Charness and Dufwenberg, 2006 and Battigalli and Dufwenberg, 2007), Player 1s have larger incentives to report state Y truthfully in treatment FINE than in COMP. The reason being that in the former of the two treatments, Player 1s can reduce their (expected) feelings of guilt only by reporting the true state of the world. In the latter treatment, conversely, lying Player 1s have the possibility to "morally cleanse" themselves. Their guilt can be reduced if they are convicted and they compensate the victim for the experienced loss.

If Player 2s anticipate how guilt aversion affects lying across FINE and COMP differently, they should act upon Player 1s' messages more often in FINE than in COMP. We summarise our behavioural predictions as follows:

Hypothesis 2: *Assuming guilt aversion on the part of Player 1s, we expect less lying and more trust in* FINE *than in* COMP *in part 1 and 2 of the experiment.*

Lastly, we assume that the sanction regimes in FINE and COMP in parts 1 and 2 are able to sustainably establish pro-social behaviour, as similarly discussed in Mulder et al. (2006). Based on this, we expect that Player 1's comparably greater truthfulness and Player 2's larger trust in FINE and COMP compared to NO S carry over to part 3 of the experiment when sanctions are lifted. Since victims cannot be compensated in part 3 anymore, Player 1s' guilt aversion does no longer influence their behaviour differently across treatments. We hence conjecture:

[10] For a recent overview on social preferences, see Drouvelis (2021). For a survey on norm-abiding behaviour, see Legros and Cislaghi (2020).

Hypothesis 3: *Assuming sustained compliance on the part of Player 1s, we expect less lying and more trust in* Fine *and* Comp *than in* No S *in part 3 of the experiment.*

4 Empirical results

We start our empirical analysis with an overview of some basic statistics on misbehaviour across treatments. We define misbehaviour in our experiment as lying about the state of the world if state Y prevails.[11] Table 2 reveals that while in No S 70.14% of all reported messages in state Y are lies, the respective figures in Fine and Comp are comparably lower (30.88% and 47.06%). Also when focusing on the number of infringers, we observe that 87.50% of Player 1s lie about state Y at least once in No S. while the respective figures amount to only 44.44% in Fine and 62.50% in Comp. Consequently, 100% of Player 2s in No S encounter at least one liar during parts 1 and 2, while this is true for only 63.89% of Player 2s in Fine and 79.17% in Comp. The respective Fisher exact test results (p-values) are presented in the three right most columns of Table 2. Note that here and throughout the paper we only report p-values from two-sided test statistics.

Table 2: Summary statistics: behaviour in parts 1 and 2 (combined)

		Fisher's exact test results (p-values)		
		No S vs Fine	No S vs Comp	Fine vs Comp
Share of lies about state y				
No S	70.14%	< 0.001	< 0.001	0.001
Fine	30.88%			
Comp	47.06%			
Share of lying Player 1s				
No S	87.50%	< 0.001	0.001	0.043
Fine	44.44%			
Comp	62.50%			
Share of Player 2s who encounter at least one liar				
No S	100%	< 0.001	< 0.001	0.064
Fine	63.89%			
Comp	79.17%			

Notes: Definition of a lie... Definition of a lying Player 1... We define Player 1 being a liar if they lie about state y at least in one of the 5 rounds in parts 1 and 2.

[11] As argued above, lying in state X is irrational given the implemented payoff structure. It is hence not relevant for our subsequent analysis. In fact, we observe only two lies in state X in total. One lie in No S and one in Fine.

As explained in detail in Section 3.1, by design every third message sent by a Player 1 is randomly checked for truthfulness. As a result, in both FINE and COMP 86.11% of Player 1s are checked at least once during parts 1 and 2, 20.83% of Player 1s (not conditional on being checked) are caught lying in FINE, 37.50% in COMP.

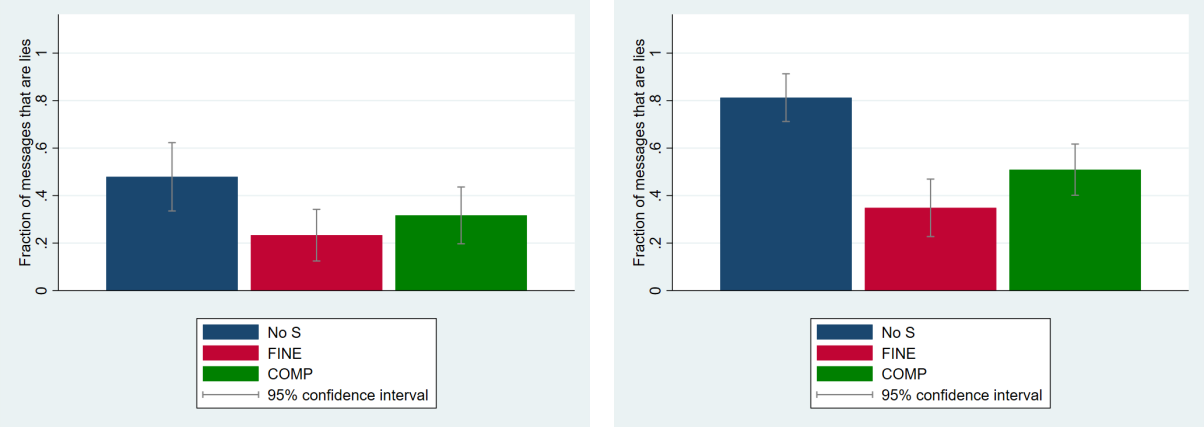
4.1 Player 1s (mis-)behaviour in parts 1 and 2

We first consider Player 1s', that is, the potential infringers' behaviour in part 1 of the experiment, the only true one-shot interaction. Next we turn to their behaviour in part 2, in which subjects play the same game again four more times with random new matching partners, allowing for learning effects, net of additional reputational effects.

Graphic (a) in Figure 2 presents the share of lying Player 1s in part 1 of the experiment. We observe most misbehaviour in NO S. Almost half of Player 1s in that treatment (48%) lie when state Y prevails. In FINE and COMP subjects are comparably more truthful. Only 23% of subjects lie in FINE, while slightly more (32%) lie in COMP (Fisher's exact test results: NO S vs. FINE: $p < 0.01$, NO S vs. COMP: $p = 0.11$). In particular the difference between NO S and FINE is highly significant. Player 1's propensity to lie in FINE and COMP, on the other hand, is not significantly different (Fisher's exact test result: $p = 0.41$). We corroborate these findings in a linear probability model, see column (1) of Table A.1 in Appendix A. The treatment difference between NO S and COMP turns out marginally significant ($p<0.1$) there. In column (2) we extend the regression by controls for subjects' gender and general risk preferences, which we measure using a scale similar to Dohmen et al.'s (2010) well-established questionnaire item.[12] Recent meta studies by Abeler et al. (2019) and Capraro (2017) show that women are more honest than men. Moreover, one could conjecture that participants' general risk preferences shape their propensity to lie differently across NO S and the sanctioning treatments. Indeed, we find a significantly positive effect of risk proneness on Player 1's propensity to lie, the coefficient of the gender dummy, conversely, turns out insignificant. However reassuringly, including these controls leaves the size and significance of the treatment coefficients virtually unchanged.[13]

[12] We asked subjects to rate their "willingness to take risks, in general" on a scale from 1 to 10, where 1 is "completely unwilling" and 10 is "completely willing".

[13] We confirm the findings from the linear probability models reported in Table A.1 in logit regressions. The findings are available from the authors upon request.



(a) Part 1 (b) Part 2

Figure 2: Player 1's misbehaviour in parts 1 and 2

The behavioural patterns from part 1 of the experiment carry over to part 2 – the treatment differences become even more pronounced, compare graphics (a) and (b) in Figure 2. Mann-Whitney ranksum tests, comparing average choices aggregated at the matching group level find significant treatment differences between all three treatments (No S vs. Fine: $p < 0.01$, No S vs. Comp: $p < 0.01$, Fine vs. Comp: $p = 0.04$).

We use a regression analysis (linear probability models) to study behaviour in part 2 in more detail. It comes with the advantage of allowing us to control for potential round effects as well as other influencing factors that we detail below. Column (1) in Table 3 reveals that also in this part of the experiment, Player 1s lie significantly less often in the two sanctioning treatments than in No S. In Fine, on average across the four rounds, 34% = 0.81–0.47 of Player 1s lie in state Y, in Comp 53% = 0.81–0.28 lie and in the benchmark treatment No S 81% = 0.81 of Player 1s lie. Furthermore, Player 1s lie significantly less often in Fine than in Comp, see the result of the corresponding Wald test reported in the bottom part of the table ($p = 0.01$).[14]

Column (2) confirms that the treatment differences remain significant (though smaller than those estimated in column (1)) if we control for whether a Player 1 has lied before and whether any of the previous lies were successful. By the latter, we mean whether a lie about the state of the world made the matched Player 2 choose C instead of D in one of the previous rounds. As evident, in particular the former of the two factors increases Player 1s' propensity to lie.

We replicate the findings from column (2) in the model presented in column (3), in which

[14]However, evidently, the propensity to lie increases in all treatments compared to part 1. This can potentially be explained by the repeated setting of part 2. In a meta study on the workhorse model for altruism and pro-sociality, the dictator game, Engel (2011) similarly finds that subjects behave less pro-socially in repeated settings than in one-shot interactions.

we add further controls for rounds, subjects' gender and risk preferences.

Finding 1: *In line with Hypothesis 1, we observe less lying in* Fine *and* Comp *than in* No S*; both sanction schemes promote compliance. The repeated setting of part 2 furthermore corroborates Hypothesis 2; Player 1s lie less in* Fine *than in* Comp*, arguably due to differences in Player 1s' treatment-specific degree of guilt aversion.*

Table 3: Player 1's decision to lie in part 2

	(1)	(2)	(3)	(4)	(5)
Fine	-0.47***	-0.30***	-0.31***	-0.33***	-0.32***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Comp	-0.28***	-0.16***	-0.19***	-0.21***	-0.23**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Player has lied before		0.44***	0.45***	0.41***	0.44***
		(0.00)	(0.00)	(0.00)	(0.00)
A previous lie was successful		0.09	0.10	0.12	0.11
		(0.27)	(0.20)	(0.17)	(0.17)
Fine× Player was caught lying before				0.14	
				(0.38)	
Comp × Player was caught lying before				0.05	
				(0.59)	
Fine× Player was checked for lying before					0.02
					(0.75)
Comp × Player was checked for lying before					0.05
					(0.56)
Gender: Female			-0.01	-0.01	-0.01
			(0.87)	(0.91)	(0.88)
Risk proneness			0.00	0.00	0.00
			(0.89)	(0.91)	(0.92)
Constant	0.81***	0.49***	0.77***	0.79***	0.78***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	384	384	374	374	374
R-squared	0.13	0.37	0.40	0.40	0.40
Comparing Fine and Comp Wald test results (p-values)	0.01	0.02	0.02	0.02	0.30

Notes: Linear probability models. Dependent variable: Player 1's decision to lie in part 2. No S serves as baseline treatment in all regressions. We add dummies for rounds in columns (3) – (5). Robust standard errors are clustered at the matching group level, p-values given in parentheses: ** p<0.05, *** p<0.01.

In models (4) and (5), we extend the regression model from column (3) and interact the sanction treatment coefficients with the dummy variables "Player was caught lying before" and "Player was checked for lying before", respectively, to find out if there are any additional

effects of having experienced the law in either way in FINE and COMP. The FINE and COMP coefficients then indicate the treatment effects on those Player 1s who have *not* experienced the law in either way in any previous round. As it turns out, those Player 1s lie significantly less often than their counterparts in the benchmark treatment, NO S. Moreover, the corresponding Wald test results confirm that those Player 1s in FINE lie significantly less often than their counterparts in COMP. This holds for both models in column (4) and (5). Lastly, as evident from the positive, but insignificant coefficients of the interaction terms in both models, having experienced the law in either way does not significantly deter Player 1s from lying again, in neither of the two sanctioning treatments.

We qualitatively and quantitatively replicate the findings from the linear probability models presented in Table 3 in additional logit regressions in Table A.2 in Appendix A.

4.2 Player 2s trust in parts 1 and 2

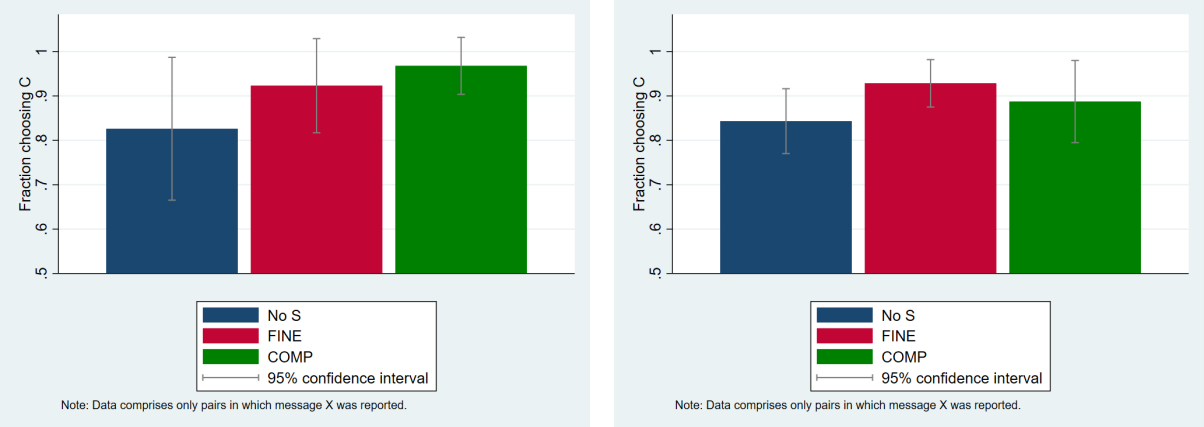
In the following, we take a closer look at Player 2s', that is, the potential victims' behaviour. We are particularly interested in treatment differences in their propensity to choose action C when their matched Player 1 reported that state X prevails. Given the structure of the experimental game, we interpret differences in the share of C choices as indicative for differences in treatment specific trust in Player 1's X messages.

Graphic (a) in Figure 3 presents the share of Player 2s who choose action C upon receiving message X in part 1 of the experiment. In COMP, almost all Player 2s (97%) follow their matched Player 1s' message X, in FINE 92% do, and even in NO S 83% do so. Behaviour across treatments is not significantly different. All Fisher's exact test results from pairwise treatment comparisons turn out insignificant (NO S vs. FINE: $p = 0.40$, NO S vs. COMP: $p = 0.15$, FINE vs. COMP: $p = 0.59$). We corroborate these findings in linear probability models that allow us to control for participants' gender and general risk preferences, see Table A.3 in Appendix A.[15]

The behavioural patterns change in part 2, consider graphics (a) and (b) in Figure 2. Mann-Whitney ranksum tests, comparing average choices aggregated at the matching group level reveal that Player 2s choose action C significantly more often in FINE than in NO S. All other results from pairwise treatment comparisons are insignificant (NO S vs. FINE: $p = 0.05$, NO S vs. COMP: $p = 0.27$, FINE vs. COMP: $p = 0.68$).

Similar to Section 4.1, we use a regression analysis (linear probability models) to study

[15] We confirm the insignificant treatment differences from the linear probability models reported in Table A.3 in additional logit regressions, results available from the authors upon request.



(a) Part 1 (b) Part 2

Figure 3: Player 2s choosing C in parts 1 and 2

behaviour in part 2 in more detail. Column (1) in Table 4 reveals that averaged over all rounds, in No S 84% (=0.84) of Player 2s choose action C if their matched Player 1's reports state X prevails. The share amounts to 92% (=0.84+0.08) in Fine and to 87% (=0.84+0.03) in Comp. Column (1) corroborate the finding that Player 2s trust their matched Player 1s' message X and choose C more often in Fine than in No S. Conversely, both the Comp coefficient as well as the Wald test result that compares both sanctioning treatment coefficients are insignificant, revealing that there do not exist any further significant treatment differences.

In column (2), we add a dummy variable that captures whether a Player 2 has been lied to in any of the previous rounds. Intriguingly, controlling for this event renders the Fine coefficient insignificant. We replicate the findings from column (2) in the model presented in column (3), in which we add further controls for rounds, subjects' gender and risk preferences.

Finding 2: *In line with Hypothesis 1, we observe more trust in* Fine *than in* No S*, at least in part 2 of the experiment. This is the only significant treatment difference – and it vanishes once one controls for having been lied to in a previous round. This suggests that as long as Player 2s have not been lied to,* Fine *creates more trust than* No S*. After they've encountered an infringer, Player 2s' trust is no longer significantly shaped by the sanction regime per se.*

In models (4) and (5), we extend the regression model from column (3) and interact the sanction treatment coefficients with the dummy variables "A previously matched Player 1 was caught lying" and "A previously matched Player 1 was checked for lying", respectively, to find out if trust can be restored after having experienced the law in either way in Fine and Comp. However, as indicated by the positive, but insignificant coefficients of the interaction terms,

Table 4: Player 2's decision to choose C in part 2

	(1)	(2)	(3)	(4)	(5)
FINE	0.08*	0.04	0.04	0.05	0.03
	(0.06)	(0.28)	(0.32)	(0.19)	(0.59)
COMP	0.03	0.01	0.01	0.02	0.01
	(0.57)	(0.85)	(0.84)	(0.65)	(0.89)
A previously matched Player 1 lied		-0.14***	-0.14***	-0.12***	-0.14***
		(0.00)	(0.00)	(0.00)	(0.00)
FINE × A previously matched				-0.11	
Player 1 was caught lying				(0.38)	
COMP × A previously matched				-0.05	
Player 1 was caught lying				(0.61)	
FINE × A previously matched					0.03
Player 1 was checked for lying					(0.44)
COMP × A previously matched					0.01
Player 1 was checked for lying					(0.91)
Gender: Female			0.03	0.03	0.04
			(0.45)	(0.43)	(0.43)
Risk proneness			-0.00	-0.00	-0.00
			(0.65)	(0.65)	(0.70)
Constant	0.84***	0.92***	0.93***	0.92***	0.93***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	586	586	581	581	581
R-squared	0.01	0.05	0.06	0.06	0.06
Comparing FINE and COMP Wald test results (p-values)	0.35	0.48	0.51	0.50	0.78

Notes: Linear probability models. Dependent variable: Player 2's decision to choose C if Player 1 reports message X in part 2. NO S serves as baseline treatment in all regressions. We add dummies for rounds in columns (3) – (5). Robust standard errors are clustered at the matching group level, p-values given in parentheses: * p<0.10, *** p<0.01.

none of these events significantly increases Player 2s' trust in Player 1s' X-messages in neither of the two sanctioning treatments compared to NO S. Having encountered an infringer in one of the previous rounds is still the only significant explanatory factor in these models.

We qualitatively and quantitatively replicate the findings from the linear probability models presented in Table 4 in additional logit regressions in Table A.4 in Appendix A.

5 Removing the sanctions

5.1 Potential infringers' (mis-)behaviour in part 3

Finally, we study Player 1s' behaviour in part 3 of the experiment in which all subjects play the experimental game under the sanction free regime of No S.

Figure 4 reveals that Player 1s' propensity to lie is relatively high in all treatments (88% in No S. 63% in Fine and 67% in Comp). Yet, both sanctioning regimes of Fine and Comp have lasting effects that spill over to part 3. Mann-Whitney ranksum tests that compare Player 1's propensity to lie across treatments (aggregated at the matching group level) reveal significant treatment differences between No S and Fine as well as between No S and Comp (p=0.02 and p=0.03, respectively). The amount of lying in Fine and Comp, conversely, is not statistically different from one another (p=0.76).

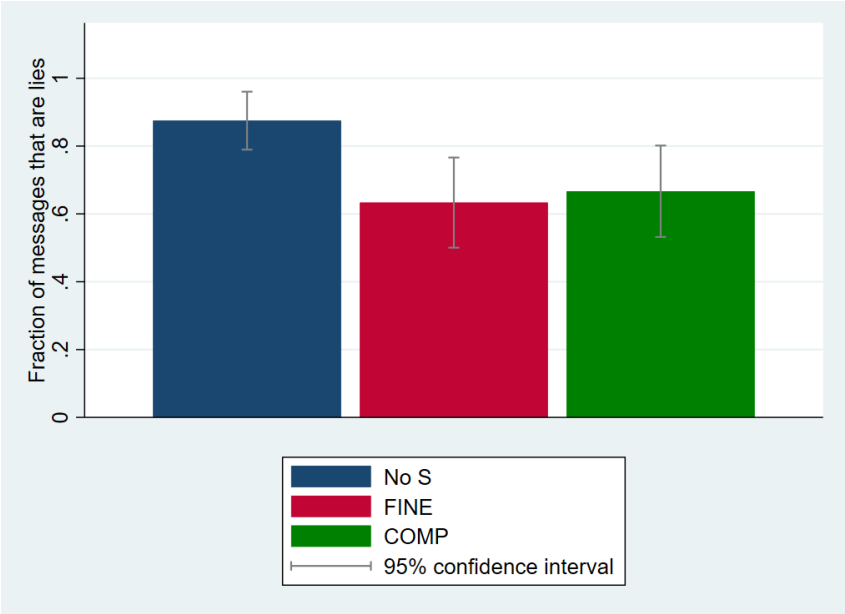


Figure 4: Player 1's misbehaviour in part 3

Linear probability model (1) in Table 5 corroborates the findings from the non-parametric tests. However, if we control for Player 1's experience with previous lies and their success (model (2)), the treatment differences between No S and the two sanctioning treatments are no longer significant. In particular having lied successfully, that is, having convinced a matched Player 2 to choose C instead of D in a previous round, significantly increases Player 1s' propensity to lie in part 3. We reproduce these findings in model (3), in which we additionally control for Player 1s' gender and general risk preferences.[16]

[16]Table A.5 in Appendix A replicates the findings from the linear probability models presented in Table 5 in additional logit regressions.

Taken together, models (1) – (3) suggest that the sanction regimes' compliance effects in Fine and Comp carry over to part 3. Their lasting effects are driven by the honesty of those Player 1s who were previously deterred from lying and keep refraining from lying once sanctions are lifted. We conclude:

Finding 3: *In line with Hypothesis 3, we observe less lying in* Fine *and* Comp *than in* No S *in part 3 of the experiment. As it turns out, these effects are driven by sustained compliance on the part of Player 1s who were successfully deterred from lying from the very beginning.*

Table 5: Player 1's decision to lie in part 3

	(1)	(2)	(3)
Fine	-0.24***	0.00	0.00
	(0.00)	(0.94)	(0.92)
Comp	-0.21**	-0.01	-0.01
	(0.01)	(0.93)	(0.82)
Player has lied before		0.26*	0.20
		(0.08)	(0.18)
A previous lie was successful		0.36***	0.41***
		(0.01)	(0.00)
Gender: Female			0.06
			(0.26)
Risk proneness			0.01
			(0.30)
Constant	0.88***	0.34***	0.23**
	(0.00)	(0.00)	(0.04)
Observations	168	168	164
R-squared	0.05	0.42	0.42
Comparing Fine and Comp Wald test results (p-values)	0.71	0.90	0.80

Notes: Linear probability models. Dependent variable: Player 1's decision to lie in part 3. No S serves as baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

5.2 Potential victims' trust in part 3

Figure 5 suggests that there are no significant treatment differences in part 3. In No S, 81% of Player 2s choose action C when their matching partner reports state X, 86% do so in Fine and 75% in Comp. Mann-Whitney ranksum tests that consider behaviour aggregated at the

matching group level, do not find any significant treatment differences in trust (No S vs Fine: p = 0.47, No S vs. Comp: p = 0.61, Fine vs. Comp: p = 0.17).

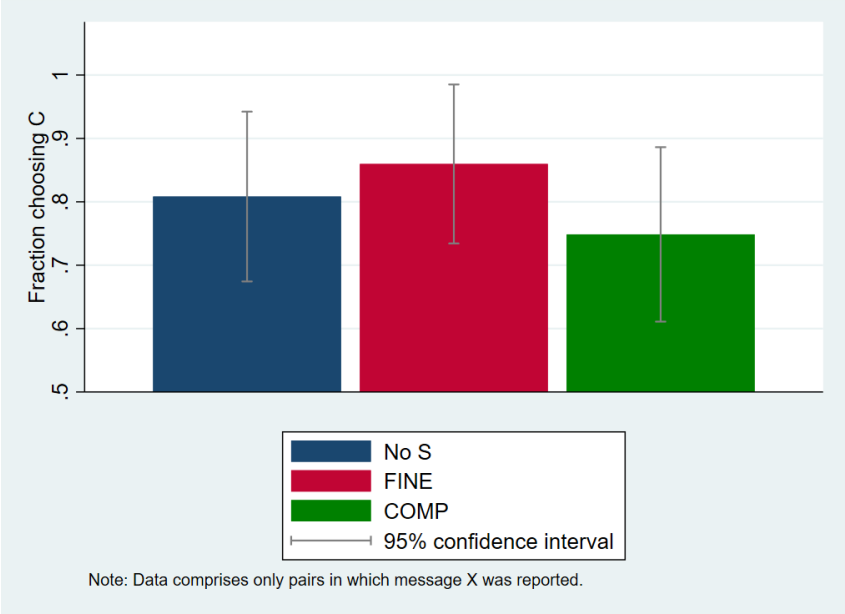


Figure 5: Player 2s choosing C in part 3

These findings are corroborated in corresponding linear probability models, see Table 6. The treatment coefficients of Fine and Comp are insignificant, also a Wald test result, comparing the two coefficients turns out insignificant (column (1)). Controlling for whether a Player 2 has been matched to a lying Player 1 in one of the previous rounds (column (2)) or taking into account Player 2s' gender or general risk preferences (column (3)) does not change these results. We conclude by stating that we do not find any indication for spill-over effects from the previous sanction schemes in Fine (and Comp). In columns (2) and (3), the only significant coefficient is that of "A previously matched Player 1 lied". Having been lied to significantly decreases Player 2s' propensity to choose C when the matched Player 1 reports that state X prevails.[17]

Finding 4: *Different from what was predicted in Hypothesis 3, Player 2s' propensity to trust in part 3 is not significantly affected by the treatment they experienced in parts 1 and 2 of the experiment. Their behaviour can rather be explained by previous experiences with lying Player 1s.*

[17]Table A.6 in Appendix A replicates the findings from the linear probability models presented in Table 6 in additional logit regressions.

Table 6: Player 2's decision to choose C in part 3

	(1)	(2)	(3)
Fine	0.05	-0.04	-0.04
	(0.54)	(0.61)	(0.67)
Comp	-0.06	-0.11	-0.10
	(0.48)	(0.26)	(0.28)
A previously matched Player 1 lied		-0.22***	-0.21***
		(0.00)	(0.01)
Gender: Female			-0.02
			(0.82)
Risk proneness			-0.02
			(0.16)
Constant	0.81***	1.03***	1.15***
	(0.00)	(0.00)	(0.00)
Observations	143	143	143
R-squared	0.01	0.06	0.07
Comparing Fine and Comp Wald test results (p-values)	0.42	0.50	0.52

Notes: Linear probability models. Dependent variable: Player 2's decision to choose C if Player 1 reports message X in part 3. No S serves as baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: *** p<0.01.

6 Discussion

Considering the observed treatment differences in Player 1s' propensity to lie across all three parts of the experiment, we conclude that the compliance-inducing capacity of fine and compensation schemes is rather robust. Focusing on part 2, we observe significantly less lying in Fine and Comp than in No S, regardless of whether Player 1s had lied (successfully) before. In the repeated setting of part 2, in particular the sanction scheme in Fine successfully induces Player 1s not to lie. At the aggregate level, both sanction schemes' compliance effects nicely carry over to part 3 of the experiment, in which misbehaviour is not punished any longer. This lasting effect is driven by the fact that both sanction regimes successfully deterred a non-negligible part of Player 1s from becoming "first offenders" and, as a result, these Player 1s are also less inclined to lie when sanctions were lifted. This confirms the predicted habit-forming effect of both fine and compensation schemes.

Focusing on treatment differences in trust, in the true one-shot interaction of part 1, Player 2s

seem to display (if at all, and only insignificantly so) most trust in their matching partners' reported X messages in COMP. Only in part 2, Player 2 seem to learn to correctly anticipate the larger compliance effects in fine regimes. Our findings, however, also indicate that once Player 2s have been lied to, it is the actual experiences rather than the prevailing sanction regime that determines their trust. In part 3 we do not observe any lasting effects of either of the sanction regimes on trust.

It is of note that Player 2s' average round payoffs in parts 1 and 2 combined are 15.50 points (s.d. 5.15) in NO S. 17.69 points (s.d. 4.41) in FINE and 17.72 points (s.d. 4.58) in COMP. That is, Player 2s' average round earnings in FINE and COMP are 14% higher than in NO S. Interestingly, they are very similar in FINE and COMP, even though Player 2s are eligible for compensation payments only in the latter treatment. Mann-Whitney ranksum tests that compare aggregated round payoffs at the matching group level reveal significant treatment differences between NO S and FINE as well as between NO S and COMP (both $p < 0.01$). The distributions of payoffs in FINE and COMP, conversely, are not statistically different from another ($p = 0.90$). These treatment differences replicate when considering part 1 payoffs and part 2 payoffs separately.[18]

We conclude that there is a mismatch between potential infringers' compliance and potential victims' trust: In part 1, Player 1s lie significantly less often in FINE and COMP than in NO S. Player 2s' propensity to trust their counterpart, however, does not differ significantly across treatments. In the repeated setting of part 2 then, Player 1s are most compliant in FINE, less so in COMP and comply least/lie most in NO S. Player 2s' trust does still not fully match these treatment differences. But we do find that they trust significantly more in FINE than in NO S. More important than the treatment-specific sanction scheme per se, are, however, "first-hand" experiences with lying/compliant Player 1s. These findings indicate that it might take quite some time to install trust through sanction schemes.

7 Conclusion

From the potential infringers' point of view compensation is not just fine: we find evidence that fines induce larger compliance, which is consistent with explanations that assume at least some degree of guilt aversion on the side of the potential infringers. Hence, our findings, firstly,

[18] Additional Mann-Whitney ranksum test results for differences in payoffs in part 1 (considering treatment comparisons of behaviour at the individual level): NO S vs FINE: p=0.07, NO S vs COMP: p=0.06, COMP vs FINE: p=0.1; analogous test results for part 2 (considering treatment comparisons of behaviour at the matching group level): NO S vs FINE: $p < 0.01$, NO S vs COMP: $p < 0.01$, COMP vs FINE: p=1.

underline that, as long as the probability of detection and size of sanction payment are kept constant, compensation and fine payments do not lead to the same compliance levels.

Secondly, our findings question the idea that having a compensation scheme that potentially safeguards some payoff in case of misbehaviour is favourable from a potential victims' point of view. In fact, findings from part 2 suggest that, if at all, the fine scheme has a positive impact on trust. This may be attributed to its larger deterrent capacity.

Third, our findings shed some interesting light on how victims and infringers behave once they have experienced the law and been confronted with sanctions. Admittedly, those results convey a rather pessimistic picture. It appears that the sanction schemes primarily deterred potential infringers from becoming first offenders. Having experienced the law by being checked for or even caught lying does not lower their propensity to lie again in a future round. Similarly, whereas the presence of sanctions did increase trust on the part of potential victims to some extent, once misbehaviour was experienced, their effects on future trust vanished.

By detailing the matches and mismatches between how, on the one side, potential infringers are induced to comply by sanctions and how, on the other side, potential victims trust under different sanction regimes, our findings underline the value of including both perspectives in the study of sanction schemes.

For policy makers, the reasons for preferring fines or compensation are manifold: fines are additional income to the state budget, different administrative costs may play a role, compensation is argued to make victims happy ex post to name just a few. Our findings provide new crucial input in the policy discussion on the implementation of different sanction regimes, namely the actual behavioural effects of a fine vs. compensation regime on the actors involved. In many legal domains the experimentation with the optimal type of sanction is still ongoing. In European consumer contract law the most traditional sanction is compensation, however, the European legislator is increasingly prescribing fines. In European competition law the traditional approach of public law enforcement by way of fines has in the last years been complemented by a right to compensation. Also the European Data protection regulation ensures that Member States have both fines and compensation in their toolkit.

More research in the lab and ultimately also in the field is needed to corroborate and extend our findings on compensation and fines. For instance, it seems worthwhile to compare the sanction specific effects for different levels of detection probabilities and sanction payments. Also, we focused on infringement as intentional acts. Recent research has suggested that compensation

regimes, more than fine regimes, may stimulate care investment to prevent unintentional harm (Baumann et al., 2020). It would be interesting to directly contrast the effectiveness of both regimes in preventing both intentional and unintentional harm. Last but not least, another interesting avenue for future research would be to consider fines and compensation payments in combination.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for truth-telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Agranov, Marina and Anastasia Buyalskaya, "Deterrence effects of enforcement schemes: An experimental study," *Management Science*, 2022, *68* (5), 3573–3589.

Andreoni, James, "Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?," *The RAND Journal of Economics*, 1991, pp. 385–395.

Battigalli, Pierpaolo and Martin Dufwenberg, "Guilt in games," *American Economic Review*, 2007, *97* (2), 170–176.

Baumann, Florian, Tim Friehe, and Pascal Langenbach, "Fines versus damages: Experimental evidence on care investments," *MPI Collective Goods Discussion Paper*, 2020, (2020/8).

Becker, Gary S, "Crime and punishment: An economic approach," in "The Economic Dimensions of Crime," Springer, 1968, pp. 13–68.

Bohnet, Iris and Yael Baytelman, "Institutions and trust: Implications for preferences, beliefs and behavior," *Rationality and Society*, 2007, *19* (1), 99–135.

_ , Bruno S Frey, and Steffen Huck, "More order with less law: On contract enforcement, trust, and crowding," *American Political Science Review*, 2001, *95* (1), 131–144.

Bottom, William P, Kevin Gibson, Steven E Daniels, and J Keith Murnighan, "When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation," *Organization Science*, 2002, *13* (5), 497–513.

Capraro, Valerio, "Gender differences in lying in sender-receiver games: A meta-analysis," *arXiv preprint arXiv:1703.03739*, 2017.

Cardi, W Jonathan, Randall D Penfield, and Albert H Yoon, "Does tort law deter individuals? A behavioral science study," *Journal of Empirical Legal Studies*, 2012, *9* (3), 567–603.

Charness, Gary and Martin Dufwenberg, "Promises and partnership," *Econometrica*, 2006, *74* (6), 1579–1601.

Cooter, Robert D, "Punitive damages for deterence: When and how much," *Alabama Law Review*, 1988, *40*, 1143.

Dari-Mattiacci, Giuseppe and Alex Raskolnikov, "Unexpected effects of expected sanctions," *The Journal of Legal Studies*, 2021, *50* (1), 35–74.

Desmet, Pieter and Franziska Weber, "Infringers' willingness to pay compensation versus fines," *European Journal of Law and Economics*, 2022, *53* (1), 63–80.

_ , David De Cremer, and Eric van Dijk, "On the psychology of financial compensations to restore fairness transgressions: When intentions determine value," *Journal of Business Ethics*, 2010, *95* (1), 105–115.

_ , _ , and _ , "In money we trust? The use of financial compensations to repair trust in the aftermath of distributive harm," *Organizational Behavior and Human Decision Processes*, 2011, *114* (2), 75–86.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "Are risk aversion and impatience related to cognitive ability?," *American Economic Review*, 2010, *100* (3), 1238–60.

Drouvelis, Michalis, *Social Preferences: An Introduction to Behavioural Economics and Experimental Research*, Agenda Publishing, 2021.

Eisenberg, Theodore and Christoph Engel, "Assuring civil damages adequately deter: A public good experiment," *The Journal of Empirical Legal Studies*, 2014, *11* (2), 301–349.

Engel, Christoph, "Dictator games: A meta study," *Experimental Economics*, 2011, *14* (4), 583–610.

_ , "Experimental criminal law: a survey of contributions from law, economics, and criminology," *Empirical Legal Research in Action*, 2018, pp. 57–108.

Garoupa, Nuno, "Optimal magnitude and probability of fines," *European Economic Review*, 2001, *45* (9), 1765–1771.

Gneezy, Uri and Aldo Rustichini, "A fine is a price," *The Journal of Legal Studies*, 2000, *29* (1), 1–17.

Khadjavi, Menusch, "On the interaction of deterrence and emotions," *The Journal of Law, Economics, & Organization*, 2015, *31* (2), 287–319.

Kurz, Tim, William E Thomas, and Miguel A Fonseca, "A fine is a more effective financial deterrent when framed retributively and extracted publicly," *Journal of Experimental Social Psychology*, 2014, *54*, 170–177.

Legros, Sophie and Beniamino Cislaghi, "Mapping the social-norms literature: An overview of reviews," *Perspectives on Psychological Science*, 2020, *15* (1), 62–80.

Lewicki, Roy J, Barbara B Bunker et al., "Developing and maintaining trust in work relationships," *Trust in Organizations: Frontiers of Theory and Research*, 1996, *114*, 139.

Malhotra, Deepak and J Keith Murnighan, "The effects of contracts on interpersonal trust," *Administrative Science Quarterly*, 2002, *47* (3), 534–559.

Mulder, Laetitia B, "When sanctions convey moral norms," *European Journal of Law and Economics*, 2018, *46* (3), 331–342.

_ , Eric Van Dijk, David De Cremer, and Henk Wilke, "Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas," *Journal of Experimental Social Psychology*, 2006, *42* (2), 147–162.

Rousseau, Denise M, Sim B Sitkin, Ronald S Burt, and Colin Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of Management Review*, 1998, *23* (3), 393–404.

Schildberg-Hörisch, Hannah and Christina Strassmair, "An experimental test of the deterrence hypothesis," *The Journal of Law, Economics, & Organization*, 2012, *28* (3), 447–459.

Slemrod, Joel, "Tax compliance and enforcement: New research and its policy implications," *Ross School of Business Paper No. 1302*, 2016.

Stigler, George J., "The Optimum Enforcement of Laws," *Journal of Political Economy*, 1970, *78* (3), 526–536.

Veljanovski, Cento G, "The economics of regulatory enforcement," *Enforcing Regulation*, 1984, *171*, 186.

Vollan, Björn, "The difference between kinship and friendship:(Field-) experimental evidence on trust and punishment," *The Journal of Socio-Economics*, 2011, *40* (1), 14–25.

A Additional analyses

Table A.1: Player 1's decision to lie in part 1

	(1)	(2)
FINE	-0.25***	-0.25***
	(0.01)	(0.01)
COMP	-0.16*	-0.18*
	(0.09)	(0.06)
Gender: Female		-0.00
		(0.99)
Risk proneness		0.03*
		(0.07)
Constant	0.48***	0.31**
	(0.00)	(0.02)
Observations	168	164
R-squared	0.04	0.07
Comparing FINE and COMP Wald test results (p-values)	0.31	0.40

Notes: Linear probability models. Dependent variable: Player 1's decision to lie in part 1. NO S serves as baseline treatment in both regressions. Robust standard errors are clustered at the individual level, p-values given in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

Table A.2: Robustness check, Player 1's decision to lie in part 2

	(1)	(2)	(3)	(4)	(5)
FINE	-2.13***	-1.86***	-2.02***	-2.08***	-2.13***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
COMP	-1.33***	-1.11***	-1.28***	-1.29***	-1.53***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Player has lied before		2.03***	2.21***	2.12***	2.18***
		(0.00)	(0.00)	(0.00)	(0.00)
A previous lie was successful		0.78	0.83	0.84	0.87
		(0.17)	(0.13)	(0.15)	(0.11)
FINE × Player was caught				0.51	
lying before				(0.60)	
COMP × Player was caught				0.07	
lying before				(0.91)	
FINE × Player was checked					0.12
for lying before					(0.70)
COMP × Player was checked					0.31
for lying before					(0.57)
Gender: Female			-0.04	-0.03	-0.03
			(0.91)	(0.93)	(0.92)
Risk proneness			0.00	0.00	0.00
			(0.96)	(0.97)	(0.99)
Constant	1.47***	0.30	1.88**	1.93**	2.00**
	(0.00)	(0.35)	(0.02)	(0.01)	(0.02)
Observations	384	384	374	374	374
Pseudo R-squared	0.10	0.31	0.34	0.34	0.34
Comparing FINE and COMP Wald test results (p-values)	0.01	0.01	0.01	0.01	0.26

Notes: Logit regressions. Dependent variable: Player 1's decision to lie in part 2. NO S serves as baseline treatment in all regressions. We add dummies for rounds in columns (3) – (5). Robust standard errors are clustered at the matching group level, p-values given in parentheses: ** p<0.05, *** p<0.01.

Table A.3: Player 2's decision to choose C in part 1

	(1)	(2)
FINE	0.10	0.09
	(0.32)	(0.39)
COMP	0.14	0.14
	(0.11)	(0.11)
Gender: Female		0.02
		(0.75)
Risk proneness		0.01
		(0.75)
Constant	0.83***	0.78***
	(0.00)	(0.00)
Observations	80	78
R-squared	0.04	0.04
Comparing FINE and COMP Wald test results (p-values)	0.48	0.52

Notes: Linear probability models. Dependent variable: Player 2's decision to choose C if Player 1 reports message X in part 1. NO S serves as baseline treatment in both regressions. Robust standard errors are clustered at the individual level, p-values given in parentheses: *** p<0.01.

Table A.4: Robustness check, Player 2's decision to choose C in part 2

	(1)	(2)	(3)	(4)	(5)
FINE	0.85**	0.47	0.46	0.61	0.26
	(0.05)	(0.23)	(0.26)	(0.15)	(0.62)
COMP	0.28	0.08	0.07	0.20	0.05
	(0.58)	(0.88)	(0.89)	(0.71)	(0.94)
A previously matched Player 1 lied		-1.39***	-1.37***	-1.25***	-1.37***
		(0.00)	(0.00)	(0.00)	(0.00)
FINE × A previously matched				-0.82	
Player 1 was caught lying				(0.19)	
COMP × A previously matched				-0.33	
Player 1 was caught lying				(0.60)	
FINE × A previously matched					0.42
Player 1 was checked for lying					(0.47)
COMP × A previously matched					0.05
Player 1 was checked for lying					(0.92)
Gender: Female			0.34	0.34	0.36
			(0.43)	(0.43)	(0.42)
Risk proneness			-0.04	-0.04	-0.04
			(0.62)	(0.60)	(0.66)
Constant	1.69***	2.64***	2.78***	2.70***	2.77***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	586	586	581	581	581
Pseudo R-squared	0.02	0.08	0.08	0.08	0.08
Comparing FINE and COMP Wald test results (p-values)	0.30	0.43	0.43	0.45	0.75

Notes: Logit regressions. Dependent variable: Player 2's decision to choose C if Player 1 reports message X in part 2. NO S serves as baseline treatment in all regressions. We add dummies for rounds in columns (3) – (5). Robust standard errors are clustered at the matching group level, p-values given in parentheses: ** p<0.05, *** p<0.01.

Table A.5: Robustness check, Player 1's decision to lie in part 3

	(1)	(2)	(3)
FINE	-1.40***	0.02	0.05
	(0.00)	(0.97)	(0.93)
COMP	-1.25***	-0.04	-0.12
	(0.01)	(0.94)	(0.84)
Player has lied before		1.07*	0.79
		(0.09)	(0.18)
A previous lie was successful		2.72***	3.02***
		(0.00)	(0.00)
Gender: Female			0.50
			(0.26)
Risk proneness			0.11
			(0.28)
Constant	1.95***	-0.64	-1.54*
	(0.00)	(0.19)	(0.05)
Observations	168	168	164
Pseudo R-squared	0.05	0.38	0.38
Comparing FINE and COMP Wald test results (p-values)	0.71	0.90	0.75

Notes: Logit regressions. Dependent variable: Player 1's decision to lie in part 3. NO S serves as baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: * p<0.10, *** p<0.01.

Table A.6: Robustness check, Player 2's decision to choose C in part 3

	(1)	(2)	(3)
Fine	0.37	-0.19	-0.18
	(0.53)	(0.72)	(0.75)
Comp	-0.37	-0.63	-0.60
	(0.47)	(0.24)	(0.26)
A previously matched Player 1 lied		-2.31**	-2.23**
		(0.04)	(0.05)
Gender: Female			-0.12
			(0.83)
Risk proneness			-0.15
			(0.16)
Constant	1.45***	3.75***	4.63***
	(0.00)	(0.00)	(0.00)
Observations	143	143	143
Pseudo R-squared	0.02	0.07	0.09
Comparing Fine and Comp Wald test results (p-values)	0.41	0.46	0.50

Notes: Logit regressions. Dependent variable: Player 2's decision to choose C if Player 1 reports message X in part 3. No S serves as baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: ** p<0.05, *** p<0.01.

B Translated instruction

Welcome to today's experiment!

You are taking part in a study on decision-making in which you can earn money. You will receive EUR 5 for showing up on time. Your further pay-out depends on your decisions and the decisions of other participants matched to you, but also on which role you are assigned. Please read and follow the instructions carefully. They contain everything you need to know for your participation. At the end of the experiment, we kindly ask you to answer a short questionnaire.

Please note that from now on and throughout the experiment, **communication is not allowed**. If you have a question, please raise your hand. One of the experimenters will then come to you. The use of mobile phones, smartphones, tablets or similar is prohibited throughout the experiment. Please note that failure to comply will result in exclusion from the experiment and all payments. All decisions will be made anonymously, i.e. none of the participants will know the identity of the other. Also the payments will be made anonymously at the end of the experiment.

Instructions

What is it about? – An overview

In this experiment, two participants – Person 1 and Person 2 – will be anonymously matched to each other. Person 1 and Person 2 will each make a choice between two *options*. Depending on the *situation*, one or the other option may be more advantageous for each Person.

Your payoff depends, firstly, on which option you choose and which option the participant matched to you chooses. Secondly, it depends on whether you have the role of Person 1 or Person 2. Thirdly, it depends on which of the possible situations – X or Y – prevails. The chart below describes what the payoffs (denoted in points) are for the different combinations of options chosen by Person 1 and Person 2 and depending on whether situation X (left table) or Y (right table) prevails.

Payoffs in situation X

		Person 2	
		Option C	Option D
Person 1	Option A	20, 20	10, 10
	Option B	10, 10	0, 0

Payoffs in situation Y

		Person 2	
		Option C	Option D
Person 1	Option A	10, 0	0, 10
	Option B	20, 10	10, 20

In situation X, the following applies:

- If Person 1 chooses option A and Person 2 chooses option C, then Person 1 and Person 2 both get paid 20 points each.
- If Person 1 chooses option A and Person 2 chooses option D, then Person 1 and Person 2 both get paid 10 points each.
- If Person 1 chooses option B and Person 2 chooses option C, then Person 1 and Person 2 both get paid 10 points each.
- If Person 1 chooses option B and Person 2 chooses option D, then Person 1 and Person 2 both get paid 0 points each.

In situation Y, the following applies:

- If Person 1 chooses option A and Person 2 chooses option C, then Person 1 gets paid 10 points and Person 2 gets paid 0 points.
- If Person 1 chooses option A and Person 2 chooses option D, then Person 1 gets paid 0 points and Person 2 gets paid 10 points.
- If Person 1 chooses option B and Person 2 chooses option C, then Person 1 gets paid 20 points and Person 2 gets paid 10 points.
- If Person 1 chooses option B and Person 2 chooses option D, then Person 1 gets paid 10 points and Person 2 gets paid 20 points.

Please note:

1. The situation is randomly determined by the computer; **both situations, X and Y are equally likely**, i.e. they are each realised with 50 percent probability. The situation determined by the computer applies to both Persons matched to each other; i.e. both Person 1's and Person 2's payoffs are determined either by the left table or by the right table. Thus, one could also say that the computer randomly draws one of the two tables for both Persons, with both tables being equally likely.

2. **Only Person 1 learns which of the two possible situations** – situation X or situation Y – actually prevails. The computer informs him or her about it at the beginning of the experiment. Afterwards, Person 1 can inform Person 2 about which situation. He or she is obliged to transmit one piece of information – X or Y.

3. In order to make a choice between the 2 options in each case, the Persons matched to each other go through a two-stage process. **At the first stage, Person 1 can inform Person 2** which of the two situations has been indicated to him or her. **At the second stage, Person 1 and Person 2 then choose** one of their two **options**.

1. Experimental procedures

The experiment consists of 3 parts. In the following we describe part 1 of the experiment. You will receive the instructions for part 2 and part 3 at the beginning of the respective part.

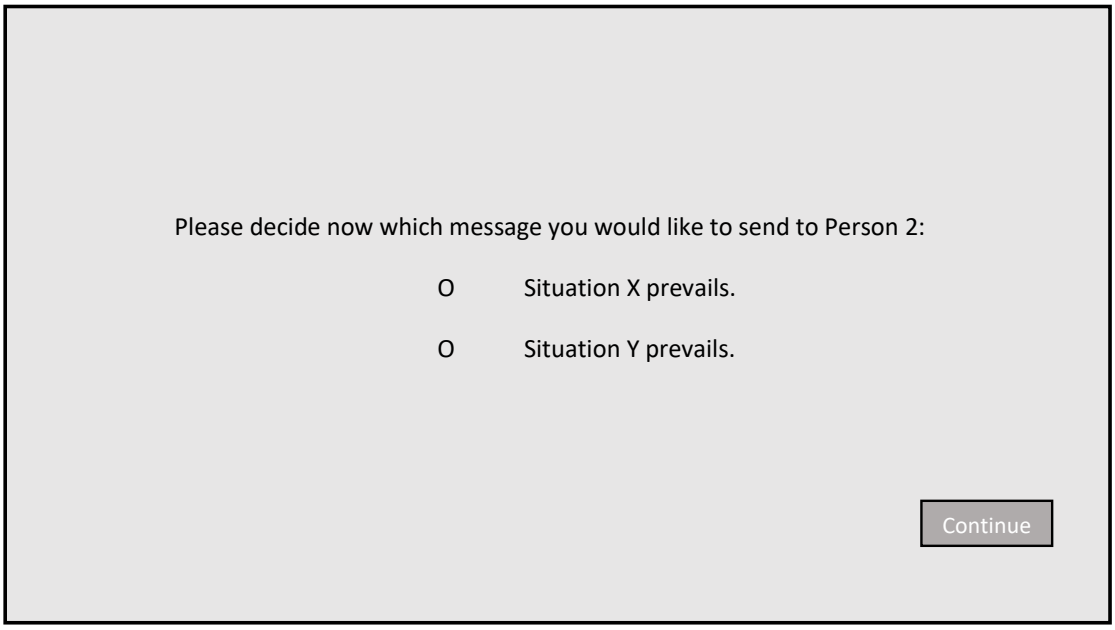
2. Allocation of roles

At the beginning of the experiment, the computer randomly assigns each participant either the role of Person 1 or Person 2. You will keep this role throughout all 3 parts of the experiment.

3. Procedure of the decision round in part 1

Person 1 receives information at the beginning of part 1 as to whether situation X or situation Y prevails. Person 2 does not receive any information.

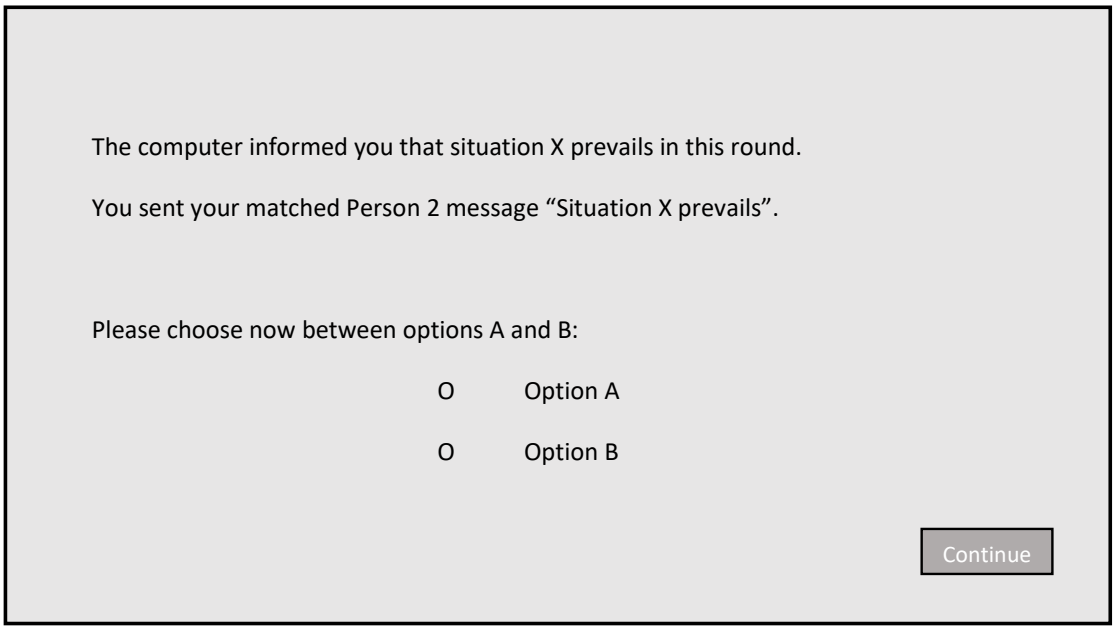
Then Person 1 can inform Person 2 about which situation prevails. He or she is obliged to transmit one piece of information – X or Y. The screen looks as follows:



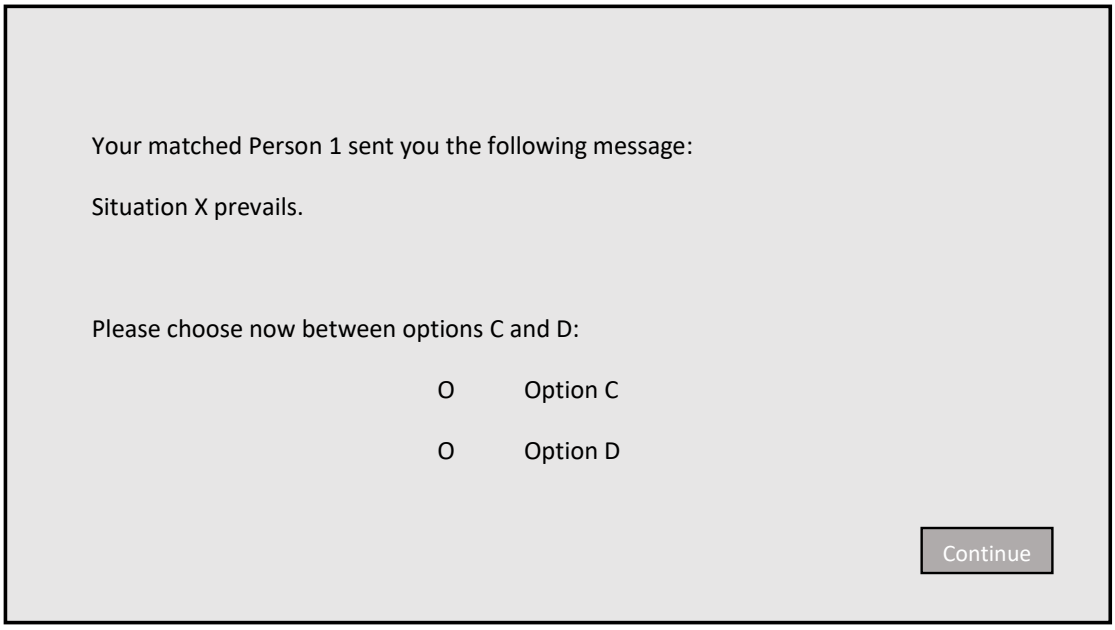
Next, Person 1 and Person 2 choose between their options. Person 1 makes a choice between Option A and Option B, Person 2 makes a choice between Option C and Option D.

Since both Persons make their choices simultaneously, at this point, they do not know yet which choice the other Person has made. Therefore they have to form expectations about which of the two possible options was chosen by the other Person.

Example: Decision screen of Person 1:



Example: Decision screen of Person 2:



4

Finally, both Persons are informed which options were chosen by Person 1 and Person 2. In addition, both Persons are told which situation – X or Y – actually prevailed and how many points Person 1 and Person 2 receive.

[In Treatment Fine additionally:]

In addition, the computer randomly checks every third participant in the role of Person 1 in this decision round:

- ***If** the computer discovers that a checked Person 1 has transmitted **false information** about the situation – X or Y – to Person 2, **10 points will be deducted from Person 1's original round payoff**.*
- ***Otherwise**, i.e. if a Person 1 has not been checked or if a checked Person 1 has transmitted correct information about the situation, **no points will be deducted from their original round payoff**.*

[In Treatment Comp additionally:]

In addition, the computer randomly checks every third participant in the role of Person 1 in this decision round:

- ***If** the computer discovers that a checked Person 1 has transmitted **false information** about the situation – X or Y – to Person 2, **10 points will be deducted from Person 1's original round payoff. This amount is then added to Person 2's round payoff**.*
- ***Otherwise**, i.e. if a Person 1 has not been checked or if a checked Person 1 has transmitted correct information about the situation, **no points will not be deducted from their original round payoff and no amount is added to Person 2's round payoff**.*

4. Pay-out from today's experiment

In part 1 of the experiment, you will make only 1 decision, in part 2 of the experiment you will make 4 decisions and in part 3 of the experiment you will again make only 1 decision. At the end of the experiment, 1 of your decisions will be randomly drawn to determine your final pay-out. All decisions from the 3 parts of the experiment are equally likely to be drawn.

The final pay-out you earn from the drawn decision is converted into Euros, with the following **exchange rate: 1 point = EUR 0.40**. The resulting amount plus the show-up fee of EUR 5 is your total pay-out from today's experiment.

Control questions

1. The computer matches one Person 1 and one Person 2 each. In general: Does the computer inform **Person 1** or **Person 2** about which situation actually prevails?

   Answer: _____

2. What is the probability that situation X prevails?

   Answer: _____%

3. Suppose you are Person 2, situation X prevails, Person 1 chose option A and you chose option C. How much do you earn when this decision round is randomly drawn to be paid out?

   In points: _____

4. Suppose you are Person 2, situation Y prevails, you chose option D and Person 1 chose option B. How much do you earn when this decision round is randomly drawn to be paid out?

   In points: _____

5. Suppose you are Person 1, situation Y prevails, you chose option B and Person 2 chose option C. How much do you earn when this decision round is randomly drawn to be paid out?

   In points: _____

6. a) The experiment consists of 3 parts. How many decisions (without knowing further details about part 2 and 3) are you going to take in parts 1, 2 and 3?

   In part 1: _____

   In part 2: _____

   In part 3: _____

   b) How many of these decisions are randomly drawn by the computer and paid out to you at the end of the experiment?

   Answer: _____

7. Will Person 1 incur financial consequences if he or she transmits false information, i.e. transmit a different situation than the actually prevailing one, to Person 2?
   O yes

   O no

[In Treatment Fine and Treatment Comp instead:]

8. How many participants are randomly checked by the computer?
   O every second

   O every third

   O every forth

[In Treatment Fine and Treatment Comp additionally:]

8. What amount will then be deducted from Person 1's round payoff if he or she transmits false information?

   In points: _____

[In Treatment Comp additionally:]

9. Who then receives the amount deducted?

   Answer: _____

[The instructions for part 2 and 3 are only displayed on participants' computer screens:]

Instructions for part 2

You continue to keep your role from part 1 in part 2. That is, if you previously had the role of Person 1, you continue to be Person 1 and are informed by the computer as to whether situation X or situation Y prevails**.**

If you previously had the role of Person 2, you continue to be Person 2 and receive no information about the situation from the computer.

Part 2 of the experiment comprises 4 decision rounds. The payoffs in a given decision round depend only on what happens in that decision round – they are independent of part 1 and of the other decision rounds in part 2. Similarly, the prevailing situation in a given decision round is independent of part 1 and of the other decision rounds in part 2.

You will be matched to a new Person in each of the 4 decision rounds. This could be any Person except the ones you were matched to before. If you have the role of Person 1, you will be matched to a new Person 2 in each decision round. If you have the role of Person 2, you will be matched to a new Person 1 in each round.

Each of the 4 decision rounds in part 2 follows basically the same procedure as the decision round in part 1.

- As a reminder: This means that, first, Person 1 receives information at the beginning of each decision-making round as to whether situation X or situation Y prevails. Person 2 does not receive information.

- Next, Person 1 can inform Person 2 which situation prevails. He or she is obliged to transmit one piece of information – X or Y.

- After that, Person 1 and Person 2 simultaneously choose between their options. Person 1 makes a choice between option A and option B, Person 2 makes a choice between option C and option D.

- Finally, both Persons are informed which options were chosen by Person 1 and Person 2. In addition, both Persons are told which situation – X or Y – actually prevailed and how many points Person 1 and Person 2 earned.

[In Treatment Fine additionally:]

In addition, the computer randomly checks every third participant in the role of Person 1 in each of the 4 decision rounds:

***If** the computer discovers that a checked Person 1 has transmitted **false information** about the situation – X or Y – to Person 2, **10 points will be deducted from Person 1's original round payoff**.*

***Otherwise**, i.e. if a Person 1 has not been checked or if a checked Person 1 has transmitted correct information about the situation, **no points will be deducted from their original round payoff**.*

[In Treatment Comp additionally:]

In addition, the computer randomly checks every third participant in the role of Person 1 in each of the 4 decision rounds:

***If** the computer discovers that a checked Person 1 has transmitted **false information** about the situation – X or Y – to Person 2, **10 points will be deducted from Person 1's original round payoff. This amount is then added to Person 2's round payoff**.*

***Otherwise**, i.e. if a Person 1 has not been checked or if a checked Person 1 has transmitted correct information about the situation, **no points will not be deducted from their original round payoff and no amount is added to Person 2's round payoff**.*

As a reminder: At the very end of the experiment, 1 of your decisions will be randomly drawn from 1 of the 3 parts of the experiment to determine your final pay-out. All decisions from the 3 parts of the experiment are equally likely to be drawn. In part 1 of the experiment, you made only 1 decision, in part 2 of the experiment you will make 4 decisions and in part 3 of the experiment you will again make only 1 decision.

If you have any questions about the instructions (now or later), please raise your hand. The experimenter will then come to you. Please do not hesitate to ask questions if you are in doubt.

Please click "continue" to start part 2 of the experiment.

Instructions for part 3

You continue to keep your role from part 1 and part 2. That is, if you previously had the role of Person 1, you continue to be Person 1 and are informed by the computer as to whether situation X or situation Y prevails**.**

If you previously had the role of Person 2, you continue to be Person 2 and receive no information about the situation from the computer.

Part 3 of the experiment comprises only 1 decision round. The payoffs in this decision round depend only on what happens in this decision round – they are independent of the other decision rounds in part 1 and part 2.

You will be matched to a new Person. This could be any Person except the ones you were matched to in part 1 or part 2. If you have the role of Person 1, you will be matched to a new Person 2. If you have the role of Person 2, you will be matched to a new Person 1.

The decision round in part 3 follows basically the same procedure as the decision rounds in part 1 and part 2.

- As a reminder: This means that, first, Person 1 receives information as to whether situation X or situation Y prevails. Person 2 does not receive information.

- Next, Person 1 can inform Person 2 which situation prevails. He or she is obliged to transmit one piece of information – X or Y.

- After that, Person 1 and Person 2 simultaneously choose between their options. Person 1 makes a choice between option A and option B, Person 2 makes a choice between option C and option D.

- Finally, both Persons are informed which options were chosen by Person 1 and Person 2. In addition, both Persons are told which situation – X or Y – actually prevailed and how many points Person 1 and Person 2 earned.

[In Treatment Fine additionally:]

Important difference to part 1 and part 2: In part 3, the **computer no longer checks whether participants in the role of Person 1 transmitted false information** about the situation – X or Y – to Person 2. So, if the information is false, Person 1 will no longer have 10 points deducted from their original round payoff.

[In Treatment Comp additionally:]

Important difference to part 1 and part 2: In part 3, the **computer no longer checks whether participants in the role of Person 1 transmitted false information** about the situation – X or Y – to Person 2. So, if the information is false, Person 1 will no longer have 10 points deducted from their original round payoff and this amount is no longer added to Person 2's round payoff.

After part 3, the experiment ends with a short questionnaire.

As a reminder: At the very end of the experiment, 1 of your decisions will be randomly drawn from 1 of the 3 parts of the experiment to determine your final pay-out. All decisions from the 3 parts of the experiment are equally likely to be drawn. In part 1 of the experiment, you made only 1 decision, in part 2 of the experiment you made 4 decisions and in part 3 of the experiment you will again make only 1 decision.

If you have any questions about the instructions (now or later), please raise your hand. The experimenter will then come to you. Please do not hesitate to ask questions if you are in doubt.

Please click "continue" to start part 3 of the experiment.

Questionnaire

You have now reached the end of the experiment. Before your screen displays the information on your pay-out from the experiment, we would like to ask you to answer the following questions as precisely as possible. Your answers will be analysed anonymously, and it will be impossible to trace your identity.

Are you...?

O male
O female
O prefer not to say

How old are you?

_____ (Free text field)

What is your nationality?

_____ (Free text field)

What subject are you studying?

_____ (Free text field)

How many experiments at WISO research lab have you already participated in?

_____ (Free text field)

On a scale from 1 to 10, are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

Not at all willing to take risks	1 – O O O O O O O O O O – 10	Very willing to take risks

On a scale from 1 to 10, would you say that, in general, most people can be trusted or that you can't be too careful?

You can't be too careful	1 – O O O O O O O O O O – 10	Most people can be trusted

On a scale from 1 to 10, how important is sustainability to you in general?

Not important at all	1 – O O O O O O O O O O – 10	Very important

On a scale from 1 to 10, how important is the existence of a fair legal system to you in general?

Not important at all	1 – O O O O O O O O O O – 10	Very important

On a scale from 1 to 10, how did you feel you were treated in your role as Person 1 or Person 2 under the experimental conditions that were in place in parts 1 and 2?

Not treated fairly at all	1 – O O O O O O O O O O – 10	Treated very fairly

On a scale from 1 to 10, how effective did you perceive the deterrent effect of the experimental condition on lying behaviour in parts 1 and 2?

No effective at all	1 – O O O O O O O O O O – 10	Very effective

Were there parts of the experiment that you found confusing? If so, we would appreciate it if you could briefly tell us about them.

_____ (free text field)