

# Is Compensation Fine?

## Sanction Schemes and their Effects on Deterrence and Trust

Pieter Desmet\*

Leonie Gerhards<sup>†</sup>

Franziska Weber<sup>‡</sup>

February, 2024

### 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. This is, however, only partly reflected in potential victim's trust. It appears that the higher levels of trust we observe in the fine regime are not necessarily a consequence of the sanction scheme itself, but rather because individuals experience less misbehaviour in that regime.

**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](mailto:desmet@law.eur.nl)

<sup>†</sup>King's Business School, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, UK. Email: [leonie.gerhards@kcl.ac.uk](mailto:leonie.gerhards@kcl.ac.uk)

<sup>‡</sup>Rotterdam Institute of Law and Economics/Erasmus School of Law, Erasmus University Rotterdam; Burg. Oudlaan 50, 3062 PA Rotterdam. Email: [weber@law.eur.nl](mailto: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<sup>1</sup> 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 have traditionally used different sanctions to remedy the very same unfair practices. In Spain, for instance, since 2010 consumers can make individual claims for compensation.<sup>2</sup> In Italy on the other hand, the provisions are traditionally enforced by way of public law fines.<sup>3</sup>

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 societal 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 (that is, to trust, Rousseau et al., 1998).<sup>4</sup>

To fill these gaps, we design a novel laboratory experiment that allows us to simultaneously compare the 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 misbehaviour to an extent that would be infeasible with field data.

We set out to compare the effect of fine and compensation schemes in a 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 believe and 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

---

<sup>1</sup>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.

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

<sup>3</sup>See Part II Art. 27 of Codice del Consumo as introduced by legislative decree No. 146 of 2007.

<sup>4</sup>A recent exception is Friehe and Do (2023), who explore the adverse impact of becoming a victim of crime on people's future trust.

respective effects, across treatments FINE and COMP, we vary the type of sanction (fines versus compensation), but keep the size of the sanction payment the same. We also include a control treatment with no sanction regime (NO S) (short for “No Sanction”).

The experiment consists of three parts. In part 1 we analyse ex-ante compliance and trust 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 and sanctioned for lying. In part 3, sanctions are lifted in order to test for lasting effects of the sanction schemes.

The main findings are as follows. 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 higher in FINE than in NO S in part 2 of the experiment. A closer inspection reveals that this increased level of trust results from the compliance that the threat of fines induces on the potential infringers. Rather than a FINE effect per se, having encountered an infringer seem to shape victims’ future trust. That is, since FINE turns out most successful in reducing misbehaviour, it also creates the highest level of trust on the potential victim’s side. Apart from the difference between FINE and NO S in part 2, we do not observe any other significant treatment differences in trust in any of the three parts.

The paper proceeds as follows. In the next Section, we review existing research on the effects of different sanction regimes on deterrence and compliance as well as research on the effects of these sanctions on trust. In Section 3 we introduce our experiment and derive behavioural hypotheses. We present our findings from experiment parts 1 and 2 in Section 4 and the results from the sanction free part 3 in Section 5. We end the paper with a discussion in Section 6 and a conclusion in Section 7.

## 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 actual infringers (see, for instance, Andreoni, 1991; Bar-Ilan and Sacerdote, 2004; Cooter, 1988; Garoupa, 2001; Miceli et al., 2022; Polinsky and Shavell, 1992 or Stigler, 1970). 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 (Miceli, 2023; 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 as a function of 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 misbehaviour, 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. Friehe et al. (2023) conduct a lab experiment to investigate how individuals update their beliefs about the probability of

detection when being exposed to either severe or mild sanctions. They find that the magnitude of the sanction – which should be irrelevant – does in fact influence how individuals update their beliefs about said probability.

Arguably, due to its simplicity and its associated methodological advantages, fine schemes have been subject of many studies. Contrarily, compensation schemes have received comparably less attention (for an exception with hypothetical vignettes, see [Cardi et al., 2012](#)). Even fewer studies compare fines and sanctions ([Mulder, 2018](#)). [Kurz et al. \(2014\)](#) 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 consider situations in which victims are actually the beneficiaries 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 imposes 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.

Most closely related to our study are the experiments by [Baumann et al. \(forthcoming\)](#), [Desmet and Weber \(2022\)](#), [Feldman and Teichman \(2008\)](#), [Kornhauser et al. \(2020\)](#) and [Metcalfe et al. \(2020\)](#). As a part of their comprehensive study, [Feldman and Teichman \(2008\)](#) examine the impact of probabilistic fines and compensations on potential infringers’ wrongdoing, similar to us. They do, however, use non-incentivised, hypothetical vignettes and do not study compliance directly, but consider how the prospect of the respective sanction scheme affects potential injurers’ economic decisions as well as their moral reasoning and perceptions of wrongdoing. Their findings suggest that wrongdoing is perceived as more unethical and less appropriate under a fine than under a compensation scheme. In a further vignette study, [Desmet and Weber \(2022\)](#) observe that infringers’ willingness to pay a sanction is higher under a compensation than under a fine scheme. Yet, infringers are similarly willing to take precautionary measures

under both sanction schemes. It is of note, however, that this study operationalises infringements as non-intentional acts and does not focus on deterrence in particular. Baumann et al. (forthcoming) conduct a lab experiment to compare the effects of fines and compensation on investments 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.

Kornhauser et al. (2020) set out to test Gneezy and Rustichini’s (2000) a-fine-is-price hypothesis by studying contract breaches in a lab experiment. In doing so, as a byproduct, they also compare the effectiveness of fines paid to the experimenter to compensation paid to the contracting party. Only when focusing on pro-social individuals, they do find more compliant behaviour under fines than under compensation. Metcalf et al. (2020), similarly, intend to replicate Gneezy and Rustichini’s (2000) findings in the original daycare as well as in a new tax reporting context, applying a vignette-based experimental survey on Amazon Mechanical Turk (MTurk). Different from the original study, they find that individuals’ compliance increases once fines are introduced – and decreases again once fines are removed.

We build on previous research and make some important contributions beyond existing studies. First of all, in contrast to Kornhauser et al. (2020) and Metcalf et al. (2020), who use certain sanctions, our experiment tests a more realistic scenario of probabilistic fines and compensation. Secondly, in contrast to Feldman and Teichman (2008) and Metcalf et al. (2020), we use an incentivised experimental game instead of hypothetical vignettes to test for the effect of these sanctions. Thirdly, we do not only employ a one-shot set-up to observe ex ante deterrence, but also include repeated interaction where participants are re-matched in a perfect stranger fashion. This allows us to study infringers’ learning after having experienced the treatment-specific sanction scheme, such as having been checked for or even been caught lying. Finally, apart from studying the (potential) infringer’s deterrence, our design also allows to study the (potential) victim’s side, enabling us to find out which sanction regime can generate and retrieve more trust.

## **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, to trust). 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). Trusting someone involves interpersonal risks based on evaluating the other person's intentions and behaviour.

The presence of a sanction regime 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. Under compensation schemes, on the other hand, potential victims' probability of loss is effectively reduced, irrespective of the perceived deterrent capacity of compensation.

Trust that 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): The introduction of a sanction scheme will increase potential victims' willingness to trust 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.

If we look at the existing literature on sanctions and trust, some critical gaps become clear. First of all, many studies that consider the effects of sanctions on trust take an ex-post perspective, focusing on *actual* victims' reactions to receiving compensation or to seeing an infringer being punished, rather than considering the ex ante inclination of *potential* victims to make themselves vulnerable under different sanction systems (see for instance 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 disregarding the spill-over effects that betrayal and sanctions may have on trust in subsequent interactions with new interaction partners.

Similar to the literature on sanctions and infringers, studies that do look into ex-ante trust have not directly compared the relative effects of compensation and fines on potential victims' trust. Vollan (2011) observes the effects of (potential) third party punishment on ex ante trust using one-shot trust games for people interacting with strangers and finds 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 set of lab experiments. 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 (for instance by a compensation scheme) will increase trust. However, these authors do not exactly study the presence of a compensation scheme, but rather look at the effects of automatic contract enforcement where the probability of enforcement is 100%.

Similarly, focusing on trust in contractual relations, Bohnet et al. (2001) study 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**

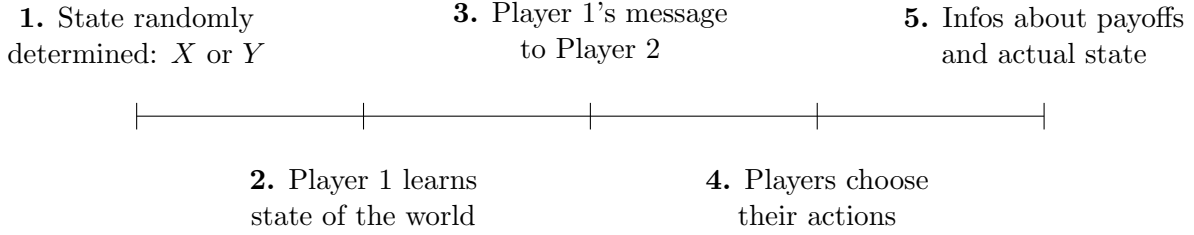
#### **3.1 Design**

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, reproduced in Appendix B, we refer to them as Player 1 and Player 2, respectively. The experimental game evolves in five stages, summarised in Figure 1.

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



Figure 1: The experimental game



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) players obtain given the actions chosen and the prevailing state of the world.<sup>5</sup> This table is similarly presented and carefully explained in the instructions. Hence, 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, 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 Player 1's incentive to lie explicit. Player 1 has no incentive to lie in state  $X$ .<sup>6</sup> When studying treatment differences in lying, we therefore focus on state  $Y$ . Analogously, we analyse treatment differences in Player 2's trust by considering their propensity to choose action C when their matched Player 1 reports that state  $X$  prevails.

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 learns if they have been lied to at the end of each round. 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' mes-

<sup>5</sup>In the experiment, one point corresponds to 0.40 Euro.

<sup>6</sup>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 versus 10.

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

Monetary payoffs in state $X$ :				Monetary payoffs in state $Y$ :			
		Player 2				Player 2	
		C	D			C	D
Player 1	A	20,20	10,10	Player 1	A	10,0	0,10
	B	10,10	0,0		B	20,10	10,20

*Notes:* Payoffs denoted in points.

sages are randomly checked for truthfulness. 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 corresponds to the victim's loss that results from reaching (B,C) instead of (B,D) in state  $Y$ .<sup>7</sup> Given the experimental parameters, the sanction is non-deterrent in expectations (which is a realistic assumption in many contexts). In expectations, 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.

### 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 and they take their first decisions. 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

---

<sup>7</sup>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.

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 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 checked for or even 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 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 ran a total of 16 sessions, with 96 participants in NO S and 144 participants each in FINE and COMP.<sup>8</sup> 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 51% of participants were female, in FINE 50% and in COMP 51%.

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.<sup>9</sup> We hence conclude that the overwhelming part of participants understands the rules of the game.

---

<sup>8</sup>We did not pre-register our study as pre-registration was not yet considered the norm at the time our experiment was conducted. Nowadays, this practice is more common as a means to prevent post-hoc theorising and the non-publication of null results. We would like to point out, however, that our experimental design and empirical analyses are clearly guided by our hypotheses, which are grounded in existing theories. To optimise resource utilisation, we conducted six sessions for each of the sanctions treatments – as we anticipated smaller differences between these treatments – and four sessions for our benchmark treatment NO S – expecting comparably larger differences between NO S and FINE and COMP, respectively. Lastly, we would like to note that we comprehensively present all our findings in this paper, including those that do not support our hypotheses.

<sup>9</sup>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.

### 3.3 Predictions

Firstly, we note that 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 believe and 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, that is, play C with probability  $p_C \in [0, 1]$ .<sup>10</sup>

Ample evidence shows that many individuals do not act in a purely selfish manner, but reveal social preferences and exhibit norm-abiding behaviour.<sup>11</sup> For the context of the present experiment, one can plausibly argue that the sanction schemes in FINE and COMP convey social norms that condemn and deter lying, even if the implemented sanctions are non-deterrent in expectations. This arguably increases average compliance on the part of Player 1s – as evidenced by the experimental findings discussed in Section 2.1 – and consequently promotes deterrence-based trust (Lewicki et al., 1996) on the part of Player 2s in FINE and COMP compared to NO S. This leads to our first behavioural prediction:

**Hypothesis 1:** *Due to compliance effects in FINE and COMP, Player 1s lie less often and*

<sup>10</sup>If Player 2s always play D, expected payoffs in NO S and FINE are  $15 = \frac{1}{2} \times 10 + \frac{1}{2} \times 20$ , in COMP:  $16\frac{2}{3} = \frac{1}{2} \times 10 + \frac{1}{2} \times 23\frac{1}{3}$ . If Player 2s mix, that is, play C with probability  $p_C \in [0, 1]$ , expected payoffs in NO S and FINE are  $15 = \frac{1}{2}(20p_C + 10(1-p_C)) + \frac{1}{2}(10p_C + 20(1-p_C))$ , 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 (that is, irrespective of state) send message “State  $X$  prevails”, play their state-specific dominant action and Player 2s always (that is, 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.

<sup>11</sup>For a recent overview on social preferences, see Drouvelis (2021). For a survey on norm-abiding behaviour, see Legros and Cislighi (2020).

*Player 2s trust more in FINE and COMP than in NO S in part 1 and 2 of the experiment.*

Besides, the experimental findings by Feldman and Teichman (2008) and Kornhauser et al. (2020) suggest that we should observe more compliance in FINE than in COMP. Baumann et al. (forthcoming) highlight the role that the infringers' guilt aversion can play in this regard. Guilt aversion considers in how far people care about others' expectations and anticipate feelings of guilt if they fall short of their expectations (cf. Charness and Dufwenberg, 2006 and Battigalli and Dufwenberg, 2007). A defining difference between fines and compensation in this respect is that only in the latter type of sanction scheme, infringers can ex-post "morally cleanse" themselves when they compensate victims for their losses. Following this reasoning, Player 1s are less inclined to lie in treatment FINE than in COMP since in the former treatment, Player 1s can reduce their (expected) feelings of guilt only by reporting the true state of the world. If Player 2s anticipate these treatment effects, they should believe and act upon Player 1s' messages more often in FINE than in COMP. We summarise the resulting behavioural predictions as follows:

**Hypothesis 2:** *Assuming some degree of guilt aversion on the part of Player 1s, we expect to see 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 discussed by Mulder et al. (2006). Based on this, we expect that Player 1s' comparably greater truthfulness and Player 2s' 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:

**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

In our analysis, we are not primarily interested in *absolute* levels of Player 1s' compliance and Player 2s' trust, but rather in the *relative* differences in these variables that emerge from the treatment-specific sanction schemes. We start our empirical analysis with an overview of some basic statistics on misbehaviour across treatments. In our experiment, we define misbehaviour

as lying about the state of the world if state  $Y$  prevails.<sup>12</sup> 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 most right columns of Table 2.<sup>13</sup>

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

		$p$ -values from Fisher's exact tests (two-sided)		
		NO S vs FINE	NO S vs COMP	FINE vs COMP
<u>Share of lies about state <math>Y</math></u>				
NO S	70.14%	] < 0.001	] < 0.001	] 0.001
FINE	30.88%			
COMP	47.06%			
<u>Share of lying Player 1s</u>				
NO S	87.50%	] < 0.001	] 0.001	] 0.043
FINE	44.44%			
COMP	62.50%			
<u>Share of Player 2s who encounter at least one liar</u>				
NO S	100%	] < 0.001	] < 0.001	] 0.064
FINE	63.89%			
COMP	79.17%			

*Notes:* We record a lie if Player 1 reports message  $X$  if in fact state  $Y$  prevails. We define Player 1 being a liar if they lie about state  $Y$  at least once.

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. Out of the respective full samples of Player 1s (not conditional on being checked), 20.83% 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

<sup>12</sup>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.

<sup>13</sup>Note that here and throughout the paper we report  $p$ -values from two-sided test statistics.

which participants play the same one-shot game another four times with random new matching partners, allowing for learning effects, net of additional reputational effects.

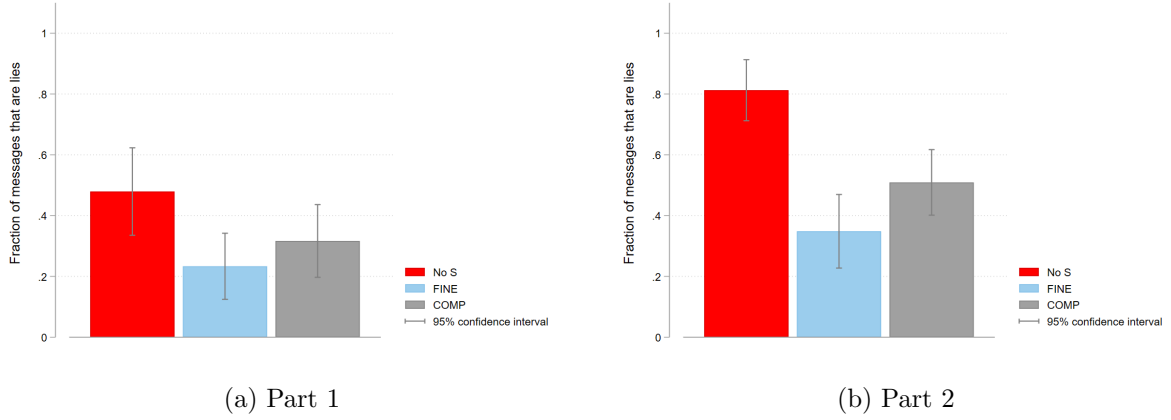


Figure 2: Player 1's misbehaviour in parts 1 and 2

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, participants are comparably more truthful. Only 23% of participants lie in FINE, while slightly more (32%) lie in COMP. In particular the difference between NO S and FINE is highly significant (Fisher's exact test results: NO S versus FINE:  $p < 0.01$ , NO S versus COMP:  $p = 0.11$ ). 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. There, also the treatment difference between NO S and COMP turns out marginally significant ( $p = 0.09$ ). In columns (2) and (3), we extend the regression and control for all items elicited in the final questionnaire.<sup>14</sup> The finding that the FINE treatment significantly reduces lying compared to NO S remains virtually unchanged. The anyway weaker treatment differences between COMP and FINE and COMP and NO S, on the other hand, are less robust to the inclusion and omission of control variables.

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-

<sup>14</sup>We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the total 16 sessions. Column (2) shows the specification that includes all control variables except “# Number participated in so far”. In column (3) we additionally control for this variable, to capture any potential effects from participants' general experience with lab experiments. This reduces the number of observations from 168 to 132. Apart from the FINE treatment effect, only participants' perception of the treatment specific deterrent effect consistently reduces their propensity to lie. Participants' experience with experiments, conversely, tend to increase their propensity to lie. We confirm the findings from the linear probability models reported in Table A.1 in further logit regressions. The findings are available from the authors upon request.

Whitney ranksum tests, comparing average choices aggregated at the matching group level now find significant treatment differences between all three treatments (NO S versus FINE:  $p < 0.01$ , NO S versus COMP:  $p = 0.01$ , FINE versus COMP:  $p = 0.04$ ).

Regression analysis (linear probability models) corroborates these findings. Column (1) in Table 3 reveals that also in part 2 of the experiment, Player 1s lie significantly less often when confronted with the sanction schemes 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 do. 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 ( $p = 0.01$ ).<sup>15</sup> Player 1s’ treatment-specific degrees of guilt aversion can readily explain this latter difference.

Columns (2) and (3) confirm that the treatment differences remain significant, though smaller than those estimated in column (1), if we control for whether a Player 1 has lied in any previous round (dummy variable “Player has lied before”) and whether any of the previous lies were successful (dummy variable “A previous lie was successful”). By the latter, we mean whether a lie about the state of the world made the matched Player 2 trust and choose C instead of D in any of the previous rounds. As evident, in particular the former of the two factors increases Player 1s’ propensity to lie again.

We replicate these findings in further regressions presented in Table A.2 in Appendix A, in which we control for rounds as well as for all items elicited in the final questionnaire.

**Finding 1:** *In line with Hypothesis 1, we observe less lying in FINE and COMP than in NO S in part 1 and 2; both sanction schemes promote compliance. The repeated setting of part 2, furthermore, corroborates Hypothesis 2: Player 1s lie less in FINE than in COMP.*

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 compliance effects of having experienced the law in either way in FINE and COMP.<sup>16</sup> However,

<sup>15</sup>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 individuals behave less pro-socially in repeated settings than in one-shot interactions.

<sup>16</sup>In these regressions, the FINE and COMP coefficients indicate the treatment effects on those Player 1s who have *not* experienced the law in any previous round. Note that our specifications neither include a dummy for whether a “Player was caught lying before” nor for whether a “Player was checked for lying before” as their coefficients



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

	(1)	(2)	(3)	(4)	(5)
FINE	-0.47*** (0.00)	-0.30*** (0.00)	-0.30*** (0.00)	-0.31*** (0.00)	-0.27*** (0.00)
COMP	-0.28*** (0.00)	-0.17*** (0.00)	-0.16*** (0.00)	-0.18** (0.01)	-0.15 (0.20)
Player has lied before		0.50*** (0.00)	0.44*** (0.00)	0.40*** (0.00)	0.44*** (0.00)
A previous lie was successful			0.09 (0.27)	0.11 (0.21)	0.09 (0.28)
FINE $\times$ Player was caught lying before				0.11 (0.46)	
COMP $\times$ Player was caught lying before				0.06 (0.52)	
FINE $\times$ Player was checked for lying before					-0.04 (0.49)
COMP $\times$ Player was checked for lying before					-0.01 (0.93)
Constant	0.81*** (0.00)	0.49*** (0.00)	0.49*** (0.00)	0.50*** (0.00)	0.49*** (0.00)
Observations	384	384	384	384	384
Independent observations	31	31	31	31	31
R-squared	0.13	0.37	0.37	0.37	0.37
Comparing FINE and COMP Wald test results (p-values)	0.01	0.02	0.02	0.02	0.38

*Notes:* Linear probability models. Dependent variable: Player 1's decision to lie in part 2. NO S serves as baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

as evident from the insignificant coefficients of the interaction terms, having experienced the law in either way does not evoke additional compliance effects, in neither of the two sanction treatments. These findings are largely robust to the inclusion of controls for rounds and all items elicited in the final questionnaire, see Table A.2<sup>17</sup>, as well as to the choice of regression model, see the results from additional logit regressions in Table A.3 in Appendix A.

would, in principle, indicate the effects of these mechanisms in our benchmark treatment NO S. However, in NO S players could by design neither be checked for lying nor get caught. The estimated specifications therefore contain all necessary variables to study potential interaction effects between the sanction treatments (FINE and COMP, respectively) and the effects of having experienced the law in the treatment-specific sanction scheme.

<sup>17</sup>Note that in the regressions in Table A.2, we additionally find that in FINE, having been caught lying previously increases Player 1's propensity to lie again, see column (4). In column (6), the COMP coefficient now also turns out significant, suggesting that Player 1s in COMP who have not yet been checked for lying, have a lower propensity to lie than their counterparts from NO S. We do not want to overemphasise these findings as these additional significances particularly emerge in regressions in which we control for participants' experience with lab experiments, which reduces the number of observations to 312.

## 4.2 Player 2s trust in parts 1 and 2

In the following, we take a closer look at the potential victims' behaviour. We are primarily interested in treatment differences in Player 2s' trust, that is, differences in their propensity to choose action C when their matched Player 1 reported state  $X$  prevails.

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 versus FINE:  $p = 0.40$ , NO S versus COMP:  $p = 0.15$ , FINE versus COMP:  $p = 0.59$ ). We corroborate these findings in linear probability models, which allow us to control for participants' personal characteristics that we elicited in the final questionnaire. Trust levels are generally high and not significantly different across treatments, see Table A.4 in Appendix A.<sup>18</sup>

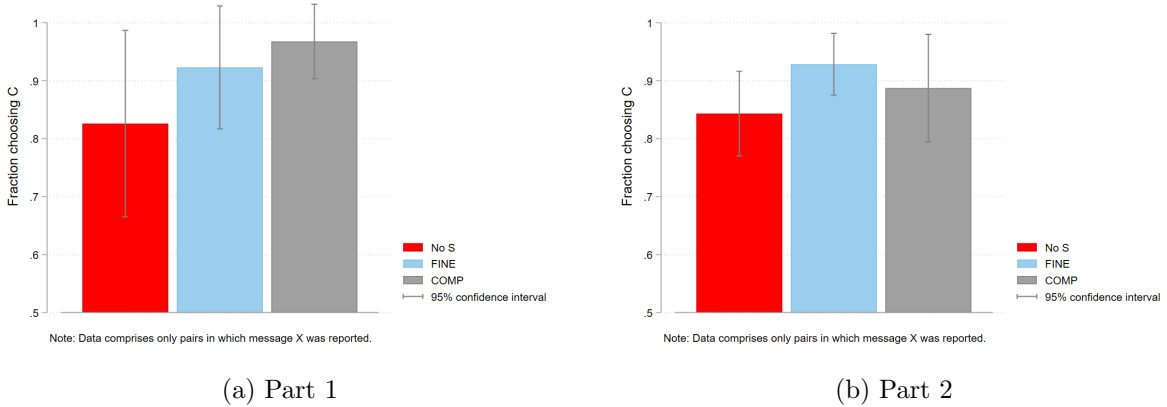


Figure 3: Player 2s choosing C in parts 1 and 2

The behavioural patterns change in part 2, consider graphics (a) and (b) in Figure 3. 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 versus FINE:  $p = 0.05$ , NO S versus COMP:  $p = 0.27$ , FINE versus COMP:  $p = 0.68$ ).

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

<sup>18</sup>Somewhat in line with the findings on Player 1s' propensity to lie increases in those participants' experience with lab experiments, in Table A.4 we similarly find that Player 2's trust reduces increases in those participants' experience with experiments. It should be noted, though, that the number of observations drops to 65 once we add this control. Further logit regressions confirm the insignificant treatment differences from the linear probability models reported in Table A.4, results available from the authors upon request.

behaviour in part 2 in more detail. Column (1) in Table 4 reveals that averaged over the four rounds, in NO S 84% ( $=0.84$ ) of Player 2s choose action C if their matched Player 1's reports state  $X$  prevails. The respective shares amounts to 92% ( $=0.84+0.08$ ) in FINE and to 87% ( $=0.84+0.03$ ) in COMP. Column (1) corroborates 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 the two sanction treatment coefficients are insignificant, revealing that there do not exist any further significant treatment differences.

However, also the general treatment effect of FINE proves rather fragile. 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, the respective coefficient turns out significantly negative: after having been lied to, Player 2's reduce their propensity to choose C by about 14 percentage points. The FINE treatment dummy is no longer significant.

**Finding 2:** *We find only limited evidence for the general treatment effects on trust that we posited in Hypothesis 1. Only in some of our analyses on part 2 we observe higher levels of trust in FINE than in NO S. Conversely, having been lied to in a previous round has a significantly negative effect on Player 2s' future trust, suggesting that Player 2s' propensity to trust is not directly shaped by the sanction regime, but rather by experience.*

We qualitatively and quantitatively replicate the findings from the linear probability models presented in Table 4 in further linear probability models in which we control for the personal characteristics that we elicited in the final questionnaire (Table A.5). Additional logit regressions further corroborate the findings from Table 4, see Table A.6, see Appendix A.

The question remains as to whether Player 2s might simply choose C less often after having experienced state  $Y$  – irrespective of having encountered a liar. One could, for instance, imagine that Player 2s overestimate the likelihood of state  $Y$  occurring or that they choose D with a higher probability after having experienced state  $Y$ , where this would have been the optimal choice. Additional robustness checks, reported in Table A.7 in Appendix A, however, confirm that Player 2s do not generally react to having experienced state  $Y$ . Instead, by choosing D they seem to genuinely respond to having encountered a liar previously.<sup>19</sup>

<sup>19</sup>In column (1) of Table A.7, we extend the original specification (2) of Table 4 by adding a dummy for whether a Player 2 has encountered state  $Y$  in any previous round. Neither the dummy's coefficient itself turns out significant, nor does the size or significance of the dummy "A previously matched Player 1 lied" decrease compared to the original specification. In column (2) of Table A.7, we keep the "State  $Y$  in any of the previous

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

	(1)	(2)	(3)	(4)
FINE	0.08*	0.04	0.05	0.03
	(0.06)	(0.28)	(0.18)	(0.49)
COMP	0.03	0.01	0.02	0.00
	(0.57)	(0.85)	(0.69)	(0.96)
A previously matched Player 1 lied		-0.14***	-0.13***	-0.14***
		(0.00)	(0.00)	(0.00)
FINE $\times$ A previously matched Player 1 was caught lying			-0.07	
			(0.51)	
COMP $\times$ A previously matched Player 1 was caught lying			-0.04	
			(0.66)	
FINE $\times$ A previously matched Player 1 was checked for lying				0.02
				(0.59)
COMP $\times$ A previously matched Player 1 was checked for lying				0.01
				(0.75)
Constant	0.84***	0.92***	0.92***	0.92***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	586	586	586	586
Independent observations	32	32	32	32
R-squared	0.01	0.05	0.06	0.05
Comparing FINE and COMP Wald test results (p-values)	0.35	0.48	0.48	0.62

*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. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \*  $p < 0.10$ , \*\*\*  $p < 0.01$ .

In models (3) and (4) of Table 4, we extend the regression model from column (2) 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 or COMP. However, as indicated by the insignificant coefficients of the interaction terms, none of these experiences significantly increases Player 2s' trust in Player 1s'  $X$ -messages in neither of the two sanction treatments. Having encountered an infringer in one of the previous rounds

“rounds” dummy and drop the “A previously matched Player 1 lied”. Also in this specification, the respective coefficient remains small and insignificant. Moreover, in columns (3) and (4) of Table A.7, we reiterate these exercises by estimating a dummy coefficient for whether a Player 2 has encountered state  $Y$  in precisely the round prior to the one under investigation (in contrast to *any* previous round), which might in principle provoke a stronger reaction. However, the findings resemble those in specifications (1) and (2). These findings strongly suggest that Player 2s' reduced propensity to trust and choose C cannot be explained by the fact that they experienced state  $Y$  previously.

remains the only significant explanatory factor in these models.

## 5 Removing the sanctions

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

Lastly, we study Player 1s' behaviour in part 3 of the experiment in which all participants play the experimental game under the sanction free regime of NO S for one final round.

Figure 4 reveals that Player 1s' propensity to lie when state  $Y$  prevails 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 exert lasting compliance effects that spill over to part 3. Mann-Whitney ranksum tests that compare Player 1s' 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  $p = 0.04$ , respectively). The propensity to lie in FINE and COMP, conversely, is not statistically different from one another ( $p = 0.76$ ).

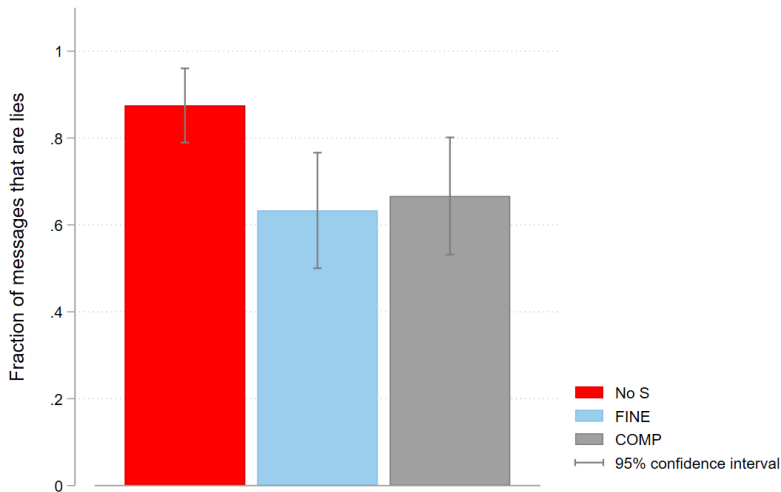


Figure 4: Player 1's misbehaviour in part 3

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 (model (2)) and their success in lying (model (3)), the treatment differences between NO S and the two sanction 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 again in part 3.

These findings are quantitatively and qualitatively unaffected by the inclusion of controls

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

	(1)	(2)	(3)
FINE	-0.24*** (0.00)	-0.02 (0.74)	0.00 (0.94)
COMP	-0.21** (0.01)	-0.05 (0.36)	-0.01 (0.93)
Player has lied before		0.57*** (0.00)	0.26* (0.08)
A previous lie was successful			0.36*** (0.01)
Constant	0.88*** (0.00)	0.37*** (0.00)	0.34*** (0.00)
Observations	168	168	168
R-squared	0.05	0.38	0.42
Comparing FINE and COMP Wald test results (p-values)	0.71	0.63	0.90

*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: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

for personal characteristics, see Table A.8. Moreover, we reproduce the findings from Table 5 in alternative logit regressions, see Table A.9 in Appendix A. This strongly suggests that the sanction schemes' lasting effects are driven by those Player 1s who were early on deterred from lying – and, therefore, remain honest once sanctions are lifted in part 3:

**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 beginning of the experiment.*

## 5.2 Potential victims' trust in part 3

Figure 5 suggests that there are no significant treatment differences in trust 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 compare 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 versus COMP:  $p = 0.61$ , FINE versus COMP:  $p = 0.17$ ).

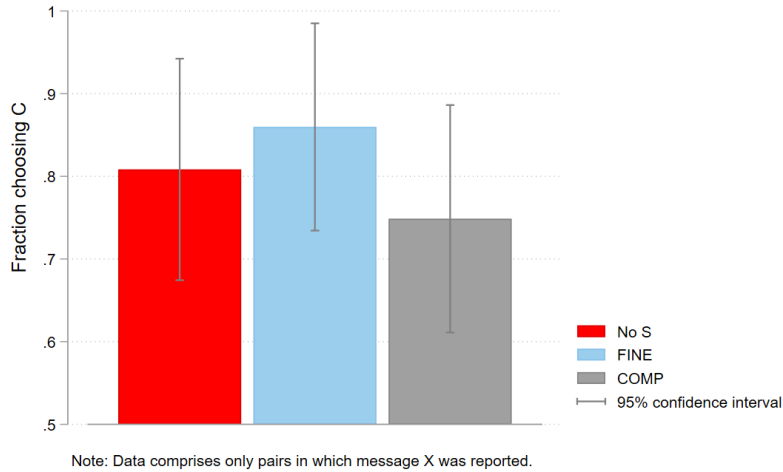


Figure 5: Player 2s choosing C in part 3

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, see column (1). Controlling for whether a Player 2 has been matched to a lying Player 1 in one of the previous rounds does not change these results, see column (2). Player 2s' trust does therefore not match the treatment differences we observed in Player 1s' honesty in part 3. Neither do we find that the previous sanction schemes lastingly affect trust after these sanctions are lifted. In fact also in part 3, the only significant finding is that having been lied to in part 1 or 2 significantly decreases Player 2s' propensity to choose C in part 3 when the matched Player 1 reports that state  $X$  prevails.

These findings are quantitatively and qualitatively unaffected by the inclusion of controls for personal characteristics, see Table A.10. Moreover, Table A.11 replicates the findings from the linear probability models presented in Table 6 in additional logit regressions, see Appendix A.

**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.*

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

	(1)	(2)
FINE	0.05 (0.54)	-0.04 (0.61)
COMP	-0.06 (0.48)	-0.11 (0.26)
A previously matched Player 1 lied		-0.22*** (0.00)
Constant	0.81*** (0.00)	1.03*** (0.00)
Observations	143	143
R-squared	0.01	0.06
Comparing FINE and COMP Wald test results (p-values)	0.42	0.50

*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 particular the sanction scheme in FINE successfully induces Player 1s not to lie. This is in line with Feldman and Teichman’s (2008) and Kornhauser et al.’s (2020) findings, who also observe that fines lead to more compliant behaviour than compensation.

At the aggregate level, both sanction schemes’ compliance effects carry over to part 3 of the experiment, in which misbehaviour is no longer punished. This lasting effect is driven by the fact that both sanction schemes 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 (cf. Mulder, 2018) which is a finding that considerably fine-tunes and extends prior experimental results.

Focusing on trust in the matching partners’  $X$  messages, we do not detect any significant



treatment differences in Player 2s' trust in the initial one-shot interaction in part 1. In the repeated setting of part 2 then, Player 2 show the highest levels of trust in FINE. Ultimately however, our findings suggest that it is pre-dominantly Player 2s experiences of having been lied to before, rather than the sanction scheme per se, 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. Intriguingly, Player 2s earn in expectations essentially the same 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.<sup>20</sup>

We conclude with noting that 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 not match these treatment differences. While Player 2s tend to trust more in FINE than in NO S, this does not reflect a treatment-specific sanction effect per se but rather a consequence of "first-hand" experiences with lying/compliant Player 1s. These findings indicate that it might take quite some time to install trust through sanction schemes as potential victims have to first-hand experience its implications on compliance, potentially for an extended period of time or in a large number of interactions.

## 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

---

<sup>20</sup>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:  $p=1$ .

some degree of guilt aversion on the side of the potential infringers. Hence, our findings, firstly, underline that, with the probability of detection and size of sanction payment being 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 *ex post* is favourable from a potential victims' point of view. More compliance, and with that less exposure to misbehaviour, can be achieved with a fine scheme. The findings from our experiment therefore suggest that if at all fine schemes have a positive impact on trust, due to their larger deterrent capacity.

Thirdly, 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 the future. 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 increase victim satisfaction *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 fine versus compensation regimes on the actors involved. In many legal domains the experimentation with the optimal type of sanction is still ongoing. For instance, in European consumer contract law, compensation has traditionally been the primary sanction, but the European legislator is now emphasising fines. Similarly, in European competition law, the conventional use of fines for public law enforcement has recently been supplemented by a right to compensation. And the European Data Protection Regulation ensures that Member States have both fines and compensation at their disposal.

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 sanc-

tion 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., forthcoming). It would be interesting to directly contrast the effectiveness of both schemes 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

- 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.
- Bar-Ilan, Avner and Bruce Sacerdote**, “The response of criminals and noncriminals to fines,” *The Journal of Law and Economics*, 2004, 47 (1), 1–17.
- 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,” *The Journal of Legal Studies*, forthcoming.
- 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.

- 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 deterrence: 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.
- 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.
- Feldman, Yuval and Doron Teichman**, “Are all legal dollars created equal,” *Northwestern University Law Review*, 2008, 102, 223.

- Friehe, Tim and Vu Mai Linh Do**, “Do crime victims lose trust in others? Evidence from Germany,” *Journal of Behavioral and Experimental Economics*, 2023, *105*, 102027.
- , **Pascal Langenbach, and Murat C Mungan**, “Does the Severity of Sanctions Influence Learning about Enforcement Policy? Experimental Evidence,” *The Journal of Legal Studies*, 2023, *52* (1), 83–106.
- 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.
- Kornhauser, Lewis, Yijia Lu, and Stephan Tontrup**, “Testing a fine is a price in the lab,” *International Review of Law and Economics*, 2020, *63*, 105931.
- 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 Cislighi**, “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.
- Metcalf, Cherie, Emily A Satterthwaite, J Shahar Dillbary, and Brock Stoddard**, “Is a fine still a price? Replication as robustness in empirical legal studies,” *International Review of Law and Economics*, 2020, *63*, 105906.
- Miceli, Thomas J**, “On Economic Theories of Criminal Punishment: Pricing, Prevention, or Proportionality?,” *American Law and Economics Review*, 2023, p. ahad003.

- , **Kathleen Segerson, and Dietrich Earnhart**, “The role of experience in deterring crime: A theory of specific versus general deterrence,” *Economic Inquiry*, 2022, 60 (4), 1833–1853.
- 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.
- Polinsky, A. Mitchell and Steven Shavell**, “Enforcement costs and the optimal magnitude and probability of fines,” *The Journal of Law and Economics*, 1992, 35 (1), 133–148.
- 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)	(3)
FINE	-0.25*** (0.01)	-0.22** (0.02)	-0.36*** (0.00)
COMP	-0.16* (0.09)	-0.14 (0.13)	-0.21* (0.05)
Gender: Female		0.03 (0.67)	-0.05 (0.58)
Risk proneness		0.00 (0.79)	-0.01 (0.48)
Age		-0.01 (0.38)	-0.01 (0.20)
Belonging to the majority in terms of nationality		-0.12 (0.24)	-0.09 (0.41)
Studies Law (2nd largest group of participants)		0.04 (0.77)	0.13 (0.29)
Studies Economics or Business (largest group of participants)		0.07 (0.39)	0.12 (0.17)
General trust		0.00 (0.79)	-0.00 (0.98)
Opinion: importance of sustainability		0.00 (0.98)	0.03 (0.38)
Opinion: importance of fair legal system		-0.04 (0.37)	-0.09** (0.04)
Feeling treated fairly as Player 1		-0.02 (0.15)	-0.02 (0.24)
Perception of treatment specific deterrent effect		-0.05*** (0.00)	-0.06*** (0.00)
# Experiments participated in so far			0.09* (0.08)
Constant	0.48*** (0.00)	1.39*** (0.00)	1.76*** (0.00)
Observations	168	168	132
R-squared	0.04	0.16	0.29
Comparing FINE and COMP Wald test results (p-values)	0.31	0.35	0.09

*Notes:* Linear probability models. Dependent variable: Player 1's decision to lie in part 1. NO S serves as baseline treatment in both regressions. We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the in total 16 sessions. 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 I, Player 1's decision to lie in part 2

	(1)	(2)	(3)	(4)	(5)	(6)
FINE	-0.31*** (0.00)	-0.36*** (0.00)	-0.34*** (0.00)	-0.42*** (0.00)	-0.36*** (0.00)	-0.40*** (0.00)
COMP	-0.17*** (0.01)	-0.21** (0.01)	-0.19** (0.01)	-0.24*** (0.01)	-0.20* (0.05)	-0.23** (0.05)
Player has lied before	0.45*** (0.00)	0.42*** (0.00)	0.41*** (0.00)	0.32** (0.01)	0.45*** (0.00)	0.42*** (0.00)
A previous lie was successful	0.05 (0.48)	0.07 (0.45)	0.07 (0.40)	0.12 (0.27)	0.06 (0.44)	0.07 (0.42)
FINE × Player was caught lying before			0.19 (0.19)	0.36** (0.01)		
COMP × Player was caught lying before			0.06 (0.56)	0.11 (0.28)		
FINE × Player was checked for lying before					0.06 (0.27)	0.05 (0.46)
COMP × Player was checked for lying before					0.04 (0.73)	0.03 (0.81)
Gender: Female	-0.01 (0.92)	-0.03 (0.62)	-0.00 (0.96)	-0.04 (0.55)	-0.00 (0.94)	-0.03 (0.64)
Risk proneness	0.00 (0.88)	0.00 (0.80)	0.00 (0.91)	0.00 (0.82)	0.00 (0.91)	0.00 (0.81)
Age	-0.00 (0.59)	-0.00 (0.46)	-0.00 (0.54)	-0.01 (0.35)	-0.00 (0.71)	-0.00 (0.54)
Belonging to the majority in terms of nationality	0.13* (0.05)	0.13** (0.05)	0.14** (0.04)	0.15** (0.03)	0.12* (0.06)	0.13* (0.05)
Studies Law (2nd largest group of participants)	0.13* (0.06)	0.16* (0.07)	0.14** (0.04)	0.18** (0.04)	0.13* (0.06)	0.16* (0.07)
Studies Economics or Business (largest group of participants)	0.07 (0.16)	0.08 (0.16)	0.08 (0.13)	0.08 (0.12)	0.07 (0.16)	0.08 (0.17)
General trust	-0.01 (0.20)	-0.01 (0.34)	-0.01 (0.22)	-0.01 (0.34)	-0.01 (0.21)	-0.01 (0.35)
Opinion: importance of sustainability	-0.02 (0.19)	-0.01 (0.81)	-0.02 (0.21)	-0.00 (0.87)	-0.03 (0.17)	-0.01 (0.76)
Opinion: importance of fair legal system	0.00 (0.85)	-0.00 (0.93)	0.00 (0.94)	-0.01 (0.80)	0.01 (0.79)	-0.00 (0.98)
Feeling treated fairly as Player 1	-0.01** (0.04)	-0.01 (0.16)	-0.01** (0.04)	-0.01 (0.14)	-0.01** (0.04)	-0.01 (0.14)
Perception of treatment specific deterrent effect	-0.01* (0.08)	-0.02** (0.03)	-0.01 (0.10)	-0.02** (0.05)	-0.01* (0.08)	-0.02** (0.04)
# Experiments participated in so far		0.03 (0.46)		0.03 (0.36)		0.03 (0.45)
Constant	1.07*** (0.00)	1.01*** (0.01)	1.10*** (0.00)	1.08*** (0.00)	1.08*** (0.00)	1.01*** (0.01)
Observations	384	312	384	312	384	312
Independent observations	31	25	31	25	31	25
R-squared	0.43	0.44	0.43	0.45	0.43	0.44
Comparing FINE and COMP Wald test results (p-values)	0.01	0.01	0.01	0.00	0.15	0.11

*Notes:* Linear probability models. Dependent variable: Player 1's decision to lie in part 2. NO S serves as baseline treatment in all regressions. Dummies for rounds 3, 4 and 5 are included in all specifications. We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the in total 16 sessions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.



Table A.3: Robustness check II, Player 1's decision to lie in part 2

	(1)	(2)	(3)	(4)	(5)
FINE	-2.13*** (0.00)	-1.85*** (0.00)	-1.86*** (0.00)	-1.90*** (0.00)	-1.67*** (0.00)
COMP	-1.33*** (0.00)	-1.17*** (0.00)	-1.11*** (0.00)	-1.14*** (0.00)	-1.05 (0.13)
Player has lied before		2.55*** (0.00)	2.03*** (0.00)	1.89*** (0.00)	2.05*** (0.00)
A previous lie was successful			0.78 (0.17)	0.85 (0.17)	0.77 (0.17)
FINE $\times$ Player was caught lying before				0.34 (0.71)	
COMP $\times$ Player was caught lying before				0.19 (0.76)	
FINE $\times$ Player was checked for lying before					-0.26 (0.43)
COMP $\times$ Player was checked for lying before					-0.08 (0.90)
Constant	1.47*** (0.00)	0.32 (0.31)	0.30 (0.35)	0.32 (0.33)	0.30 (0.35)
Observations	384	384	384	384	384
Independent observations	31	31	31	31	31
Pseudo R-squared	0.10	0.30	0.31	0.31	0.31
Comparing FINE and COMP Wald test results (p-values)	0.01	0.01	0.01	0.01	0.37

*Notes:* Logit regressions. Dependent variable: Player 1's decision to lie in part 2. 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.

Table A.4: Player 2's decision to choose C in part 1

	(1)	(2)	(3)
FINE	0.10 (0.32)	0.10 (0.44)	0.03 (0.80)
COMP	0.14 (0.11)	0.15 (0.27)	0.10 (0.46)
Gender: Female		0.01 (0.93)	0.01 (0.95)
Risk proneness		-0.00 (0.94)	-0.02 (0.45)
Age		0.00 (0.37)	0.01 (0.25)
Belonging to the majority in terms of nationality		0.04 (0.76)	-0.07 (0.26)
Studies Law (2nd largest group of participants)		0.02 (0.84)	0.00 (1.00)
Studies Economics or Business (largest group of participants)		-0.08 (0.42)	-0.13 (0.15)
General trust		-0.00 (0.93)	0.01 (0.45)
Opinion: importance of sustainability		-0.01 (0.70)	-0.01 (0.56)
Opinion: importance of fair legal system		-0.01 (0.83)	0.04 (0.50)
Feeling treated fairly as Player 2		-0.00 (0.83)	-0.02 (0.45)
Perception of treatment specific deterrent effect		-0.02 (0.34)	-0.01 (0.56)
# Experiments participated in so far			-0.08* (0.08)
Constant	0.83*** (0.00)	1.02 (0.11)	0.87 (0.23)
Observations	80	79	65
R-squared	0.04	0.11	0.18
Comparing FINE and COMP Wald test results (p-values)	0.48	0.51	0.40

*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. We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the in total 16 sessions. Robust standard errors are clustered at the individual level, p-values given in parentheses: \*  $p < 0.1$ , \*\*\*  $p < 0.01$ .

Table A.5: Robustness check I, Player 2's decision to choose C in part 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FINE	0.05 (0.29)	0.05 (0.22)	0.02 (0.58)	0.01 (0.84)	0.04 (0.42)	0.02 (0.52)	0.02 (0.61)	0.01 (0.80)
COMP	0.00 (1.00)	0.01 (0.92)	-0.01 (0.86)	-0.02 (0.69)	0.00 (0.96)	-0.01 (0.93)	-0.00 (0.95)	-0.02 (0.79)
A previously matched Player 1 lied			-0.12*** (0.00)	-0.14*** (0.01)	-0.10** (0.02)	-0.12** (0.03)	-0.12*** (0.01)	-0.14*** (0.01)
FINE × A previously matched Player 1 was caught lying					-0.08 (0.26)	-0.10 (0.13)		
COMP × A previously matched Player 1 was caught lying					-0.07 (0.51)	-0.07 (0.55)		
FINE × A previously matched Player 1 was checked for lying							0.00 (0.96)	-0.01 (0.87)
COMP × A previously matched Player 1 was checked for lying							-0.01 (0.79)	-0.01 (0.84)
Gender: Female	0.04 (0.42)	0.06 (0.29)	0.04 (0.47)	0.07 (0.25)	0.04 (0.44)	0.07 (0.24)	0.04 (0.47)	0.07 (0.27)
Risk proneness	-0.01 (0.16)	-0.01 (0.31)	-0.01 (0.25)	-0.01 (0.54)	-0.01 (0.25)	-0.01 (0.52)	-0.01 (0.24)	-0.01 (0.51)
Age	-0.00 (0.70)	-0.00 (0.51)	-0.00 (0.54)	-0.00 (0.37)	-0.00 (0.50)	-0.00 (0.34)	-0.00 (0.54)	-0.00 (0.36)
Belonging to the majority in terms of nationality	-0.08* (0.10)	-0.04 (0.44)	-0.08* (0.07)	-0.04 (0.41)	-0.08* (0.07)	-0.04 (0.37)	-0.08* (0.07)	-0.04 (0.42)
Studies Law (2nd largest group of participants)	0.10* (0.09)	0.13** (0.01)	0.10 (0.12)	0.14** (0.02)	0.10 (0.13)	0.15** (0.02)	0.10 (0.12)	0.14** (0.02)
Studies Economics or Business (largest group of participants)	-0.01 (0.75)	-0.01 (0.88)	-0.00 (0.93)	0.01 (0.93)	-0.00 (1.00)	0.01 (0.88)	-0.00 (0.94)	0.01 (0.92)
General trust	0.01 (0.33)	0.01 (0.46)	0.01 (0.32)	0.01 (0.36)	0.01 (0.29)	0.01 (0.33)	0.01 (0.31)	0.01 (0.34)
Opinion: importance of sustainability	-0.02* (0.05)	-0.02 (0.13)	-0.02** (0.04)	-0.02 (0.14)	-0.02* (0.06)	-0.02 (0.17)	-0.02** (0.04)	-0.02 (0.13)
Opinion: importance of fair legal system	0.03 (0.20)	0.02 (0.39)	0.03 (0.15)	0.03 (0.34)	0.03 (0.15)	0.03 (0.33)	0.03 (0.16)	0.03 (0.34)
Feeling treated fairly as Player 2	0.02** (0.02)	0.02** (0.02)	0.02* (0.06)	0.02* (0.07)	0.02* (0.06)	0.02* (0.08)	0.02* (0.06)	0.02* (0.07)
Perception of treatment specific deterrent effect	0.01* (0.09)	0.01 (0.17)	0.01 (0.16)	0.01 (0.28)	0.01 (0.17)	0.01 (0.29)	0.01 (0.16)	0.01 (0.27)
# Experiments participated in so far		-0.02 (0.61)		-0.02 (0.69)		-0.01 (0.70)		-0.02 (0.69)
Constant	0.69*** (0.00)	0.75*** (0.00)	0.76*** (0.00)	0.81*** (0.00)	0.74*** (0.00)	0.79*** (0.00)	0.76*** (0.00)	0.81*** (0.00)
Observations	584	464	584	464	584	464	584	464
Independent observations	32	26	32	26	32	26	32	26
R-squared	0.07	0.10	0.10	0.13	0.11	0.14	0.10	0.13
Comparing FINE and COMP Wald test results (p-values)	0.38	0.38	0.46	0.50	0.48	0.51	0.65	0.66

*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. Dummies for rounds 3, 4 and 5 are included in all specifications. We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the in total 16 sessions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

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

	(1)	(2)	(3)	(4)
FINE	0.85** (0.05)	0.47 (0.23)	0.60 (0.15)	0.32 (0.53)
COMP	0.28 (0.58)	0.08 (0.88)	0.17 (0.75)	-0.00 (1.00)
A previously matched Player 1 lied		-1.39*** (0.00)	-1.29*** (0.00)	-1.40*** (0.00)
FINE $\times$ A previously matched Player 1 was caught lying			-0.67 (0.32)	
COMP $\times$ A previously matched Player 1 was caught lying			-0.25 (0.68)	
FINE $\times$ A previously matched Player 1 was checked for lying				0.33 (0.58)
COMP $\times$ A previously matched Player 1 was checked for lying				0.15 (0.74)
Constant	1.69*** (0.00)	2.64*** (0.00)	2.56*** (0.00)	2.65*** (0.00)
Observations	586	586	586	586
Independent observations	32	32	32	
Pseudo R-squared	0.02	0.08	0.08	0.08
Comparing FINE and COMP Wald test results (p-values)	0.30	0.43	0.44	0.61

*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. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.7: Robustness check III, Player 2's decision to choose C in part 2

	(1)	(2)	(3)	(4)
FINE	0.05 (0.24)	0.07* (0.09)	0.04 (0.26)	0.08* (0.06)
COMP	0.02 (0.80)	0.03 (0.66)	0.01 (0.84)	0.03 (0.58)
A previously matched Player 1 lied	-0.15*** (0.00)		-0.15*** (0.00)	
State $Y$ in any of the previous rounds	0.03 (0.57)	-0.04 (0.48)		
State $Y$ in the previous round			0.03 (0.19)	-0.01 (0.75)
Constant	0.90*** (0.00)	0.88*** (0.00)	0.91*** (0.00)	0.85*** (0.00)
Observations	586	586	586	586
Independent observations	32	32	32	32
R-squared	0.05	0.01	0.06	0.01
Comparing FINE and COMP Wald test results (p-values)	0.47	0.37	0.46	0.35

*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. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \*  $p < 0.10$ , \*\*\*  $p < 0.01$ .

Table A.8: Robustness check I, Player 1's decision to lie in part 3

	(1)	(2)	(3)	(4)	(5)	(6)
FINE	-0.21** (0.02)	-0.30** (0.02)	-0.01 (0.81)	-0.06 (0.40)	0.01 (0.89)	-0.03 (0.69)
COMP	-0.19** (0.02)	-0.27*** (0.01)	-0.07 (0.30)	-0.13* (0.05)	-0.02 (0.73)	-0.08 (0.18)
Player has lied before			0.56*** (0.00)	0.58*** (0.00)	0.27* (0.06)	0.37** (0.02)
A previous lie was successful					0.33*** (0.01)	0.26** (0.03)
Gender: Female	0.02 (0.76)	0.03 (0.72)	0.04 (0.53)	0.09 (0.23)	0.05 (0.47)	0.09 (0.21)
Risk proneness	0.03 (0.14)	0.04 (0.12)	0.02 (0.16)	0.03* (0.07)	0.02 (0.10)	0.03** (0.04)
Age	-0.00 (0.58)	0.00 (0.80)	-0.00 (0.96)	0.01 (0.39)	-0.00 (0.91)	0.01 (0.41)
Belonging to the majority in terms of nationality	0.11 (0.31)	0.16 (0.13)	0.06 (0.48)	0.07 (0.37)	0.04 (0.60)	0.07 (0.44)
Studies Law (2nd largest group of participants)	0.06 (0.62)	0.06 (0.73)	-0.03 (0.78)	-0.07 (0.69)	-0.03 (0.76)	-0.06 (0.69)
Studies Economics or Business (largest group of participants)	0.05 (0.48)	0.08 (0.28)	-0.03 (0.58)	-0.05 (0.50)	-0.05 (0.38)	-0.06 (0.42)
General trust	-0.02 (0.21)	-0.03 (0.10)	-0.01 (0.35)	-0.02 (0.15)	-0.01 (0.42)	-0.02 (0.17)
Opinion: importance of sustainability	-0.06*** (0.01)	-0.06** (0.05)	-0.06*** (0.00)	-0.08*** (0.00)	-0.05*** (0.01)	-0.07*** (0.00)
Opinion: importance of fair legal system	0.00 (0.96)	-0.02 (0.64)	0.01 (0.55)	0.01 (0.69)	0.01 (0.71)	0.01 (0.76)
Feeling treated fairly as Player 1	-0.02 (0.14)	-0.02 (0.19)	-0.00 (0.77)	-0.00 (0.74)	-0.00 (0.73)	-0.01 (0.71)
Perception of treatment specific deterrent effect	-0.00 (0.87)	-0.00 (0.96)	0.00 (0.73)	0.01 (0.51)	0.00 (0.67)	0.01 (0.48)
# Experiments participated in so far		-0.02 (0.75)		-0.04 (0.56)		-0.04 (0.56)
Constant	1.37*** (0.00)	1.45*** (0.00)	0.63** (0.03)	0.71** (0.05)	0.58** (0.03)	0.63* (0.06)
Observations	168	132	168	132	168	132
R-squared	0.14	0.17	0.43	0.47	0.46	0.49
Comparing FINE and COMP Wald test results (p-values)	0.87	0.72	0.46	0.36	0.67	0.45

*Notes:* Linear probability models. Dependent variable: Player 1's decision to lie in part 3. NO S serves as baseline treatment in all regressions. We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the in total 16 sessions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

Table A.9: Robustness check II, Player 1's decision to lie in part 3

	(1)	(2)	(3)
FINE	-1.40*** (0.00)	-0.24 (0.59)	0.02 (0.97)
COMP	-1.25*** (0.01)	-0.48 (0.33)	-0.04 (0.94)
Player has lied before		3.10*** (0.00)	1.07* (0.09)
A previous lie was successful			2.72*** (0.00)
Constant	1.95*** (0.00)	-0.34 (0.43)	-0.64 (0.19)
Observations	168	168	168
Pseudo R-squared	0.05	0.33	0.38
Comparing FINE and COMP Wald test results (p-values)	0.71	0.62	0.90

*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.10: Robustness check I, Player 2's decision to choose C in part 3

	(1)	(2)	(3)	(4)
FINE	0.06 (0.49)	-0.02 (0.87)	-0.02 (0.85)	-0.08 (0.48)
COMP	-0.05 (0.58)	-0.12 (0.31)	-0.07 (0.46)	-0.13 (0.28)
A previously matched Player 1 lied			-0.29*** (0.01)	-0.28** (0.01)
Gender: Female	-0.04 (0.65)	-0.06 (0.57)	-0.05 (0.57)	-0.06 (0.57)
Risk proneness	-0.03 (0.12)	-0.02 (0.41)	-0.03 (0.15)	-0.01 (0.58)
Age	0.01 (0.18)	0.01 (0.28)	0.01 (0.33)	0.01 (0.51)
Belonging to the majority in terms of nationality	0.14 (0.22)	0.16 (0.23)	0.13 (0.25)	0.15 (0.25)
Studies Law (2nd largest group of participants)	0.15 (0.15)	0.20** (0.04)	0.18* (0.10)	0.23** (0.02)
Studies Economics or Business (largest group of participants)	0.06 (0.51)	0.03 (0.81)	0.04 (0.64)	0.01 (0.90)
General trust	0.02 (0.40)	-0.00 (0.92)	0.02 (0.30)	0.00 (0.94)
Opinion: importance of sustainability	0.02 (0.42)	0.01 (0.73)	0.01 (0.69)	0.00 (0.88)
Opinion: importance of fair legal system	-0.02 (0.67)	-0.01 (0.78)	-0.02 (0.70)	-0.01 (0.81)
Feeling treated fairly as Player 2	-0.00 (0.69)	0.00 (0.78)	-0.02 (0.12)	-0.02 (0.38)
Perception of treatment specific deterrent effect	0.01 (0.68)	0.01 (0.75)	0.00 (0.93)	-0.00 (0.95)
# Experiments participated in so far		-0.07 (0.15)		-0.08* (0.09)
Constant	0.56 (0.22)	0.80 (0.12)	1.02** (0.04)	1.25** (0.02)
Observations	142	113	142	113
R-squared	0.08	0.10	0.13	0.16
Comparing FINE and COMP Wald test results (p-values)	0.38	0.44	0.71	0.54

*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. We elicited the number of experiments participants have taken part in prior to the present experiment only in the last 14 of the in total 16 sessions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .



Table A.11: Robustness check II, Player 2's decision to choose C in part 3

	(1)	(2)
FINE	0.37 (0.53)	-0.19 (0.72)
COMP	-0.37 (0.47)	-0.63 (0.24)
A previously matched Player 1 lied		-2.31** (0.04)
Constant	1.45*** (0.00)	3.75*** (0.00)
Observations	143	143
Pseudo R-squared	0.02	0.07
Comparing FINE and COMP Wald test results (p-values)	0.41	0.46

*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			Payoffs in situation Y				
		Person 2				Person 2	
		Option C	Option D			Option C	Option D
Person 1	Option A	20, 20	10, 10	Person 1	Option A	10, 0	0, 10
	Option B	10, 10	0, 0		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.

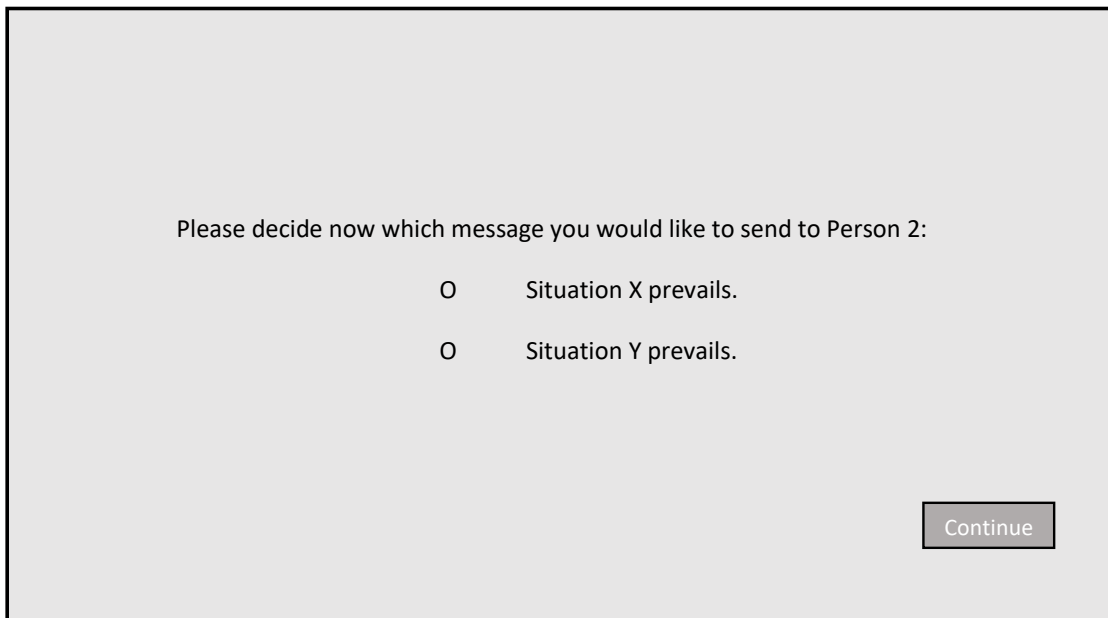
### 2. *Allocation of roles*

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:



The screenshot shows a light gray rectangular box with a black border. Inside the box, the text reads: "Please decide now which message you would like to send to Person 2:". Below this text are two radio button options: "O Situation X prevails." and "O Situation Y prevails.". In the bottom right corner of the box, there is a button labeled "Continue".

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:

The computer informed you that situation X prevails in this round.

You sent your matched Person 2 message "Situation X prevails".

Please choose now between options A and B:

- Option A
- Option B

Example: Decision screen of Person 2:

Your matched Person 1 sent you the following message:

Situation X prevails.

Please choose now between options C and D:

- Option C
- 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 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?

yes

no

*[In Treatment Fine and Treatment Comp instead:]*

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

every second

every third

every fourth

*[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...?

- male
- female
- 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 –             – 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 –             – 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 –             – 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 –             – 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 –             – 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 –             – 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)