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

Pieter Desmet* Leonie Gerhards† Franziska Weber‡

May, 2026

Abstract

Both fines and compensation payments are widely used as sanctions to address contract breaches and other forms of opportunistic behaviour. Yet, their application varies substantially across jurisdictions and their comparative effectiveness is not well understood. We experimentally compare two enforcement schemes – one based on fines and one based on compensation – while holding constant the probability of detection and the sanction amount. The setting captures a situation akin to a buyer–seller relationship with asymmetric information, in which one player may misrepresent the state of the world to their own advantage. We examine how the two sanction schemes affect both misconduct and counterpart trust. We find that fines induce higher compliance than compensation. However, this difference translates only partially into higher trust. Trust appears to be shaped less by the enforcement scheme in place than by participants' actual experiences with misconduct. Ultimately, participants in the role of potential victims earn virtually identical payoffs under both sanction schemes, despite receiving compensation payments only under the compensation scheme. Overall, our findings suggest that compensation-based sanctions may weaken deterrence relative to fines without improving outcomes for potential victims. More broadly, the results highlight the importance of considering both compliance incentives and trust when designing enforcement schemes for contractual and consumer relationships.

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

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

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

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

1 Introduction

Effective enforcement institutions are essential for deterring misconduct and sustaining trust in market interactions. It is, therefore, crucial to identify and implement the sanction scheme that best promotes compliance while also ensuring that individuals can trust others not to take advantage of them. Two classical sanction schemes are fines and compensation payments. While both schemes require the offender to make a monetary payment following misconduct, the key difference lies in the recipient of that payment: under fines, the payment goes to the state, whereas under compensation it is transferred to the victim. Both fines and compensation payments are widely used as enforcement tools in consumer and competition law. However, even for identical infringements, their application varies considerably across jurisdictions.

Unfair commercial practices law illustrates this inconsistency. It regulates how sellers may market their services and products to buyers, aiming to reduce information asymmetries between the two sides. Such asymmetries resemble the classic "market for lemons" problem described by Akerlof (1978), where sellers may conceal relevant information or product defects from buyers. To address these issues, laws prohibit sellers from making false or deceptive claims about the nature or main characteristics of a product or service and requires the disclosure of material information needed for informed decision-making.

Consider, for example, a car dealer selling a used vehicle while falsely advertising it as accident-free and concealing serious mechanical defects. Such deception may induce consumers to purchase a car they otherwise would not have bought, or to pay a substantially inflated price. Jurisdictions have historically differed markedly in how they sanction this same underlying misconduct. Italy, until recently, relied primarily on public enforcement through administrative fines for unfair commercial practices.[1] In such cases, the monetary payment was made to the state budget rather than to the consumer harmed by the deceptive practice. Other countries, such as the Netherlands, have historically placed greater emphasis on compensation claims for the same consumer law violation, allowing consumers to recover their financial losses directly.[2]

Given the central role of enforcement in maintaining fair market behaviour, it is striking how little empirical evidence exists on the comparative effectiveness of fines and compensation payments. Moreover, prior research has largely focused on how enforcement deters misconduct by potential infringers, whereas much less is known about how different sanction schemes shape

[1] See Part II, Art. 27 of the Codice del Consumo, as introduced by Legislative Decree No. 146 of 2007. The article was in 2023 extended to allow for compensation claims as well.

[2] Articles 6:193a-j Dutch Civil Code.

the willingness of potential victims to trust others in environments characterised by asymmetric information and the possibility of opportunistic behaviour.

To study these mechanisms directly, we conduct a laboratory experiment comparing a fine-based and a compensation-based sanction scheme while holding constant both the detection probability and sanction magnitude. The controlled environment allows us to isolate how the two sanction schemes affect both compliance behaviour on the part of potential infringers and trust on the part of potential victims.

A large literature shows that individuals respond to sanctions for at least two broad reasons. First, sanctions may signal prevailing social norms and socially appropriate behaviour. In this regard, Drouvelis (2021) and Villeval (2020) review a wide range of laboratory and field experiments examining how social norms and punishment affect cooperation in social dilemma settings. Second, sanctions may deter misconduct through the anticipated monetary consequences emphasised in Becker's (1968) economic model of crime and deterrence. However, despite this extensive literature on sanctions and deterrence, studies directly comparing different sanction schemes – especially fine-based versus compensation-based schemes – remain scarce.

There are clear reasons to expect the two schemes to differ systematically. While both schemes impose monetary consequences and signal socially appropriate behaviour, only compensation allows infringers to partially repair the harm caused by misconduct by making the victim whole ex post. This distinction may matter in the presence of guilt aversion, i.e. when individuals anticipate feelings of guilt from violating others' expectations. Under compensation schemes, these anticipated moral costs may be lower because the victim can potentially be compensated after the misconduct occurs. By contrast, under fine schemes the only way to avoid such anticipated guilt is to refrain from misconduct altogether. Compensation may therefore weaken deterrence relative to an otherwise equivalent fine-based scheme.

On the victim side, the effects are inherently ambiguous because trust may be shaped through multiple channels. The presence and design of an enforcement scheme can influence perceived risk, either by reducing the likelihood of misconduct through deterrence or, in the case of compensation, by providing victims with some insurance against losses. In repeated settings, individuals may furthermore update their trust based on the compliance behaviour they actually observe and experience. Consequently, trust may depend not only on the formal characteristics of the enforcement scheme, but also on the behavioural patterns it generates. Compensation may therefore increase trust through its insurance effect, but it may also indirectly reduce trust

if weaker deterrence leads to lower compliance and more frequent experiences of misconduct.

We review the relevant literatures in Section 2 and, building on this, derive our behavioural hypotheses for both potential infringers and victims in Section 3.3.

Our two-player experimental game captures a strategic interaction in which one player may misrepresent the state of the world to their own advantage. In this sense, the setting resembles a cheap-talk game with partially conflicting interests between sender and receiver (Crawford and Sobel, 1982; see also Gneezy, 2005 for a related experimental implementation). We define misconduct as lying for personal gain at the expense of the other player and interpret the counterpart's willingness to believe and act on potentially false information as a measure of trust.

Lies are detected with a fixed, commonly known probability. Depending on the treatment, detected misconduct triggers either a fine or a compensation payment (Fine and Comp), with the payment amount held constant across the two. A third treatment with no sanction (No S) serves as a control.

The experiment consists of three parts. In part 1, we measure ex-ante compliance and trust in a one-shot game. Part 2 uses perfect-stranger matching to examine learning effects once participants have experienced the treatment-specific sanction scheme. This setting allows us to observe how infringers adjust their behaviour after being checked for or caught lying and how potential victims update their trust after experiencing treatment-specific patterns of compliance and misconduct. In part 3, all sanctions are removed, enabling us to assess whether behavioural changes persist once external incentives are lifted.

Our main results are as follows. In parts 1 and 2, lying is least frequent under the Fine scheme, more common under Comp and most prevalent in NoS, with the differences particularly pronounced in the repeated setting of part 2. Notably, even after sanctions are removed in part 3, lying remains lower in both Fine and Comp than in NoS, suggesting that both sanction schemes generate carry-over effects on compliance.

Trust partly mirrors these compliance patterns. In part 2, trust is significantly higher in Fine than in NoS. However, closer analysis indicates that this difference is driven primarily by the higher compliance induced by fines rather than by the enforcement scheme itself. In other words, victims' trust responds more strongly to their experience with compliant counterparts than to the formal sanction scheme in place. Apart from the Fine–NoS difference in part 2, we find no significant treatment differences in trust across the three parts.

The remainder of the paper is structured as follows. Section 2 reviews the literature on compliance and trust under different sanction schemes. Section 3 outlines the experimental design and hypotheses. Section 4 presents the results from parts 1 and 2, while Section 5 reports on the sanction-free part 3. Section 6 discusses the findings and Section 7 concludes.

2 Related literature

2.1 Sanctions and infringers – Deterrence

The literature on enforcement has so far been rather infringer-centred, with a primary focus on the ex ante perspective – that is, the extent to which sanction schemes 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 central to incentivising compliance. According to deterrence theory, the interaction between substantive laws and their enforcement determines the incentives and, therefore, the deterrent backbone of compliance (Miceli, 2023; Veljanovski, 1984). The 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.

Many empirical findings support the predictions of deterrence theory in settings in which sanctions typically take the form of fines. See for example Engel's (2018) review of empirical and experimental research in criminal law or Slemrod's (2016) survey on tax compliance. However, the matter can be more complex. Dari-Mattiacci and Raskolnikov's (2021) model extends the basic deterrence model by relaxing some of the original assumptions and discusses contexts in which compliance does not necessarily increase with expected sanctions and showing that equally sized rewards and punishments can have different incentive effects. Gneezy and Rustichini (2000) famously demonstrate how a newly introduced sanction can even lead to an increase in unwanted behaviour, when the monetary sanction is perceived as a price.[3]

Schildberg-Hörisch and Strassmair (2012) test deterrence theory in the laboratory and find that misconduct only (weakly) decreases under very high fines. They argue that these deter-

[3]In Gneezy and Rustichini's (2000) field experiment, a group of daycare centres introduced a sanction payment for parents who arrive late to collect their children. Although the sanction in their setting is labelled a fine, it shares an important structural feature with compensation schemes: the payment accrues to the harmed party (the daycare, though not the individual teachers) rather than to the state. In this sense, one could argue that the sanction operates more like a compensation payment than a classical fine.

rent incentives can crowd out intrinsic motivation to act pro-socially – an explanation Khadjavi (2015) corroborates and further links to the emotional state of decision-makers. Agranov and Buyalskaya's (2022) experiment shows that sanction schemes providing only partial information – in particular, the minimum fine – 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 punishments. They find that the magnitude of the fine – which should be irrelevant – does in fact influence how individuals update their beliefs about said probability.

The deterrence of compensation-based schemes has received less attention (for an exception using hypothetical vignettes, see Cardi et al., 2012). Comparative studies between fine- and compensation-based schemes are even scarcer (Mulder, 2018), despite clear behavioural reasons to expect systematic differences between the two. In particular, guilt aversion – i.e. individuals' desire to meet others' expectations and avoid anticipated guilt from falling short (see Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007) – is likely to operate differently across the two schemes. A key distinction is that only compensation allows infringers to mitigate such feelings ex post by making the harmed party whole. This may in turn weaken deterrence under compensation relative to fines.

Kurz et al. (2014) study the effects of identical penalties, framed either as retributive (fine-like) or compensatory, on punctuality in a lab experiment. Participants are more punctual when the sanction is framed retributively, suggesting that compensation-based schemes may indeed be less effective at deterring. However, in their setting both payments go to the same beneficiary (the experimenter), which misses the core distinction between fines and compensation. Other work goes beyond framing to examine settings where the harmed party actually receives the payment – the defining feature of compensation. Eisenberg and Engel (2014) compare three types of damages, including a treatment where the punisher can forfeit some or all of the infringer's income without benefiting themselves – essentially a fine. In their study, however, this option is primarily symbolic. Adding the forfeit option does not make a difference in terms of deterrence and moreover, the option is also only rarely chosen.

The studies most closely related to ours are Feldman and Teichman (2008), Desmet and Weber (2022), Baumann et al. (2024), Kornhauser et al. (2020) and Metcalf et al. (2020). As a part of their comprehensive study, Feldman and Teichman (2008) examine probabilistic fines and compensation 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. Desmet and Weber (2022) also use vignettes and find higher willingness to pay a sanction under compensation than under fines, but no difference in precautionary behaviour. 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. (2024) use a lab experiment to compare fines and compensation in accident prevention, observing greater prevention under compensation. However, in their study, too, harm is not intentionally inflicted by the infringers themselves.

By contrast, Kornhauser et al. (2020) and Metcalf et al. (2020) focus on intentional breaches. 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 this literature in several ways. First, unlike Kornhauser et al. (2020) and Metcalf et al. (2020), we compare probabilistic, rather than deterministic, fines and compensation, better reflecting real-world enforcement. Second, while Feldman and Teichman (2008) and Metcalf et al. (2020) rely on hypothetical vignettes, we use an incentivised experimental game to examine behavioural responses under controlled but behaviourally rich conditions. Third, our repeated-game design with perfect-stranger matching allows us to measure learning effects after participants experience the enforcement scheme through being checked for or caught lying. Finally, our framework measures not only deterrence among potential infringers but also trust among victims, allowing us to assess which sanction scheme better fosters or restores cooperative exchange.

2.2 Sanctions and victims – Trust

While it is essential to know to what extent different sanction schemes can induce compliance among potential infringers, from a societal perspective an equally important question is under which scheme 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 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 an enforcement schemes can reduce risk by decreasing the probability of loss. Yet not all schemes achieve this in the same way. The introduction of a compensation scheme *directly* affects the probability of loss for victims by creating a possibility to recoup some of the losses and increasing the expected payoff in case of betrayal. Compensation in this sense functions as an insurance mechanism that (potentially) safeguards payoffs. Fines, on the other hand, are primarily punitive measures aimed at infringers. As a result, 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 fines therefore depends on a subjective appraisal of how others will likely respond to them. Under compensation schemes, by contrast, potential victims' probability of loss is also effectively reduced independent of the perceived deterrent capacity of compensation.

Trust that depends on the perception of the deterrents faced by a potential infringer 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 others to the extent that the scheme is perceived as deterring. If potential victims assume that infringers behave consistently with the classic deterrence framework, they will expect them to be mainly deterred by the size and probability of the sanction. Crucially, this perceived deterrence need not be formed purely ex ante, from abstract knowledge of the scheme: it will also be updated through direct experience of how infringers actually behave under it.

In repeated interactions with changing partners (as in the setting we study), potential victims observe, round by round, whether they have been lied to or dealt with honestly. As a result, they may update their expectation of the deterrent capacity of the enforcement scheme based on the compliance behaviour they observe. Deterrence-based trust in this sense is shaped not only by the formal characteristics of the sanction scheme, but also by accumulated experience

with the actual compliance behaviour it generates. A sanction scheme that reduces lying more effectively may therefore sustain higher trust over time because potential victims encounter fewer instances of misconduct.

If we look at the existing literature on sanctions and trust, several critical gaps become clear. First, 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 punished, rather than the ex-ante inclination of *potential* victims to make themselves vulnerable under different schemes (see for instance Bottom et al., 2002, Desmet et al., 2010, Desmet et al., 2011). Moreover, these studies typically examine the repair of trust and cooperation within the same relationship – that is, the decline and restoration of trust between the same interaction partners – overlooking the one-shot nature of many interactions where betrayal occurs, and disregarding the spillover effects that betrayal and enforcement may have on trust in subsequent interactions with new partners. A recent exception is Friehe and Do (2023), who explore the adverse impact of becoming a victim of crime on people's future trust in others in general.

Similar to the literature on enforcement and infringers, studies on ex-ante trust have not directly compared the relative effects of compensation and fines on potential victims' trust. Vollan (2011) examine the effects of (potential) third-party punishment on ex-ante trust in one-shot trust games with strangers and find that such punishment significantly increases trust. Malhotra and Murnighan (2002) study how contracts guaranteeing a certain payoff for trustors affect initial trust and trust building between two players in a set of lab experiments. They find that the certainty of receiving a guaranteed payoff increases potential victims' trust, supporting the idea that reducing the risk of lower payoffs (for instance through compensation) will increase trust. However, these authors do not directly study a compensation scheme, but rather the effects of automatic contract enforcement with a 100% probability of execution.

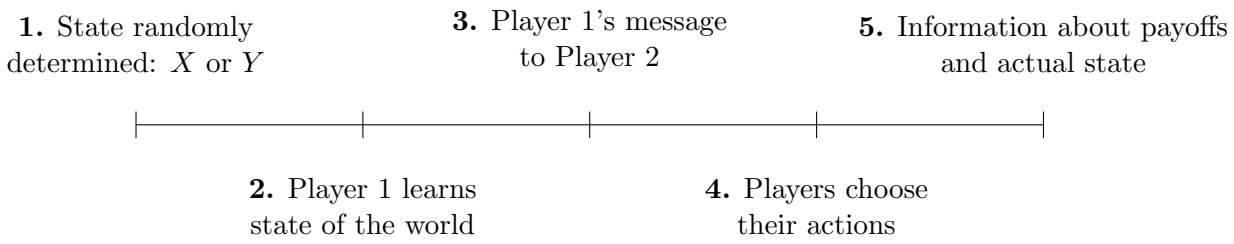
Similarly, focusing on trust in contractual relations, Bohnet et al. (2001) study the behaviour of first movers who have to decide whether to enter a contract without knowing whether the second mover will perform. They find that contractual stipulation of damages in case of breach can stimulate trust. 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 settings with repeated decision-making, which would allow the study of 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.

Figure 1: The experimental game



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

Table 1 summarises the monetary payoffs (in points) players obtain given the actions chosen and the state of the world.[4] 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

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

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.[5] When studying treatment differences in lying, we therefore focus on state Y. Analogously, we analyse treatment differences in Player 2's trust which we measure through their propensity to choose action C when their matched Player 1 reports that state X prevails.

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

Monetary payoffs in state X:

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

Monetary payoffs in state Y:

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

Notes: Payoffs denoted in points.

In the final stage, payoffs are realised and both players are informed about the actual state of the world. Hence, irrespective of treatment, every Player 2 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 punished for lying.[6]

In treatment No S (short for "No Sanction") Player 1 is never punished for lying. In treatments Fine and Comp (short for "Compensation"), conversely, a third of Player 1s' messages are randomly checked for truthfulness. If caught lying, Player 1 is punished: 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 corresponds to the loss the victim incurs when the outcome (B,C) occurs instead of (B,D) in state Y. Given the experimental parameters, the sanction is non-deterrent in expectation.[7] Specifically, lying generates an expected sanction of

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

[6] At first glance, our design resembles Gneezy's (2005) deception game. However, it differs in several key respects. In Gneezy's setup, receivers are unaware of the sender's incentives and must decide without knowing how the message maps to payoffs. In contrast, our Player 2s are fully informed about Player 1's payoffs following a lie, their own monetary consequences of being lied to, the probability of lie detection and the size of the resulting sanction. This allows them to more reliably assess the extent to which Player 1s are tempted to lie as well as the extent to which the enforcement schemes are perceived as deterrent or compensatory.

[7] Imperfect enforcement is common in many real-world contexts. For example, a freelance contractor who accepts payment but delivers substandard work may formally face legal sanctions, yet actual prosecution or

only 3.33 points in both FINE and COMP.

We intentionally chose an imperfect monitoring environment and a non-deterrent sanction level to preserve substantial scope for misconduct within the experiment while still maintaining meaningful incentives generated by the sanction schemes.[8] This design also ensured a sufficient number of lies for analysis, while allowing victims in COMP to receive meaningful expected compensation when misconduct occurred.[9]

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, repeated four more times. Player 1s and Player 2s are randomly re-matched in every round in a perfect-stranger matching fashion. Thus, while part 1 allows us to observe participants' decisions in a true one-shot game, part 2 enables us to study potential learning effects, while reputational effects are ruled out by design. By analysing behaviour of Player 1s who have been caught lying in previous rounds (in part 1 or 2), part 2 permits us to test for effects of having experienced the sanction scheme on future compliance. Similarly, we can study which scheme 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 obtain a first indication for carry-over effects of the sanction schemes 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,

successful claims may be unlikely because enforcement is costly, time-consuming, or difficult to pursue.

[8] The one-in-three detection probability is also consistent with related laboratory studies using probabilistic enforcement. For example, Agranov and Buyalskaya (2022) use a detection probability of 20%, DeAngelo and Charness (2012), Engel and Nagin (2015), and Baumann et al. (2024) include conditions with 33% detection probabilities, while Schildberg-Hörisch and Strassmair (2012) and Khadjavi (2015) study probabilities between 50% and 60%.

[9] Lying Player 1s are sanctioned regardless of whether Player 2 ultimately suffers harm. Thus, if Player 2 chooses D instead of C in state Y, the lying Player 1 is still punished even though Player 2 avoids the loss.

making 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.[10] Due to the implemented perfect-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 2s choose their payoff-maximising action D if their matched Player 1 reports that state Y prevails.[11] We hence conclude that the overwhelming majority of participants understands the rules of the game.

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 expectation: a lie can yield an additional 10 points if successful – and costs an infringer at most 3.33 points in expectation (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

[10] 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 a common 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 enforcement 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.

[11] 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.

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 $p_C \in [0, 1]$.[12]

Ample evidence shows that many individuals do not act in a purely selfish manner, but reveal social preferences and exhibit norm-abiding behaviour.[13]

In the context of the present experiment, one can plausibly argue that the enforcement schemes in Fine and Comp signal that lying is socially inappropriate, even if the implemented sanctions are under-deterrent in expectation. Relative to No S, both schemes should therefore increase the expected material costs of misconduct as well as the anticipated moral costs associated with violating prevailing social expectations, thereby reducing the propensity of Player 1s to lie.[14]

In the Comp treatment, compensation creates a possibility for victims to recoup part of their losses in case of betrayal and therefore functions as an insurance mechanism that partially safeguards payoffs. Since compensation occurs only when misconduct is detected – which happens with probability one third in our experiment – we expect this direct insurance effect on trust to be limited.

More importantly, as discussed in Section 2.2, trust may also arise through deterrence-based considerations (Lewicki et al., 1996). Player 2s should become more willing to trust to the extent that they perceive the enforcement scheme as effectively discouraging misconduct. In repeated interactions with changing matching partners, these perceptions are shaped not only by abstract knowledge of the sanction scheme, but also by accumulated experience with actual compliance behaviour. To the extent that Fine and Comp generate higher levels of compliance than No S,

[12] 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 exist 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.

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

[14] This prediction assumes that in the Comp treatment, the deterrence and norm-signalling effects of introducing sanctions outweigh any reduction in anticipated guilt arising from the possibility of compensating the victim ex post.

Player 2s should therefore become more willing to trust under both sanction schemes. This leads to our first behavioural prediction:

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

Moreover, the experimental findings by Feldman and Teichman (2008) and Kornhauser et al. (2020) suggest that compliance may differ systematically between Fine and Comp. While both sanction schemes impose monetary consequences and signal that lying is socially inappropriate, they differ in an important behavioural dimension. Under compensation, infringers can repair the harm caused by their misconduct ex post by compensating the victim, whereas under fines the sanction payment accrues to a third party and does not directly benefit the harmed party.

This distinction may matter in the presence of guilt aversion. If sanctions signal that lying violates prevailing social expectations, guilt-averse individuals may anticipate psychological costs from misconduct. Guilt-averse individuals care about others' expectations and anticipate feelings of guilt when they fail to meet those expectations (cf. Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007). In our setting, compensation may weaken deterrence because it allows infringers to reduce these anticipated moral costs by subsequently compensating the victim. Under fines, by contrast, infringers can reduce anticipated guilt only by refraining from misconduct altogether, i.e., by reporting the true state of the world. Consistent with this reasoning, Baumann et al. (2024) highlight the role of guilt aversion in shaping compliance.

Taken together, these arguments suggest that Player 1s should be less inclined to lie in Fine than in Comp. If Player 2s anticipate these behavioural differences, they should also be more willing to believe and act upon Player 1s' messages 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 less lying and more trust in* Fine *than in* Comp *in part 1 and 2 of the experiment.*

Lastly, we turn to part 3, where sanctions are removed. Based on standard deterrence theory, behaviour in both Fine and Comp should converge to the level observed in No S. Since lying is no longer punished, Player 1s face no direct monetary incentive to report truthfully when state Y occurs, implying no reason for differential compliance across treatments once enforcement is lifted.

However, prior work suggests that exposure to sanctioning institutions may generate behavioural carry-over effects by shaping norms or temporarily stabilising patterns of pro-social conduct (Mulder et al., 2006). To the extent that such effects arise, the comparably higher truthfulness of Player 1s and the higher trust of Player 2s observed under FINE and COMP in parts 1 and 2 may partially carry over into part 3, where sanctions are lifted.

Importantly, these general carry-over effects are likely to affect both sanction schemes in a broadly similar way, independent of scheme-specific incentives. By contrast, the mechanism contributing to the differences between FINE and COMP in part 2 is tied to enforcement: under compensation, infringers could ex post reduce guilt by compensating the harmed party, weakening deterrence relative to fines. Once sanctions are removed in part 3, this ex post guilt-mitigation channel is no longer available, removing an important source of asymmetry between the two schemes.

Taken together, this implies that FINE and COMP may still exhibit more truth-telling and trust than NO S in part 3 due to short-run behavioural carry-over effects, but there is no clear mechanism that sustains the treatment-specific difference between FINE and COMP. We therefore expect any remaining differences between the two sanction schemes in part 3 to be attenuated.

Hypothesis 3: *Assuming behavioural carry-over effects from prior exposure to the sanction schemes, we expect less lying and more trust in* FINE *and* COMP *than in* NO S *in part 3 of the experiment, while any differences between* FINE *and* COMP *are expected to be small.*

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 begin with an overview of basic statistics on misbehaviour across treatments. In our experiment, we define misbehaviour as lying about the state of the world if state Y prevails.[15] Table 2 shows that while in NO S 70.14% of all reported messages in state Y are lies, the respective figures in FINE and COMP are lower (30.88% and 47.06%). Focusing on the number of infringers, 87.50% of Player 1s lie about state Y at least once in NO S, compared to 44.44% in FINE and 62.50% in COMP. As a result, 100% of Player 2s

[15] As argued earlier, lying in state X is irrational given the implemented payoff structure and therefore not relevant for our analysis. In fact, we observe only two lies in state X in total: one in NO S and one in FINE.

in No S encounter at least one liar during parts 1 and 2, whereas this is true for 63.89% in Fine and 79.17% in Comp. The corresponding Fisher's exact test p-values are reported in the three rightmost columns of Table 2.[16]

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
Share of lies about state Y				
No S	70.14%	< 0.001	< 0.001	0.001
Fine	30.88%			
Comp	47.06%			
Share of lying Player 1s				
No S	87.50%	< 0.001	0.001	0.043
Fine	44.44%			
Comp	62.50%			
Share of Player 2s who encounter at least one liar				
No S	100%	< 0.001	< 0.001	0.064
Fine	63.89%			
Comp	79.17%			

Notes: 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.

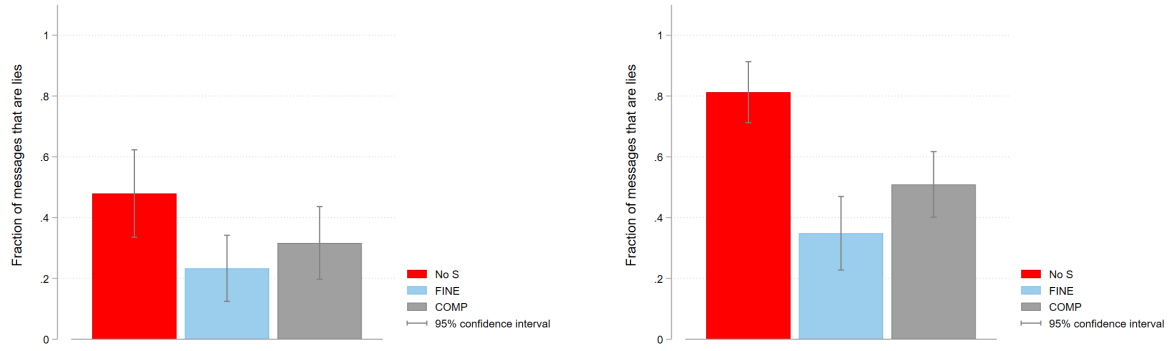
As explained in detail in Section 3.1, by design every third message sent by a Player 1 is randomly checked for truthfulness. As a result, in both Fine and Comp 86.11% of Player 1s are checked at least once during parts 1 and 2. Out of the respective full samples of Player 1s (not conditional on being 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 the behaviour of Player 1s (the potential infringers) in part 1 of the experiment, the only true one-shot interaction. We then turn to part 2, in which participants play the same one-shot game four additional times with new, randomly matched partners, allowing for learning effects net of additional reputational effects.

Panel (a) of Figure 2 presents the share of lying Player 1s in part 1. We observe most misbehaviour in No S: almost half (48%) lie when state Y prevails. In Fine and Comp, participants are more truthful: 23% lie in Fine and 32% in Comp. The difference between No S and Fine is highly significant (Fisher's exact test: No S vs Fine: $p < 0.01$, No S vs Comp: $p = 0.11$). The difference between Fine and Comp is not significant ($p = 0.41$).

[16] Note that here and throughout the paper we report p-values from two-sided test statistics.



(a) Part 1 (b) Part 2

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

We corroborate these findings in a linear probability model, see column (1) of Table A.3 in Appendix A. There, the treatment difference between No S and Comp is marginally significant ($p = 0.09$). In columns (2) and (3), we extend the regression and control for all items elicited in the final questionnaire.[17,18] The finding that Fine significantly reduces lying relative to No S remains virtually unchanged. The weaker differences between Comp and the other treatments are less robust to the inclusion/omission of controls.

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

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 under both 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 $53\% = 0.81 - 0.28$ and in the benchmark treatment No S $81\% = 0.81$. Moreover, Player 1s lie significantly less often in Fine than in Comp, as indicated by the Wald test at the bottom of the table ($p = 0.01$).[19] These treatment

[17] See Tables A.1 and A.2 in Appendix A for balance checks.

[18] We elicited the number of experiments participants had taken part in prior to the present experiment only in the last 14 of the 16 sessions. Column (2) includes all controls except "# Number participated in so far." Column (3) additionally controls for this variable, capturing potential effects of general lab experience and reducing N from 168 to 132. Apart from the Fine treatment effect, only participants' perceived treatment-specific deterrent effect consistently reduces their propensity to lie. Participants' experiment experience, conversely, tends to increase it. Logit regressions confirm the results from the linear probability models in Table A.3, results available on request.

[19] Across treatments, the propensity to lie increases from part 1 to part 2. This can be explained by the repeated setting of part 2. In a meta-study on the workhorse model for altruism and pro-sociality, the dictator

differences are consistent with treatment-specific differences in guilt aversion among Player 1s.

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

	(1)	(2)	(3)	(4)	(5)
FINE	-0.47***	-0.30***	-0.30***	-0.31***	-0.27***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
COMP	-0.28***	-0.17***	-0.16***	-0.18**	-0.15
	(0.00)	(0.00)	(0.00)	(0.01)	(0.20)
Player has lied before		0.50***	0.44***	0.40***	0.44***
		(0.00)	(0.00)	(0.00)	(0.00)
A previous lie was successful			0.09	0.11	0.09
			(0.27)	(0.21)	(0.28)
FINE × Player was caught lying before				0.11	
				(0.46)	
COMP × Player was caught lying before				0.06	
				(0.52)	
FINE × Player was checked for lying before					-0.04
					(0.49)
COMP × Player was checked for lying before					-0.01
					(0.93)
Constant	0.81***	0.49***	0.49***	0.50***	0.49***
	(0.00)	(0.00)	(0.00)	(0.00)	(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 the 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.

Columns (2) and (3) confirm that the treatment differences remain significant, though smaller than those estimated in column (1), when controlling 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.

We replicate these findings in further regressions presented in Table A.4 in Appendix A, in which we control for rounds as well as for all items elicited in the final questionnaire.

game, Engel (2011) similarly finds less pro-social behaviour in repeated than in one-shot interactions.

Finding 1: *In line with Hypothesis 1, lying is less frequent in* Fine *and* Comp *than in* No S *in parts 1 and 2; both sanction schemes promote compliance. In the repeated setting of part 2, the results furthermore support Hypothesis 2: Player 1s lie less in* Fine *than in* Comp.

In models (4) and (5), we extend the specification from column (3) by interacting the sanction treatment dummies with the indicators "Player was caught lying before" and "Player was checked for lying before", respectively, to test whether experiencing the law in either way further boosts compliance in Fine or Comp.[20] The interaction terms are insignificant, indicating no additional compliance effects from these experiences. These findings are largely robust to controlling for rounds and all items elicited in the final questionnaire (Table A.4)[21], as well as to the choice of regression model, see the results from additional logit regressions in Table A.5 in Appendix A.

4.2 Player 2s' trust in parts 1 and 2

We now turn to the potential victims. We are primarily interested in treatment differences in Player 2s' trust, which we measure through their propensity to choose action C when their matched Player 1 reports that state X prevails.

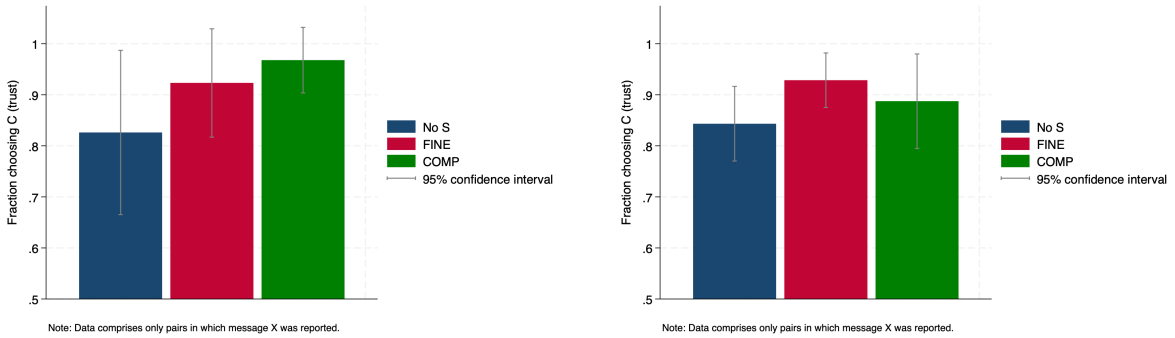
Panel (a) of Figure 3 presents the share of Player 2s who choose C upon receiving message X in part 1. In Comp, almost all (97%) follow message X, in Fine, 92% and even in No S, 83%. Behaviour across treatments is not significantly different (Fisher's exact tests: No S vs Fine: $p = 0.40$, No S vs Comp: $p = 0.15$, Fine vs Comp: $p = 0.59$). Linear probability models that control for replies to questionnaire items reach the same conclusion. Behavioural trust is generally high and not significantly different across treatments, see Table A.6 in Appendix A.[22]

Behaviour changes in part 2; compare panels (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 (No S vs Fine:

[20] In these regressions, the Fine and Comp coefficients capture treatment effects among 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, by design, these mechanisms do not exist in our benchmark treatment No S.

[21] In Table A.4, we additionally find that having been caught lying previously increases lying in Fine (column (4)). In column (6), the Comp coefficient turns significant, suggesting that Player 1s in Comp who have not yet been checked for lying have a lower propensity to lie than their No S counterparts. We do not overemphasise these findings, as they emerge in specifications controlling for lab experience, which reduces the number of observations to 312.

[22] Somewhat mirroring the pattern for Player 1s, we find suggestive evidence that Player 2s' trust decreases with experiment experience (Table A.6). Note, however, that adding this control reduces the number of observations to 65. Logit regressions confirm the insignificant treatment differences from the linear probability models reported in Table A.6 (results available on request).



(a) Part 1 (b) Part 2

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

$p = 0.05$). All other results from pairwise treatment comparisons are insignificant (No S vs Comp: p = 0.27, Fine vs Comp: p = 0.68).

Similar to Section 4.1, we use a regression analysis (linear probability models) to study 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 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 significantly 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 for whether a Player 2 has been lied to in any previous round. Intriguingly, the respective coefficient turns out significantly negative: after having been lied to, Player 2s 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 trust effects predicted in Hypothesis 1. In some specifications for part 2, trust is higher in* Fine *than in* No S*. By contrast, having been lied to in a previous round robustly and significantly reduces subsequent trust, suggesting that trust is shaped more by prior experiences of misconduct than by the sanction scheme itself.*

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.7). Additional logit regressions further corroborate the findings from Table 4, see Table A.8 in 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.9 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.[23]

In models (3) and (4) of Table 4, we extend the regression model from column (2) and interact the sanction treatment dummies 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 a liar in one of the previous rounds remains the only robust explanatory factor in these models.

[23]In column (1) of Table A.9, 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.9, we keep the "State Y in any of the previous 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.9, we repeat 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.

Table 4: Player 2's decision to choose C (trust) 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 × A previously matched			-0.07	
Player 1 was caught lying			(0.51)	
COMP × A previously matched			-0.04	
Player 1 was caught lying			(0.66)	
FINE × A previously matched				0.02
Player 1 was checked for lying				(0.59)
COMP × A previously matched				0.01
Player 1 was checked for lying				(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 the 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.

5 Removing the sanctions

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

Lastly, we study Player 1s' behaviour in part 3, in which all participants play one final round under the sanction-free scheme NO S.

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 schemes of FINE and COMP, exert compliance effects that carry over into 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 ($p = 0.76$).

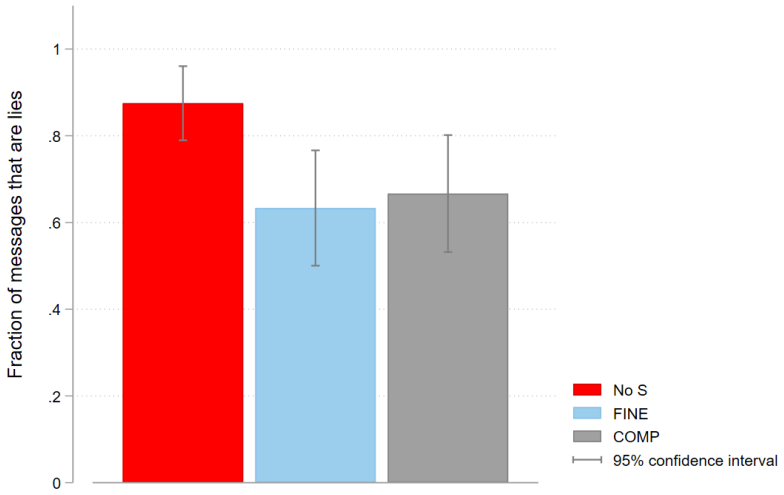


Figure 4: Player 1's misbehaviour in part 3

The absence of significant differences between FINE and COMP in part 3 is consistent with the idea that the enforcement-specific asymmetry between the two schemes disappears once sanctions are removed. Under compensation, infringers can no longer mitigate anticipated guilt ex post by compensating the victim, eliminating the key behavioural distinction between the two regimes.

Linear probability model (1) in Table 5 corroborates the findings from the non-parametric tests. However, once 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 for personal characteristics, see Table A.10. Moreover, we reproduce the findings from Table 5 in alternative logit regressions, see Table A.11 in Appendix A. This strongly suggests that the sanction schemes' carry-over effects are driven by those Player 1s who were deterred from lying early on and subsequently continue to behave honestly once sanctions are lifted:

Finding 3: *In line with Hypothesis 3, lying is less frequent in* FINE *and* COMP *than in* NO S *in part 3 of the experiment. These effects are driven by sustained compliance among Player 1s who were successfully deterred from lying from the beginning of the experiment.*

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

	(1)	(2)	(3)
Fine	-0.24***	-0.02	0.00
	(0.00)	(0.74)	(0.94)
Comp	-0.21**	-0.05	-0.01
	(0.01)	(0.36)	(0.93)
Player has lied before		0.57***	0.26*
		(0.00)	(0.08)
A previous lie was successful			0.36***
			(0.01)
Constant	0.88***	0.37***	0.34***
	(0.00)	(0.00)	(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 the baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: * p<0.10, ** p<0.05, *** p<0.01.

5.2 Potential victims' trust in part 3

Figure 5 suggests 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 vs Comp: $p = 0.61$, Fine vs Comp: $p = 0.17$).

These findings are corroborated in corresponding linear probability models, see Table 6. The treatment coefficients for Fine and Comp are insignificant; a Wald test comparing them is also insignificant, see column (1). Controlling for whether a Player 2 has been matched to a lying Player 1 in a previous round 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 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.

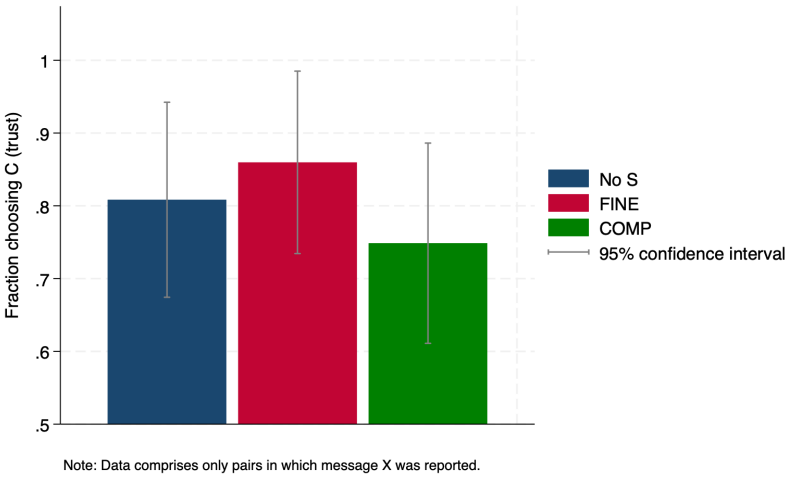


Figure 5: Player 2s choosing C (trust) in part 3

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

	(1)	(2)
FINE	0.05	-0.04
	(0.54)	(0.61)
COMP	-0.06	-0.11
	(0.48)	(0.26)
A previously matched Player 1 lied		-0.22***
		(0.00)
Constant	0.81***	1.03***
	(0.00)	(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 the baseline treatment in all regressions. Robust standard errors are clustered at the matching group level, p-values given in parentheses: *** p<0.01.

These findings are quantitatively and qualitatively unaffected by the inclusion of controls for personal characteristics, see Table A.12. Moreover, Table A.13 replicates the findings from the linear probability models presented in Table 6 in additional logit regressions, see Appendix A.

Finding 4: *Contrary to Hypothesis 3, Player 2s' propensity to trust in part 3 is not significantly affected by the treatment they experienced in parts 1 and 2. Their behaviour is better explained by previous encounters with lying Player 1s.*

6 Discussion

Considering the observed treatment differences in Player 1s' propensity to lie across all three parts of the experiment, we conclude that both sanction schemes exert robust compliance-enhancing effects relative to No S, consistent with Hypothesis 1. In part 2, we furthermore observe significantly less lying in Fine than in Comp, in line with Hypothesis 2 and consistent with the idea that compensation may weaken deterrence by allowing infringers to mitigate anticipated guilt ex post.

At the aggregate level, the compliance effects of both enforcement schemes extend into part 3 of the experiment, in which misconduct is no longer punished. These short-run carry-over effects appear to be driven by the fact that both sanction schemes deterred a non-negligible proportion of Player 1s from becoming "first offenders", making them less inclined to lie once sanctions are lifted. At the same time, the absence of significant differences between Fine and Comp in part 3 is consistent with our argument that the compensation-specific guilt-mitigation channel ceases to operate once sanctions are removed.

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 trust is shaped less by the formal presence of a particular sanction scheme per se than by accumulated experience with actual compliance behaviour under that scheme. This is consistent with the deterrence-based trust mechanism developed in Section 2.2, according to which individuals update their trust based on experienced misconduct and compliance over repeated interactions. In part 3 we do not observe any carry-over effects of either of the sanction schemes 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 expectation 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 each other (p = 0.90). These treatment differences replicate when considering part 1 payoffs and part 2 payoffs separately.[24]

We conclude by 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 only partially mirrors these treatment differences. While Player 2s tend to trust more in FINE than in NO S, this difference does not reflect a treatment-specific sanction effect per se, but rather the consequence of "first-hand" experiences with lying or compliant Player 1s. This pattern is consistent with our theoretical argument that deterrence-based trust evolves through accumulated experience with compliance behaviour rather than adjusting immediately to the formal introduction of a sanction scheme.

These findings indicate that it might take quite some time to install trust through enforcement schemes as potential victims need to experience their impact on compliance – potentially over an extended period or across a large number of interactions.

7 Conclusion

From the potential infringers' point of view, compensation is not just fine: our evidence shows that fines induce higher compliance, consistent with explanations that assume at least some degree of guilt aversion on the side of the potential infringers. Firstly, even when the probability of detection and the sanction amount are held constant, fine-based and compensation-based enforcement schemes do not yield the same behavioural outcomes.

Secondly, our results question the idea that compensation schemes are necessarily more attractive from the perspective of potential victims. While compensation may provide some financial redress ex post, our findings indicate that greater protection against misbehaviour ex ante is achieved under fine-based enforcement. In this sense, fines reduce exposure to misconduct more effectively and, through this deterrent effect, can foster higher levels of trust.

Thirdly, our study provides new insights into how both infringers and victims adapt once they have experienced the law and been confronted with sanctions. Admittedly, those results

[24] 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.

convey a rather pessimistic picture. It appears that the sanction schemes primarily deterred potential infringers from becoming first offenders, but having experienced the law by being checked for or even caught lying does not lower their propensity to lie (again) in the future. Similarly, although the presence of sanctions initially increased trust, its effect is fragile and quickly eroded once misbehaviour was experienced.

Taken together, our findings underline the need to consider both compliance and trust when designing enforcement schemes for consumer protection and contractual relationships.

For policymakers, the choice between fines and compensation is rarely only about compliance incentives. Fines generate public revenue, compensation may raise victim satisfaction, and administrative costs differ across approaches. Our findings add crucial behavioural evidence to this debate by showing how fines and compensation affect actual interactions between potential infringers and potential victims.

Further research is needed to expand on these results. Future work could examine how detection probabilities and sanction sizes interact with the two enforcement schemes, or explore whether fines and compensation differ in their ability to prevent unintentional harm as opposed to deliberate wrongdoing. It would also be valuable to incorporate belief elicitation in order to better understand the extent to which different sanction schemes shape perceived social norms and expectations about compliant behaviour. Finally, it would be worthwhile to consider enforcement designs that combine fines and compensation, potentially capturing the strengths of both approaches.

Acknowledgments

We thank Kompetenzzentrum Nachhaltige Universität (KNU) and the Dutch Royal Foundation KNAW for financial support to carry out this experiment. We are grateful for helpful comments from seminar and conference participants at the 2019 Lüneburg Workshop on Microeconomics, the 14th Nordic Conference on Behavioral and Experimental Economics, the 2019 FLEX 10-Year Anniversary Conference, the 2019 European Meeting of the Economic Science Association, participants of 2019 German association in Law and Economics meeting in Hannover, the 2022 Research workshop at King's College London, the 2023 conference of the European Society for Empirical Legal Studies, the 2024 Conference of the European Economic Association and the 2024 Annual Meeting of the German Economic Association.

References

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

Akerlof, George A, "The market for "lemons": Quality uncertainty and the market mechanism," in "Uncertainty in economics," Elsevier, 1978, pp. 235–251.

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 Investments in Care," *The Journal of Legal Studies*, 2024, *53* (1), 209–236.

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 deterence: When and how much," *Alabama Law Review*, 1988, *40*, 1143.

Crawford, Vincent P and Joel Sobel, "Strategic information transmission," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1431–1451.

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

DeAngelo, Gregory and Gary Charness, "Deterrence, expected cost, uncertainty and voting: Experimental evidence," *Journal of Risk and Uncertainty*, 2012, *44* (1), 73–100.

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.

_ **and Daniel Nagin**, "Who is afraid of the stick? Experimentally testing the deterrent effect of sanction certainty," *Review of Behavioral Economics*, 2015, *2* (4), 405–434.

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, "Deception: The role of consequences," *American Economic Review*, 2005, *95* (1), 384–394.

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

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

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

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.

Villeval, Marie Claire, "Public goods, norms and cooperation," in "Handbook of experimental game theory," Edward Elgar Publishing, 2020, pp. 184–212.

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 tables and analyses

Table A.1: Balance of Player 1 Characteristics by Treatment

	Means (SD)			p-values		
	No S	Fine	Comp	No S vs Fine	No S vs Comp	Fine vs Comp
Gender: Female	0.44	0.50	0.49	0.506	0.605	0.869
	(0.50)	(0.50)	(0.50)			
Risk proneness	6.33	5.67	6.33	0.099	1.000	0.073
	(1.97)	(2.26)	(2.18)			
Age	24.62	25.47	25.10	0.348	0.532	0.614
	(4.46)	(5.05)	(3.74)			
Belonging to the majority in terms of nationality	0.81	0.85	0.76	0.621	0.531	0.209
	(0.39)	(0.36)	(0.43)			
Studies Law (2nd largest group)	0.06	0.11	0.07	0.370	0.882	0.387
	(0.24)	(0.32)	(0.26)			
Studies Economics or Business (largest group)	0.50	0.38	0.42	0.178	0.373	0.612
	(0.51)	(0.49)	(0.50)			
General trust	5.04	5.39	5.44	0.427	0.360	0.882
	(2.48)	(2.24)	(2.26)			
Opinion: importance of sustainability	8.06	8.07	7.88	0.983	0.524	0.499
	(1.54)	(1.83)	(1.60)			
Opinion: importance of fair legal system	9.02	9.25	9.14	0.374	0.577	0.607
	(1.21)	(1.48)	(1.08)			
Feeling treated fairly as Player 1	6.33	7.22	7.06	0.067	0.143	0.694
	(2.71)	(2.50)	(2.58)			
Perception of deterrent effect	5.35	4.82	5.26	0.321	0.874	0.332
	(3.27)	(2.59)	(2.88)			
# Experiments participated in so far	2.12	2.38	2.31	0.114	0.286	0.512
	(0.74)	(0.64)	(0.70)			

Notes: Means are reported with standard deviations in parentheses.
p-values are from pairwise t-tests across treatment groups.

Table A.2: Balance of Player 2 Characteristics by Treatment

	Means (SD)			p-values		
	No S	Fine	Comp	No S vs Fine	No S vs Comp	Fine vs Comp
Gender: Female	0.51	0.50	0.51	0.875	0.958	0.907
	(0.50)	(0.50)	(0.50)			
Risk proneness	6.19	5.74	6.24	0.119	0.844	0.051
	(1.71)	(1.79)	(1.70)			
Age	25.25	25.88	25.04	0.377	0.742	0.168
	(4.21)	(4.09)	(4.10)			
Belonging to the majority in terms of nationality	0.81	0.84	0.79	0.577	0.694	0.289
	(0.39)	(0.37)	(0.41)			
Studies Law (2nd largest group of participants)	0.06	0.09	0.07	0.437	0.823	0.527
	(0.24)	(0.29)	(0.25)			
Studies Economics or Business (largest group of participants)	0.42	0.34	0.43	0.232	0.797	0.105
	(0.49)	(0.48)	(0.50)			
General trust	5.41	5.58	5.57	0.594	0.606	0.979
	(1.31)	(1.28)	(1.30)			
Opinion: importance of sustainability	7.95	8.06	8.07	0.606	0.571	0.972
	(1.55)	(1.49)	(1.52)			
Opinion: importance of fair legal system	9.19	9.34	9.32	0.336	0.333	0.875
	(1.08)	(0.97)	(1.01)			
Feeling treated fairly as Player 2	5.26	6.80	6.69	0.000	0.000	0.741
	(2.10)	(1.95)	(2.00)			
Perception of deterrent effect	5.62	5.28	5.51	0.358	0.767	0.473
	(1.67)	(1.72)	(1.70)			
# Experiments participated in so far	2.04	2.32	2.26	0.023	0.076	0.486
	(1.10)	(1.25)	(1.18)			

Notes: Means are reported with standard deviations in parentheses.
p-values are from pairwise t-tests across treatment groups.

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

	(1)	(2)	(3)
Fine	-0.25***	-0.22**	-0.36***
	(0.01)	(0.02)	(0.00)
Comp	-0.16*	-0.14	-0.21*
	(0.09)	(0.13)	(0.05)
Gender: Female		0.03	-0.05
		(0.67)	(0.58)
Risk proneness		0.00	-0.01
		(0.79)	(0.48)
Age		-0.01	-0.01
		(0.38)	(0.20)
Belonging to the majority in terms of nationality		-0.12	-0.09
		(0.24)	(0.41)
Studies Law		0.04	0.13
(2nd largest group of participants)		(0.77)	(0.29)
Studies Economics or Business		0.07	0.12
(largest group of participants)		(0.39)	(0.17)
General trust		0.00	-0.00
		(0.79)	(0.98)
Opinion: importance of sustainability		0.00	0.03
		(0.98)	(0.38)
Opinion: importance of fair legal system		-0.04	-0.09**
		(0.37)	(0.04)
Feeling treated fairly as Player 1		-0.02	-0.02
		(0.15)	(0.24)
Perception of treatment specific deterrent effect		-0.05***	-0.06***
		(0.00)	(0.00)
# Experiments participated in so far			0.09*
			(0.08)
Constant	0.48***	1.39***	1.76***
	(0.00)	(0.00)	(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 the 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.4: Robustness check I, Player 1's decision to lie in part 2

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

	(1)	(2)	(3)	(4)	(5)
FINE	-2.13***	-1.85***	-1.86***	-1.90***	-1.67***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
COMP	-1.33***	-1.17***	-1.11***	-1.14***	-1.05
	(0.00)	(0.00)	(0.00)	(0.00)	(0.13)
Player has lied before		2.55***	2.03***	1.89***	2.05***
		(0.00)	(0.00)	(0.00)	(0.00)
A previous lie was successful			0.78	0.85	0.77
			(0.17)	(0.17)	(0.17)
FINE × Player was caught lying before				0.34	
				(0.71)	
COMP × Player was caught lying before				0.19	
				(0.76)	
FINE × Player was checked for lying before					-0.26
					(0.43)
COMP × Player was checked for lying before					-0.08
					(0.90)
Constant	1.47***	0.32	0.30	0.32	0.30
	(0.00)	(0.31)	(0.35)	(0.33)	(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 the 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.6: Player 2's decision to choose C (trust) in part 1

	(1)	(2)	(3)
FINE	0.10	0.10	0.03
	(0.32)	(0.44)	(0.80)
COMP	0.14	0.15	0.10
	(0.11)	(0.27)	(0.46)
Gender: Female		0.01	0.01
		(0.93)	(0.95)
Risk proneness		-0.00	-0.02
		(0.94)	(0.45)
Age		0.00	0.01
		(0.37)	(0.25)
Belonging to the majority in terms of nationality		0.04	-0.07
		(0.76)	(0.26)
Studies Law		0.02	0.00
(2nd largest group of participants)		(0.84)	(1.00)
Studies Economics or Business		-0.08	-0.13
(largest group of participants)		(0.42)	(0.15)
General trust		-0.00	0.01
		(0.93)	(0.45)
Opinion: importance of sustainability		-0.01	-0.01
		(0.70)	(0.56)
Opinion: importance of fair legal system		-0.01	0.04
		(0.83)	(0.50)
Feeling treated fairly as Player 2		-0.00	-0.02
		(0.83)	(0.45)
Perception of treatment specific deterrent effect		-0.02	-0.01
		(0.34)	(0.56)
# Experiments participated in so far			-0.08*
			(0.08)
Constant	0.83***	1.02	0.87
	(0.00)	(0.11)	(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 the 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.7: Robustness check I, Player 2's decision to choose C (trust) in part 2

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

	(1)	(2)	(3)	(4)
Fine	0.85**	0.47	0.60	0.32
	(0.05)	(0.23)	(0.15)	(0.53)
Comp	0.28	0.08	0.17	-0.00
	(0.58)	(0.88)	(0.75)	(1.00)
A previously matched Player 1 lied		-1.39***	-1.29***	-1.40***
		(0.00)	(0.00)	(0.00)
Fine × A previously matched Player 1 was caught lying			-0.67	
			(0.32)	
Comp × A previously matched Player 1 was caught lying			-0.25	
			(0.68)	
Fine × A previously matched Player 1 was checked for lying				0.33
				(0.58)
Comp × A previously matched Player 1 was checked for lying				0.15
				(0.74)
Constant	1.69***	2.64***	2.56***	2.65***
	(0.00)	(0.00)	(0.00)	(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 the 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.9: Robustness check III, Player 2's decision to choose C (trust) in part 2

	(1)	(2)	(3)	(4)
FINE	0.05	0.07*	0.04	0.08*
	(0.24)	(0.09)	(0.26)	(0.06)
COMP	0.02	0.03	0.01	0.03
	(0.80)	(0.66)	(0.84)	(0.58)
A previously matched Player 1 lied	-0.15***		-0.15***	
	(0.00)		(0.00)	
State Y in any of the previous rounds	0.03	-0.04		
	(0.57)	(0.48)		
State Y in the previous round			0.03	-0.01
			(0.19)	(0.75)
Constant	0.90***	0.88***	0.91***	0.85***
	(0.00)	(0.00)	(0.00)	(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 the 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 1's decision to lie in part 3

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

	(1)	(2)	(3)
Fine	-1.40***	-0.24	0.02
	(0.00)	(0.59)	(0.97)
Comp	-1.25***	-0.48	-0.04
	(0.01)	(0.33)	(0.94)
Player has lied before		3.10***	1.07*
		(0.00)	(0.09)
A previous lie was successful			2.72***
			(0.00)
Constant	1.95***	-0.34	-0.64
	(0.00)	(0.43)	(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 the 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.12: Robustness check I, Player 2's decision to choose C (trust) in part 3

	(1)	(2)	(3)	(4)
FINE	0.06	-0.02	-0.02	-0.08
	(0.49)	(0.87)	(0.85)	(0.48)
COMP	-0.05	-0.12	-0.07	-0.13
	(0.58)	(0.31)	(0.46)	(0.28)
A previously matched Player 1 lied			-0.29***	-0.28**
			(0.01)	(0.01)
Gender: Female	-0.04	-0.06	-0.05	-0.06
	(0.65)	(0.57)	(0.57)	(0.57)
Risk proneness	-0.03	-0.02	-0.03	-0.01
	(0.12)	(0.41)	(0.15)	(0.58)
Age	0.01	0.01	0.01	0.01
	(0.18)	(0.28)	(0.33)	(0.51)
Belonging to the majority in terms of nationality	0.14	0.16	0.13	0.15
	(0.22)	(0.23)	(0.25)	(0.25)
Studies Law	0.15	0.20**	0.18*	0.23**
(2nd largest group of participants)	(0.15)	(0.04)	(0.10)	(0.02)
Studies Economics or Business	0.06	0.03	0.04	0.01
(largest group of participants)	(0.51)	(0.81)	(0.64)	(0.90)
General trust	0.02	-0.00	0.02	0.00
	(0.40)	(0.92)	(0.30)	(0.94)
Opinion: importance of sustainability	0.02	0.01	0.01	0.00
	(0.42)	(0.73)	(0.69)	(0.88)
Opinion: importance of fair legal system	-0.02	-0.01	-0.02	-0.01
	(0.67)	(0.78)	(0.70)	(0.81)
Feeling treated fairly as Player 2	-0.00	0.00	-0.02	-0.02
	(0.69)	(0.78)	(0.12)	(0.38)
Perception of treatment specific deterrent effect	0.01	0.01	0.00	-0.00
	(0.68)	(0.75)	(0.93)	(0.95)
# Experiments participated in so far		-0.07		-0.08*
		(0.15)		(0.09)
Constant	0.56	0.80	1.02**	1.25**
	(0.22)	(0.12)	(0.04)	(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 the 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.13: Robustness check II, Player 2's decision to choose C (trust) in part 3

	(1)	(2)
FINE	0.37	-0.19
	(0.53)	(0.72)
COMP	-0.37	-0.63
	(0.47)	(0.24)
A previously matched Player 1 lied		-2.31**
		(0.04)
Constant	1.45***	3.75***
	(0.00)	(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 the 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 instructions

Welcome to today's experiment!

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

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

Instructions

What is it about? – An overview

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

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

Payoffs in situation X

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

Payoffs in situation Y

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

In situation X, the following applies:

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

In situation Y, the following applies:

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

Please note:

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

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

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

1. Experimental procedures

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

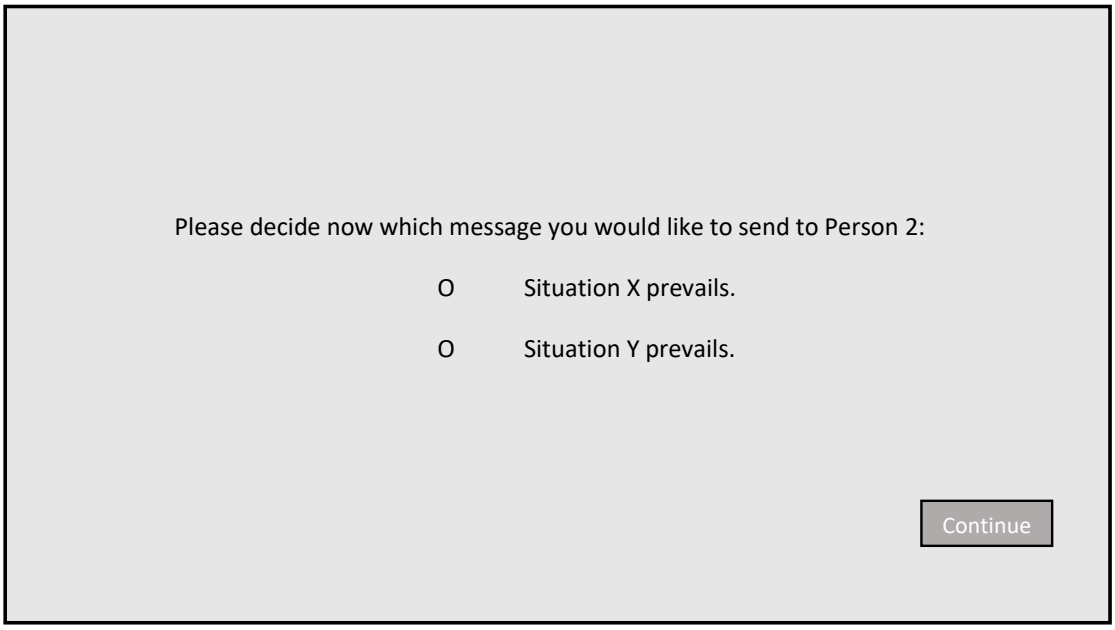
2. Allocation of roles

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

3. Procedure of the decision round in part 1

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

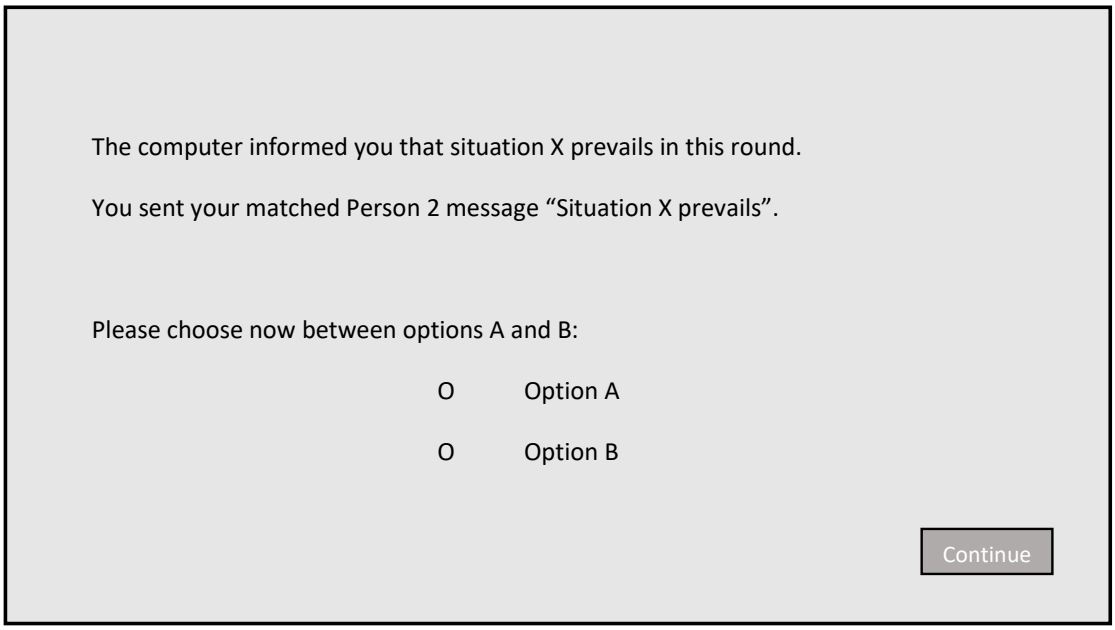
Then Person 1 can inform Person 2 about which situation prevails. He or she is obliged to transmit one piece of information – X or Y. The screen looks as follows:



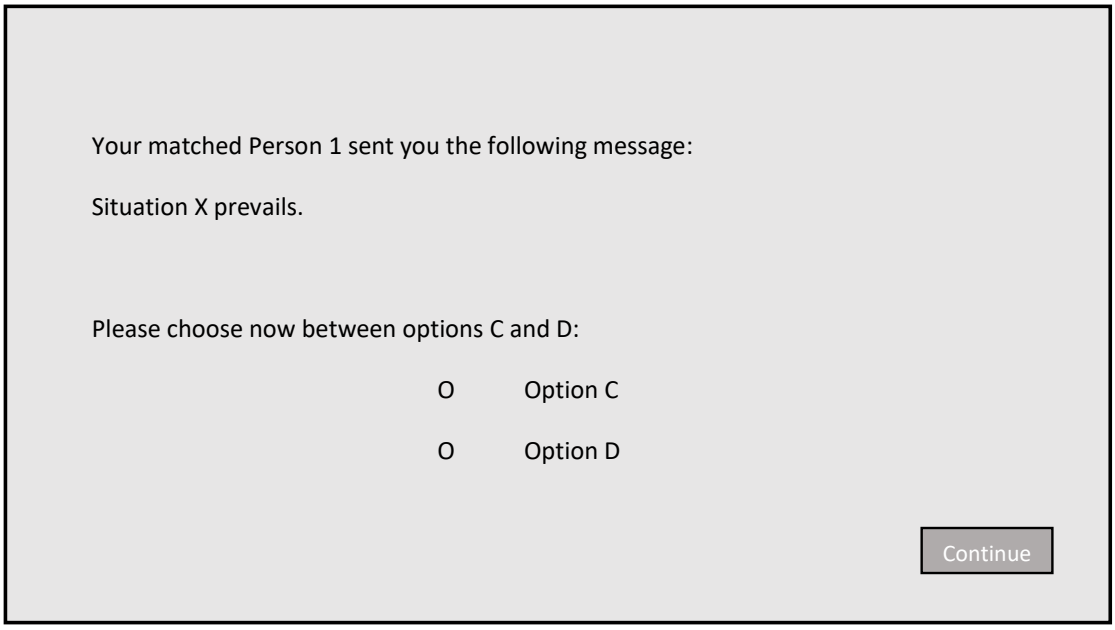
Next, Person 1 and Person 2 choose between their options. Person 1 makes a choice between Option A and Option B, Person 2 makes a choice between Option C and Option D.

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

Example: Decision screen of Person 1:



Example: Decision screen of Person 2:



4

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

[In Treatment Fine additionally:]

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

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

[In Treatment Comp additionally:]

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

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

4. Pay-out from today's experiment

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

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

Control questions

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

   Answer: _____

2. What is the probability that situation X prevails?

   Answer: _____%

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

   In points: _____

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

   In points: _____

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

   In points: _____

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

   In part 1: _____

   In part 2: _____

   In part 3: _____

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

   Answer: _____

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

   O no

[In Treatment Fine and Treatment Comp instead:]

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

   O every third

   O every forth

[In Treatment Fine and Treatment Comp additionally:]

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

   In points: _____

[In Treatment Comp additionally:]

9. Who then receives the amount deducted?

   Answer: _____

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

Instructions for part 2

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

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

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

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

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

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

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

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

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

[In Treatment Fine additionally:]

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

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

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

[In Treatment Comp additionally:]

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

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

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

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

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

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

Instructions for part 3

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

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

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

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

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

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

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

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

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

[In Treatment Fine additionally:]

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

[In Treatment Comp additionally:]

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

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

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

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

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

Questionnaire

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

Are you...?

O male
O female
O prefer not to say

How old are you?

_____ (Free text field)

What is your nationality?

_____ (Free text field)

What subject are you studying?

_____ (Free text field)

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

_____ (Free text field)

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

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

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

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

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

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

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

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

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

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

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

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

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

_____ (free text field)