Lying, Spying, Sabotaging
– Procedures and Consequences –

Nadine Chlaß* and Gerhard Riener‡

November 23, 2023

Abstract

We conduct experiments in which one party chooses the rules of a two-party winner-takes-it-all competition. This party can either choose 'fair' rules which grant herself and her opponent equal decision and information rights, and equal chances to win all payoff. She can also choose 'unfair' rules which allow her to cheat – that is, to fabricate, sabotage, or spy the opponent's actions. Resorting to fabrication, sabotage, or spying allows her to win all payoff for sure. Our results show, first, that a large share of individuals do not wish to win competition by cheating. They show, second, that this preference for fair competition springs from an ethical ideal purely about the equality of decision rights – that everybody should enjoy equal rights to pursue their own self-interest. They show, third, that all reservation against unfair competition disappears, if the party can cheat her opponent without taking away the latter's decision rights.

JEL: D02,D03,D63,D64

Keywords: institutional design, moral judgement, cheating, competition, distribution of decision rights

*Friedrich Schiller University Jena, Carl-Zeiss-Str. 3, D-07743 Jena, Germany. Email: nadine.chlass@uni-jena.de, Phone: 0049 (0)3641 943243. *Chlaß gratefully acknowledges financial support from Leverhulme Visiting Fellowship DKA 7200 at the Economics Department, University of Essex, UK.*

‡DICE, Heinrich Heine University Düsseldorf, Universitätsstraße 1, D-45372 Düsseldorf, Germany. Email: riener@dice.hhu.de. We thank audiences at the EEA/ESEM, the Royal Economic Society, and research seminars at the Universities of Cologne, Düsseldorf, Jena, Mannheim, Maastricht, and Turku for their valued input. Markus Prasser and Martin Schneider helped conducting the experiments. We thank Marie-Claire Villeval, Kaisa Kotakorpi, and Topi Miettinen for commenting the manuscript. This research was funded by the German Research Foundation under grant RTG 1411.

1 Introduction

In 2013, E. Snowden's leaks of classified information about global surveillance activities by the U.S. secret service led to an international diplomatic crisis. The leaks documented that – in pursuit of preventing terrorist attacks – the U.S. secret service had systematically and pre-emptively intercepted and stored private communications and information on U.S. citizens, foreign governments, heads of friendly nations, and sabotaged internet encryption as a means to this end.[1] In his interviews with the Guardian, Snowden stated that *'he was willing to sacrifice all [...] because he could not in good conscience allow the destruction of privacy and basic liberties [...]'* (Greenwald et al. 2013). Similarly, D. Ellsberg risked a 115 years sentence under the Espionage Act of 1917 cost by leaking the Pentagon Papers to reinstate the U.S. public's and congress's right of information about the government's evaluation of the Vietnam war (Sheehan 1971; Cooper and Roberts 2011).

Individuals' attitudes toward cheating are of fundamental relevance to economics. Mar- ket agents who reduce prices or seek innovation in their competition for revenue and gain, ultimately benefit the welfare of a society by pursuing their self-interest (Smith 1904). Yet, competitive pressure may also induce some agents to manipulate a competitor's cost, to fab- ricate information about her solvency to an investor, or to spy her business secrets to improve their competitive situation. If some agents resort to such activities whenever the opportu- nity presents itself while others avoid them whatever the gain implied, the self-regulating behaviour of the market place – Smith's invisible hand – is at stake. Understanding if and why individuals avoid cheating activities, or pursue them instead, is the key to understanding where competition fosters the common good, and where it does not.

The evidence on individuals' attitudes toward cheating, is, however, highly controversial. While individuals can hold strong reservations against some forms of cheating such as lying (Abeler et al. 2018) and sabotage (Harbring and Irlenbusch 2011), they also seem to inherently enjoy these activities at times (Charness et al. 2014; Abbink and Sadrieh 2009; Abbink and Herrmann 2011); spying appears to be seen as a largely legitimate activity in pursuit of individual self-interest (Beresford et al. 2012). The question why individuals dislike cheating is an ongoing controversy which has focused on lying to date. Is truth-telling a focal point for intuitive decision makers (Lightle 2014; Cappelen et al. 2013) who do not understand the mone- tary benefits from lying? Is lie aversion dis-

[1] Comments by NSA officials do not deny these activities and state they are 'hardly surprising' (Larson et al. 2013). Similarly, insiders broke practices of 'parallel construction' in the U.S. Drug Enforcement Administration to Reuter's journalists Shiffman and Cooke (2013): the 'fabrication' of investigative trails to cover up that trails are actually based on inadmissible evidence from NSA warrantless surveillance.

guised self-interest because one expects the truth to be mistaken for a lie anyway (Sutter 2009)? Do people suffer a psychological cost when lying which they trade off against the potential gains (Gneezy 2005; Erat and Gneezy 2012; Miettinen 2013)? Could guilt aversion, i.e. an aversion against disappointing others' expectations or against violating a social norm trigger this psychological cost (Battigalli et al. 2013; Miettinen 2013) Is, in the end, perhaps none of these motives at play (Villeval and van den Ven 2015) and is lie aversion for its larger part an innate concern (López-Pérez and Spiegelman 2013; Hurkens and Kartik 2009; Abeler et al. 2018)?

This paper shows that the lion's share of cheating aversion when individuals compete for payoff[2] ultimately springs from a common source, from an ethical ideal not unlike Edward Snowden's concern about basic liberties and rights: that all parties enjoy equal freedom to pursue their own self-interest, or put differently, that competition be fair. Where individuals can respect this ethical ideal and cheat at the same time, cheating aversion disappears and individuals enjoy cheating, a result which may provide the key to understanding why evidence on cheating aversion and joy-of-cheating coexist in the literature.

Our analysis rests on four cornerstones. First, we design a framework which allows us to observe three main forms of cheating from the literature – lying, spying, and sabotaging – in an identical setup and to measure the cheating aversion associated with each. In this setup, one party chooses how to compete with her opponent for all payoff. The party can opt for a constant sum game in which both competitors have equal decision and information rights, or she can – unbeknownst to her opponent – grant herself either an option to fabricate, spy, or sabotage the opponent's move. By lying, spying, or sabotaging, she transforms the constant sum game into a dictator game and can take all payoff for sure. We retrieve the distribution of cheating aversion from the literature: many parties deviate substantially from rational self-interest when they must lie or sabotage to win all payoff, but nearly all parties spy in pursuit of their own self-interest. Thereby, our setup accommodates various types of cheating aversion known to date: a

[2]A strand of literature in its own right (Abeler et al. 2014; Gibson et al. 2013; Fischbacher and Föllmi-Heusi 2013; Abeler et al. 2018) studies lies which do not affect any opponent other than the experimenter, mostly by means of Fischbacher and Föllmi-Heusi's (2013) die-roll task. In this task, subjects report the result of a die-roll; the reported roll determines a subject's payoff whereby the die roll cannot be monitored by the experimenter. Subjects' overall honesty unfolds from a comparison of the actual distribution of reports with the theoretical distribution of die rolls. For our results to apply in this setting, subjects must deem that they compete against the experimenter for payoff. Indeed, Fischbacher and Föllmi-Heusi (2013) find that lying and lie aversion do not depend on whether the opponent is a subject, or an experimenter. If so, subjects' degree of cheating aversion could spring from the choice of rules through the experimenter (rather than through a subject opponent, as in our own setup): in the die-roll task, the experimenter imposes a disadvantage in information and decision rights on herself. Subjects who are honest might depart from rational self-interest because of their advantageous position relative to the experimenter.

party can commit to the constant sum game and avoid any exposure to the cheating option. She can grant herself the option to cheat, feel remorse about doing so, and reduce her guilt by giving all payoff away. She can decide to cheat 'whitely' in order to give all payoff to the opponent. She can also weigh the psychological cost against her gain from cheating and set the likelihood of the cheating option accordingly. Parties can, however, not disguise their self-interest as cheating aversion.

Second, we take these different ways to depart from rational self-interest and analyze the ethical ideal(s) – if any – underlying each. To that end, we first elicit the ethical ideals which individuals actually employ to derive the right course of action. In the twentieth century, Piaget (1948) and Kohlberg (1984) conducted large-scale field studies to see which criteria individuals consult when they derive what they deem ethically right. We elicit individuals' preferences over the ethical ideals documented in this field work by way of a moral judgement test (Lind 1978; Lind 2008)[3] to identify how much individuals refer to punishment or reward, how much to the intention behind an action, to others' expectations and approval, to social norms and image, or to legal rules, when they derive the right course of action; how much to basic liberties and rights stipulated in a social contract, or to general ethical principles of conscience valid even beyond this contract. The ethical criteria in this taxonomy do, there- fore, cover the main ethical ideals put forth to explain cheating aversion to date. Surprisingly, however, we find that all ways to depart from rational self-interest link to one and the same ethical ideal – the equality of basic liberties and rights.[4]

The third cornerstone consists of two treatments which remove these ethical grounds for cheating aversion. In a first setup, the party who chooses the rules of competition can grant herself the option to cheat without changing the opponent's position of rights. In this case, we do indeed hardly observe any cheating aversion and individuals seem to inherently enjoy cheating as in (Charness et al. 2014). A second setup equalizes both

[3]The test minimizes bias from ex-post rationalization, for an extensive discussion of the issue, see (Chlaß and Moffatt 2012); in addition, our setup presents every subject with a variety of tasks such that attempts at ex-post rationalizing these manifold decisions through answering the test in a particular way becomes unfeasible. According to this result, the driving force behind cheating aversion is that one party is privileged over

[4]According to this result, the driving force behind cheating aversion is that one party is privileged over another by the rules of the game – be it in terms of information, or freedom to choose – and seeks to compensate her opponent for their disadvantage. This mechanism could, in principle, also be at play in other frameworks for the study of cheating aversion: In sender-receiver games, for instance, the sender defines the payoff consequence of the receiver's actions (by sending a message about the state of the world unknown to the latter). Tournament games typically endow subjects with different costs of effort, or different productivity; subjects with high cost of effort have lesser freedom of choice (fewer effort levels they can afford) to choose from than subjects with low cost of effort as in (Harbring et al. 2007; Harbring and Irlenbusch 2011).

parties' rights by granting the opponent an option to punish the choice of rules. Here, we observe that those individuals who refrained from cheating in the original setup, pursue their rational self-interest once rights are equally distributed.

The fourth cornerstone is a formal discussion of all treatments. We show theoretically that preferences other than for the equality of decision rights do not predict how parties' cheating aversion varies across our experimental setups. In particular beliefs about others' expectations and social norms which we deemed powerful predictors of cheating aversion, imply a variation of cheating aversion into a different direction than we observe.

Apart from showing that there is promising scope to integrate various strands and setups of the vast literature on cheating aversion, which implications do our findings have? For one, lying-, spying-, and sabotaging-like activities are part of many people's work lives (Abratt and Penman 2002). Online shops collect, analyse, and complete information on clients' buying behaviour to develop comprehensive customer profiles, personnel managers screen social media to obtain information about the social life, and the character of job candidates (Brown and Vaughn 2011), credit reference agencies collect and analyse information on financial incidents in people's lives[5], employees who develop or maintain software for cyber-security seek to exploit weaknesses in firms' or nations' security systems. Little is known about how indi- viduals react to the nature of such work. The introductory examples imply that, even after self-selecting into a workplace, people's reactions differ.

Thereby, the purposes of lying, spying, or sabotage may be altruistic ones. The desire to prevent terrorist attacks aims at saving lives; paying attention to the person-organization fit when hiring new employees may foster job satisfaction, a harmonious work atmosphere, and reduce moral hazard; matching clients with the products they wish to buy saves them time and cost. If, however, employees feel that the activities which they carry out to achieve these ends infringe others' rights and are wrong per se, employees may not succeed in justifying their work through its purpose. We observe intruiging behavioral implications of such con- flicts. Some individuals – 'procedural' – types avoid cheating altogether and nudge themselves into fair rules. Others – 'compensatory' – types opt into cheating and divert cheating from its intended use to benefit the potential victim. The shares of both types vary substantially across lying, sabotage, or spying. The 'procedural' type is most prevalent when unfair com- petition involves lying. The 'compensatory' type occurs most often with sabotage, and never with spying. Yet, as outlined above, both types depart from rational self-interest for the same ethical

[5]The German Schufa credit reference agency for example, holds and sells information about purchases, credit demand and credit worthiness of roughly 75% of the German population.

reason – the opponent's unprotected right to pursue her own self-interest.[6]

Ultimately, we find econometric evidence that both types differ along the well-known materalism-postmaterialism taxonomy. Materialists value hierarchy, duty, and power, post- materialists value individuality, the emancipation from authorities, and autonomy (Inglehart 1977; Baker and Inglehart 2000; Klages and Gensicke 2006). 'Procedural' types score higher on postmaterialist values than 'compensatory' types and the latter higher on materialist values. Indeed, the 'procedural' type foregoes all power by reinstating her opponent's information and decision rights. The 'compensatory' type exerts her power 'for good', trading the opponent's rights off against a monetary compensation.

In the next section we illustrate our main setup, section 3 outlines our experimental design in more detail and presents two modifications to the main setup. Section **??** presents the results, section 5 analyzes to what extent individuals' ways to make moral judgements and their values can organize those. Section 6 discusses our results and which economic preference models might explain them, and Section 7 concludes. In the next section we illustrate our main setup, section 3 outlines our experimental design in more detail and presents two modifications to the main setup. Section **??** presents the results, section 5 analyzes to what extent individuals' ways to make moral judgements and their values can organize those. Section 6 discusses our results and which economic preference models might explain them, and Section 7 concludes.

2 Lying, spying, and sabotaging: rules and payoffs

This section briefly illustrates which notions and payoff consequences of lying, spying, and sabotage we study in this paper. Table 1a) shows the spy-, lie-, and sabotage-free set of rules how two parties A and B can interact to allocate one ex-post non-zero payoff. Neither party has *information* about the opponent's move and hence, both parties are equally well off in terms of information. Parties also have the same *freedom of choice*: each party has two pure actions L and R each of which can be preferred by the same degree over the other given *some* circumstance: each action allows the individual to take all payoff for exactly one specific choice of the opponent (Jones and Sugden 1982).

B can choose the set of rules; she can either opt for this 'fair' set of rules, or she can

[6]Given that only this particular ethical ideal can be confirmed to be at play, the preference type which best explains the altruism under different rules, are Chlaß et al.'s (2019) purely procedural preferences: inequity aversion over decision and information rights. Note that in this paper, choices of rules and altruism could have linked to all main ethical criteria around which economics has formulated preferences: desires to comply with social norms, others' expectations, or others' intentions, maintaining one's social image, the status quo, seeking reward or avoiding punishment. We use individuals' propensity to invoke this entire set of moral criteria but only the concern about an equal position of (civic) rights shows an effect.

opt for a second set of rules where she spies, sabotages, or fabricates A's decision. Under this second 'unfair' set of rules, B transforms payoff matrix 1a) into payoff matrix 1b) where L^A and R^A denote the spied[7], fabricated, or sabotaged versions of A's actions L and R. This way, B obtains two identical dominant strategies LR^A and RL^A which secure all payoff for sure and A's choice becomes payoff-irrelevant.

Table 1: How does party B profit from spying, sabotaging, or fabricating A's decisions? Normal forms of the fair, and the unfair set of rules.

1a) the 'fair' set of rules

		party A: L	party A: R
party B	L	100 0	0 100
	R	0 100	100 0

1b) the 'unfair' set of rules

		party A: L	party A: R
party B	LL^A	100 0	100 0
	RL^A	0 100	0 100
	LR^A	0 100	0 100
	RR^A	100 0	100 0

We study three different activities through which B can transform payoff matrix 1a) into 1b). First, B can opt for a set of rules where she *spies*, that is, looks up A's decision while A cannot see B's choice. We describe spying more accurately in the extensive form game of Fig. 2 and describe the 'unfairness' of this set of rules in section 6 by the inequality in parties' information partitions over the outcomes – i.e. over the terminal histories – of the game at the time when parties choose their actions[8].

Second, B can opt for a set of rules where she *sabotages* A, that is, replaces A's decision and chooses in A's stead. Thus, if A chooses L, she may suddenly encounter the consequences of action R and vice versa. To date, sabotage has been conceptualized as increasing an opponent's cost of producing output (Harbring et al. 2007), as directly reducing others' output (Harbring and Irlenbusch 2011), as destroying others' output (Falk et al. 2008), or as manipulating how others' output performance is evaluated (Carpenter et al. 2010). In each formulation, sabotage redefines the link between the sabotaged party's action and the consequence – or utility – attached to this action, see e.g. appendix E. When B sabotages, she does not necessarily acquire information about

[7]Note that for the spying case, the normal form in table 1b) is not completely accurate since it suggests that A and B choose simultaneously. For B to be able to spy A's decision, however, A must already have made her choice. We capture these differences more accurately in section 3.1 by means of the extensive game form.

[8]The ideas used to express the unfairness of rules by the inequality in the distribution of information or decision rights and the corresponding quantitative measures are taken from (Chlaß et al. 2019).

what A has, or would have chosen; rather, she infringes A's freedom of choice. We capture sabotage in the extensive form game of Fig. 3 and describe the unfairness of this set of rules by the inequality in decision rights across parties A and B in section 6.

Third, B can transform payoff matrix 1a) into 1b) by anonymously reporting a *fabricated* decision for A which – upon reaching a third party – becomes payoff-relevant. Here, we think about planting or spreading rumours about an opponent which upon reaching a superior, become payoff-relevant while nobody observes whether the rumour was intentionally planted or just an innocent or failed guess. In this paper, the fabricated action always becomes payoff-relevant such that fabrication is always 'successful'.

Throughout, we study fabrication, spying, and sabotage as *clandestine* activities. Party A never learns whether B opted for the fair, or for the unfair set of rules, that is, whether B spied, sabotaged, or fabricated A's decisions. Hence, A does not know whether the payoff matrix is 1a) or 1b). B can cheaply arrive or 'nudge' herself into the spy-, lie-, or sabotage-free set of rules, or into the set of rules which allows for fabrication, spying, or sabotage. This nudge could be a party's choice to walk to her own desk without passing her colleague's (or deliberately passing that desk, respectively) in order to forego (or obtain) the chance to spy or manipulate that colleague's progress. Similarly, it could be avoiding the coffee corner to prevent being part in creating or spreading rumours about others.

More formally, we can measure A's freedom of choice in Jones's and Sugden's (1982) and Sugden's (1998) *metric of opportunity*. Actions L and R do not expand A's freedom of choice in 1b) since no economic preference type would predict that $R \nsim L$. If R and L are identical then A does not prefer choice set $\{L, R\}$ to choice set $\varnothing$ in 1b). In 1a), however, $R \nsim L$ in some circumstances and hence A may prefer $\{L, R\}$ to $\varnothing$. Therefore, when B chooses the 'unfair' set of rules, she reduces A's choice set compared to 1a), and compared to her own choice set. If B deemed that both parties should have equal decision rights, she would hold reservations against doing so. These reservations should crowd out when B can secure all payoff under *both* sets of rules and cannot reduce A's freedom of choice. These reservations should also lessen as soon as A exerts control about how much L and R expand B's freedom of choice via punishment or reward, see appendix D. Finally, such reservations should exist under fabrication and sabotage which attach new consequences to A's actions, but not under spying which affects A's relative position of information rights but not her freedom of choice.

3 Experiment

The experiment proceeded in three parts as shown in table 2. At the outset, subjects were seated at visually isolated computer terminals, and handed a hard copy of the German instructions for our baseline treatment.[9] Instructions for two upcoming parts 2 and 3 were shown on screen, once the experiment had proceeded this far; at no point in time did subjects have information about any upcoming parts. Once participants had confirmed on screen they had read the instructions, the experiment started automatically by a set of control questions which all participants answered successfully. Subsequently, participants were randomly assigned either role A, B, or C. A and B participants were randomly matched into pairs and two C participants assigned to each session of treatment LIE. Next, Bs chose between two situations S_1 and S_2 at their own discretion, S_1 offering symmetric decision and information rights, and S_2 affording B the opportunity to either LIE, SPY, or SABOTAGE. A and B next submitted their choices for the situation determined by B – A's options being identical across S_1 and S_2 such that the situation remained B's private information always. No feedback was given after part 1. Part 2 proceeded the same way, except that A was announced to have an option to punish or reward B's choice of the situation. Again, B chose between S_1 and S_2 after which both A and B made their decisions for the situation selected by B. No feedback was given after part 2. Part 3 elicited risk and envy preferences, demographics, administered a pen-and-paper moral judgement test, and the pen-and-paper ranking scales for materialist and postmaterialist values. Only one of the first two parts was paid out, part 3 was always paid; average payments included a show-up fee of €2.50 and amounted to €7.94 (min: €3.60, max: €12.10) where €1 $\widehat{=}$ \$1.28 at the time. 630 subjects participated, 49% of them female.[10]

3.1 Part 1: Baseline Treatments LIE, SPY, and SABOTAGE

Figure 1 formalizes part 1. A and B have an initial endowment of 50 ECU, the show-up fee of €2.50. B moves first and chooses the probability Prob (S_2) for situation S_2 which is initialized at 50%. This default has two purposes: first, it portrays an unintentional choice and second, does not point subjects toward either S_1 or S_2. Each one percent change to this default costs B 0.1 ECU where 1 ECU = €0.05. B may therefore select one situation for sure at the relatively small cost of 5 ECU or 25 Euro Cents. Next, situations S_1 and S_2 are drawn according to B's choice of Prob (S_2). A neither knows Prob (S_2), nor the situation which is drawn. She chooses between L (left), R (right), and the toss of a fair coin between the two. B's choices in turn depend on the situation which is drawn. If S_1 is drawn, B's choices are the same as A's, and neither A nor B know the opponent's choice. If S_2 is drawn in treatment SPY, B's choices are the same as A's, but B *sees* A's choice. If S_2 is drawn in treatment SABOTAGE, B overrides

[9] See appendix B for translations of these instructions into English for our three baseline treatments LIE, SPY, and SABOTAGE.

[10] A session lasted approximately 50 minutes including payment. Subjects were undergraduate students and native German speakers at Friedrich-Schiller-University Jena, randomly recruited from all fields of study via ORSEE (Greiner 2004). At the time of the experiment, the subject pool counted around 3000 students. The experiment was programmed in z-Tree (Fischbacher 2007). Payouts were distributed in sealed envelopes; receipts did not match subjects' names with their client numbers.

Treatment	*Spy*		*Sabotage*		*Lie*	
Payoff regime	Neutral	Competitive	Neutral	Competitive	Neutral	Competitive
B-participants	# 52	# 54	# 53	# 53	# 47	# 44
Part 1	**Baseline** B chooses probability Prob(S_2) of situation S_2 A chooses L or R In S_2, B learns A's choice | In S_2, B overrules A's choice | In S_2, B transmits A's choice to C B chooses L or R					
Part 2	**Reward and Punishment** B chooses probability Prob(S_2) of situation S_2 A chooses L or R In S_2, B learns A's choice | In S_2, B overrides A's choice | In S_2, B transmits A's choice to C B chooses L or R A chooses punishment/reward schedule without knowing Prob(S_2), the situation, or B's choice. B submits 1st order beliefs about A's punishment and reward schedule.					
Part 3	**Covariates** Risk Preferences Envy Moral Judgement Test (pen and paper) Materialist and Postmaterialist values (pen and paper) Demographics					

Table 2: Experimental design

A's unknown choice by L (left) or R (right), and chooses for herself between L (left), R (right), and the toss of a fair coin betweeen the two. If S_2 is drawn in treatment LIE, B transmits some choice for A and her own choice to participant C who implements the choices transmitted. Throughout SPY, SABOTAGE and LIE, A and B have an equal ex-ante chance to obtain the payout of a constant sum game if B selects S_1; in S_2, B has all allocation power and can secure this payout. Fabrication, spying, and sabotage therefore turn the constant sum game S_1 shown in table 3a into dictator game S_2. Thereby, SABOTAGE and LIE allow B to take decision rights from A, whereas SPY allows B to increase her own information rights. In a second variant *payoff neutrality*, S_1, too, is a dictator game such that B's power to take A's decision rights is removed from all treatments and B may fabricate and sabotage without affecting A's decision rights in any way. Table 3b shows S_1 in variant *payoff neutrality*: since A can no longer prefer either L over R, or vice versa, she has zero decision rights and B dictates the allocation also in S_1, without, however, resorting to fabrication, sabotage, or spying.

competitive payoffs

		A: L	A: R
B	L	$u_B^* = 0$, $u_A = 100$	$x_B^* = 100$, $x_A = 0$
	R	$v_B^* = 100$, $v_A = 0$	$y_B^* = 0$, $y_A = 100$

payoff neutrality

		A: L	A: R
B	L	$u_B^* = 0$, $u_A = 100$	$x_B^* = 0$, $x_A = 100$
	R	$v_B^* = 100$, $v_A = 0$	$y_B^* = 100$, $y_A = 0$

Table 3: Payoffs in S_1.

Note: Table 3a on the left reviews A's and B's payoffs in S_1 for treatment *competitive payoffs*, table 3b on the right reviews A's and B's payoffs in S_1 for treatment *payoff neutrality*. Thereby, u_B^* disregards the cost B has incurred from choosing Prob(S_2), that is, $u_B^* - 0.1 \cdot |50\% - \text{Prob}(S_2)| = u_B$ where u_B denotes B's actual payout.

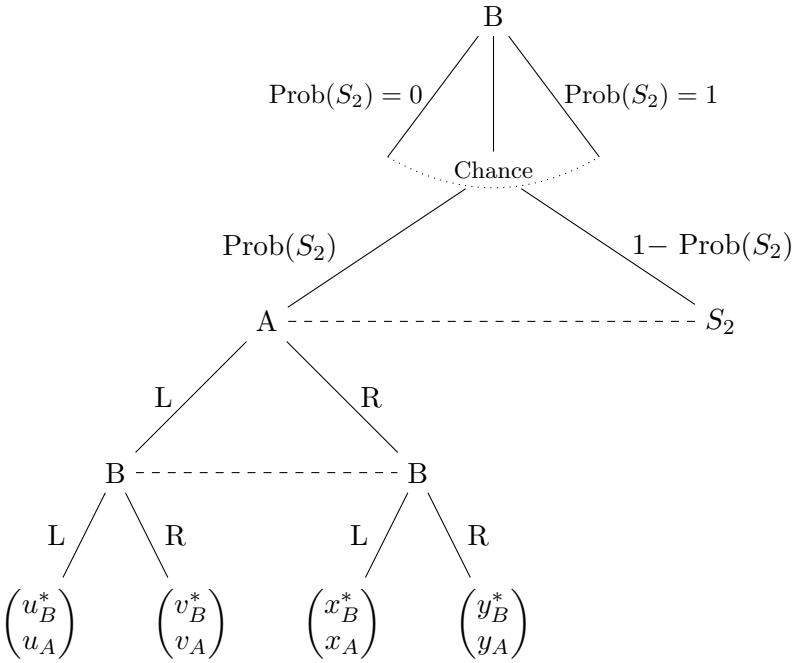


Figure 1: Basic Game Structure

Note: This tree illustrates our baseline treatments from table 2. S_2 is a place holder for Figure 2 in treatment SPY, for Figure 3 in treatment SABOTAGE and for Figure 4 in treatment LIE.

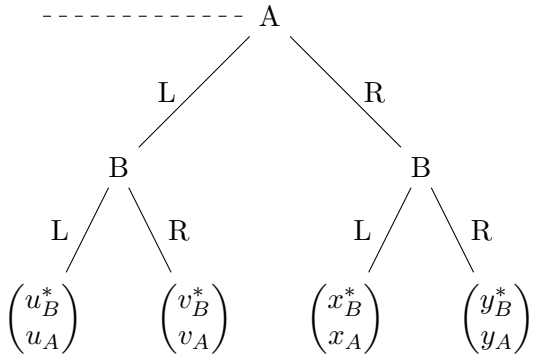


Figure 2: S_2 in treatment SPY.

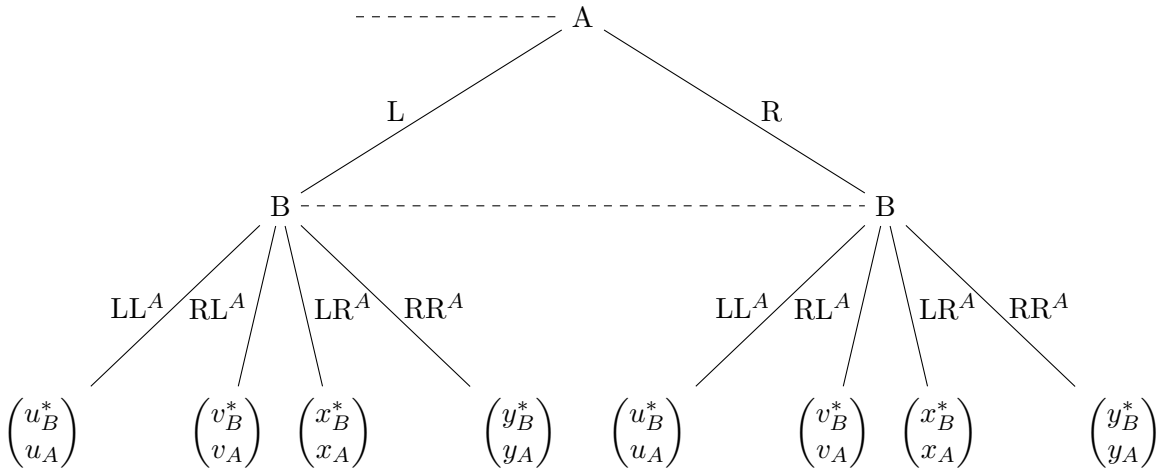


Figure 3: S_2 in treatment SABOTAGE.

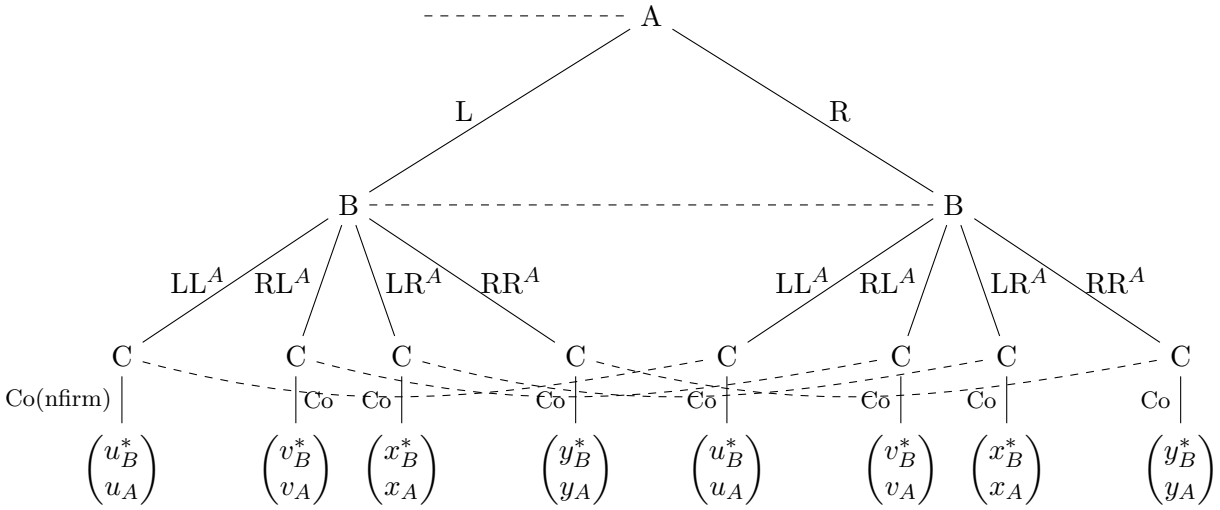


Figure 4: S_2 in treatment LIE.

If B's concern for A's decision rights makes her averse to fabrication and sabotage, this aversion will disappear in *payoff neutrality* and LIE, SPY, and SABOTAGE yield similar results. *Payoff neutrality* is worded identical to *competitive payoffs* such that any difference in wording between LIE, SPY, and SABOTAGE is preserved along with its potential effects. Appendix A shows screen shots for B's choice of Prob(S_2) in Fig. A1, for situation S_1 in Figure A2, situation S_2 SPY in Figure A3, S_2 SABOTAGE in Figure A4, and S_2 LIE in Figure A5. Note that throughout S_1 and S_2, B, in addition to her choices L (left) and R (right), is given the explicit option to toss a fair coin. This way, B can always equalize A's and B's chances of obtaining all payoff which, coincidentally, is also a feature of the equilibrium solution for S_1 as we discuss in theory section 6. Having already had the opportunity to randomize between S_1 and S_2, B participants, contrary to our concerns, never use this option in the experiment.

3.2 Part 2: Giving A a symbolic punishment or reward option

In part 2 of each session, A and B repeat part 1 with a new opponent, knowing that A can punish or reward B's choice of Prob(S_2). This affords A new decision rights which grant A some control over B's freedom of choice in that she can magnify or reduce the degree by which B prefers S_1 over S_2. Again, we expect B's concern about A's lack of decision rights to crowd out. In particular, A submits a punishment and reward schedule in which she may subtract up to 30 ECU, or may add up to 30 ECU to B's payoff, depending on whether B chooses S_1 (1) for sure, (2) with Prob(S_1) $\in [75\%, 99\%]$, (3) with Prob(S_1) $\in]50\%, 75\%[$, (4) with Prob(S_1) $= 50\%$, or chooses S_2 with (5) Prob(S_2) $\in]50\%, 75\%[$, with (6) Prob(S_2) $\in [75\%, 99\%]$, or (7) for sure. Any 1 ECU change to B's payoff costs A 1 ECU. B participants submit their beliefs about A's punishment and reward schedule. The correct guess of A's entire schedule earned B 35 ECU, the correct guess for any of the seven cases above, earned B 5 ECU. For each ECU by which B misguessed A's actual plan, B earned 0.08 ECU less. Figure A6 in appendix A shows the corresponding screen shot. Appendix D shows S_1 and S_2 from B's point of view: In S_1, B has suddenly lesser decision rights than A whereas in S_2, B still retains greater decision rights but cannot reduce A's to zero.

3.3 Part 3/ Controls and Instrumental Variable

Part 3 began by eliciting *envy* (Kirchsteiger 1994) to see how much B participants dislike being materially worse off than others. To this end, subjects were randomly rematched with a new opponent and submitted their choice between "10 ECU for themselves and 10 ECU for the other" or "10 ECU for themselves and 20 ECU for the other". A fair coin determined whether their own, or their opponent's decision would be payoff-relevant (Bartling et al. 2009). Part 3 also elicited *risk preferences* in a Holt-Laury price list format (Holt and Laury 2002) with subjects choosing ten times between a lottery and a sure payoff of 25 ECU. Each lottery paid either 10 or 35 ECU whereby lotteries systematically increased the chance of paying 10 ECU by 10%.

Next, an on-screen announcement pointed to a copy of Lind's (1978, 2008) standardized moral judgement test (M-J-T) placed upside down at the side of each desk. All information pertaining to the name or purpose of this test[11] had been removed. The test draws upon an inventory by Jean Piaget and Lawrence Kohlberg (Piaget 1948; Kohlberg 1969; Kohlberg 1984) who, in the 20th century, conducted extensive field research to observe and classify which criteria individuals use to make moral judgements. The test elicits Bs' preferences over these criteria: if, and by how much she uses a given criterion to judge whether a course of action is ethically right.

As by *Kohlberg class 1* and *2*, individuals deem those actions ethically right which are either not punished in material terms, or are rewarded instead. By *Kohlberg class 3*, individuals judge actions ethically right if the latter comply with a social norm, with others' expectations, were done with a good intention, or assist their social image with their peers. By *Kohlberg class 4*, individuals resort to the law, and to the idea of maintaing the status quo and the social order to judge whether an action is ethically right. By *Kohlberg class 5*, an action is deemed right if it respects parties' equality rights granted by a democratic social contract, and by *Kohlberg class 6*, if it satisfies some universal principle of conscience such as parties' human rights, parties' right to state their own will, or their human dignity. Chlaß et al. (2019) show in particular, that *purely procedural preferences* link to subjects' *Kohlberg class 5* scores and point out which demographic data might intercept this link.

The test introduces two vignettes, a first portraying workers who break into a factory in order to find and steal evidence that management was listening in on them, and a second, portraying a woman who is fatally ill and asks a doctor to medically assist her suicide. After each vignette, subjects are asked for their opinion whether or not the respective protagonists' behaviour was right or wrong. Next, the test lists 24 arguments (12 arguments after each vignette, six to judge the behaviour in question was wrong, each pertaining to one *Kohlberg class*; another six to judge it was right) and asks subjects how much they would agree or disagree on a nine-point Likert scale to judge the protagonists' course of action by each argument. In sum, we obtain four ratings per subject for each of the six *Kohlberg classes*, and a set of six preferences. Thereby, the test is constructed such that subjects who do not give their actual opinion in the test, answering, for instance, in what they deem a socially acceptable way, do not succeed in biasing the sample distribution of scores but add noise to the latter.

The experiment resumed with a payoff screen after which subjects submitted their age, gender, field of study, semester, and the type of degree they were studying for. Thereof, relevant controls for *Kohlberg class 5* scores are *field of study: Law*, and *gender*; relevant controls for *Kohlberg class 6* are *age*, *gender*, and *fields of study: Law*, *IT*, *Education*, and *Medicine*.

Finally, subjects filled in a questionnaire to elicit their *materialism* and *postmaterialism* values (Inglehart 1977; Klages and Gensicke 2006) where materalists appreciate

[11] Freely available for research purposes from Georg Lind's webpage at http://moralcompetence.net. Appendix H reproduces a standardized English version. See also:

power, order, obedience, and hierarchy, whereas postmaterialists value individualism, autonomy, and self-fulfillment. Some people may seek and condone power to put to rights what they see as ethically wrong, trading off monetary value against power, whereas others may deem that some individual rights are inalienable and must be reinstated; if such attitudes exist, they may explain why some subjects amend their opponents' rights whereas others seek additional rights to compensate the opponent materially. In (Chlaß et al. 2019), both behaviours were observed and linked to *Kohlberg class 5* and would, in our setup, imply postmaterialist B participants to opt into S_1, and materialist B participants to opt into S_2 and give all payoff away.[12]

3.4 Summary of Treatments

purely procedural aspects:	competitive payoffs[11]			payoff neutrality			competitive pun/rew[12]			payoff neutral pun/rew		
decision rights ↓	LIE	SPY	SAB	LIE	SPY	SAB	LIE	SPY	SAB	LIE	SPY	SAB
A has decision rights	+	+	+	-	-	-	+	+	+	+	+	+
B can take some of A′s decision rights	+	-	+	-	-	-	+	-	+	-	-	-
B can take all of A′s decision rights	+	-	+	-	-	-	-	-	-	-	-	-
wording of instructions is identical between treatments:	⟷						⟷					
		⟷						⟷				
			⟷						⟷			

[11] LIE + SAB competitive. S_1 – A and B have equal decision rights. S_2 – B has *greater* decision rights.
[12] LIE + SAB competitive pun/rew: S_1 – B has *lesser* decision rights than A. S_2 – B has *greater* decision rights than A.

Hypothesis 1 – B's concern for A's decision rights causes high levels of altruism. In LIE and SABOTAGE with competitive payoffs, many Bs therefore depart from rational self-interest, but not in SPY where B exerts no influence over A's decision rights.

Hypothesis 2 – These results by the Rubin causal model are confirmed by an instrumental variable: B's altruism links to B's Kohlberg class 5 scores *after controlling for latent correlates of the latter which might intercept the link.*

Hypothesis 3 – As B's influence over A's decision rights declines, so does her altruism, dropping significantly in LIE/SABOTAGE payoff neutrality, and in LIE/SABOTAGE with punishment/reward. Residual altruism does not link to B's Kohlberg class 5 scores.

[12] We elicit these value groups by the 'Speyerer value inventory' (Klages and Gensicke 2006) which consists of 12 items to be rated on a seven point Likert scale (1 – not important at all, to 7 – very important). Three items load on a first scale *'duty and acceptance values'*, four on a second *'hedonistic and materialist values'*, and three on a third, *'idealistic values and political participation'*. Typically, five value groups (clusters) emerge; amongst them *'conventionalists'* – Inglehart's original materialists, and so-called *'idealists'* – Inglehart's original postmaterialists. We use individuals' absolute ratings of all three scales for our analysis. Klages and Gensicke's measurement instrument has three main advantages over Inglehart's in our setup: first, the items being directly validated on German samples, second, the use of separate scales for materialism and postmaterialism values (Inglehart obtains these as opposite ends of the same scale; they are therefore by construction consistent and cannot be used to check the other) and third, the possibility of hybrid value groups which, in Inglehart's measurement, need to be post-assigned to the only two value groups allowed. For details, see appendix I; a concise review of Klages' research in English is found in (Borg et al. 2019) who also show that Klages' three scales emerge as the first three principal components of the popular Schwartz' portrait value scales.

4 Results

4.1 Descriptives: B's choice of situation and allocation

4A)[13] LIE, SPY, SABOTAGE – competitive payoffs

treatment (obs.) →	LIE (44)		SPY (53)		SAB (54)	
situation →	S_1	S_2	S_1	S_2	S_1	S_2
Bs who pay for S_1 or S_2, respectively...	9 20%	5 11%	5 9%	36 68%	2 4%	37 69%
...set Prob(S_1) or Prob(S_2) to median	0.6	0.7	1	0.8	0.7	0.8
Bs in each situation	19	25	13	40	26	28
Bs who give A all payoff in S_2		17 68%		0 0%		20 71%

4B) LIE, SPY, SABOTAGE – payoff neutrality

treatment →	LIE (47)		SPY (53)		SAB (52)	
situation →	S_1	S_2	S_1	S_2	S_1	S_2
Bs who pay for S_1 or S_2, respectively...	8 17%	3 6%	2 4%	19 36%	4 8%	18 35%
...set Prob(S_1) or Prob(S_2) to median	0.6	0.8	0.8	0.7	0.7	0.7
Bs in each situation	25	22	20	33	22	30
Bs who give A all payoff in S_1 or S_2	5 20%	5 23%	2 10%	2 6%	4 18%	4 13%

4C) LIE, SPY, SABOTAGE – competitive payoffs punishment/reward

treatment →	LIE (44)		SPY (53)		SAB (54)	
situation →	S_1	S_2	S_1	S_2	S_1	S_2
Bs who pay for S_1 or S_2, respectively...	7 16%	13 30%	2 4%	37 70%	4 7%	36 67%
...set Prob(S_1) or Prob(S_2) to median	0.6	0.7	0.7	0.9	0.6	0.7
Bs in each situation	20	24	16	37	14	40
Bs who give A all payoff in S_2		12 50%		0 0%		19 48%

4D) LIE, SPY, SABOTAGE – payoff neutrality punishment/reward

treatment →	LIE (47)		SPY (53)		SAB (52)	
situation →	S_1	S_2	S_1	S_2	S_1	S_2
Bs who pay for S_1 or S_2, respectively...	6 13%	6 13%	4 8%	15 28%	10 19%	11 21%
...set Prob(S_1) or Prob(S_2) to median	0.7	0.6	0.6	0.7	0.7	0.8
Bs in each situation	30	17	25	28	29	23
Bs who give A all payoff in S_1 or S_2	5 17%	1 6%	1 4%	4 14%	7 24%	0 0%

Table 4: B's choice of situation and her choice of situation by allocation

Tables 4 list, how many B participants pay for situation S_1, how many for S_2, which probability they set for their preferred situation, and which allocation B participants impose if given the opportunity. Table 4A summarizes our baseline treatments with *competitive payoffs*. In LIE, 20% (9 of 44) B participants pay for S_1, compared with 9% (5 of 53) in SPY and 4% (2 of 54) in SABOTAGE. 11% (5 of 44) B participants pay for S_2, compared with seven times as many, i.e. 68% (36 of 53), in SPY and 69% (37 of 54) in SABOTAGE. In sum, significantly *fewer* B participants fabricate than spy or sabotage by Fisher's exact tests, all *p-values* < 0.02. Turning to altruism, 68% (17 of 25) B participants in S_2 give all payoff to A in LIE, *none* of the 40 B participants in S_2 does so in SPY, and 71% (20 of 28) do so in SABOTAGE. Most altruism – most departures from rational self-interest – does therefore occur, when rational self interest requires B to impair A's decision rights.

Result 1. In LIE and SABOTAGE with competitive payoffs, significantly more B participants give all payoff to A than in SPY with competitive payoffs where B's only source of power is her advantage in information (Fisher's Exact tests, p-value < 0.01).

[13] Reading example: In treatment LIE, there are 44 B participants, 5 of which (11%) pay for S_2 and set Prob(S_2) to median 0.7. 25 B participants arrive in S_2, 17 of which (68%) give give all payoff to A.

Table 4B summarizes LIE, SPY, and SABOTAGE under *payoff neutrality*. Roughly as many B participants as before pay for S_1, but only half as many for S_2. 21% (S_1: 5 + S_2: 5 = 10 of 47) give all payoff to A in LIE and 15% (S_1: 4 + S_2: 4 = 8 of 52) do so in SABOTAGE which are significantly fewer than before by Fisher's exact tests, all *p-value* < 0.001. Treatment SPY remains unchanged by Fisher's exact test, *p-value* $= 0.136$ with 8% (S_1: 2 + S_2: 2 = 4 out of 53) giving all payoff to A. Again, altruism decreases where self-interest does not impair A's decision rights.

Turning to tables 4C and D, symbolic punishment and reward sustains a considerable level of altruism, maintained, however, by a largely different set of individuals. Roughly 40% of Bs opt for a different situation in LIE and SABOTAGE, some 30% do so under *payoff neutrality*. Altruism among altruists from part 1 drops by one third in LIE, by two thirds in SABOTAGE, and more strongly so under *payoff neutrality*, i.e. by 80% in LIE and 88% in SABOTAGE. Throughout, behaviour in SPY is least affected. Contingency tables in appendix F report the exact absolute and relative numbers. Symbolic punishment and reward might therefore indeed crowd out B's concern for A's decision rights, if, in addition, a new ethical criterion – preferably referring to *reward and punishment* – were at play.[14] Figure 6 illustrates Bs' choice of the situation as by the allocation they impose, for all treatments.

Result 3. As B's influence over A's decision rights decreases, so does her altruism: in LIE/SABOTAGE with payoff neutrality and LIE/SABOTAGE with punishment/reward.

4.2 Descriptives: Bs' beliefs

In this section, we look at whether Bs do indeed believe that A deems S_2 undesirable. Figure 6A illustrates that Bs believe to be rewarded for opting into S_1, and less so as this choice tends toward the toss of a fair coin. At this point, they expect neither reward nor punishment. Bs believe As to punish S_2, and increasingly so as S_2 becomes certain. Expected average punishment is 9.77 ECU for Prob(S_2) = 100%, 7.60 ECU for Prob(S_2) $\in [75\%, 99\%]$, and 5.27 ECU for Prob(S_2) $\in]50\%, 75\%[$, each category significantly greater than the next.[15] Bs therefore believe that S_2 – fabrication, spying, and sabotage – is undesirable in As' eyes. Appendix O shows that this pattern is strongest in SPY where Bs opt most frequently into S_2, less strong in SABOTAGE, and least so in LIE where Bs hardly opt into S_2. Bs' choice of Prob(S_2) and their beliefs about what As wish them to do therefore seem to vary at odds with each other across LIE, SPY, and SABOTAGE. This is apparent from individuals' beliefs pertaining to

[14] If no ethical criterion at all were at play, Bs might simply have adopted new behaviours to keep the task interesting. If previously selfish Bs felt guilt, *others' expectations*, i.e. *Kohlberg class 3*, would explain Bs' choices. In section 5, we show that Bs' choices link to *Kohlberg class 1* which derives the right course of action from material punishment and reward.

[15] Bs expect punishment to be highest for Prob(S_2)=100 (SPY: Wilcoxon Signed Rank *p-value* < 0.001, SABOTAGE: *p-value* < 0.032, LIE: *p-value* < 0.043), second highest for Prob(S_2) $\in [75\%, 99\%[$ (SPY: *p-value* < 0.001, SABOTAGE: *p-value* < 0.005, LIE: *p-value* $= 0.23$), third highest for Prob(S_2) $\in [50\%, 75\%[$, and least for Prob(S_2) = 50% (SPY: *p-value* < 0.001, SABOTAGE: *p-value* < 0.001, LIE: *p-value* < 0.09). In LIE where the order is least pronounced, average aggregate punishment for S_2 is weakly significantly larger than for the toss of a fair coin, *p-value* < 0.059.

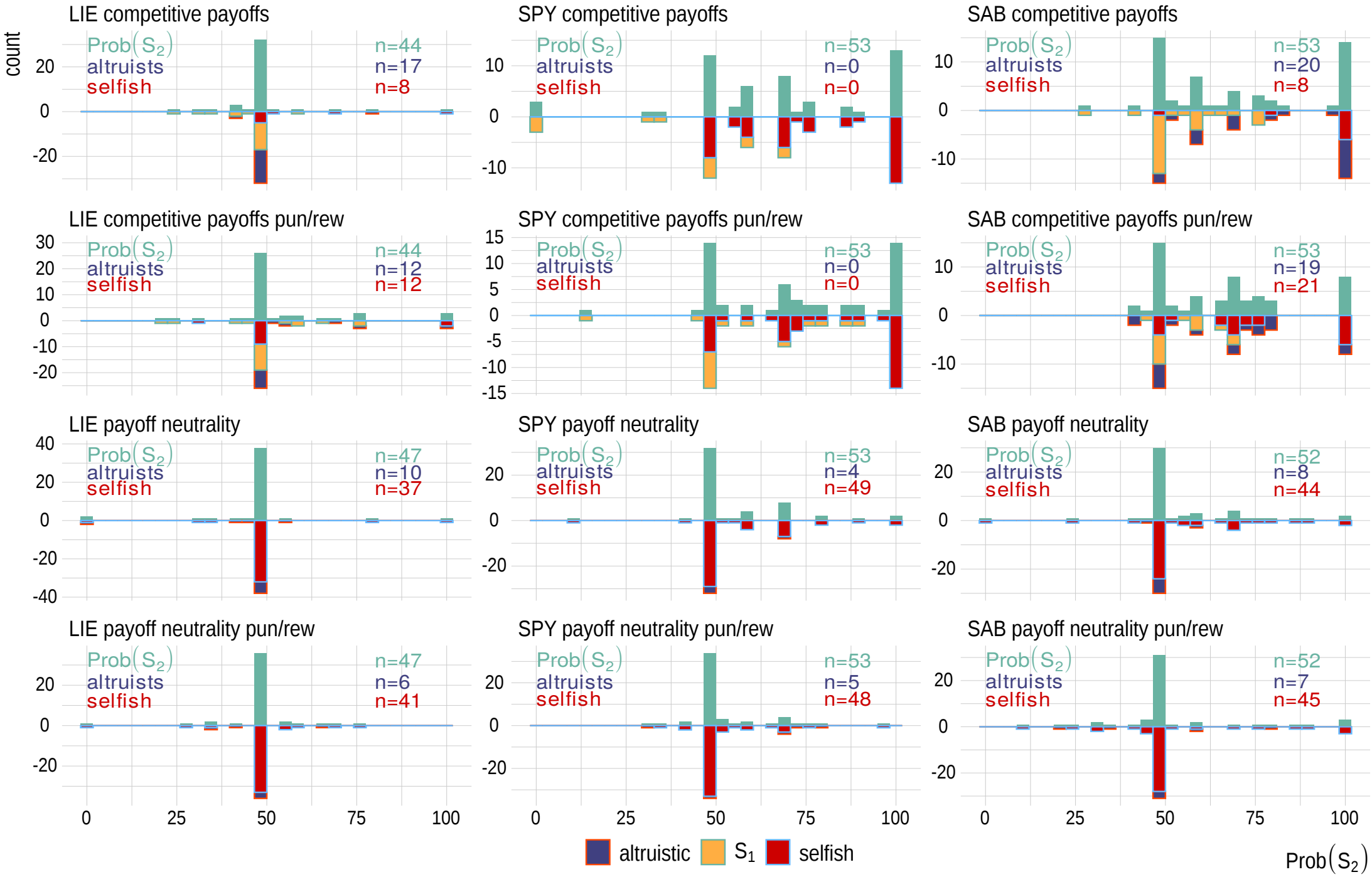
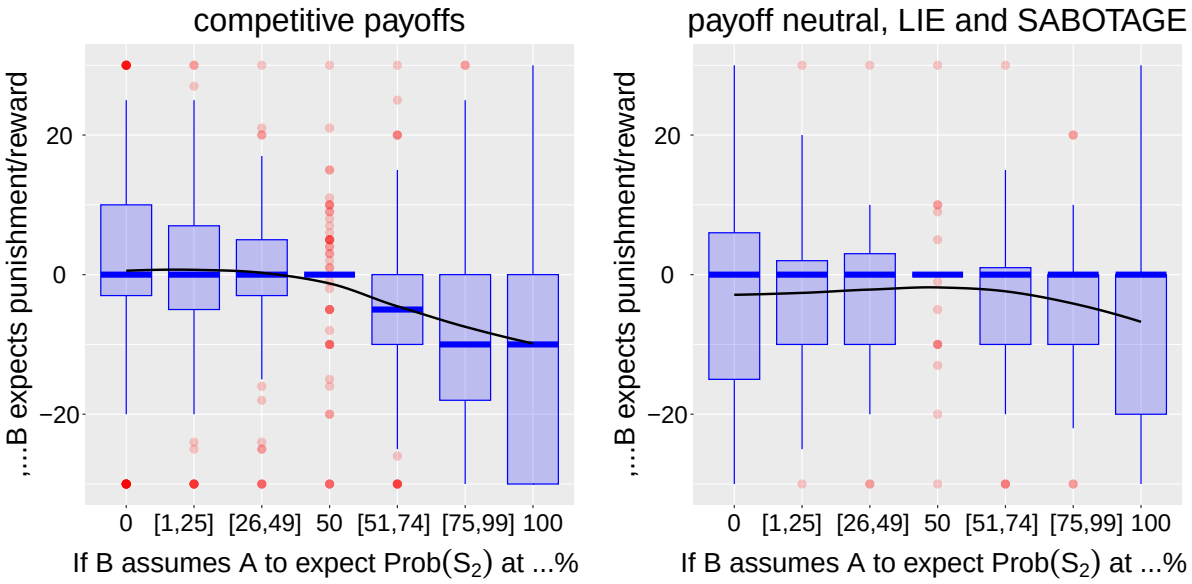


Figure 5: B's choice of situation, and her choice of situation by allocation imposed.

their own actual choice of Prob(S_2), as well as from their beliefs about the entire (hypothetical) choice set.[16] Arguably, punishment beliefs also provide access, however imperfect, to social norms which might regulate lying, spying, and sabotage differently. If social norms guide Bs' beliefs about As' punishment, the social norm against spying turns out strongest, followed by the norm against sabotage, and then lying. Beliefs and choices are logically linked. Bs who opt into S_2 and take all payoff, make S_2 as likely as possible while keeping punishment at a reasonable level. Bs who opt into S_1 or toss a fair coin, expect As to punish S_2 significantly more than their actual choice. Bs who give all payoff to A show no such belief patterns.[17] Thereby, beliefs and moral judgement are not linked, and seem to describe what Bs believe As *actually* do, rather than *should* do. Figure 6B shows that when A has zero decision rights always, Bs expect to be punished for every intentional choice, increasing in its intentionality. In these cases, exerting one's rights to choose the procedure when one dictates the allocation always, increases the asymmetry in decision rights even further. Bs expect to be punished for Prob(S_2) = 0 (p-$value < 0.034$), Prob(S_2) $\in]0\%, 25\%]$ (p-$value < 0.02$), Prob(S_2) $\in]25\%, 50\%[$ (p-$value < 0.008$), not for the toss of a fair coin (p-$value = 0.350$), and again for Prob(S_2) $\in]50, 75[$ (p-$value < 0.007$), Prob(S_2) $\in$ $[75\%, 99\%[$ (p-$value < 0.002$) and Prob(S_2) = 100% (p-$value < 0.001$). In SPY payoff neutrality, the original punishment belief pattern remains intact.

Figure 6: B's beliefs about A's decision to punish or reward B's choice of $Prob(S_2)$. Left: competitive payoffs; Right: payoff neutrality.



[16] Bs expect more punishment for Prob(S_2) > 50% in SPY than in LIE (Wilcoxon Rank Sum, p-$value < 0.001$) or SABOTAGE (p-$value < 0.034$). In LIE where Bs rarely opt into S_2, Bs expect *less* punishment for S_2 than in SABOTAGE (p-$value < 0.036$) where 69% opt into S_2. Similarly, Bs expect more severe punishment for their actual choice in SPY than in LIE (p-$value < 0.016$) or SABOTAGE (p-$value < 0.026$).

[17] In SABOTAGE, Bs who opt into S_1 set Prob(S_2) to an average 47.5% (SPY: 47.5%), expecting greater punishment for $Prob(S_2) = 100$ (Wilcoxon Signed Rank tests, p-$value$= 0.002, SPY: p-$value$= 0.002), for Prob(S_2) $\in [75\%, 99\%]$ (p-$value$< 0.001, SPY: p-$value$= 0.006), and for Prob(S_2) $\in]50\%, 75\%[$ (p-$value$< 0.001, SPY: p-$value$= 0.06) than for their own actual choice. Bs who take all payoff, set Prob(S_2) to an average 79.94% (SPY: 85.77%), expecting greater punishment for Prob(S_2) = 100 (p-$value < 0.021$, SPY: p-$value$= 0.001) and for Prob(S_2) $\in]75\%, 99\%]$ (p-$value < 0.3114$, SPY: p-$value < 0.011$) than for their actual choice, but lesser punishment for Prob(S_2) $\in]50\%, 75\%]$ (p-$value$< 0.0625, SPY: p-$value$< 0.177) and all categories Prob(S_2) < 50% (p-$values$< 0.01). Bs who give all payoff to A, set Prob(S_2) > 50% to an average Prob(S_2) = 76.08% and do not expect greater or lesser punishment for other choices.

5 Ethical criteria at play

Next, we study which ethical criteria – if any – underlie B's decision not to opt into S_2 and secure all payoff. B might, for instance, avoid the option out of concern for her social image, in order not to disappoint A's expectations (Battigalli and Dufwenberg 2007), not to violate some, or several, social norms[18], or in order to signal her own generous intentions (Falk and Fischbacher 2006).[19] In the previous section, we saw that Bs behave particularly selfish where they expect A to punish this selfishness most: a desire to avoid letting A down or to comply with a social norm would imply a different pattern of altruism across LIE, SPY, and SABOTAGE. Finally, B might deem that fabrication, sabotage and spying violate the opponent's civil rights granted by the social contract (Chlaß et al. 2019), or that stripping the opponent of any freedom to choose violates her human rights and dignity (Chlaß and Moffatt 2012).

If indeed, Bs' altruism arose from a concern purely about As' decision rights, Bs' decision to opt into S_2 and give all payoff to A must link to *Kohlberg class 5* which regroups criteria around the equality of rights as stipulated by a democratic social contract. Chlaß et al. (2019) identify a link between the latter and individuals' willingness to pay for changes in the information and decision structure of a formally defined game when these changes are either of no, or against individuals' material self-interest. The link at hand was intercepted by two demographic variables, i.e. *field of study: Law*, and *gender*. Any link between Bs' altruism and *Kohlberg class 5* must therefore be robust to including these as well as the complete set of six Kohlbergian classes.[20]

In a series of Logit models, we contrast each variant of altruism: I) paying for S_1, II) paying for S_2 and giving away all payoff, and III) tossing a fair coin, against IV) opting into S_2 and taking all payoff. To account for the entirety of the data set, we assign altruists who arrive in S_2 by dint of a fair coin or by paying for S_1, to II. B participants who pay for S_2 and end up in S_1 are also assigned to this group such that they may operate most effectively *against* a potential effect of *Kohlberg class 5*.[21] Appendix J shows the actual count of Bs' behaviours per treatment. We regress the resulting pairs of behaviour on B's average rankings over all six Kohlbergian classes[22], and a treatment Dummy. To avoid omitted variable bias and, at the same time, preserve the estimator's efficiency, models are tested downward, removing insignificant variables which do not affect the goodness-of-fit.

[18] More precisely, if B were guilt averse, she would prefer to avoid feeling guilt. She would feel guilty, if she opted into S_2 and took all payoff while expecting A to expect her not to do so (Battigalli and Dufwenberg 2007) or knowing that a social norm (Miettinen 2013) bans the actions in question.

[19] Note that if B simply tried to ex-ante allocate outcomes in a fair (Bolton and Ockenfels 2000; Bolton et al. 2005) or kind (Dufwenberg and Kirchsteiger 2004; Sebald 2010) way, she would also seek to comply with a social norm (for payoff equality), or seek to signal her intentions.

[20] The current paper uses the same subject pool on which the instrument was tested; recruiting measures and success, the size of the subject pool, and the influx of students to the university of Jena did not change during the time which elapsed in between. Indeed, the distribution of scores in both papers is similar.

[21] Suppose these B participants opted for S_2 to take all payoff. In this case, their *Kohlberg class 5 scores* – if the latter does explain altruism – would be smaller than those of the actually observed altruists who form this group, weakening the effect. If they intended to give all payoff, the effect simply remains intact.

[22] B's average *Kohlberg class 1 (2,3,4,5,6)* ranking is the average over her (four) ratings of the (four) arguments pertaining to *Kohlberg class 1 (2,3,4,5,6)* in the moral judgement test, divided by the difference between the largest, and the smallest rating B ever ticks in the entire test, accounting for B's personal use of the Likert scale. Average ratings are standardized by subtracting their sample mean and dividing by their standard deviation. All moral judgement variables are computed analogously to (Chlaß et al. 2019).

DEPENDENT VARIABLE: VARIANT OF ALTRUISM VS. RATIONAL SELF-INTEREST			
	S_1 (1) VS. SELFISH (0)	S_2 + GIVE ALL (1) VS. SELFISH (0)	FAIR COIN (1) VS. SELFISH (0)
nr. of obs.	19 (10 vs. 9)	83 (73 vs. 10)	16 (6 vs. 10)
Kohlberg class 1	-0.051 (0.088)	-0.140^a (0.055)	-0.315^b (0.142)
Kohlberg class 3	-0.126 (0.093)	0.050 (0.042)	-0.063^c (0.034)
Kohlberg class 5	0.451^a (0.087)	0.081^b (0.032)	0.567^a (0.147)
Kohlberg class 6	-0.253^a (0.086)	0.027 (0.038)	-0.068 (0.097)
Dummy lie		0.086^b (0.039)	0.123 (0.136)
postmaterialism	0.160^c (0.094)		
materialism		-0.054^b (0.021)	-0.062^b (0.028)
Count R^2	0.90	0.90	0.88

Table 5: Ethical determinants of B participants' departures from rational self-interest, and the type of altruistic behavior they adopt (marginal effects).

Note: Significance levels of z-tests are indicated by $a : p < .01$, $b : p < .05$, $c : p < .10$

In order to clearly see whereto likelihood is shifted away from rational self-interest, we specify independent binary Logits with robust errors and return a trifle too conservative *p-values* (Agresti 2002). Table 5 shows our results. Estimated Logits yield a Count R^2 beyond 88%. Results are robust to the inclusion of demographics and to an increase in sample size as shown in appendix L.

→ Kohlberg class 1. The more strongly B deems that an action which is not punished, cannot be wrong, the more likely she opts into S_2 and takes all payoff. Per one-unit increase in the strength of this conviction, she is 14%, *p-value* = 0.01, less likely to give all payoff to A and 31.5%, *p-value* = 0.026, less likely to toss a fair coin.

→ Kohlberg classes 2,3, and 4. *Kohlberg classes 2* and *4* are not significant in any binary comparison, neither on the reduced, nor on the full model – see appendix L – and, for the sake of efficiency and fit, left out from table 5. Note that if our results were caused by this omission, both variables would either separately, or jointly have needed to turn out significant themselves. Ethical criteria of *Kohlberg class 3* do not seem to increase the likelihood of B's altruism either. That is, the extent to which B refers to her social image, others' expectations, social norms, or intentions to derive the right course of action does not make her less inclined to behave selfishly.

→ Kohlberg class 5. The more strongly B resorts to the social contract and the civil rights granted therein to derive the right course of action, the more likely she opts into S_1, i.e. 45.1%, *p-value* = 0.000, the more likely she opts into S_2 and gives all payoff to A, i.e. 8.1%, *p-value*[23] = 0.011, and the more likely she tosses a fair coin, i.e. 56.7%, *p-value* = 0.000, rather than being selfish.

[23] In the fully reduced model with variables significant at the 10% level only, the effect becomes 10.7%, *p-value* = 0.005; the effect also reaches a 1% significance level on the full model in appendix L.

→ Materialism and postmaterialism which capture B's attitudes toward authority and autonomy, significantly influence her choice of S_2. The more B values power ('materialism score'), the more likely she seeks S_2 and exerts her allocation power in S_2 to take all payoff rather than giving all payoff to A (effect: 5.4%, *p-value*=0.011), or tossing a fair coin (6.2%, *p-value* = 0.028). The more she values autonomy ('postmaterialism'), the more likely she opts into S_1[24] and reinstates A's decision rights.

V) Sample size, false positives & omitted variable bias. Appendix L shows that the results from this section hold if we augment table 5 by treatment SPY, and add critical demographic data which might intercept the link between *Kohlberg class 5* and Chlaß et al.'s (2019) *purely procedural preferences*. Appendix D analyzes how a punishment and reward option accorded to A improves her decision rights, a situation where B's ethical concern about A's lack of decision rights should crowd out. Table 15 in appendix M confirms that B no longer refers to *Kohlberg class 5*. As the distribution of rights evens out, B's motive to evade punishment, fueled by *Kohlberg class 1*, becomes predominant instead.

Result V: As long as rational self-interest implies that B must strip A of all decision rights, Kohlberg class 5 explains all forms of altruism.

6 Underlying Preferences & Discussion

In this section, we discuss theoretically which preferences can explain the variation of B participants' behaviour across treatments, and whether or not our empirical results are consistent with each preference type. We restrict our attention to the competitive payoff setting where B must lie, spy, or sabotage to secure all payoff for sure.

Self-interested opportunism. If B only cares about her own material payoff, she spies, lies, or sabotages for sure to take all payoff. She pays 5 ECU to set $\text{Prob}(S_2) = \alpha = 1$ and in S_2, opts for strategy combination $\{B : RL^A, A : \{\cdot\}\}$, or $\{B : LR^A, A : \{\cdot\}\}$ to achieve allocation (B: 100, A: 0). Altogether, B receives $100 - 5 = 95$ ECU, and A receives 0 ECU in treatments LIE, SPY, and SABOTAGE[25]. Self-interested opportunism can therefore neither explain the variation in B participants' procedural choices across treatments LIE, SPY, and SABOTAGE, nor the empirical link between B's behaviour and her ethical preferences over *Kohlberg class 5* documented in section 5 and appendix L.

Altruism. If B cares more about A's material payoff than about her own – in Charness

[24] Removing insignificant variables *Kohlberg class 1* and *3*, the effect turns significant at the 1% level: if *postmaterialism* increases by one-unit, B is 20.8%, *p-value* = 0.001 more likely to opt into S_1. Similarly, removing *Kohlberg class 6* from specification 3, *materialism* turns significant at the 1% level, i.e. −7.1%, *p-value* = 0.010. Reductions of both models increase their goodness-of-fit; *Kohlberg class 3* never turns significant.

[25] 95 ECU is the maximal payout as can be seen from comparing the following cases: If B opts into S_1 for sure, she pays 5 ECU to set $\alpha = 0$ and receives an expected equilibrium payout of 50 ECU in S_1, overall $50 - 5 = 45$ ECU. If B leaves the default $\alpha = 0.5$, she receives an equilibrium payout of 50 ECU from S_1 which occurs with 50% probability, and a payoff of 100 ECU from S_2 which also occurs with 50% probability. Hence, her overall expected payoff from not influencing the set of rules is $0.5 \cdot 50$ ECU $+0.5 \cdot 100 = 75$ ECU. Making S_2 one per cent more likely costs 0.1 ECU, but yields an expected payoff increase of $0.01 \cdot (95 - 75) = 0.2$ ECU. Hence, the 95 ECU which B earns from making S_2 sure are her maximal payoff.

and Rabin's (2002) notation, for instance, B weights A's payoff by σ and her own payoff by $1-\rho$ where $\sigma > 1-\rho$, – she prefers allocation (B: 0, A: 100) to (B: 100, A: 0). To achieve this allocation, she pays 5 ECU for setting Prob $(S_2) = \alpha = 1$ and arrives for sure in S_2 where she imposes allocation (B: 0, A: 100) either via strategy combination $\{B : LL^A, A : \{\cdot\}\}$ or $\{B : RR^A, A : \{\cdot\}\}$. B receives -5 ECU and A 100 ECU in LIE, SPY, and SABOTAGE. Altruistic preferences therefore do not explain the variation in B participants' procedural and allocation choices across treatments LIE, SPY, and SABOTAGE, or the empirical link between B's behaviour and her ethical preferences over *Kohlberg class 5*.

Preferences for equal expected payoffs. B may be willing to forego some of her payoff in order to ex-ante grant A more equal chances on the one ex-post nonzero payoff. Formally, if B is inequity-averse over expected payoffs (Bolton et al. 2005), she has utility $u_B = a_B \cdot E(y_B) - 0.5b_B \left(E(y_B) \cdot 100^{-1} - 0.5\right)^2$. Thereby, y_B denotes her own expected payoff, $a_B \geq 0$ her aversion against disadvantageous inequality, and $b_B \geq 0$ her aversion against advantageous inequality, both forms of aversion being driven by a social norm of payoff equality. In S_1, two perfectly selfish players would each choose to toss the fair coin between L and R which, coincidentally, also guarantees ex-ante equality in payoffs. B's corresponding utility is $a_B \cdot 50$ with no disutility from advantageous inequality. In S_2, B can implement any distribution of chances she prefers with the explicit option of tossing a fair coin. If B has a_B, b_B such that she cannot reach her preferred distribution of chances in S_1, she prefers S_2. This decision is identical in LIE, SPY and SABOTAGE. Preferences for equal expected payoffs do therefore not explain the variation in B participants' procedural and allocation choices across said treatments. Similarly, we could not confirm that B participants predominantly resort to social norms, a criterion located in *Kohlberg class 3*.

Preferences for kind procedures (Sebald 2010). A and B may care for the *kindness* of a procedural choice (whereby the kindness of a person who chooses a procedure is equal to the kindness of the distribution of outcomes which this procedure is expected to induce) and, upon observing a kind (unkind) procedural choice, be kind (unkind) in return. In our setting, it is commonly known that A never observes B's procedural choice. However, A may hold expectations about B's procedural choice, and B may expect A to have such expectations. *a) suppose B expects A to expect S_2.* In this case, A expects to have no opportunity to reciprocate and she is always neutral toward B. This implies that B's payoff from reciprocity is zero and her preferences in S_2 coincide with self-interest: B chooses either $\{B : RL^A, A : \{\cdot\}\}$, or $\{B : LR^A, A : \{\cdot\}\}$ which earn her $100 - 5 = 95$ ECU. *b) suppose instead that B expects A to expect S_1.* When B is called upon to choose in S_1, she only considers her efficient strategies: yet, all are efficient since neither L nor R destroy the pie. If B believes A plays L with probability q_L and R with $1-q_L$, B's kindness in choosing L equals $q_L \cdot 100 + (1-q_L) \cdot 0 - (q_L \cdot 100 + (1-q_L) \cdot 0 + q_L \cdot 0 + (1-q_L) \cdot 100)/2$,[26] and her kindness in choosing R equals $q_L \cdot 0 + (1-q_L) \cdot 100 - (q_L \cdot 100 + (1-q_L) \cdot 0 + q_L \cdot 0 + (1-q_L) \cdot 100)/2$. If B believes that A tosses the fair coin, i.e. $q_L = 0.5$ which is the only equilibrium in S_1, then B's choice of L and R is exactly neutral toward A. Since B is not unkind in equilibrium,

[26] $q_L \cdot 100 + (1 - q_L) \cdot 0$ is A's payoff from B choosing L when B believes A plays L with probability q_L. This payoff is compared to the average payoff for A over all pure strategies which are still available to B at a given node: since B can still choose between L and R, this average payoff for A over B's pure strategies L and R is: $(q_L \cdot 100 + (1 - q_L) \cdot 0 + q_L \cdot 0 + (1 - q_L) \cdot 100)/2$. A payoff for A equal to this average payoff is neutral, payoffs for A greater than this average are kind (Dufwenberg and Kirchsteiger 2004).

A need not reciprocate, and the payoffs from reciprocity in S_1 are zero. Hence, A and B implement the selfish solution and each tosses a fair coin which yields both players 50 ECU. Therefore, B participants who prefer kind over unkind procedures opt into S_2 which earns them $100 - 5$ ECU. Even if B held off-equilibrium beliefs in S_1, any reciprocation she expects in S_1 would be identical across SPY, LIE, and SABOTAGE. No variation in B's procedural or allocation choice should occur. In terms of ethical criteria, A and B assess their own and each others' choices in terms of *intentions*, and the degree to which the intended outcomes comply with a social norm of payoff equality. We could not confirm that B participants strongly invoke social norms or intentions which are both located in *Kohlberg class 3*.

Guilt aversion. If B is guilt-averse, she seeks to avoid disappointing A's payoff expectation and seeks to avoid being blamed by A for doing so (Battigalli and Dufwenberg 2007). In part two – see section 3.2 – B submits her expectations about A's symbolic punishment and reward plan[27], a plan which lists by how much A increases or decreases B's payoff for any given choice of procedure Prob (S_2). This plan fuses information about how much A disapproves of a given procedural choice along with the allocation A expects this choice to entail. B participants expect more symbolic punishment for choosing S_2 in SPY than in LIE (one-sided Wilcoxon Rank Sum tests, *p-value* < 0.01 for $\alpha \in]0.5, 0.75[$, for $\alpha \in]0.75, 0.99[$, and for $\alpha = 1$), expect similar punishment for S_2 across LIE and SABOTAGE and also across SPY and SABOTAGE. Since B participants choose S_2 often in SPY and rarely in LIE, their procedural choices run contrary to their beliefs about what A approves them to do. Similarly, guilt aversion does not explain why, given that Bs expect the same punishment for S_2 in SPY and SABOTAGE, we observe substantial altruism in SABOTAGE, but none in SPY. In terms of ethical criteria, we could not confirm that B participants predominantly resort to others' expectations which are located in *Kohlberg class 3*.

Purely Procedural Preferences. B participants may have ethical reservations against being favored by the rules of the game, noteably in terms of decision or information rights (Chlaß et al. 2019). Suppose B's utility function includes some element similar to: $-\beta_B \max\{\#S_B - \#S_A, 0\} - \alpha_B \max\{\#S_A - \#S_B, 0\}$ where $\#S_B - \#S_A$ and $\#S_A - \#S_B$ count the difference between A's and B's number of effective pure strategies: strategies which induce genuinely different outcomes and therefore add to their freedom of choice – see section 2; where β_B denotes B's dislike of having greater, and α_B her dislike of having lesser rights. For LIE and SABOTAGE, B has two such pure strategies in S_1, and two in S_2 whereas A has two in S_1 but none in S_2. In S_1, therefore, B has no disutility from the rules of the game themselves whereas in S_2, her disutility is $\beta_B \cdot 2$. If this disutility is larger than the utility from her payoff advantage in S_2, then $S_1 \underset{B}{\succ} S_2$. In SPY, on the other hand, A and B always have equal decision rights: two effective pure strategies in S_1, and two in S_2. B's allocation power in S_2 arises from an advantage

[27] A's expectations about B's choice of the interaction structure, and B's choice of the allocation may differ across LYING, SPYING, and SABOTAGING, for instance, because there are different social norms regarding lying, spying, or sabotaging which may in turn imply that the shares of individuals in the population who lie, spy, and sabotage differ, or because individuals also hold expectations whether or not others lie, spy, or sabotage, and expect others to have such expectations, too.

in information.[28] Summing up, LIE, SABOTAGE and SPY put different rights at stake and also differ in how B brings her advantage in S_2 about: in LIE and SABOTAGE, B takes decision rights *away from* A whereas in SPY, she assigns *herself* more information rights. We conclude that, if B has preferences over A's decision rights, she may prefer S_1 over S_2 in LIE and SABOTAGE, but not in SPY. In terms of ethical criteria, we can statistically confirm that B participants predominantly resort to the notion of civic rights as granted by a democratic social contract located in *Kohlberg class 5*, the very criterion underlying Chlaß et al. (2019)'s *purely procedural preferences*, after controlling for all known potential confounds for this link. Looking at B's giving all payoff to A in S_2, note that already Chlaß et al. (2019) find individuals who value decision rights, and yet reduce the opponent's rights while paying that opponent off, thus trading off monetary payoff and rights. B participants who opt into S_2, give all payoff to A and are motivated by *Kohlberg class 5* belong to this group. If B cares for A's decision rights, we therefore expect altruism in LIE and SABOTAGE but not in SPY.

Preferences for power & control. If B prefers to maintain power and control (Bartling et al. 2014), she opts for interaction structure S_2, thereby avoids any interference from A and implements whatever allocation she prefers. Preferences for power and control therefore do not predict variation in B participants' procedural or allocation choices across LIE, SPY, and SABOTAGE. Similarly, B participants would not resort to any ethical (fairness) criterion; yet, we observe such a link with *Kohlberg class 5*.[29] Preferences for power might, however, explain why some B participants *within* the same treatment, opt for S_1 whereas others opt into S_2 and give all payoff away, both motivated by the same ethical criterion *Kohlberg class 5*. If B dislikes power, she might prefer to reinstate A's decision rights by opting into S_1; if she values power, she might seek and exert her power to compensate A for her lack of rights. Indeed, we find such a correspondence between B's choice and her materialism and postmaterialism values who, amongst other aspects, measure B's attitudes toward power, hierarchy, and autonomy.

Risk attitudes. In S_1 and S_2, B can achieve the same payoffs ex-post: 100 ECU, and 0 ECU. In S_2, however, B can obtain the 100 ECU for sure which is why a risk-averse B prefers S_2 where she takes all payoff.[30] Indeed, B's risk aversion slightly correlates

[28] We can express this advantage by the cardinalities (the fineness) of A's and B's information partitions over all possible terminal histories $z \in Z$. Again, B's utility function might include some element similar to $-b_B \max\{\#\mathcal{I}_B^z - \#\mathcal{I}_A^z, 0\} - a_B \max\{\#\mathcal{I}_A^z - \#\mathcal{I}_B^z, 0\}$ where $\#\mathcal{I}_B^z - \#\mathcal{I}_A^z$ and $\#\mathcal{I}_A^z - \#\mathcal{I}_B^z$ measure the difference between the cardinalities of A's and B's information partitions over all possible terminal histories, and a_B and b_B express B's aversion against having greater, or lesser, information rights. In S_1, B knows her own, but not A's choice and B's partition over the four terminal histories of S_1 has cardinality two. In S_2, B's partition over the four terminal histories has cardinality four: at the time of her decision, she knows which terminal history she will reach. A's information partition over the terminal histories in turn has cardinality one always, since she does not know whether she operates in S_1 or S_2. B's choice of S_2 therefore increases her own information rights, but does not reduce A's. In LIE and SABOTAGE, information rights are distributed the same way in S_1 and S_2: B's information partition has cardinality two always, A's partition always cardinality one.

[29] A preference for power would be a preference for maximizing one's own rights. The *purely procedural preferences* above build this idea into a framework of inequity aversion over decision rights (Chlaß et al. 2019) [one feels the infringement of one's own rights more immediately than one feels the infringement of another individual's rights], a preference for power would imply a disutility from losing control over the payoff distribution to other individuals, but no disutility at all from taking decision rights from others.

[30] Since B cannot obtain a higher ex-post payoff than these 100 ECU through incurring additional risk, also risk-loving or risk-neutral Bs prefer S_2 and take all payoff, but they prefer S_2 to a lesser extent than a risk-averse B.

with B's choice of S_2 and her taking all payoff, see table L. Risk attitudes do not predict varying degrees of altruism across LIE, SPY, or SABOTAGE.

Experimenter demand effects. Other than having addressed any of these preferences, we might— despite a neutral framing — have induced a cognitive or a social experimenter demand effect (Zizzo 2010) in that the existence of an experimenter, or the awareness of participating in an experiment affected B participants' behaviour. If so, B participants should be strongly motivated by a desire to satisfy our expectations and to behave in a socially acceptable way. We could not confirm that B participants strongly resort to ethical criteria such as others' expectations, social norms, social image concerns or a desire to be taken as a nice person – all located in *Kohlberg class 3* – when choosing either procedure or allocation.

7 Conclusion

We show, for the first time, that individuals value fair competition for its own sake and prefer to compete with opponents who are in a position to look after their own self-interest. In particular, individuals prefer to forego *all* payoff rather than fabricate information about their opponent or sabotage the latter, when doing so would win a constant sum game but at the same time, also encroach on the opponent's decision rights.

We begin with an intervention study, and design three different ways to compete unfairly within the same setup. Two of them affect the opponent's decision rights – fabrication and sabotage –, and one does not – spying. Substantial amounts of altruism occur in the first two treatments, and little to none in the third. We formally discuss at length that the only preference to produce this difference must be one *purely* over the rules of the game, such as Chlaß et al.'s (2019) preference for the equality of decision and information rights. In particular social norms or guilt aversion should, according to individuals' actual beliefs, produce either no, or the exact opposite difference in altruism.

To cross-check our theoretical analysis empirically, we supplement the intervention study by an instrumental variable approach. Chlaß et al. (2019) build individuals' purely procedural preferences around the ethical criterion of equal civic rights – equal freedom of choice, opportunity, and participation – as granted by the social contract. The same psychometric analysis which elicits individuals' preferences over *all* ethical criteria upon which economics has built preferences to date, shows that a preference for this ethical criterion explains the altruism we observe, controlling for all relevant potential latent correlates known for the sample from which individuals are drawn.

In a second set of interventions, worded identical to the first, we remove all influence individuals hold over their opponent's decision rights in all treatments, such that fabrication, sabotage, and spying become merely different frames for the same action. All treatments show similarly low occurences of altruism.

A third set of interventions reinforces the opponents' decision rights by an option to reward or punish fabrication, sabotage, and spying, which downweighs the ethical concern at work and, at the same time, provides the connecting dots with the literature. We observe that altruism now exclusively aims at avoiding punishment and earning reward. This

change in the nature of altruism is signalled by a significant drop in its amount which is, however, still substantial. Punishment and reward now defining the right course of action, individuals sabotage and fabricate to further their own ends as soon as they expect none. Opponents do not punish or reward where they have no other decision rights, such that the new ethical criterion at play condones power to be exploited at will.

Our results indicate that altruism is caused *purely* by the rules of the game. In this light, preferences over distributions of payoffs appear as behavioural strategies to compensate aspects of these rules which inidviduals deem unethical. These unethical aspects out of the way, we observe *exact* rational self-interest.

References

Abbink, K. and B. Herrmann (2011). "The Moral Costs of Nastiness". In: *Economic Inquiry* 49.2, pp. 631–633.

Abbink, K. and A. Sadrieh (2009). "The Pleasure of Being Nasty". In: *Economics Letters* 105, pp. 306–308.

Abeler, J., A. Becker, and A. Falk (2014). "Representative Evidence on Lying Costs". In: *Journal of Public Economics* 113, pp. 96–104. URL: http://www.sciencedirect.com/science/article/pii/S0047272714000061.

Abeler, J., D. Nosenzo, and Collin B. Raymond (2018). "Preferences for truth-telling". In: *Econometrica* forthcoming.

Abratt, R. and N. Penman (2002). "Understanding Factors Affecting Salespeople's Perceptions of Ethical Behavior in South Africa". In: *Journal of Business Ethics* 35 (4). 10.1023/A:1013872805967, pp. 269–280. ISSN: 0167–4544.

Agresti, Alan (2002). *Categorical Data Analysis*. 2nd. Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons, Inc.

Baker, D. E. and R. Inglehart (2000). "Modernization, Cultural Change, And the Persistence of Traditional Values". In: *American Sociological Review* 65.1, pp. 19–51.

Bartling, B., E. Fehr, and H. Herz (2014). "The Intrinsic Value of Decision Rights". In: *Econometrica* 82.6, pp. 2005–2039.

Bartling, B. et al. (2009). "Egalitarianism and Competitiveness". In: *The American Economic Review* 99.2, pp. 93–98. DOI: 10.1257/aer.99.2.93. URL: http://www.ingentaconnect.com/content/aea/aer/2009/00000099/00000002/art00016.

Battigalli, P., G. Charness, and M. Dufwenberg (2013). "Deception: The Role of Guilt". In: *Journal of Economic Behavior & Organization* 93, pp. 227–232. URL: http://www.sciencedirect.com/science/article/pii/S0167268113000784;%20http://www.igier.unibocconi.it/files/BattigalliCharnessDufwenberg_JEBO2013-1.pdf.

Battigalli, P. and M. Dufwenberg (2007). "Guilt in Games". In: *The American Economic Review* 2.97, pp. 170–176.

Beresford, A., D. Kübler, and S. Preibusch (2012). "Unwillingness to Pay for Privacy: a Field Experiment". In: *Economics Letters* 112, pp. 25–27.

Bolton, G. E., J. Brandts, and A. Ockenfels (2005). "Fair Procedures: Evidence from Games Involving Lotteries*". In: *The Economic Journal* 115.506, pp. 1054–1076. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1468--0297.2005.01032.x/full;%20http://digital.csic.es/bitstream/10261/1924/1/48301.pdf.

Bolton, G. and A. Ockenfels (2000). "ERC - A Theory of Equity, Reciprocity and Competition". In: *American Economic Review* 90, pp. 166–193.

Borg, I. et al. (2019). "Do the PVQ and the IRVS scales for personal values support Schwartz's value circle model or Klages' value dimensions model?*". In: *Measurement Instruments for the Social Sciences* 2/2019.3, pp. 2–14. URL: https://doi.org/10.1186/s42409-018-0004-2.

Brown, V. R. and E. D. Vaughn (2011). "The Writing On the (Facebook) Wall: The Use of Social Networking Sites in Hiring Decisions". In: *Journal of Business Psychology* 26, pp. 219–225.

Busch, W. (1906). *Max und Moritz: Eine Bubengeschichte in sieben Streichen*. 53rd ed. Braun und Schneider.

Cappelen, Alexander W, Erik Ø Sørensen, and Bertil Tungodden (2013). "When Do We Lie?" In: *Journal of Economic Behavior and Organization* 103.2013, pp. 258–265.

Carpenter, J., P.H. Matthews, and J. Schirm (2010). "Tournaments and Office Politics: Evidence from a real effort experiment". In: *The American Economic Review* 100.1, pp. 504–517.

Charness, G., D. Masclet, and M. C. Villeval (2014). "The Dark Side of Competition for Status". In: *Management Science* 60.1, pp. 38–55.

Charness, G. and M. Rabin (2002). "Understanding Social Preferences with Simple Tests". In: *The Quarterly Journal of Economics* 117.3, pp. 817–869.

Chlaß, N., W. Güth, and T. Miettinen (2019). "Purely procedural preferences – Beyond procedural equity and reciprocity." In: *European Journal of Political Economy* 59, pp. 108–128.

Chlaß, N. and P. Moffatt (2012). *Giving in Dictator Games, Experimenter Demand Effect, or Preference over the Rules of the Game?* Tech. rep. 2012–44. Jena Economic Research Papers.

Cooper and Roberts (2011). "After 40 Years, the Complete Pentagon Papers". In: *The New York Times* 2011-06-07. URL: %7Bhttp://www.nytimes.com/interactive/us/2011_PENTAGON_PAPERS.html%7D.

Dufwenberg, M. and G. Kirchsteiger (2004). "A Theory of Sequential Reciprocity". In: *Games and Economic Behavior* 47.3, pp. 268–98.

Erat, S. and U. Gneezy (2012). "White Lies." In: *Management Science* 58.4, pp. 723–733. URL: http://dblp.uni-trier.de/db/journals/mansci/mansci58.html#EratG12.

Falk, A., E. Fehr, and D. Huffman (2008). *The Power and Limits of Tournament Incentives*. Tech. rep. 2008. University of Zurich.

Falk, A. and U. Fischbacher (2006). "A Theory of Reciprocity". In: *Games and Economic Behavior* 54, pp. 293–315.

Fischbacher, U. (2007). "z–Tree: Zurich Toolbox for Ready–Made Economic Experiments". In: *Experimental Economics* 10.2, pp. 171–178. URL: http://link.springer.com/article/10.1007/s10683--006--9159--4;%20http://link.springer.com/content/pdf/10.1007_252Fs10683--006--9159--4.pdf.

Fischbacher, U. and F. Föllmi-Heusi (2013). "Lies in Disguise: An Experimental Study on Cheating". In: *Journal of the European Economic Association* 11.3, pp. 525–547.

Gibson, R., C. Tanner, and A. F. Wagner (2013). "Preferences for Truthfulness: Heterogeneity Among and Within Individuals". In: *American Economic Review* 103.1, pp. 532–548.

Gneezy, U. (2005). "Deception: The Role of Consequences". In: *The American Economic Review* 95.1, pp. 384–394. DOI: 10.1257/0002828053828662.

Greenwald, G., E. MacAskill, and L. Poitras (June 2013). "Edward Snowden: the Whistleblower Behind the NSA Surveillance Revelations". In: *The Guardian* 11. URL: http://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance?guni=Podcast:in%20body%20link.

Greiner, B. (2004). "An online recruitment system for economic experiments". In: *Forschung und wissenschaftliches Rechnen* 63. Ed. by K. Kremer and V. Macho. URL: http://mpra.ub.uni--muenchen.de/13513/;%20http://mpra.ub.uni--muenchen.de/13513/1/MPRA_paper_13513.pdf.

Harbring, C. and B. Irlenbusch (2011). "Sabotage in Tournaments: Evidence from a Laboratory Experiment." In: *Management Science* 57.4, pp. 611–627.

Harbring, C. et al. (2007). "Sabotage in Corporate Contests–An Experimental Analysis". In: *International Journal of the Economics and Business* 14, pp. 201–223.

Holt, C. A. and S. K. Laury (2002). "Risk Aversion and Incentive Effects". In: *American Economic Review* 92.5, pp. 1644–1655.

Hurkens, S. and N. Kartik (2009). "Would I Lie to You? On Social Preferences and Lying Aversion". In: *Experimental Economics* 12, pp. 180–192.

Inglehart, R. (1977). *The Silent Revolution: Changing Values and Political Styles Among Western Publics*. Princeton University Press.

Jones, P. and R. Sugden (1982). "Evaluating Choice". In: *International Review of Law and Economics* 2, pp. 47–65.

Kirchsteiger, G. (1994). "The Role of Envy in Ultimatum Games". In: *Journal of Economic Behavior & Organization* 25.3, pp. 373–389. URL: http://www.sciencedirect.com/science/article/pii/0167268194901066;%20http://www.ecares.org/ecare/personal/kirchsteiger/publications/1994jebo--envy.pdf.

Klages, H. and Th. Gensicke (2006). "Wertesynthese - Funktional oder Dysfunktional?" In: *Kölner Zeitschrift für Soziologie and Sozialpsychologie* 58.2, pp. 332–351.

Kohlberg, L. (1969). "Stage and Sequence: the Cognitive–Developmental Approach to Socialization". In: ed. by D.A. Goslin. Handbook of Socialization and Endash; Theory and research. Chicago: McNally.

— (1984). *The Psychology of Moral Development*. San Francisco: Harper & Row.

Larson, J., N. Perlroth, and S. Shane (Sept. 2013). "Revealed: The NSA's Secret Campaign to Crack, Undermine Internet Security". In: *New York Times* 5. URL: http://www.propublica.org/article/the-nsas-secret-campaign-to-crack-undermine-internet-encryption.

Lightle, J. P. (2014). "The Paternalistic Bias of Expert Advice". In: *Journal of Economics & Management Strategy* 23.4, pp. 876–898. DOI: 10.1111/jems.12070. URL: http://www.ingentaconnect.com/content/bpl/jems/2014/00000023/00000004/art00006.

Lind, G. (1978). "Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts". In: *Sozialisation und Moral*. Ed. by G. Portele. Weinheim: Beltz, pp. 1215–1259.

Lind, G. (2008). "The Meaning and Measurement of Moral Judgment Competence Revisited – A Dual–Aspect Model". In: *Contemporary Philosophical and Psychological Perspectives on Moral Development and Education*. Ed. by D. Fasko and W. Willis. Cresskill. NJ: Hampton Press, pp. 185–220.

López-Pérez, Raúl and Eli Spiegelman (2013). "Why Do people Tell the Truth? Experimental Evidence for Pure Lie Aversion". In: *Experimental Economics* 16.3, p. 233. ISSN: 1573-6938. DOI: 10.1007/s10683-012-9324-x. URL: http://dx.doi.org/10.1007/s10683-012-9324-x.

Miettinen, T. (2013). "Promises and Conventions – An Approach to Pre-Play Agreements". In: *Games and Economic Behaviour* 80.80, pp. 68–84.

Piaget, J. (1948). *The Moral Judgment of the Child*. Glencoe, Illinois: Free Press.

Sebald, A. (2010). "Attribution and Reciprocity." In: *Games and Economic Behavior* 68.1, pp. 339–352. URL: http://dblp.uni--trier.de/db/journals/geb/geb68.html#Sebald10.

Sheehan, N. (1971). "Vietnam Archive: Pentagon Study Traces 3 Decades of Growing U.S. Involvement". In: *New York Times* 1971.13.06.1971.

Smith, A. (1904). *An Inquiry into the Nature and Causes of the Wealth of Nations*. 5th ed. London: Methuen & Co., Ltd. URL: http://www.econlib.org/library/Smith/smWN.html.

Sutter, M. (2009). "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams". In: *The Economic Journal* 119.534, pp. 47–60. DOI: 10.1111/j.1468--0297.2008.02205.x.

Villeval, M. C. and J. van den Ven (2015). "Dishonesty under scrutiny". In: *Journal of the Economic Science Association* 1, pp. 86–99.

Zizzo, D. J. (2010). "Experimenter Demand Effects in Economic Experiments". In: *Experimental Economics* 13.1, pp. 75–98. URL: http://link.springer.com/article/10.1007/s10683--009--9230--z.

A Screenshots

A.1 B's choice Prob(S_2) of the situation

[You are participant B.
You may now influence whether situation 1 or situation 2 occurs.
For the time being, situation 1 or situation 2 occur randomly with equal probability (50%). You may, however, change these initial probabilities, depending on whether you prefer to encounter situation 1 or situation 2.
To change the initial probabilities, move the slider below by clicking on the arrows on its left (rather situation 1) or on its right (rather situation 2). For every stepwise change of the slider, your initial endowment decreases by 0.10 ECU.
Your decision:
Cost in ECU
Please click OK

Figure A1: *B*'s probability choice Prob(S_2) of the situation.

A.2 B's choices in situation S_1

[You are participant B.
The situation drawn is situation 1.
You now have three possibilities to make a decision.
You can choose L directly, R directly, or choose randomly between L and R (L and R then occur with 50% probability each.)

Your choice:
○ L
○ R
○ random choice

To confirm, please click OK. (You can no longer change your decision afterwards).]

Figure A2: B's choices in situation S_1.

A.3 B's choices in situation S_2 SPY

[You are participant B.
The situation drawn is situation 2.
You now have three possibilities to make a decision. You can choose L directly, R directly, or choose randomly between L and R (L and R then occur with 50% probability each.)

A has chosen:
○ L
○ R
○

My choice:
○ L
○ R
○ random choice

To confirm, please click OK. (You can no longer change your decision afterwards).]

Figure A3: B's DECISION SCREEN IN S_2, TREATMENT SPY.

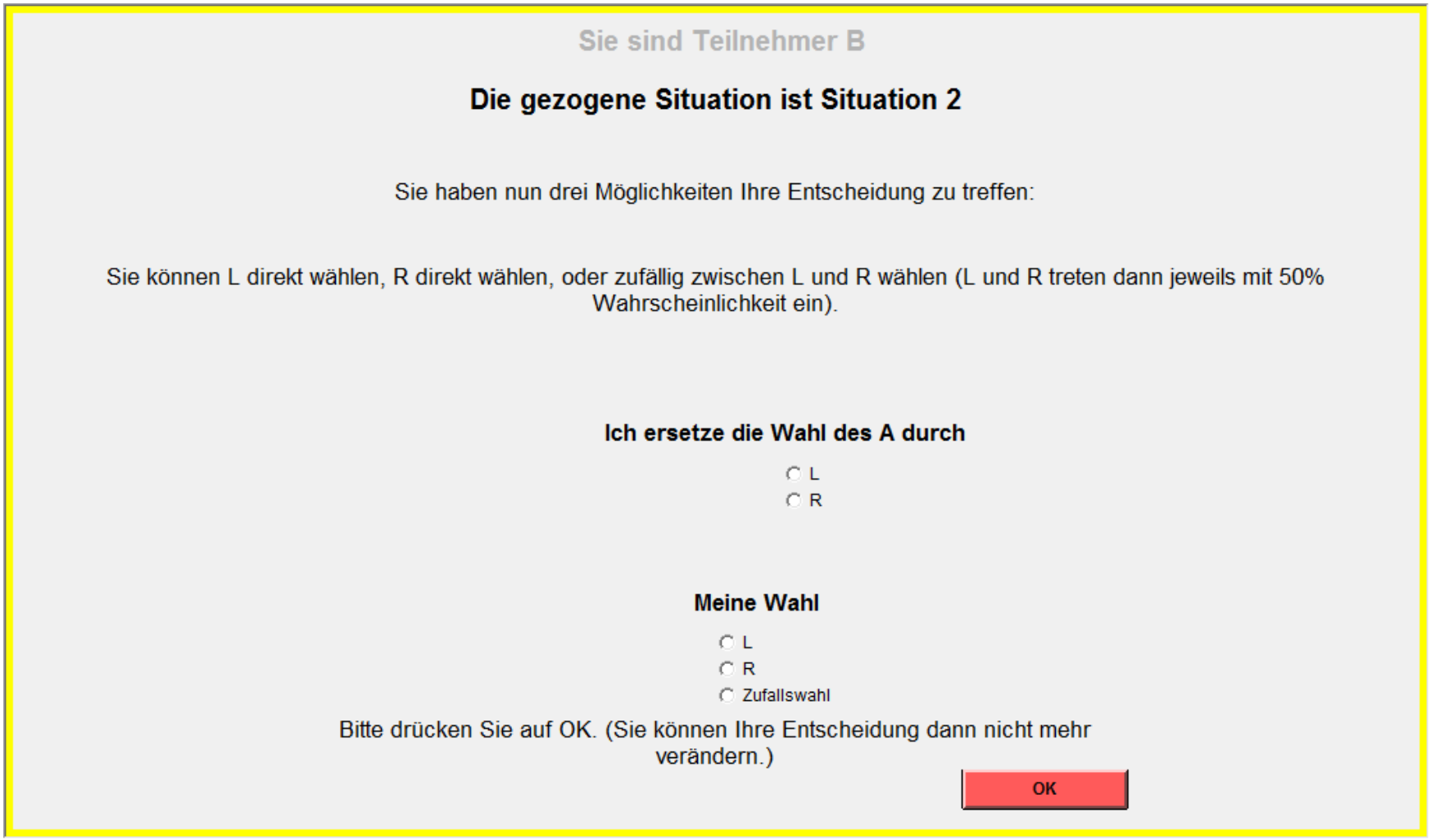


Figure A4: B's decision screen in S_2, treatment SABOTAGE.

[You are participant B.
The situation drawn is situation 2.
You now have three possibilities to make a decision. You can choose L directly, R directly, or choose randomly between L and R (L and R then occur with 50% probability each.)

I replace A's choice by:
○ L
○ R

My choice:
○ L
○ R
○ random choice

To confirm, please click OK. (You can no longer change your decision afterwards).]

A.5 B's choices in situation S_2 LIE

[You are participant B.
The situation drawn is situation 2.
You now have three possibilities to make a decision. You can choose L directly, R directly, or choose randomly between L and R (L and R then occur with 50% probability each.)

I transmit as A's choice:
○ L
○ R

My choice:
○ L
○ R
○ random choice

To confirm, please click OK. (You can no longer change your decision afterwards).]

Figure A5: B's DECISION SCREEN IN S_2, TREATMENT LIE.

A.6 B's punishment beliefs in the punishment/reward stage

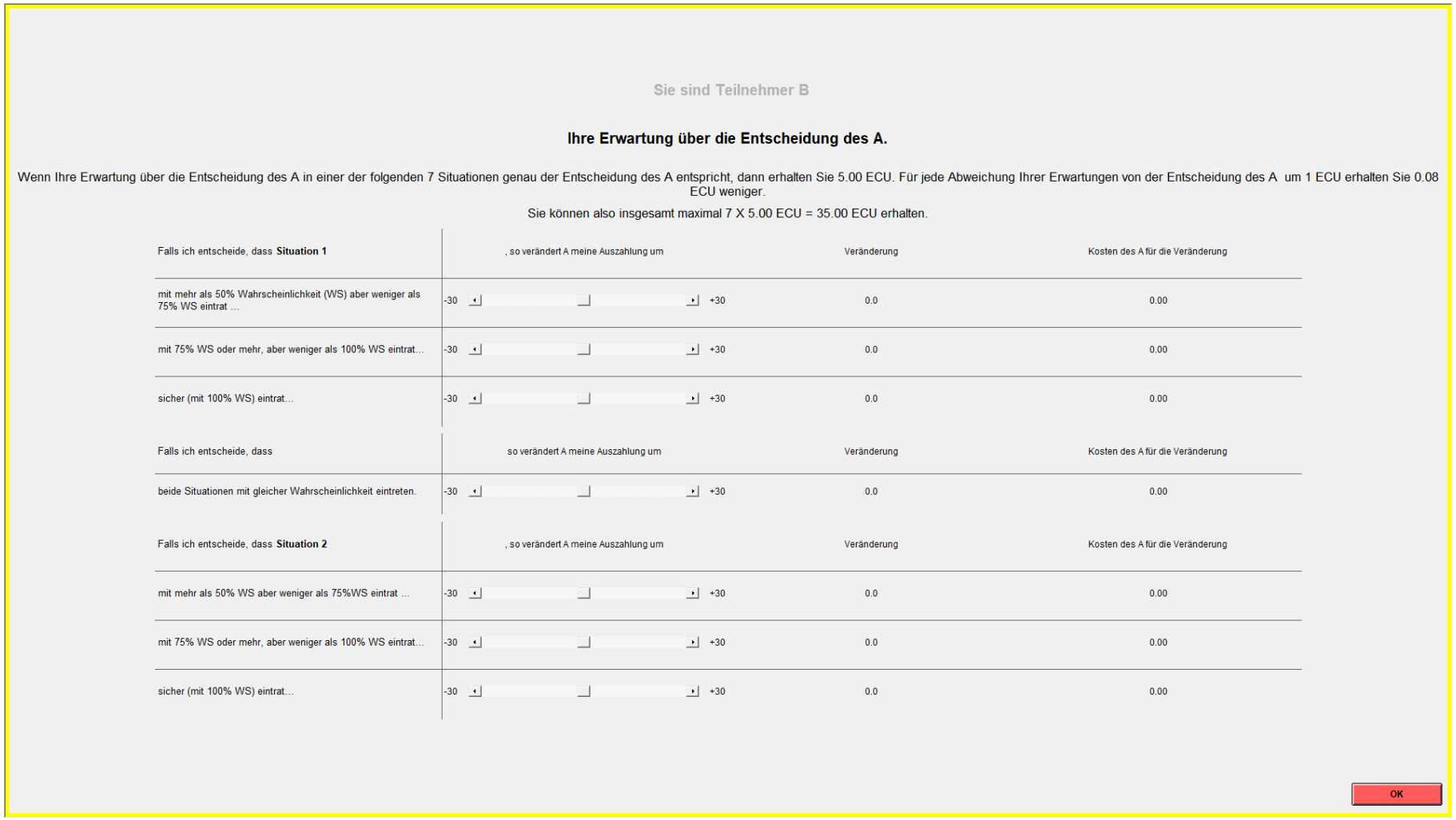


Figure A6: *B*'s decision screen to submit her 1st order beliefs about *A*'s punishment and reward of Prob(S_2).

[You are participant B.

Your expectation regarding *A*'s choice.

If your expectation about *A*'s decision is exactly equal to *A*'s decision in one of the following 7 situations, you receive 5.00 ECU. For every ECU by which your expectations differ from *A*'s actual decision, you receive 0.08 ECU less. That is, you can altogether maximally earn 7x5=35 ECU.

If I decide that **situation 1**	, *A* changes my payoff by	change	*A*'s cost for the change
occured with more than 50% but less than 75% probability (Pr)...	-30 ⟷ +30	0.0	0.00
occured with 75% Pr or more but less than 100% Pr...	-30 ⟷ +30	0.0	0.00
occurred for sure (with 100% Pr)...	-30 ⟷ +30	0.0	0.00
If I decide that	, *A* changes my payoff by	change	*A*'s cost for the change
both situations occured with equal probability...	-30 ⟷ +30	0.0	0.00
If I decide that **situation 2**	, *A* changes my payoff by	change	*A*'s cost for the change
occured with more than 50% Pr but less than 75% Pr...	-30 ⟷ +30	0.0	0.000
occured with 75% Pr or more, but less than 100% Pr...	-30 ⟷ +30	0.0	0.00
occured for sure (with 100% Pr)...	-30 ⟷ +30	0.0	0.00

B Experimental Instructions

B.1 Instructions[31]

Instructions

Welcome and thank you for participating in this experiment. Please read the following instructions carefully. The instructions are identical for all participants. Communication with other participants must cease from now on. Please turn off your mobile phone. If you have any questions, please raise your hand - we will answer them individually at your seat. Do not ask your questions aloud.

During the experiment, monetary amounts are denoted in ECU (Experimental Currency Units). The sum of your payoffs from all rounds will be disbursed to you in cash at the end of the experiment (exchange rate 1 ECU=0.05 Euro). Your initial endowment is 50 ECU.

Information about the experiment

In this experiment, you interact with other anonymous participants. Participants take on different roles **A** and **B** [TREATMENT LIE: and **C**]. Roles are randomly determined at the beginning and remain the same throughout the experiment. The experiment consists of several rounds. In each round, you are matched with a new participant. In each round, you encounter two situations. These situations are initialized to occur with probability 50%. At the beginning of each round, B can decide which situation actually occurs, and can make one situation more likely than the other. Making one situation 10 percent more likely costs 1 ECU. The two situations are characterized as follows.

Situation 1. Participant A chooses between two options L and R. Participant B does not see which option A has chosen. B then also chooses between options L and R. Both participants can also choose options L and R with equal probability.

Situation 2. Participant A chooses between two options L and R. Participant B does not see which option A has chosen. B sets *A*'s choice to either L or R. B then also chooses between options L and R. Both participants can also choose options L and R with equal probability.

[IN TREATMENT LIE, SITUATION 2 READ AS FOLLOWS:
Situation 2. Participant A chooses between two options L and R. Participant B does

[31]Instructions of the experiment were written in German. This appendix produces a translation into English for treatment SABOTAGE with competitive payoffs. Instructions for treatments SPY and LIE differed by the text in square brackets. TEXT IN CAPITAL LETTERS WAS NOT PART OF THE ORIGINAL INSTRUCTIONS. Emphases in bold or italic font are taken from the original text. Instructions for the payoff neutral treatment were worded identically, the only difference being the respective numbers in the payoff table: If *B* chose R and *A* chose L, *A* received 0, and *B* 100 ECU. If *B* chose R and *A* chose R, *A* received 0 ECU and *B* 100 ECU. If *B* chose L and *A* chose L, *A* received 100 ECU and *B* 0 ECU. If *B* chose L and *A* chose R, *A* received 100 ECU and *B* 100 ECU.

not see which option A has chosen. B transmits *A*'s choice to participant **C**. B then also chooses between options L and R. Both participants can also choose options L and R with equal probability.]

[In Treatment Spy, situation 2 read as follows:
Situation 2. Participant A chooses between two alternatives L and R. Participant B sees which option A has chosen. B then also chooses between options L and R. Both participants can also choose options L and R with equal probability.]

Your Payoff

The table below shows which payoffs A and B receive for their choices in a given round. At the end of the experiment, one round will be paid out (exchange rate 1 ECU=0.05 Euro). The computer selects this round randomly and with equal probability. [treatment lie: Participant C receives a fixed payoff of 125 ECU.]

Decisions	Payoffs for these decisions
B chooses L, **A** chooses L	**A** receives 100 ECU, **B** receives 0 ECU
B chooses L, **A** chooses R	**A** receives 0 ECU, **B** receives 100 ECU
B chooses R, **A** chooses L	**A** receives 0 ECU, **B** receives 100 ECU
B chooses R, **A** chooses R	**A** receives 100 ECU, **B** receives 0 ECU
B chooses 'randomly' and/or **A** chooses 'randomly'	Chance decides with equal probability whether **A** receives 100 ECU and**B** receives 0 ECU, or whether **A** receives 0 ECU and **B** receives 100 ECU.

If B chooses L, B receives 0 ECU (and A 100 ECU) if A also chooses L. If B chooses R, B receives 0 ECU (and A 100 ECU) if A also chooses R. If B chooses L, B receives 100 ECU (and A 0 ECU) if A chooses R. If B chooses R, B receives 100 ECU (and A 0 ECU) if A chooses L. If B chooses 'randomly' and/or A chooses 'randomly', B receives with 50% probability 100 ECU (and A 0 ECU), and with 50% probability 0 ECU (and A 100 ECU.).

Please be patient until all participants have read the instructions. Before the experiment starts, please answer the following comprehension questions.

B.2 Comprehension Questions

Comprehension Questions

Question 1 Assume B chooses 'L'. What are A's and B's payoffs in situation 2?

Participant A's payoff is:

Participant B's payoff is:

Question 2 What are A's and B's payoffs if B chooses 'L' in situation 1?

Participant A's payoff is:

Participant B's payoff is:

Question 3 Assume B chooses 'R'. What are A's and B's payoffs?

Participant A's payoff is:

Participant B's payoff is:

Question 4 If B chooses 'random choice',...

...both participants receive 100 ECU: ○ false ○ true

...both participants receive with equal probability either 0 or 100 ECU: ○ false ○ true

Question 5 Please answer the following true/false statements.

In situation 2, participant B can determine A's choice, irrespective of what A has chosen: ○ false ○ true

In situation 1, participant B cannot influence A's decision: ○ false ○ true

[In treatment LIE, Question 5 read as follows:

Question 5 Please answer the following true/false statements.

In situation 2, participant B transmits A's and B's decisions to participant C without learning A's actual decision: ○ false ○ true

In situation 1, participant C does not learn either A's or B's decision: ○ false ○ true

]

[In treatment SPY, Question 5 read as follows:

Question 5 Please answer the following true/false statements.

In situation 2, participant B learns participant A's decision ○ false ○ true

In situation 1, no participant learns the other participant's decision: ○ false ○ true

]

C Normal form representation of the payoff neutral regime.

Table 6: Payoff neutrality: party B does not gain additional freedom of choice through spying, sabotaging, or fabricating A, and does not infringe A's freedom of choice.

6a) the 'fair' set of rules

		party A: L	party A: R
party B	L	100 0	100 0
	R	0 100	0 100

6b) the 'unfair' set of rules

		party A: L	party A: R
party B	LL^A	100 0	100 0
	RL^A	0 100	0 100
	LR^A	100 0	100 0
	RR^A	0 100	0 100

D Normal form representation of the competitive payoffs regime with symbolic reward and punishment.

Table 7: A's symbolic punishment and reward option increases her decision decision rights: A can reduce (or increase) the extent to which B prefers L over R by 30 ECU in S_1, and the extent to which B prefers RL^A or LR^A over LL^A and RR^A by 30 ECU in S_2.

7a) the 'fair' set of rules

		A: $L^{Nopunish/Reward}$	A: $R^{Nopunish/Reward}$	A: $L^{Punish/Reward}$	A: $R^{Punish/Reward}$
B	L	100 0	0 100	$100 + [-30, 30] \setminus 0$ $0 + [-30, 30] \setminus 0$	$0 + [-30, 30] \setminus 0$ $100 + [-30, 30] \setminus 0$
	R	0 100	100 0	$0 + [-30, 30] \setminus 0$ $100 + [-30, 30] \setminus 0$	$100 + [-30, 30] \setminus 0$ $0 + [-30, 30] \setminus 0$

7b) the 'unfair' set of rules

		A: $L^{NoPunish/Reward}$	A: $R^{NoPunish/Reward}$	A: $L^{Punish/Reward}$	A: $R^{Punish/Reward}$
B	LL^A	100 0	100 0	$100 + [-30, 30] \setminus 0$ $0 + [-30, 30] \setminus 0$	$100 + [-30, 30] \setminus 0$ $0 + [-30, 30] \setminus 0$
	RL^A	0 100	0 100	$0 + [-30, 30] \setminus 0$ $100 + [-30, 30] \setminus 0$	$0 + [-30, 30] \setminus 0$ $100 + [-30, 30] \setminus 0$
	LR^A	100 0	100 0	$100 + [-30, 30] \setminus 0$ $0 + [-30, 30] \setminus 0$	$100 + [-30, 30] \setminus 0$ $0 + [-30, 30] \setminus 0$
	RR^A	0 100	0 100	$0 + [-30, 30] \setminus 0$ $100 + [-30, 30] \setminus 0$	$0 + [-30, 30] \setminus 0$ $100 + [-30, 30] \setminus 0$

E Defining sabotage: Max and Moritz (Busch 1906).

Figure A7: Max and Moritz fill their teacher's pipe with black powder.

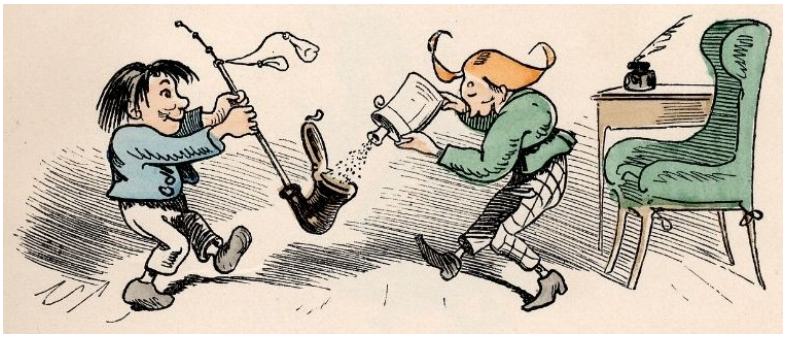


Figure A8: Lighting the pipe has a new consequence for the teacher.

F Results: Bs' behaviour across part 1 and part 2.

F.1 Competitive payoffs

LIE competitive ($n = 44$)

Prob(S_2) part 1 \ Prob(S_2) part 2	$< 50\%$	50%	$> 50\%$
$< 50\%$	3	1	5
50%	4	20	6
$> 50\%$	0	3	2

Notes: 43% within]24%, 63%] (19 of 44) Bs opt for a different situation in part 2.

SPY competitive ($n = 53$)

Prob(S_2) part 1 \ Prob(S_2) part 2	$< 50\%$	50%	$> 50\%$
$< 50\%$	2	0	3
50%	0	8	4
$> 50\%$	0	6	30

Notes: 25% within]11%, 42%] (13 of 53) Bs opt for a different situation in part 2.

SAB competitive ($n = 54$)

Prob(S_2) part 1 \ Prob(S_2) part 2	$< 50\%$	50%	$> 50\%$
$< 50\%$	0	1	1
50%	3	5	7
$> 50\%$	1	8	28

Notes: 39% within]22%, 57%] (21 of 54) Bs opt for a different situation in part 2.

LIE competitive

part 1 \ part 2	S_1	alt	self
S_1	10	4	5
alt	5	8	4
self	5	0	3

Notes: 33%]6%, 73%] (4 of 12) altruists who arrive in S_2 again, are selfish in part 2.

SPY competitive

part 1 \ part 2	S_1	alt	self
S_1	6	0	7
alt	0	0	0
self	10	0	30

Notes: No altruism occurs either in part 1 or part 2.

SAB competitive

part 1 \ part 2	S_1	alt	self
S_1	6	11	9
alt	6	5	9
self	2	3	3

Notes: 64%]27%, 91%] (9 of 14) altruists who arrive in S_2 again, are selfish in part 2.

F.2 Payoff neutrality

LIE neut ($n = 47$)

Prob(S_2) part 1 \ Prob(S_2) part 2	$< 50\%$	50%	$> 50\%$
$< 50\%$	2	4	2
50%	4	29	3
$> 50\%$	0	2	1

Notes: 32% within]15%, 51%] (15 of 47) Bs opt for a different situation in part 2.

SPY neut ($n = 53$)

Prob(S_2) part 1 \ Prob(S_2) part 2	$< 50\%$	50%	$> 50\%$
$< 50\%$	1	1	0
50%	2	25	5
$> 50\%$	1	8	10

Notes: 32% within]16%, 50%] (17 of 53) Bs opt for a different situation in part 2.

SAB neut ($n = 52$)

Prob(S_2) part 1 \ Prob(S_2) part 2	$< 50\%$	50%	$> 50\%$
$< 50\%$	3	0	1
50%	2	26	2
$> 50\%$	5	5	8

Notes: 29% within]14%, 47%] (15 of 52) Bs opt for a different situation in part 2.

LIE neut

part 1 \ part 2	alt	self
alt	2	8
self	4	33

Notes: 80%]35%, 98%] (8 of 10) altruists who arrive in S_2 again, are selfish in part 2.

SPY neut

part 1 \ part 2	alt	self
alt	1	3
self	4	45

Notes: 25%]0%, 89%] (1 of 4) altruists who arrive in S_2 again, are selfish in part 2.

SAB neut

part 1 \ part 2	alt	self
alt	1	7
self	6	38

Notes: 88%]35%, 99%] (7 of 8) altruists who arrive in S_2 again, are selfish in part 2.

G Kohlberg's six ways of moral argumentation

Table 8: Six ways of moral argumentation (summary by Ishida 2006, examples from the authors).

argumentation	Classes of motivation for moral behavior	It is good not to lie/spy/sabotage the opponent because...
preconventional way	***Class 1.*** Orientation to punishment and obedience, physical and material power. Rules are obeyed to avoid punishment. ***Class 2.*** Naïve hedonistic orientation. The individual conforms to obtain rewards.	...I can be punished If do; ...because I'll get a reward if I do not.
conventional way	***Class 1.*** "Good boy/girl" orientation to win approval and maintain expectations of one's immediate group. The individual conforms to avoid disapproval. One earns approval by being "nice". ***Class 2.*** Orientation to authority, law, and duty, to maintain a fixed order. Right behavior consists of doing one's duty and abiding by the social order.	...recipient or experimenter expect me to/will think I am a nice person ...because it is the norm not to do so; ... because it is against the law; ... because doing so would endanger all order in our society
postconventional way	***Class 1.*** Social contract orientation. Duties are defined in terms of the social contract and the respect of others' rights. Emphasis is upon equality and mutual obligation within a democratic order. ***Class 2.*** The morality of individual principles of conscience, such as the respect for the individual will, freedom of choice etc. Rightness of acts is determined by conscience in accord with comprehensive, universal and consistent ethical principles.	...the opponent's civic rights to privacy, and to democratic participation must be respected, or else be compensated; ... the opponent must as an equal human being be free to choose, to state her own will or else be compensated.

H An Excerpt of the Moral Judgement Test by Georg Lind (1976, 2008)

Doctor

A woman had cancer and she had no hope of being saved. She was in terrible pain and so weak that a large dose of a pain killer such as morphine would have caused her death. During a temporary period of improvement, she begged the doctor to give her enough morphine to kill her. She said she could no longer stand the pain and would be dead in a few weeks anyway. The doctor decided to give her a overdose of morphine.

Do you agree or disagree with the doctor's action ...

I strongly disagree — I strongly agree

-3	-2	-1	0	1	2	3

How acceptable do you find the following arguments *in favor* of the doctor's actions?
Suppose someone argued he acted *rightly*...

...because the doctor had to act according to his conscience. The woman's condition justified an exception to the moral obligation to preserve life

I strongly reject — I strongly accept

-4	-3	-2	-1	0	1	2	3	4

. . .

...because the doctor was the only one who could fulfill the woman's wish; respect for her wish made him act as he did.

I strongly reject — I strongly accept

-4	-3	-2	-1	0	1	2	3	4

How acceptable do you find the following arguments *against* the doctor's actions?
Suppose someone argued he acted *wrongly*

. . .

...because he acted contrary to his colleagues' convictions. If they are against mercy-killing the doctor shouldn't do it.

I strongly reject — I strongly accept

-4	-3	-2	-1	0	1	2	3	4

. . .

...because one should be able to have complete faith in a doctor's devotion to preserving life even if someone with great pain would rather die

I strongly reject — I strongly accept

-4	-3	-2	-1	0	1	2	3	4

NOTE: This excerpt of the moral judgement test MJT is reprinted with kind permission by Georg Lind. It does not faithfully reproduce the formatting of the original test. For ease of readability, the original test numbers each item, and the alignment slightly differs from this excerpt. The dots represent items which have been left out. The full test cannot be published due to copyright protection.

I Klages's and Gensicke's (2006) materialism - postmaterialism scales[32]

Table 9: Questionnaire items for each of Klages's and Gensicke's three value dimensions (categories) to identify materialists, postmaterialists, and mixed value types in the German population (Klages and Gensicke 2006).

value category I duty and acceptance values	**value category II hedonistic and materialistic values**	**value category III idealistic values and public participation**[33]
✓ respect law and order	✓ have a high living standard	✓ develop one's fantasy and creativity
✓ need and quest for security	✓ hold power and influence	✓ help socially disadvantaged and socially marginal groups
✓ be hard-working and ambitious	✓ enjoy life to the fullest	✓ also tolerate opinions with which one actually cannot really agree
	✓ assert oneself, and one's needs against others	✓ be politically active

conventionalists	high scores on value category I (Inglehart's classic materialist values). Intermediate scores for value categories II and III. Clear hierarchy between value category I and II/III → approximate Inglehart's 'materialists' but Inglehart classifies value category II as 'materialist' values (with the exception of item 3) and not as a separate dimension.
idealists	high scores on value category III. Intermediate scores for value category II. Clear hierarchy between both value categories. Lower scores on value category I than conventionalists → approximation of Inglehart's postmaterialists.
hedonic materialists	score lower than conventionalists in value category I and lower than idealists in value category III. No hierarchy between value categories (all similarly important).
resigned without perspective	lower scores on category I than conventionalists and lower scores on value category III than idealists. Lowest scores in value category II. One of Inglehart's 'mixed types'.
realists	second lowest value hierarchy after hedonists, high scores on category I and relatively high scores on category II; 'synthesis' of values. One of Inglehart's 'mixed types'.

[32] Klages and Gensicke (2006) use these value categories to obtain the clusters (types) below: conventionalists, resigned people, realists, hedo-materialists, and idealists. In this paper, we do not cluster people into these groups; we use each individuals' average rating for all three value categories to model B participants' choice of the fair rules (type i)), or their altruism (type ii) under the unfair rules as opposed to the selfish type (type iv). The average rating is the mean rating over all questionnaire items pertaining to the same value category. Individuals rate each item from 1 to 7.

[33] Category III corresponds to Ingelhart's postmaterialism value scale. Higher mean ratings on value category III make the procedural type i) in section 5 more likely. Category II mostly belongs to Inglehart's materialist values. Higher mean ratings of this value category makes the altruistic type ii) in section 5 more likely. Value category I does not significantly influence B participants' choices in the experiment.

J Classification of B participants' choices of situation and allocation

J.1 Competitive payoffs

altruistic allocation

	Prob(S_2)	
	< 50	≥ 50
LIE	2	15
SPY	0	0
SAB	0	20

Notes. S_2 + GIVE ALL: 2+15+20+12+8+24.

situation 1

	Prob(S_2)	
	$< 50\%$	$\geq 50\%$
LIE	7	12
SPY	5	8
SAB	2	24

Notes. S_1: 7+5+2=14

selfish allocation

	Prob(S_2)		
	$< 50\%$	50%	$> 50\%$
LIE	0	5	3
SPY	0	8	32
SAB	0	1	7

Notes. SELFISH: 3+32+7=42.
FAIR COIN: 5+8+1=14

K B participants' demographics and their ethical preferences.

	Estimate	Std. Error	t value	Pr(>|z|)
Intercept	1.6496	0.9988	1.6516	0.1009
Age	-0.0177	0.0332	-0.5337	0.5944
Gender: female	0.0739	0.1851	0.3992	0.6904
Envy	-0.2775	0.1685	-1.6465	0.1020
Risk aversion	-0.0201	0.0507	-0.3954	0.6932
Field of Study: Education	-0.4390	0.4307	-1.0194	0.3098
Field of Study: Law	-1.3273	0.4821	-2.7534	0.0067***
Field of Study: IT	-1.2939	0.8587	-1.5067	0.1342
Field of Study: Philosophy	-1.2071	0.4207	-2.8692	0.0048***
Field of Study: Social and Behavioral Sciences	-0.8330	0.4351	-1.9142	0.0577*
Field of Study: Medicine	-0.7349	0.4884	-1.5047	0.1347
Field of Study: Business and Economics	-1.0720	0.4771	-2.2467	0.0263**
Field of Study: Engeneering	-0.6268	0.5284	-1.1862	0.2376
Field of Study: Languages	-0.2515	0.4462	-0.5636	0.5739
Field of Study: Sciences	-0.4058	0.5238	-0.7747	0.4399

Table 10: Link between Kohlberg class Five and B participants' demographics.

Notes. Linear regression with robust standard errors of *Kohlberg class 5* from table 5 on variables displayed. **Age** – B participants' age in whole years, ranges from 18 to 35 with a median of 23; **Envy** – Dummy taking on a value of One if B allocated 10 ECU to herself and 10 ECU to A, rather than 10 ECU to herself and 20 ECU to A; **Risk aversion** – ordinal variable ranging from 1 and 10; indicates at which lottery on a ten-item Holt-Laury lottery list B starts to prefer the sure payoff of 25 ECU to a binary lottery of 10 ECU to 35 ECU, see section 3.3.; **Fields of study** – miscellaneous category is Field of Study: Arts.

	Estimate	Std. Error	t value	Pr(>|z|)
Intercept	2.5290	0.8425	3.0017	0.0032***
Age	-0.0483	0.0297	-1.6266	0.1061
Gender: female	0.0182	0.1523	0.1192	0.9053
Envy	-0.1543	0.1539	-1.0025	0.3179
Risk aversion	-0.0911	0.0426	-2.1381	0.0343**
Field of Study: Education	-0.3372	0.2133	-1.5809	0.1162
Field of Study: Law	-1.3515	0.2827	-4.7811	0.0000***
Field of Study: IT	-1.9186	1.0218	-1.8776	0.0626*
Field of Study: Philosophy	0.3053	0.2176	1.4029	0.1629
Field of Study: Social and Behavioral Sciences	-0.6556	0.2277	-2.8797	0.0046***
Field of Study: Medicine	-0.7300	0.3923	-1.8610	0.0649*
Field of Study: Business and Economics	-0.9653	0.3079	-3.1348	0.0021***
Field of Study: Engeneering	-0.2598	0.3386	-0.7673	0.4443
Field of Study: Languages	-0.0312	0.2882	-0.1083	0.9139
Field of Study: Sciences	-0.3261	0.2846	-1.1458	0.2539

Table 11: Correlation of Kohlberg class Six and B participants' demographics.

Notes. Linear regression with robust standard errors of *Kohlberg class 6* from table 5 on variables displayed. **Age** – B participants' age in whole years, ranges from 18 to 35 with a median of 23; **Envy** – Dummy taking on a value of One if B allocated 10 ECU to herself and 10 ECU to A, rather than 10 ECU to herself and 20 ECU to A; **Risk aversion** – ordinal variable ranging from 1 to 10; indicates at which lottery on a ten-item Holt-Laury lottery list B starts to prefer the sure payoff of 25 ECU to a binary lottery of 10 ECU and 35 ECU, see section 3.3.; **Fields of study** – miscellaneous category is Field of Study: Arts.

	Estimate	Std. Error	t value	Pr(>|z|)
Intercept	2.2201	1.0983	2.0214	0.0452**
Age	-0.0259	0.0327	-0.7941	0.4285
Gender: female	0.0870	0.1645	0.5286	0.5979
Envy	-0.3127	0.1590	-1.9669	0.0512*
Risk aversion	-0.0290	0.0494	-0.5874	0.5579
Field of Study: Education	-0.6190	0.6963	-0.8890	0.3756
Field of Study: Law	-1.7125	0.7151	-2.3948	0.0180**
Field of Study: IT	-1.6366	1.0011	-1.6347	0.1044
Field of Study: Philosophy	0.0196	0.6937	0.0283	0.9775
Field of Study: SBS	-1.0214	0.6892	-1.4822	0.1406
Field of Study: Medicine	-1.2214	0.7753	-1.5755	0.1175
Field of Study: Business and Economics	-1.2348	0.7316	-1.6877	0.0938*
Field of Study: Engeneering	-0.8794	0.7672	-1.1462	0.2537
Field of Study: Languages	-0.7270	0.7243	-1.0037	0.3173
Field of Study: Sciences	-0.8355	0.7334	-1.1392	0.2566

Table 12: Correlation of Kohlberg class Three and B participants' demographics.

Notes. Linear regression with robust standard errors of *Kohlberg class 3* from table 5 on variables displayed. **Age** – B participants' age in whole years, ranges from 18 to 35 with a median of 23; **Envy** – Dummy taking on a value of One if B allocated 10 ECU to herself and 10 ECU to A, rather than 10 ECU to herself and 20 ECU to A; **Risk aversion** – ordinal variable ranging from 1 to 10; indicates at which lottery on a ten-item Holt-Laury lottery list B starts to prefer the sure payoff of 25 ECU to a binary lottery of 10 ECU and 35 ECU, see section 3.3; **Fields of study** – miscellaneous category is Field of Study: Arts.

L Bs' altruism: sample size and demographic controls

Dependent Dummy variable →	S_1 (1) VS. SELFISH (0)	S_2 + GIVE ALL (1) VS. SELFISH (0)	S_1 (1) VS. S_2 + GIVE ALL (0)
nr. of obs.	56 (14 vs. 42)	123 (81 vs. 42)	94 (14 vs. 81)
Kohlberg class 1	-0.132^{b}	-0.101^{a}	0.084
	(0.063)	(0.035)	(0.063)
Kohlberg class 2	0.289^{c}	−0.015	0.004
	(0.156)	(0.045)	(0.056)
Kohlberg class 3	0.105^{c}	0.067^{b}	−0.093
	(0.063)	(0.034)	(0.060)
Kohlberg class 4	−0.118	0.055	0.060
	(0.072)	(0.056)	(0.063)
Kohlberg class 5	0.228^{a}	0.101^{a}	0.028
	(0.071)	(0.035)	(0.057)
Kohlberg class 6	-0.297^{a}	0.000	-0.139^{a}
	(0.073)	(0.033)	(0.052)
Dummy LIE	0.340^{a}	0.519^{a}	
	(0.093)	(0.040)	
Dummy SABOTAGE	-0.135^{c}	0.467^{a}	
	(0.077)	(0.050)	
Risk aversion	-0.058^{b}	0.017	-0.030^{c}
	(0.028)	(0.018)	(0.015)
Envy	0.148^{c}	0.018	−0.043
	(0.080)	(0.065)	(0.078)
Age	0.002	0.000	−0.002
	(0.011)	(0.010)	(0.015)
Gender:female	-0.280^{a}	−0.050	−0.030
	(0.073)	(0.077)	(0.066)
Economics	0.094	0.129	0.620^{a}
Medicine	NA	0.119	NA
Law	0.389^{b}	0.128	0.599^{a}
Social and Behavioral Sciences	0.310^{a}	0.015	0.548^{a}
Sciences	NA	0.192^{c}	0.641^{a}
Philosophy	NA	NA	NA
IT	NA	0.139	NA
Engeneering	0.470^{a}	0.160	0.643^{a}
Languages	NA	−0.053	NA
Education	0.322^{a}	0.026	0.635^{a}
Count R^2	0.89	0.86	0.86

Table 13: Ethical determinants of B participants' departures from rational self-interest, and the type of altruistic behavior they adopt (marginal effects).

Note: Significance levels of z-tests are indicated by $a : p < .01$, $b : p < .05$, $c : p < .10$.

Notes. Logit regressions with robust standard errors of – column 1: S_1 vs rational self-interest, – column 2: S_2 and give all vs rational self-interest, and – column 3: S_1 vs S_2 and give all, on variables displayed for all data with competitive payoffs. Materialism and Postmaterialism Value scores were not collected in treatment SPY where we did not expect any altruism. These variables are therefore left out if all observations are to be used. Controls include all variables from www.chlass.de/Research.html with a link to *Kohlberg class 5* to ensure there is no variable which intercepts the link between *Kohlberg class 5* and *purely procedural preferences* (Chlaß et al. 2019). **Age** – B participants' age in whole years, ranges from 18 to 35 with a median of 23; **Envy** – Dummy taking on a value of One if B allocated 10 ECU to herself and 10 ECU to A, rather than 10 ECU to herself and 20 ECU to A; **Risk aversion** – ordinal variable ranging from 1 and 10; indicates at which lottery on a ten-item Holt-Laury lottery list B starts to prefer the sure payoff of 25 ECU to a binary lottery of 10 ECU to 35 ECU, see section 3.3.; **Fields of study** – miscellaneous category is Field of Study: Arts.

M Improving $A's$ relative position of decision rights: treatment punishment/reward

Dependent Dummy variable →	S_1 (1) VS. SELFISH (0)	S_2 + GIVE ALL (1) VS. SELFISH (0)	FAIR COIN (1) VS. SELFISH (0)
nr. of obs.	57 (8 vs. 49)	119 (70 vs. 49)	68 (19 vs. 49)
Kohlberg class 1	-0.137^a (0.045)	-0.104^c (0.054)	-0.178^a (0.036)
Kohlberg class 2	-0.162^a (0.048)	0.018 (0.071)	0.062 (0.060)
Kohlberg class 3	0.249^a (0.048)	0.045 (0.061)	-0.088^c (0.050)
Kohlberg class 4	0.081 (0.064)	0.046 (0.071)	0.116^a (0.044)
Kohlberg class 5	-0.133 (0.085)	-0.131^b (0.059)	-0.092 (0.068)
Kohlberg class 6	0.057 (0.061)	0.091 (0.059)	0.098^c (0.057)
expected punishment	-0.012^a (0.004)	-0.017^a (0.004)	-0.040^a (0.005)
expected reward	0.007 (0.006)	-0.005 (0.006)	-0.030^a (0.004)
Risk aversion	-0.059 (0.062)	0.009 (0.025)	0.052 (0.039)
Envy	-0.157^b (0.079)	-0.085 (0.083)	0.039 (0.080)
Age	0.008 (0.012)	-0.005 (0.017)	-0.007 (0.018)
Gender:female	0.024 (0.099)	-0.084 (0.095)	-0.239^c (0.128)
Count R^2	0.90	0.67	0.88

Table 14: Ethical determinants of B participants' departures from rational self-interest, and the type of altruistic behavior they adopt (marginal effects).

Note: Significance levels of z-tests are indicated by $a : p < .01$, $b : p < .05$, $c : p < .10$.

Notes. Logit regressions with robust standard errors of – column 1: S_1 vs rational self-interest, – column 2: S_2 and give all vs rational self-interest, and – column 3: S_1 vs S_2 and give all, on variables displayed for all data with competitive payoffs. Materialism and Postmaterialism Value scores were not collected in treatment SPY where we did not expect any altruism. **Expected punishment** – B participants' beliefs by how much A will punish their actual choice of $Prob(S_2)$, ranges from 0 to 30 ECU since A may reduce B's payoff by up to 30 ECU; **Expected reward** – B participants' beliefs by how much A will reward their actual choice of $Prob(S_2)$, ranges from 0 to 30 ECU since A may increase B's payoff by up to 30 ECU. Controls for **fields of study** are left out since full models failed to converge and we deemed the belief data more relevant.

N Taking A's decision rights out of B's hands: treatment payoff neutrality

Note: In treatment payoff neutrality, the rational self-interested choice for B is to toss a fair coin, i.e. to leave Prob$(S_2) = 50\%$ (the default value), and to take all payoff always, since A has no decision rights in either S_1 or S_2.

Dependent Dummy→	S_1 (1) VS. FAIR COIN (0)	S_2 (1) VS. FAIR COIN (0)	S_2 (1) VS. FAIR COIN (0)
nr. of obs.	111 (14 vs. 98)	138 (40 vs. 98)	68 (19 vs. 49)
Kohlberg class 1	-0.107^b (0.049)	-0.112^b (0.051)	-0.178^a (0.036)
Kohlberg class 2	0.122^b (0.053)	0.083 (0.054)	0.062 (0.060)
Kohlberg class 3	0.000 (0.046)	0.048 (0.049)	-0.088^c (0.050)
Kohlberg class 4	0.068 (0.054)	0.054 (0.058)	0.116^a (0.044)
Kohlberg class 5	−0.048 (0.044)	−0.052 (0.050)	−0.092 (0.068)
Kohlberg class 6	−0.051 (0.052)	−0.014 (0.052)	0.098^c (0.057)
Risk aversion	−0.021 (0.022)	−0.027 (0.027)	0.052 (0.039)
Envy	−0.108 (0.064)	−0.072 (0.080)	0.039 (0.080)
Age	−0.012 (0.014)	−0.010 (0.013)	−0.007 (0.018)
Gender:female	−0.045 (0.059)	-0.133^c (0.070)	-0.239^c (0.128)
Count R^2	0.88	0.75	

Table 15: Ethical determinants of B participants' departures from rational self-interest, and the type of altruistic behavior they adopt (marginal effects).

Note: Significance levels of z-tests are indicated by $a : p < .01$, $b : p < .05$, $c : p < .10$.

O How does B expect A to punish or reward B's procedural choice?

Figure A9: B's beliefs about A's decision to punish or reward B's choice of $Prob(S_2)$. Left: payoff neutrality – B cannot impair A's decision rights; Right: competitive payoffs – B impairs A's decision rights in S_2 in treatments LIE and SABOTAGE.

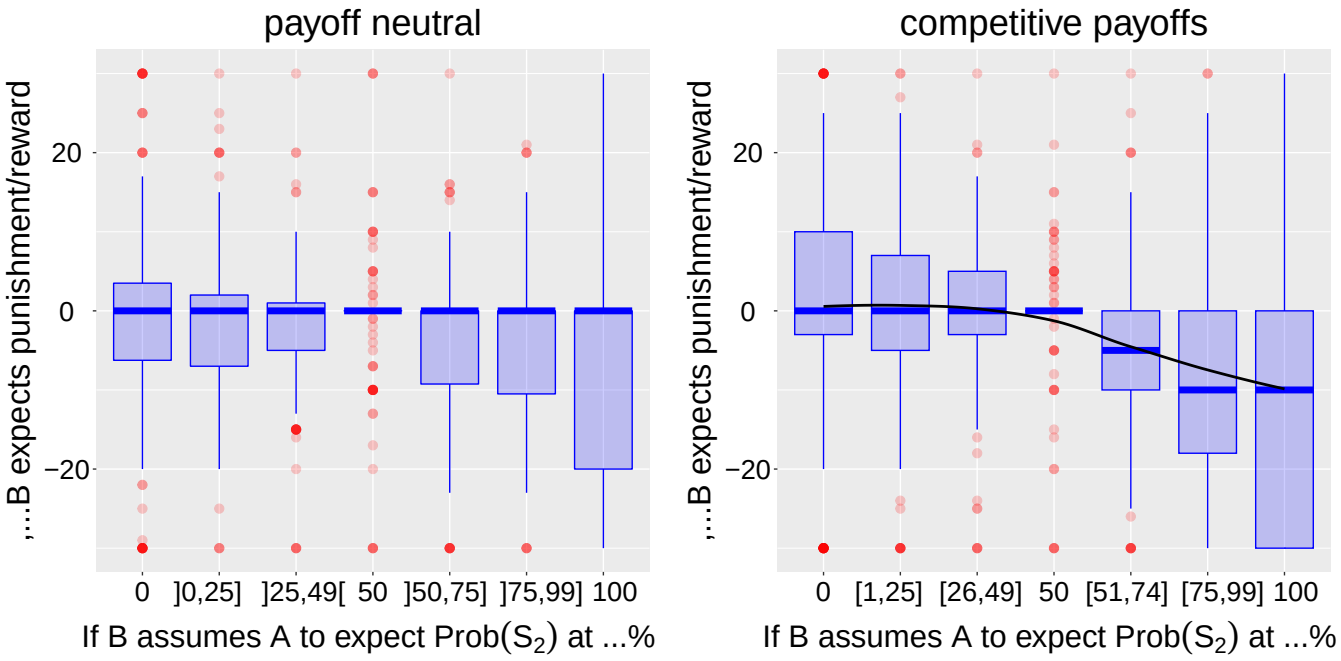


Figure A10: Treatment LIE: B's beliefs about A's decision to punish or reward B's choice of $Prob(S_2)$. Left: payoff neutrality – B cannot impair A's decision rights; Right: competitive payoffs – B impairs A's decision rights in S_2.

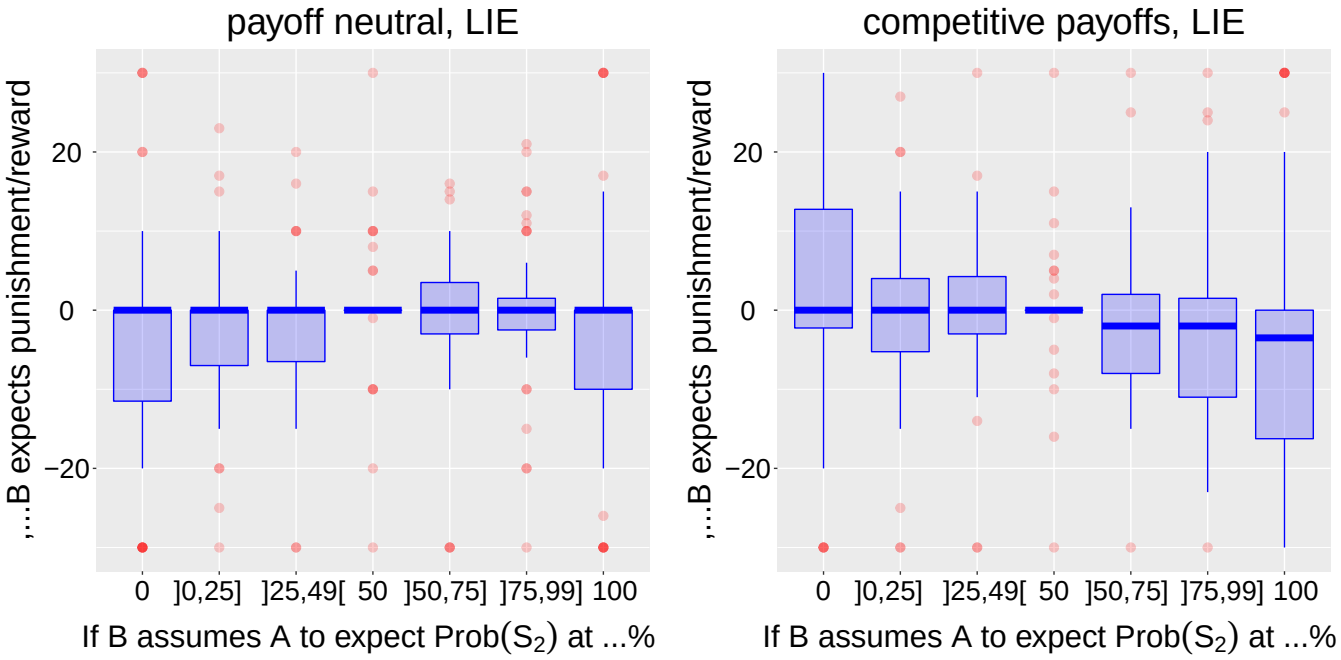


Figure A11: Treatment SPY: B's beliefs about A's decision to punish or reward B's choice of $Prob(S_2)$. Left: payoff neutrality – B cannot impair A's information or decision rights; Right: competitive payoffs – B impairs A information (but not her decision) rights.

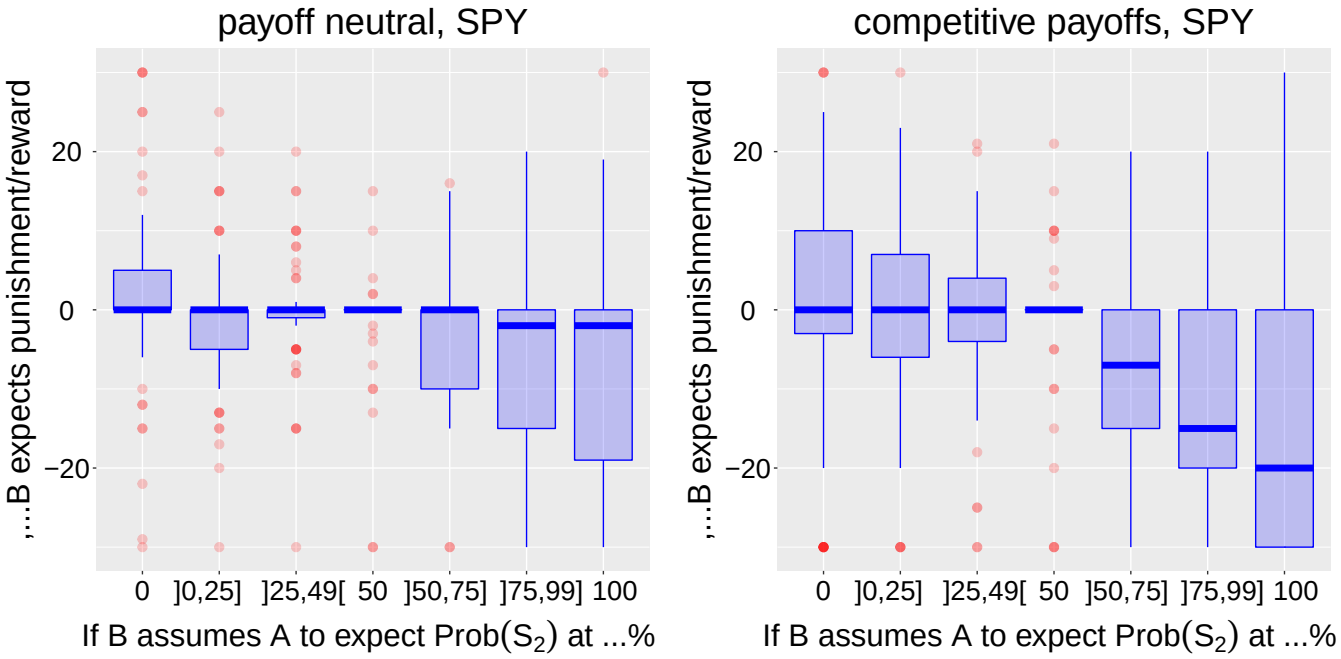
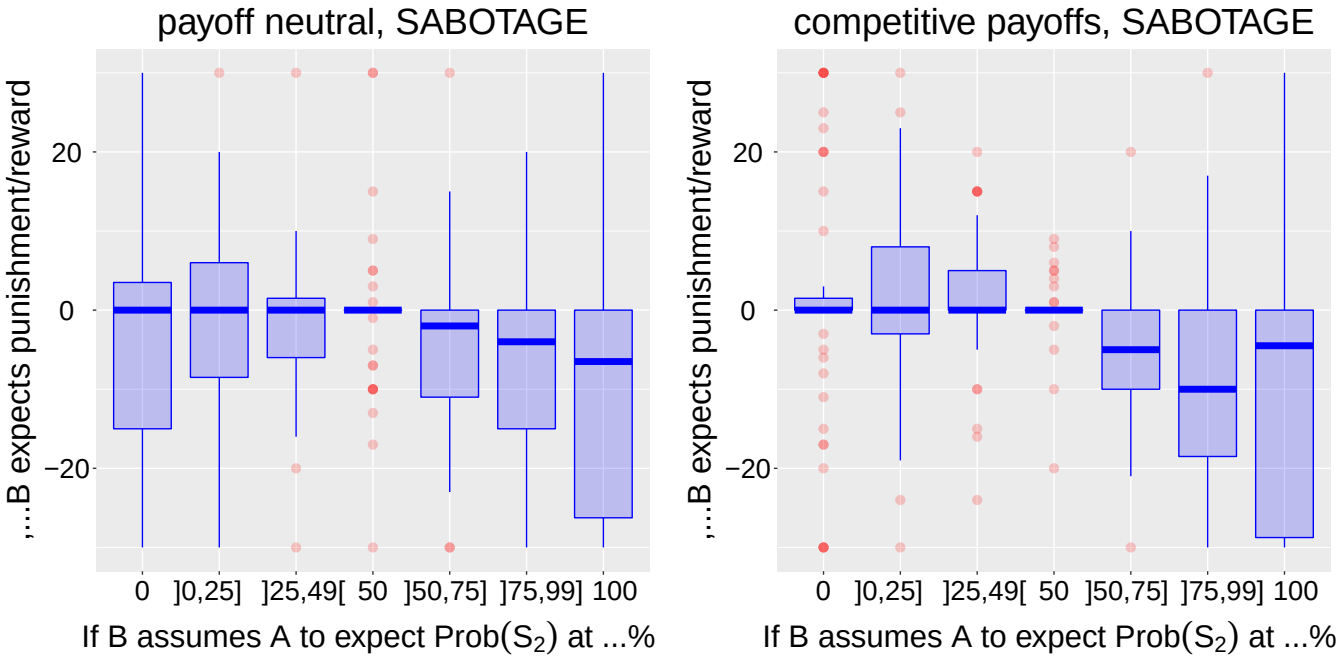


Figure A12: Treatment SABOTAGE: B's beliefs about A's decision to punish or reward B's choice of $Prob(S_2)$. Left: payoff neutrality – B cannot impair A's decision rights; RIGHT: competitive payoffs – B impairs A's decision rights in S_2..



P B's procedural choice, $Prob(S_2)$, and choice of allocation by treatment

Figure A13: B's choice of Prob(S_2) and the allocation she chooses in S_2. LEFT: competitive payoffs – A has no decision rights in S_2; RIGHT: competitive payoffs with punishment/reward – A can punish B for Prob(S_2).

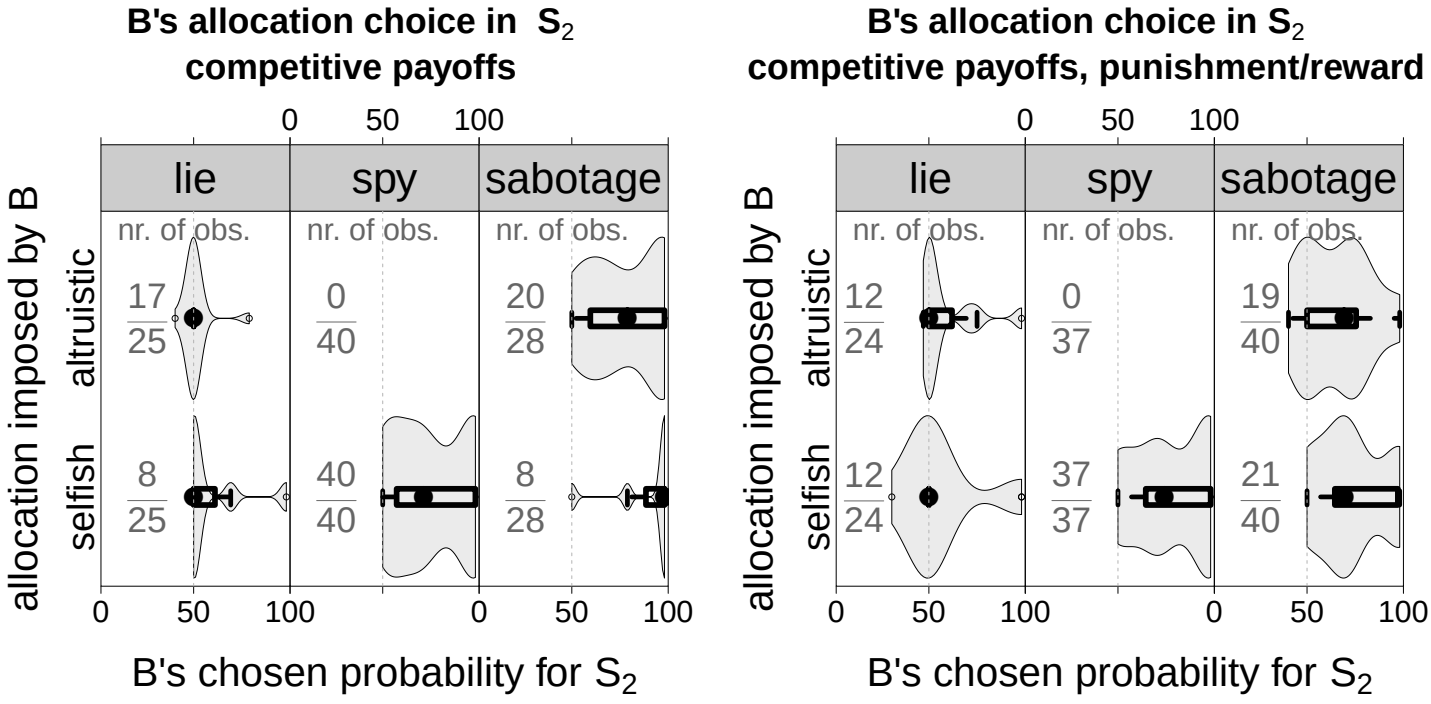


Figure A14: B's choice of Prob(S_2) and the allocation she chooses. LEFT: payoff neutrality [A has no decision rights in either S_1 or S_2]; RIGHT: payoff neutrality with punishment/reward [A has the same rights to punish and reward in S_1 and S_2].

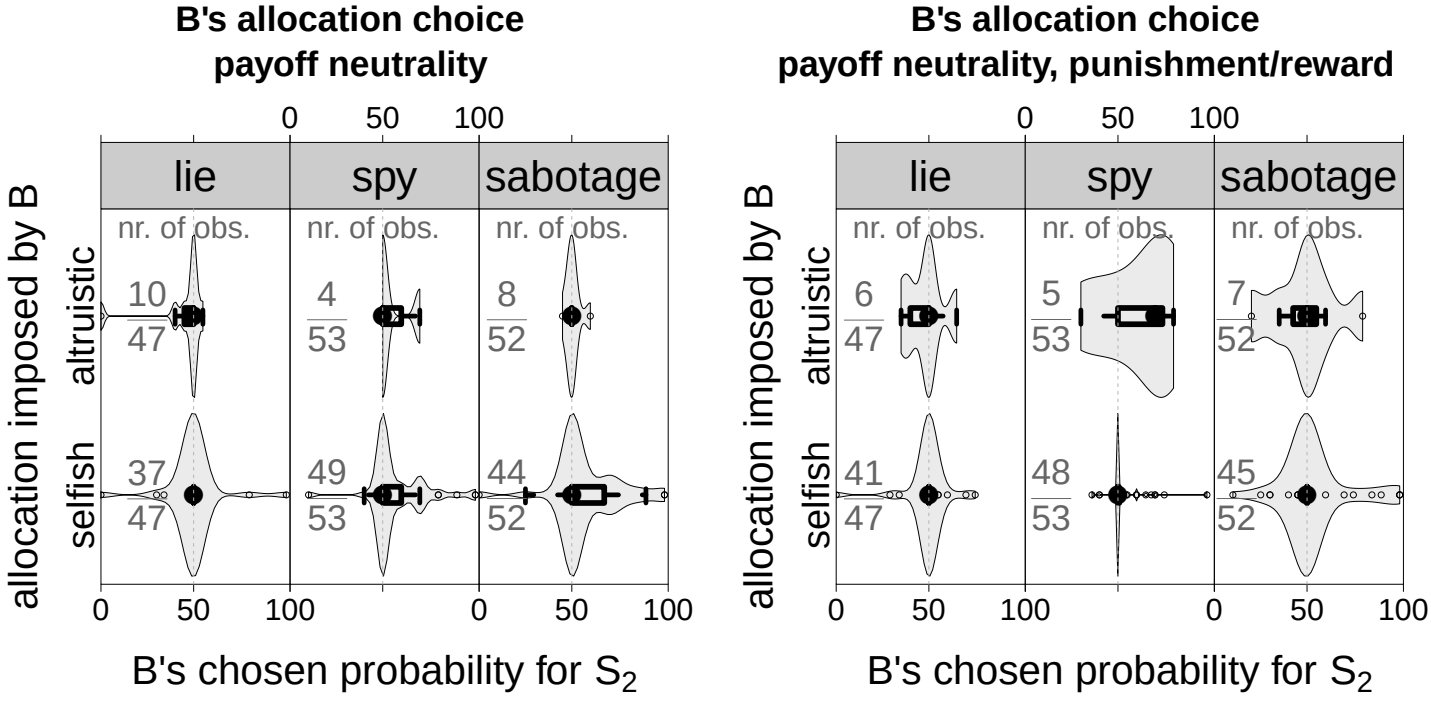


Figure A15: B's choice of Prob(S_2) and the allocation she chooses under payoff neutrality [A has no decision rights either in S_1 or S_2]; LEFT: S_1. RIGHT: S_2.

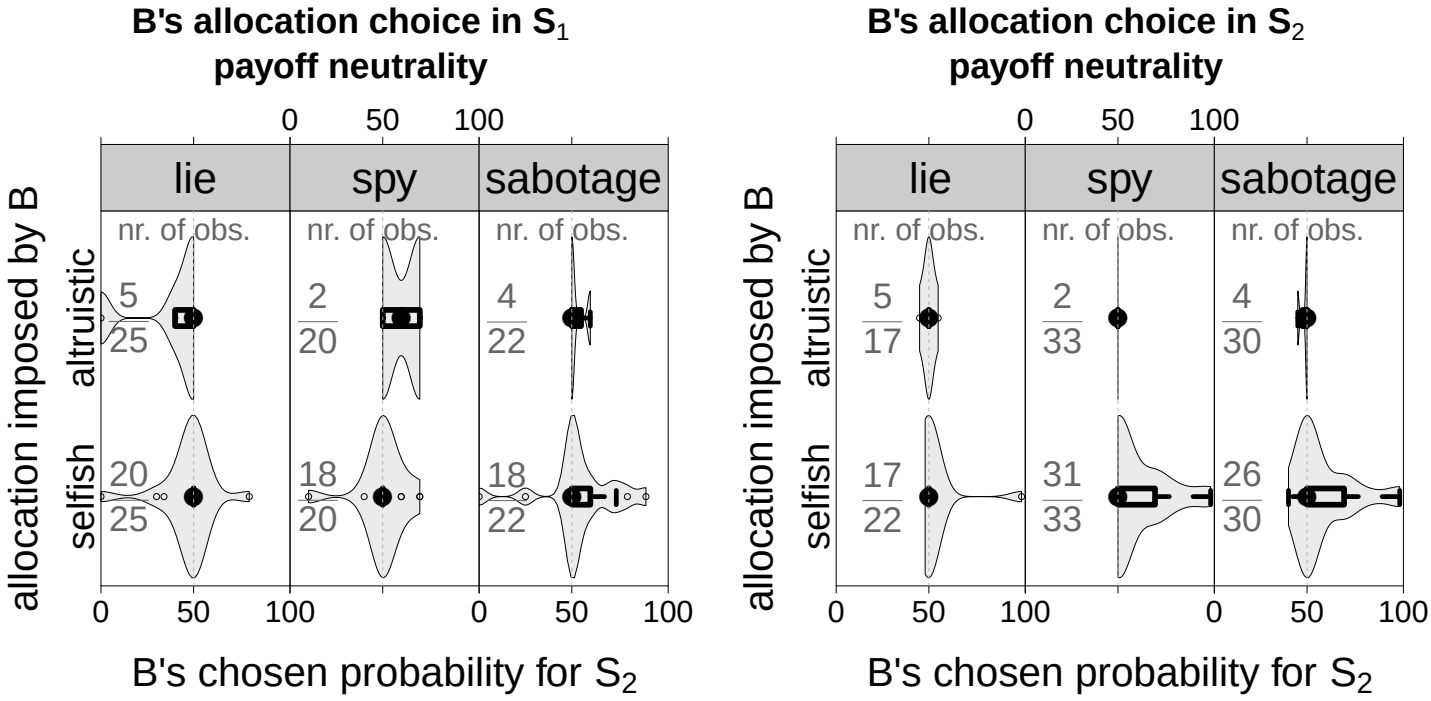
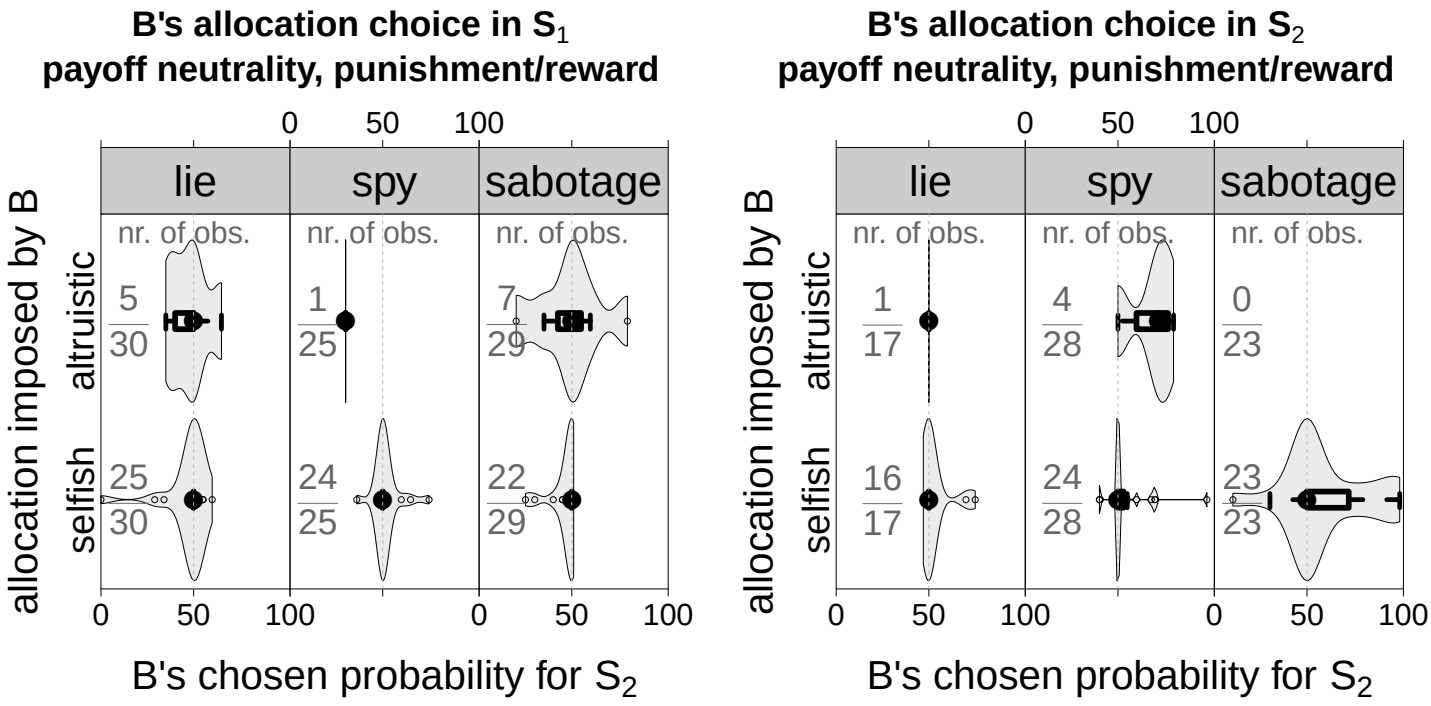


Figure A16: B's choice of Prob(S_2) and the allocation she chooses under payoff neutrality with punishment/reward [A has the same decision rights to punish and reward in S_1 and S_2]; LEFT: S_1. RIGHT: S_2.



Q Predictions: payoff neutral treatment

			BEHAVIOURAL PREDICTIONS						
			LIE		SPY		SABOTAGE		
			make selfish proposal in S_1	make selfish proposal in S_2	make selfish proposal in S_1	make selfish proposal in S_2	make selfish proposal in S_1	make selfish proposal in S_2	**same outcomes across LIE, SPY, SABOTAGE**
SOCIAL PREFERENCE MODELS	Outcome based	Self Interest	+	+	+	+	+	+	+
		Inequity Aversion	+	+	+	+	+	+	+
		Altruism	depends on degree of altruism	depends on degree of altruism	depends on degree of altruism	depends on degree of altruism	depends on degree of altruism	depends on degree of altruism	+
	Reciprocity - based	Falk & Fischbacher (2006)	+	+	+	+	+	+	+
		Dufwenberg& Kirchsteiger (2004)	+	+	+	+	+	+	+
	Guilt based	Battigalli & Dufwenberg (2007)	depends on sensitivity to guilt	depends on sensitivity to guilt	depends on sensitivity to guilt	depends on sensitivity to guilt	depends on sensitivity to guilt	depends on sensitivity to guilt	+
OUTCOME-BASED PROCEDURAL FAIRNESS MODELS	Inequity based	e.g. Bolton et al. (2005)	+	+	+	+	+	+	+
	Reciprocity - based	Sebald (2010)	+	+	+	+	+	+	+
PURELY PROCEDURAL FAIRNESS MODELS	equal decision rights	Chlaß et al. (2019)	+	+	+	+	+	+	+
	equal information rights	Chlaß et al. (2019)	+	+	depends on sensitivity to unequal information	depends on sensitivity to unequal information	+	+	−

Notes. *1) Inequity aversion.* Denote B's earnings by x, and $A's$ earnings by y. An inequity averse B has utility $x - a \cdot \max\{(y-x), 0\} - b \cdot \max\{(x-y), 0\}$ where $a, a \leq 4$ and $b, b \leq 1$ are non-negative individual parameters. Allocation ($x = 100$, $y = 0$) yields B utility $100 - b \cdot 100$, allocation ($x = 0$, $y = 100$) utility $-a \cdot 100$, respectively. An inequity averse B with $b \leq 1$ therefore always prefers ($x = 100$, $y = 0$) over ($x = 0$, $y = 100$).

2a) Reciprocity, Falk and Fischbacher (2006). B chooses between an intentionally weakly kind, i.e. $(x = 0, y = 100)$, and an intentionally selfish (unkind) allocation $(x = 100, y = 0)$. A has no decision rights at all; she cannot reject (must accept) all allocations, can hence not be unkind to B, and B need hence not be kind to induce kindness. B therefore chooses the selfish allocation $(x = 100, y = 0)$.

2b) Dufwenberg and Kirchsteiger (2004). There are only efficient strategies in the game (no strategy destroys the pie). Since A cannot reject (must accept) all allocations, she cannot be unkind to B, and B therefore chooses the selfish allocation $(x = 100, y = 0)$.

3) Guilt aversion (Battigalli and Dufwenberg (2007)). Guilt matters only if B harms A and lets her down (disappoints A's expectations). A cannot harm B and her guilt payoff is therefore always Zero. A very guilt averse B who very much expects A to expect the generous allocation, might indeed offer $(x = 0, y = 100)$. As long, however, as B's beliefs about A's payoff expectations are identical in S_1 and S_2, B makes the same choice in both situations. Looking at B's empirical punishment expectations, B

players expect A players to expect identical payoffs in S_1 and S_2.

4) Preferences for equal expected payoffs (Bolton and Ockenfels (2005). Ex-ante, B chooses between S_1 where she opts for allocation $(x = 100, y = 0)$ for sure, and S_2 where she also opts for allocation $(x = 100, y = 0)$ for sure. She therefore has no choice between more or less equal expected payoffs, the expected payoffs are degenerate in each situation, and she is indifferent between S_1 and S_2.

5) Preferences for kind procedures (Sebald 2010). Since A cannot reciprocate in either S_1 or S_2, B therefore chooses the selfish allocation $(x = 100, y = 0)$ always and S_1 and S_2 are therefore equally unkind, B is indifferent between S_1 and S_2.

6a) Purely procedural preferences for equal decision rights (Chlaß et al. 2019). In S_1 and in S_2, for any contingency of the game – that is, whether B chooses either L, or R, A's choices always leave her equally well off: if B chooses L, A's choices L and R both yield her Zero payoff and hence, she cannot prefer L over R or vice versa; if B chooses R, A's choices both yield her 100 ECU and hence, she cannot prefer L over R or vice versa either. She therefore has no decision rights and cannot look after her own self-interest. Therefore, it is not within B's power to either grant A, or impair the latter's decision rights. B is therefore indifferent between S_1 and S_2 and chooses $(x = 100, y = 0)$.

6b) Purely procedural preferences for equal information rights (Chlaß et al. 2019). In S_1, neither A nor B knows which action the opponent has chosen. Only B knows that the interaction structure is S_1. B can therefore distinguish two out of the four terminal nodes of the game. The same holds for S_2 in treatments LIE and SABOTAGE. A's cardinality over the terminal nodes of S_1 and S_2 is always One, since she does not know the interaction structure. In LIE and SABOTAGE therefore, B has no power to either grant A, or impair the latter's information rights. This does not hold for treatment SPY where in S_2, B knows A's choice, but A does not know how B has chosen and B can therefore distinguish all four terminal nodes of the game. If B dislikes having greater information rights than A in SPY, she prefers S_1 over S_2, or chooses S_2 and compensates A by opting for allocation $(x = 0, y = 100)$. If to the contrary, B prefers greater information rights, she prefers S_2 over S_1, and opts for allocation $(x = 100, y = 0)$. Note that B never takes away information rights from A; she always improves her own relative position in information rights. Note, too, that it is not within her power to grant A exactly *equal* information rights.

R Treatments: Overview

purely procedural aspects	competitive payoffs[37]			payoff neutrality			competitive punish/reward[38]			payoff neutral punish/reward		
	LIE	SPY	SAB	LIE	SPY	SAB	LIE	SPY	SAB	LIE	SPY	SAB
A has decision rights	+	+	+	-	-	-	+	+	+	+	+	+
B can take some of A′s decision rights	+	-	+	-	-	-	+	-	+	-	-	-
B can take all A′s decision rights	+	-	+	-	-	-	-	-	-	-	-	-
wording of instructions is identical between treatments:	⟷						⟷					
		⟷						⟷				
			⟷						⟷			

[37] LIE and SAB competitive: B grants A exactly equal decision rights in S_1 and in S_2, grants herself *greater* decision rights than A.
[38] LIE and SAB competitive punish/reward: B grants A *greater* decision rights in S_1 and in S_2, grants A *lesser* decision rights.