

University of Mannheim / Department of Economics

Working Paper Series

---

***Lying, Spying, Sabotaging:  
Procedures and Consequences***

Nadine Chlaß      Gerhard Riener

Working Paper 15-17

November 2015

---

---

# Lying, Spying, Sabotaging: Procedures and Consequences

Nadine Chlaß\* and Gerhard Riener‡

November 23, 2015

## Abstract

Do individuals prefer to compete fairly, or unfairly with an opponent? We study individuals who can choose how to compete for one ex-post nonzero payoff. They can either nudge themselves into a fair set of rules where they have the same information and actions as their opponent, or into unfair rules where they spy, sabotage or fabricate their opponent's action. In an experiment, we observe significant altruism under rules which allow for fabrication and sabotage, but not under rules which allow for spying. We provide direct evidence that this altruism emanates from an ethical concern purely about the rules of the game. How individuals deal with this concern – whether they nudge themselves into fabrication-free, spying-free, or sabotage-free rules, or whether they assume the power to fabricate or sabotage to compensate their opponent by giving all payoff away – varies along with individuals' attitudes towards power.

**JEL:** D02,D03, D63,D64

**Keywords:** psychological games, moral judgement, institutional design, lying aversion, sabotage aversion, spying aversion, unfair competition

---

\*Departments of Economics, University of Jena, Carl-Zeiss-Str. 3, D-07743, Germany, and University of Turku, FI-20014 Turku, Finland. Email: nadine.chlass@utu.fi. *Chlaß gratefully acknowledges financial support from Leverhulme Visiting Fellowship DKA 7200.*

‡Department of Economics, University of Mannheim and DICE, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 45372 Düsseldorf, Germany. Email: riener@mail.uni-mannheim.de. We thank participants of research seminars at the Universities of Cologne, Düsseldorf, Jena, Mannheim, and Turku for comments and discussions. Markus Prasser and Martin Schneider helped conducting the experiment. The project 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 preemptively 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.<sup>1</sup> 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, MacAskill, and Poitras 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 (Cooper and Roberts 2011; Sheehan 1971).

Lying, spying, and sabotaging are of fundamental relevance to economics. Market 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 fabricate information about her solvency to an investor, or to spy her business secrets to improve their competitive situation. If opportunities to do so arise fairly evenly in the market and all market participants unanimously exploit them, competition may still serve societal welfare. Yet, if some agents systematically resort to such activities while others do not, the self-regulating behaviour of the market place – Smith’s invisible hand – is at stake.

Lying, spying, and sabotaging therefore share the common aspect that they erode the nature and welfare implications of competition. If firms who compete spy on and sabotage each other or their customers, fabricate and plant rumours on tax non-compliance or financial difficulties of their competitors<sup>2</sup>, competition may ultimately select the most ruthless, but no longer the most cost-efficient or innovative market agent. A tournament incentive scheme implemented to detect and qualify high ability employees for a promotion, may no longer fulfil that purpose if employees spy, lie, or sabotage. Therefore, central economic concepts seem to rely on the idea that self-interested agents compete equally (un)fairly – or that market institutions successfully preclude unfair competition.

At the same time, fabrication-, 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

---

<sup>1</sup>Comments by NSA officials do not deny these activities and state they are ‘hardly surprising’ (Larson, Perlroth, and Shane 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.

<sup>2</sup>In 2002, Deutsche Bank head Breuer’s public statement about Kirch media group’s creditworthiness partly led to the group’s bankruptcy a month later. Deutsche Bank was sued for have planned on bringing down the group to profit from their customer’s insolvency. Several top managers were convicted of obstructing the course of justice by giving false evidence during the lawsuit, and in 2015, the bank paid 775 million € in settlement.

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<sup>3</sup>, employees who develop or maintain software for cyber-security seek to exploit weaknesses in firms’ or nations’ security systems. Little is known about how individuals react to the nature of such work. The introductory examples imply that, even after self-selecting to a workplace, people’s reactions differ.

Our paper studies this heterogeneity. Which psychological cost – if any – do fabrication, spying, and sabotaging induce, and what is the nature and source of this cost? Which behavioural strategies do individuals employ to deal with it? Since field data are scarce given the secretive nature of the activities at hand, we construct a laboratory experiment. In a *fabrication game*, individuals fabricate and submit – unbeknownst to their opponent – information about the opponent’s decision to a third party who validates both parties’ decisions. In a *spying game*, individuals *look up* their opponent’s decision – unbeknownst to their opponent. In a *sabotaging game*, individuals override – unbeknownst to their opponent – the opponent’s choice. Throughout, spying, fabrication, and sabotage enable the individual to implement either a selfish, or an altruistic payoff allocation. Individuals can also opt out of these activities and nudge themselves (Thaler and Sunstein 2008) into a fabrication-free, spy-free, or sabotage-free interaction with their opponent.

The *purposes* of fabrication, 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 and suffer a psychological cost.

To see whether such concerns are at play in our *fabrication*, *spying*, and *sabotaging* games, we elicit how individuals make moral judgements, that is, how they typically arrive at the conclusion that an action is either right or wrong. Relying upon Piaget’s (1948) and Lawrence Kohlberg’s (1969; 1984) field research and G. Lind’s (1978; 2008; 2002) methodological work, we elicit which moral ideals individuals employ when judging the right or wrong of an action, and use these to model the behaviour we observe.

We find three types of behaviour. A first type complies with rational self interest: she *fabricates*, *spies*, and *sabotages* to take all payoff. A second – ‘procedural’ – type avoids either activity and nudges herself into a fair set of rules. A third – ‘compensatory’ – type opts into these activities to give all payoff away. The shares of these types differ substantially depending on whether unfair competition involves fabrication, sabotage, or spying. The ‘procedural’ type is most prevalent in the fabrication game. The ‘compensatory’ type occurs most often in the sabotaging game, and *never* in the spying game. Looking into the ethical ideals which these

---

<sup>3</sup>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.

types invoke when judging whether an action is right or wrong, we observe that individuals are the more likely to be of the procedural and also of the compensatory type, the more they refer to individual rights and the democratic social contract when making a moral judgement. Therefore, both departures from rational self-interest emanate from the same ethical ideal, and the foregone payoff is linked to the infringement of the opponent's unprotected decision and information rights.<sup>4</sup>

Why would some individuals who are concerned about an opponent's rights opt out of fabrication, spying, or sabotage while others expressly use these activities to give all payoff away? The 'procedural' type reinstates her opponent's information and decision rights while the 'compensatory' type trades these rights off against a monetary compensation. We speculate that the 'procedural' type may have scruples against exerting power – be it to whatever ends – as opposed to the 'compensatory' type who assumes power and gives all payoff away. To test this idea, we classify individuals on a sociological taxonomy of *materialists* who value hierarchy, duty, and power, and *postmaterialists* who value individuality, the emancipation from authorities, and autonomy (Baker and Inglehart 2000; Inglehart 1977; Klages and Gensicke 2006). Indeed, 'procedural' types score significantly higher on postmaterialist values than 'compensatory' types and 'compensatory' types score significantly higher on materialist values. These values seem to govern how individuals who deem that fabrication, spying, or sabotage infringe a second party's unprotected rights, 'correct' or 'compensate' this infringement of their ethical ideals behaviourally.

Our results imply several challenges for the procedural design of organizations. Rules and processes which allow or require fabrication, spying, and sabotage may severely deplete individuals' work motivation, effort and productivity, endanger team cohesion and employees' psychological health, cause absenteeism and can severely affect the success of an organisation (Carpenter, Matthews, and Schirm 2010; Korsgaard, Schweiger, and H. J. Sapienza 1995). Our paper shows that not only the victims of these activities, but also the people who carry them out suffer from doing so. Similarly, national or regional cultures with traditions and norms that foster or do not prevent unfair competition, may hinder an efficient market and the economic development of entire countries (Guiso, P. Sapienza, and Zingales 2006; Tabellini 2010). In the light of our findings, firms do – before introducing even weak competitive incentives – need to design and implement institutions which effectively prevent unfair competition. Since control is usually imperfect, this goal is not easily achieved.

This paper provides a comparative study of fabrication, spying, and sabotaging, of the behavioural reactions these activities induce, and a detailed investigation into the sources of

---

<sup>4</sup>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 Güth and Miettinen's (2009) *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 moral criteria (ideals) 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.

these reactions. Our study links in particular to the literature on selfish black lies which harm others, and on altruistic white lies which benefit others (Erat and Gneezy 2012)<sup>5</sup>. This literature currently focuses on a controversy about why people are lying averse. Is truth-telling a focal point for intuitive decision makers (Cappelen, Sørensen, and Tungodden 2013; Lightle 2014) who do not understand the monetary benefits from lying? Is lie aversion disguised 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 (Erat and Gneezy 2012; Gneezy 2005; Miettinen 2013)? Could guilt aversion, i.e. an aversion against disappointing others' expectations trigger this psychological cost (Battigalli, Charness, and Dufwenberg 2013), or is there pure lie aversion which does not depend on expectations or consequences at all (Hurkens and Kartik 2009; López-Pérez and Spiegelman 2013)? In our setting, we find the latter and provide evidence that lie aversion emanates from a more general preference about the equality of rights.

Sabotaging has received less attention, mainly in the framework of tournament games, in effort choice, or real effort games. Therein, individuals can increase others' costs of effort, destroy others' outcomes, or manipulate others' performance evaluations. Sabotaging becomes the more frequent, the higher the monetary benefit entailed (Harbring and Irlenbusch 2011) and decreases if an explicit label emphasizes the nature of the activity. This raises the question whether i) similarly to lying, sabotage is sensitive to outcomes and induces ii) some psychological cost which individuals trade off against the gains. At the same time, other studies document an intrinsic pleasure to sabotage, e.g. (Abbink and Herrmann 2011; Abbink and Sadrieh 2009; Charness, Masclet, and Villeval 2014). Spying has, so far, hardly been studied at all – despite the massive media coverage in the aftermath of the NSA leaks, and recurrent public debates on privacy, information security and data protection in our highly digitalized life<sup>6</sup>. This paper studies fabrication, spying, and sabotage in a unifying framework which allows individuals to assess either activity in terms of its consequences but also allows them to avoid these activities entirely if they are felt to be innately wrong. Throughout, we find that individuals who nurture ethical ideals about the equality of *rights* – not payoffs – derive disutility from competing unfairly with their opponent. This ideal has not yet been discussed in the context of lie and sabotage aversion, or fair competition.

In the next section we illustrate our setup, section 3 outlines our experimental design in more detail. Section 4 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.

---

<sup>5</sup>Another strand of research (Abeler, Becker, and Falk 2014; Fischbacher and Föllmi-Heusi 2013; Gibson, Tanner, and Wagner 2013) studies lies which do not affect the opponent's payoff and only harm the experimenter. The authors document both payoff-dependent and payoff-independent (pure) lie aversion. Fischbacher and Föllmi-Heusi (2013) compare lying which harms the experimenter or the opponent and find no difference. Conrads et al. (2014) study parties who can lie about their own production costs in a tournament.

<sup>6</sup>Solan and Yariv (2004) study the cost of information acquisition on spying activities in a theoretical model assuming expected payoff maximization. In another context, Whitfield (2002) and Milinski and Rockenbach (2007) show that spying might be pervasive in the mammal world for evolutionary reasons, i.e. type detection.

## 2 Lying, spying, and sabotaging: rules and payoffs

This section briefly illustrates which notions and payoff consequences of fabrication, 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 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<sup>7</sup>, 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			1b) the ‘unfair’ set of rules				
		party A			party A		
		$L$	$R$		$L$		
party B	$L$	0	100	party B	$LL^A$	100	100
	$R$	100	0		$RL^A$	0	0
		0	100		$LR^A$	100	100
		100	0		$RR^A$	100	100
		100	0		0	0	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<sup>8</sup>.

Second,  $B$  can opt for a set of rules where she *sabotages*  $A$ , that is, replaces  $A$ ’s decision

<sup>7</sup>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.

<sup>8</sup>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, Güth, and Miettinen 2009).

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, Irlenbusch, et al. 2007), as directly reducing others' output (Harbring and Irlenbusch 2011), as destroying others' output (Falk, Fehr, and Huffman 2008), or as manipulating how others' output performance is evaluated (Carpenter, Matthews, and Schirm 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 D. When  $B$  sabotages, she does not necessarily acquire information about 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  $R \approx L$ . If  $R$  and  $L$  are identical then  $A$  does not prefer choice set  $\{L, R\}$  to choice set  $\emptyset$  in 1b). In 1a), however,  $R \approx L$  in some circumstances and hence  $A$  may prefer  $\{L, R\}$  to  $\emptyset$ . 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 C. 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.



Treatment	<i>Spy</i>		<i>Sabotage</i>		<i>Lie</i>	
Payoff regime	Neutral	Competitive	Neutral	Competitive	Neutral	Competitive
B-participants	# 52	# 54	# 53	# 53	# 47	# 44
<b>Part 1</b>	<b>Baseline</b>					
	B chooses probability $\alpha$ of interaction structure $S_2$					
	A chooses her strategy					
	In $S_2$ , B learns A's strategy		In $S_2$ , B overrules A's strategy		In $S_2$ , B reports A's strategy to C	
			B chooses her own strategy			
<b>Part 2</b>	<b>Reward and Punishment</b>					
	B chooses probability $\alpha$ of interaction structure					
	A chooses her strategy					
	In $S_2$ , B learns A's strategy		In $S_2$ , B overrules A's strategy		In $S_2$ , B reports A's strategy to C	
			B chooses her own strategy			
	A chooses punishment/reward schedule without knowing $\alpha$ , the situation, or B's strategy					
	B submits 1st order beliefs about punishment and reward schedule.					
<b>Part 3</b>	<b>Covariates</b>					
	Risk Aversion					
	Envy					
	Moral Judgement Test (pen and paper)					
	Materialist and Postmaterialist values (pen and paper)					
	Demographics					

**Table 2:** EXPERIMENTAL DESIGN

### 3 Experimental Setup

The experimental design consists of three parts in each session. For each part, new instructions were shown on screen.<sup>9</sup> In part 1, there are two parties  $A$  and  $B$ , and  $B$  chooses between a 'fair' ( $S_1$ ), and an 'unfair' ( $S_2$ ) interaction structure at her own discretion. Part 1 also elicits  $A$ 's and  $B$ 's behaviour within the chosen interaction structure. Part 2 proceeds the same way except that  $A$  now has a symbolic punishment and reward option to express her (dis)agreement with  $B$ 's potential choices of the interaction structure. Part 3 elicits a variety of preferences, values, and demographics to better understand the nature of individuals' decisions. We describe parts 1 and 3 in more detail below, and explore part 2 in a companion paper (Chlaß and Riener 2015). Only one of the first two parts was paid out, part 3 was always paid, and no feedback was given in between parts. Table 2 summarizes this paper's  $3 \times 2$  between subjects factorial design which studies three pairs of 'unfair' and 'fair' interaction structures under two payoff regimes. It was common knowledge that the experiment proceeded in a *perfect stranger design*. All sessions were roughly balanced on gender.

#### 3.1 Part 1: Choosing between two Sets of Rules

Figure 1 describes the structure of part 1 in all experimental sessions<sup>10</sup>. There are two players labelled  $A$  and  $B$  who have an initial endowment of 50 ECU (€ 2.50). At the root of the game

<sup>9</sup>In a given part, participants had no information about the contents of potentially upcoming parts. Instructions are available from the authors upon request.

<sup>10</sup>Appendix A provides participant  $B$ 's decision screen for her choice of the interaction structure, and her decision screen in  $S_2$  for treatments SPY, SABOTAGE, and LIE.

tree, player  $B$  always chooses how she wishes to interact with player  $A$ . More particularly,  $B$  chooses the probability  $\alpha$  that the 'fair' interaction structure  $S_1$  occurs rather than the 'unfair' interaction structure  $S_2$ . This likelihood  $\alpha$  has a default value of  $\alpha = 50\%$ . If  $B$  wishes to change  $\alpha$ , she incurs cost  $c(\alpha) = 0.1 \cdot |50 - \alpha|$  ECU<sup>11</sup> which is deducted from her payoff. Hence, player  $B$  can make one interaction structure certain at the relatively small cost of 5 ECU (25 Euro Cents). Once  $B$  has submitted her choice of  $\alpha$ , chance draws the interaction structure accordingly. Player  $A$  neither knows  $B$ 's choice of  $\alpha$ , nor the actual interaction structure which is drawn. She always chooses between left ( $L$ ), right ( $R$ ), and the toss of a fair coin between the two. Only player  $B$ 's choices depend on the actual interaction structure which is drawn. If interaction structure  $S_1$  occurs,  $A$  and  $B$  play a constant sum game, but only  $B$  knows it. In  $S_1$ ,  $B$  has the same choices as player  $A$  – namely ( $L$ ), ( $R$ ), and the fair coin – and neither player has information about her opponent's choice. Interaction structure  $S_1$  is the same in all treatments. Interaction structure  $S_2$ , however, differs across treatments. In treatment SPY, interaction structure  $S_2$  grants  $B$  *information about player A's choice* but is otherwise identical to constant sum game  $S_1$ . In treatment SABOTAGE,  $S_2$  grants  $B$  the option to *replace player A's choice* and is otherwise identical to  $S_1$ . In treatment LIE finally,  $S_2$  requires that player  $B$  *report her own choice, and a choice for A* – without actually knowing  $A$ 's choice – to player  $C$  who confirms the reported choices and makes them payoff-relevant. Apart from  $B$ 's reporting,  $S_2$  is identical to  $S_1$ .

Whereas the *rules* of the 'unfair' interaction structure  $S_2$  differ across our three treatments,  $B$  always faces the same allocation choices in  $S_2$ . In all treatments,  $B$  can exploit her privilege in  $S_2$  to obtain the exact same material advantage. We vary the size of this material advantage later on in two payoff regimes, see section 3.2. To summarize the different treatments:

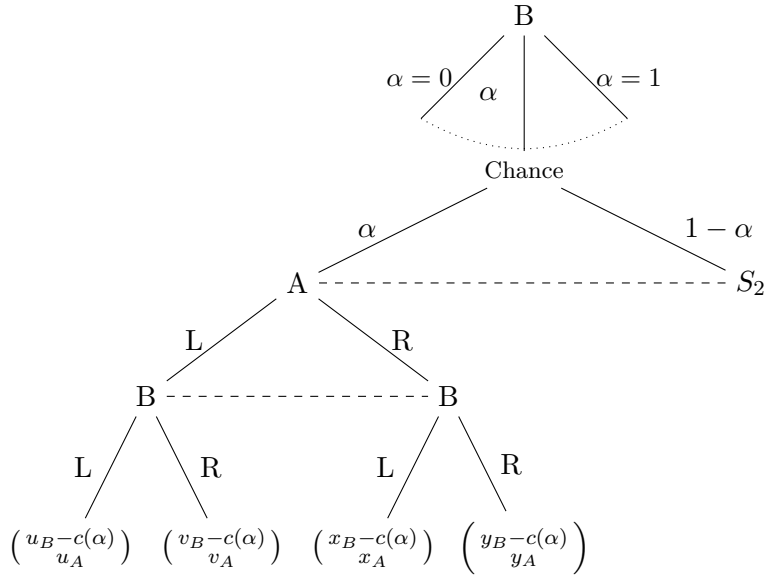
**Treatment SPY** In the spying treatment, interaction structure  $S_2$  was designed such that  $B$  sees player  $A$ 's choice and can therefore condition her decision on  $A$ 's choice whereas  $A$  does not know  $B$ 's choice, see Figure 2. If  $B$  chooses  $S_2$ , we say that she decides to spy on  $A$  since  $B$  acquires information about  $A$  unbeknownst to the latter.  $B$  grants herself a privilege in information about  $A$ 's choice.

**Treatment SABOTAGE** In the sabotaging treatment,  $S_2$  was designed such that player  $B$  cannot see  $A$ 's choice. However,  $B$  must set  $A$ 's decision to either  $L$  or  $R$ , and thereby "override"  $A$ 's choice thus making  $A$ 's choice payoff-irrelevant. In choosing  $S_2$ , we say that  $B$  decides to sabotage  $A$  because  $B$  decides to impair  $A$ 's freedom of choice unbeknownst to  $A$ <sup>12</sup>. In Figure 3, a replaced action is denoted by superscript  $A$ , e.g.  $LR^A$  means that player  $B$  chooses  $L$  herself and sets  $A$ 's choice to  $R$ .

---

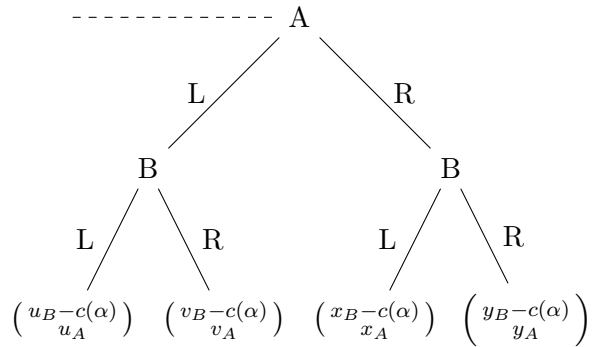
<sup>11</sup>This corresponds to 0.1 ECU for each percentage point by which  $B$  changes the default probability  $\alpha = \text{Prob}(S_1) = 50\%$  where 1 ECU = 0.05 Euro Cents.

<sup>12</sup>Another way to describe this paper's sabotaging notion is that unbeknownst to  $A$ ,  $B$  changes the meaning of  $A$ 's actions, or redefines the relation between  $A$ 's actions and their outcomes. Appendix D shows Busch's

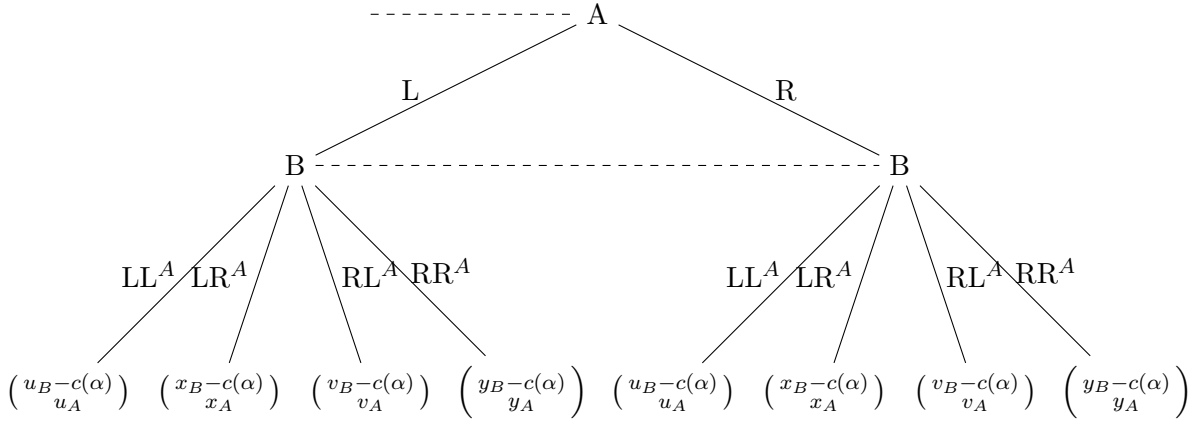


**Figure 1: BASIC GAME STRUCTURE**

*Note:* This tree illustrates the structure of part 1 (see table 2) for all treatments.  $S_2$  is a place holder for the 'unfair' interaction structure  $S_2$  which differs across treatments LIE SPY and SABOTAGE, see Figs. 2 (SPY), 3 (SABOTAGE), and 4 (LIE).



**Figure 2: THE 'UNFAIR' INTERACTION STRUCTURE  $S_2$  IN TREATMENT SPY.**



**Figure 3:** THE 'UNFAIR' INTERACTION STRUCTURE  $S_2$  IN TREATMENT SABOTAGE.

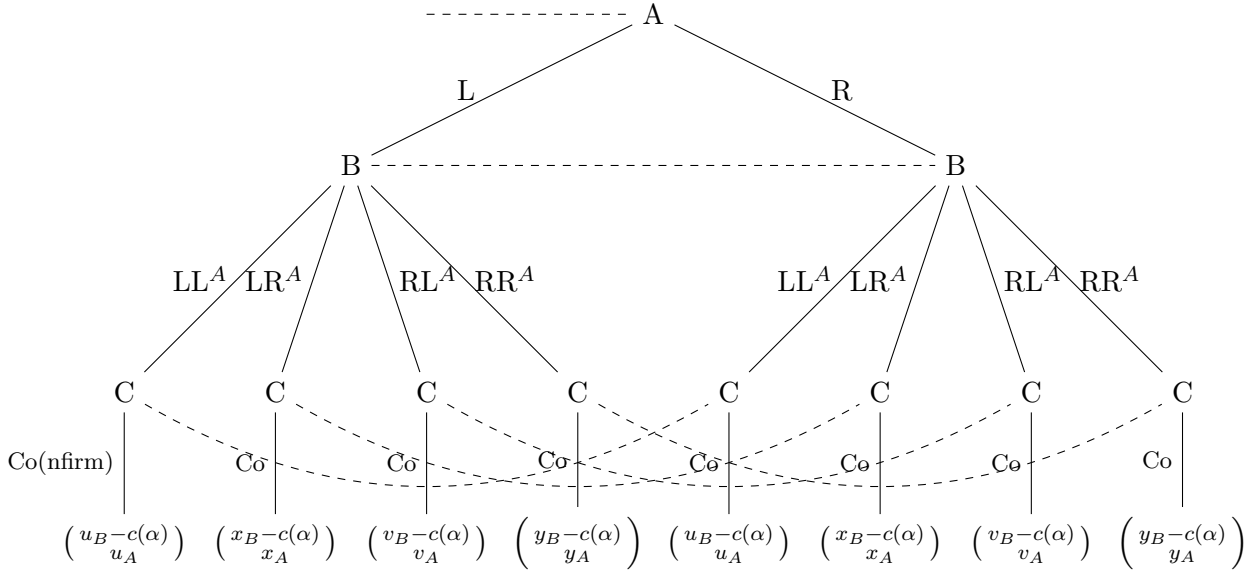
**Treatment LIE** In the fabrication treatment,  $S_2$  was designed such that  $B$  cannot see  $A$ 's choice. Instead,  $S_2$  requires  $B$  to report choices for  $A$  and  $B$  to an additional player  $C$  who has no other function than to confirm the choices reported to her, thus making them payoff-relevant. Through choosing  $S_2$  we say that  $B$  decides to fabricate information about  $A$  since she must make up a choice for  $A$  when reporting to  $C$ .<sup>13</sup> In figure 4, superscript A indicates the action  $B$  chooses to report for  $A$ . Player  $C$ 's trivial task to confirm the decision is labelled *co*, i.e. 'confirm'.

### 3.2 Part 1: which advantage can $B$ secure by lying, spying, or sabotaging?

We implement all three treatments LIE, SPY, and SABOTAGE in two payoff regimes. In a *payoff-neutral* regime,  $B$  cannot secure a material advantage through opting for  $S_2$ . In both interaction structures  $S_1$  and  $S_2$   $B$  always fully controls her own payoff, and  $A$ 's choice has no payoff consequences in any interaction structure. Therefore,  $B$  does not make  $A$ 's choices any more payoff-irrelevant by opting for  $S_2$  rather than  $S_1$  and  $B$  does therefore not infringe  $A$ 's freedom of choice through choosing  $S_2$ . Since  $B$  can for sure obtain all payoff in all interaction structures, we call  $B$ 's choice of the interaction structure *payoff neutral*. Table 3b) illustrates parties' payoffs in  $S_1$  for this payoff regime: through choosing  $L$ ,  $B$  obtains  $100 - c(\alpha)$  ECU for sure; the payoff table for  $S_2$  is identical. In the payoff neutral setting, we observe  $B$ 's attitudes toward fabrication and sabotage when these do not infringe  $A$ 's freedom to choose,

(1906) cartoon of pupils Max and Moritz who replace the tobacco in their teacher's pipe with blackpowder. When the teacher lights the pipe, it explodes to his surprise rather than starting to smoke.

<sup>13</sup>If we informed  $B$  about  $A$ 's choice in  $S_2$ ,  $B$  could decide to truthfully or untruthfully report this choice. We do not give  $B$  this option in order to prevent lying by telling the truth (see e.g. Sutter 2009). For an illustration, take a  $B$  participant who is lie-averse because she does not wish to be taken for a liar by others. She could safely opt into  $S_2$  and lie about  $A$ 's choice if she thinks  $C$  will interpret her message as the truth. In our setting, we prevent this: if  $B$  does not wish to lie, she has to avoid  $S_2$  since player  $C$  knows the rules of the game and knows that  $B$  cannot know  $A$ 's choice. This way, we keep the three treatments SPY, SABOTAGE, and LIE comparable.  $C$  observes whether a report was made or not, but she does not observe whether  $B$  actively opted into  $S_2$ , or arrived there 'accidentally'.



**Figure 4:** INTERACTION STRUCTURE  $S_2$  IN TREATMENT LIE.

and  $B$ 's desire to satisfy her curiosity through spying when she cannot acquire payoff-relevant information. Note, however, that in absence of any payoff implications,  $B$  participants might no longer grasp the meaning of opting for  $S_2$ , see also (Abbink and Hennig-Schmidt 2006).

In the *competitive payoff* regime, we study the same winner-takes-it-all scenario but this time,  $B$  can only secure all payoff for sure if she competes unfairly and opts into  $S_2$ . Again,  $A$  and  $B$  play a constant-sum game in which either  $A$  or  $B$  obtains the entire payoff, and the other player obtains nothing. Table 3c) shows how parties' payoffs depend on their own and on their opponent's payoffs in the 'fair' interaction structure  $S_1$ . We see that player  $B$  does not fully control her payoff in  $S_1$  which always varies along with  $A$ 's choice. If in such a payoff constellation,  $B$  enables herself to spy or to manipulate  $A$ 's choice, she obtains full control over her own payoff. This is exactly what  $B$  can achieve by opting for  $S_2$  rather than  $S_1$ . Thereby,  $B$  transforms payoff table 3c) into table 1b) and – in contrast to the neutral setting – takes some payoff-relevant options away from  $A$ , thus clearly reducing  $A$ 's freedom to choose.

*The mixing possibility.* As mentioned before, players  $A$  and  $B$  also have the explicit option to toss a fair coin between  $L$  and  $R$  in interaction structures  $S_1$  and  $S_2$ .<sup>14</sup> Hence,  $B$  has an ex-ante fair (Bolton, Brandts, and Ockenfels 2005) and kind (Sebald 2010) option in all interaction structures. We will, however, see that  $B$  never uses her mixing option in the experiment – which is not surprising since  $B$  can also mix over the two interaction structures before arriving in any.

<sup>14</sup>Specifically when payoffs are competitive, player  $B$  is likely to randomize in  $S_1$  since she does not know or cannot set  $A$ 's choice. In  $S_2$ , where she either knows, or can set  $A$ 's choice,  $B$  is less likely to randomize. To make both interaction structures as similar as possible in all aspects apart from the spying, lying, or sabotaging feature, we always offer subjects a button to explicitly mix in  $S_1$  and  $S_2$  for all treatments.

**Table 3:** PAYOFFS IN INTERACTION STRUCTURE  $S_1$ .

		<b>3a) General</b>			
		A			
		L	R		
B	L	$u_B - c(\alpha), u_A$	$x_B - c(\alpha), x_A$		
	R	$v_B - c(\alpha), v_A$	$y_B - c(\alpha), y_A$		
		<b>3b) Payoff Neutral</b>		<b>3c) Competitive</b>	
		A		A	
		L	R	L	R
B	L	$0 - c(\alpha), 100$	$0 - c(\alpha), 100$	$0 - c(\alpha), 100$	$100 - c(\alpha), 0$
	R	$100 - c(\alpha), 0$	$100 - c(\alpha), 0$	$100 - c(\alpha), 0$	$0 - c(\alpha), 100$

*Note:* Table 3a) presents the general payoff structure of interaction structure  $S_1$ , table 3b) the respective payoff values for the payoff neutral regime, and 3c) for the competitive regime.

### 3.3 Part 2: Giving $A$ a symbolic punishment or reward option

In part 2 of each session, subjects repeat part 1 knowing that  $A$  can punish or reward  $B$ 's choice of the interaction structure. More specifically, it is commonly known that  $A$  submits a punishment and reward schedule in which she decides whether to subtract up to 30 ECU, or to add up to 30 ECU to  $B$ 's payoff, if  $B$  chooses  $S_1$  1) for sure, 2) with  $\text{Prob}(S_1) \in [75\%, 99\%]$ , 3) with  $\text{Prob}(S_1) \in ]50\%, 75\%[$ , 4) with  $\text{Prob}(S_1) = 50\%$ , and if she chooses  $S_2$  with 5)  $\text{Prob}(S_2) \in ]50\%, 75\%[$ , with 6)  $\text{Prob}(S_2) \in [75\%, 99\%]$  and 7) if she chooses  $S_2$  for sure. Each ECU punishment or reward costs  $A$  the same amount.  $B$  submits which punishment and reward schedule she expects  $A$  to submit. If  $B$  guessed the entire schedule correctly, she earned 35 ECU. Guessing  $A$ 's punishment or reward correctly for one of the seven cases outlined above earned  $B$  5 ECU. For each ECU by which  $B$  deviated from  $A$ 's actual punishment or reward,  $B$  earned 0.08 ECU less. The resulting changes to the payoff matrices of  $S_1$  and  $S_2$  can be found in appendix C, and a detailed analysis of part 2 in (Chlaß and Riener 2015).

### 3.4 Part 3/ Moral preferences, Envy, Risk Attitudes and Values

In part 3 of each session, we elicit several subject characteristics and preferences to better understand the nature of subjects' choices in our experiment. These controls are described below. Finally, subjects also submit a variety of demographics, i.e. their field of study, their semester, age, and gender.

**Envy & Risk Preferences** We briefly elicit envious preferences (Kirchsteiger 1994) to see how much participants dislike being materially worse off than others. To that end, subjects were randomly rematched in a perfect strangers design and submitted their choice between the options "10 ECU for themselves and 10 ECU for the other" or "10 ECU for themselves

and 20 ECU for the other” knowing that a fair coin would decide whether their own decision, or their opponent’s decision determined their payoff from this task (see for example Bartling, Fehr, Maréchal, et al. 2009). Subsequently, subjects chose between lotteries in a price list format (Holt and Laury 2002) and a sure payoff.<sup>15</sup>

**Moral Judgement Test** Subsequently, subjects completed the standardized moral judgement test (M-J-T) developed by G. Lind (1978, 2008), see appendix G for an excerpt. As we have mentioned before, subjects might deem it unethical that the rules of the game grant *B* the privilege to obtain more information than *A*, or the privilege to override *A*’s choice (‘it is unfair/immoral to favour one person over another by granting her more rights or greater privileges’). There could be many other moral ideals motivating *B*’s choice of the interaction structure such as a social norm that parties should have equal chances to obtain the one nonzero payoff, or a social norm not to lie, spy, or sabotage, or a desire to satisfy some expectation of *A*, or the desire to show kind intentions toward *A*.

To test which – if any – of these motivations are at play, we first need a means to describe how *B* participants typically derive whether an action is right or wrong – for instance, which arguments or moral ideals they employ to do so. An individual typically feels comfortable to use only some of the many moral arguments which exist: each individual therefore has preferences over ways of moral argumentation (see e.g. Kohlberg 1984; Piaget 1948). L. Kohlberg studied extensively which arguments individuals in the field actually use to judge the right or wrong of an action and classified the many types of argumentation he encountered into six classes of argumentation (Kohlberg (1969, pp. 375), see appendix F) which we discuss in more detail in section 5. Lind’s moral judgement test elicits individuals’ preferences over precisely these ways of argumentation. To that end, the test presents subjects with two stories. The first story describes how workers break into a factory in order to find and steal evidence that management was listening in on them. The second story describes that a woman who is fatally ill asks a doctor to medically assist her suicide. After each story, subjects first state whether they deem the respective protagonists’ behaviour right or wrong. Subsequently, the test lists arguments one might put forth to judge the workers’ (or doctor’s) actions. Each argument represents one way of moral argumentation from appendix F. Subjects can agree or disagree to employ each argument for judging the protagonists’ behaviour on a nine-point Likert Scale. Four test items (arguments) are used to characterize an individual’s preference over each of Kohlberg’s six ways of argumentation. The test was administered in pen and paper format to keep its design and structure fully intact. Section 6 uses these results to identify the economic preference types underlying the behaviour which we observe.

**Materialist & Postmaterialist Values** In the extensive form games of Fig. 1, *B* can express her dislike of interaction structures  $S_2$  – of spying, fabrication, or sabotage – in two

---

<sup>15</sup>The lottery payoffs are 10 ECU and 35 ECU, the sure payoff is 25 ECU, respectively. The probability of the low lottery payoff increased in steps of 10% such that subjects submitted ten choices between a lottery and a sure payoff.

ways. She can either pay for  $S_1$  and thus grant  $A$  the same autonomy as herself, or she can pay for  $S_2$ , assume power over  $A$  and exert it to  $A$ 's advantage, i.e. be paternalistic toward  $A$ . If  $B$  participants have preferences over power (Bartling, Fehr, and Herz 2014) or paternalism, their compensation strategy will vary along with  $B$ 's attitudes toward exerting power. We elicit individuals' values along the well-known sociological *materialism-postmaterialism* taxonomy (Inglehart 1977; Klages and Gensicke 2006) where materialists – amongst other things – appreciate power, order, obedience, and hierarchy whereas postmaterialists value individualism, autonomy, and self-fulfillment. Instead of using Ronald Inglehart's two-dimensional ranking approach, we use a three dimensional rating approach developed by Klages and Gensicke for Germany (see e.g. 2006)<sup>16</sup>. We elicit individuals' scores on these scales by means of a twelve item questionnaire continuously re-validated on the German population. The items which individuals rate on a scale from 1 to 7 can be found in appendix H. We only elicited materialism-postmaterialism values in LIE and SABOTAGE with competitive payoffs, since these are the two treatments in which  $B$  can infringe  $A$ 's autonomy.

### 3.5 Implementation

The experiment was conducted at the experimental laboratory of the Chair for Empirical and Experimental Economics at the University of Jena, Germany. In total, we ran 36 sessions with altogether 630 participants (303  $B$ -participants, 303  $A$ -participants, and 24  $C$ -participants) in 2012 resulting in roughly 50 decisions per treatment. 309 (49%) of all participants were male. The average payment which included a show-up fee of €2.50 was €7.94 ( $B$ -participants: €9.65,  $A$ -participants: €6.35,  $C$ -participants: €6.30) with a minimum of €3.60 and a maximum of €12.10. A session lasted approximately 50 minutes including payment (10 minutes). Subjects were undergraduate students from the University of Jena which were randomly recruited from all fields of study via the opt-in web based online platform ORSEE (Greiner 2004). The experiment was programmed in z-Tree (Fischbacher 2007). We did not elicit any information that would allow us to identify subjects. Payouts were distributed in sealed envelopes.

## 4 Results

### 4.1 Which interaction structure do $B$ participants choose?

Table 4 details how many  $B$  participants paid for interaction structure  $S_1$ , and how many paid for interaction structure  $S_2$  which – depending on the treatment – would either allow them to spy, sabotage, or lie.  $B$  participants' procedural choices differ significantly across

---

<sup>16</sup>Klages and Inglehart worked in parallel. Inglehart stipulated a shift from materialist to postmaterialist societies whereas Klages (e.g. 1984) predicted a value synthesis leading e.g. to so-called *realists*: individuals who combine a postmaterialist desire for autonomy with a desire to compete and perform (classic materialist value). Klages's and Gensicke's (e.g. 2006) inventory includes a third measurement scale, and does not force individuals to prefer either one or the other value category – both features allowing to classify Inglehart's 'mixed (synthesized) value types'. The third scale appeared as an independent value dimension in the 1980-1990ies among the German population, see appendix H; it captures attitudes toward power particularly well.



Percentages of  $B$ -participants paying for interaction structure  $S_1$  ('fair') and  $S_2$  ('unfair') per treatment

treatment	LIE <sup>18</sup>				SPY				SABOTAGE			
	payoff neutral		competitive		payoff neutral		competitive		payoff neutral		competitive	
payoff regime	#47		#44		#53		#53		#52		#54	
#nr. of $B$ players												
interaction structure	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$
% who pays	17%	6%	20%	11%	4%	36%	9%	68%	8%	35%	4%	69%
median change of $\alpha$	13%	30%	10%	20%	25%	20%	50%	25%	17.50%	20%	16%	25%
% who does not pay	77%		68%		60%		23%		58%		28%	

**Table 4:** CHOICES OVER PROCEDURES FOR ALL TREATMENTS.

treatments. There are significantly more  $B$  participants who choose the spying (Fisher’s Exact test,  $p$ -value  $< 0.01$ ) or the sabotaging procedure (Fisher’s Exact test,  $p$ -value  $< 0.01$ ), than there are  $B$  participants who choose the lying procedure<sup>17</sup>. We do not observe such a difference between the spying and sabotaging procedures (Fisher’s Exact test,  $p$ -value = 1). These findings are the same for the payoff neutral, and the competitive payoffs regime. Also, the percentage of  $B$  participants who prefers the default – a fair chance draw between the interaction structures – is significantly higher in treatment LIE (77% and 68%, respectively) than in treatment SPY (Fisher’s Exact test,  $p$ -value  $< 0.02$ ) or in treatment SABOTAGE (Fisher’s Exact test,  $p$ -value  $< 0.02$ ).

*RESULT 1: Subjects opt less often into fabricating information than they opt into spying or sabotage.*

The payoff neutral setting shows that many  $B$ -participants nudge themselves into interaction structures which allow them to spy (36%) and sabotage (35%) even when there is no material advantage to be gained. Subjects do therefore seem to intrinsically enjoy gathering information and replacing others’ actions when this does not affect the other party’s payoff. Interestingly, median payments for  $S_2$  are qualitatively larger than those for  $S_1$  in both LIE and SABOTAGE. Only in treatment SPY,  $B$  players who prefer the spying-free procedure  $S_1$  are willing to pay more than those who prefer the spying procedure  $S_2$ .

## 4.2 Which allocation do $B$ participants choose?

$B$  players’ choices of the interaction structure cannot be fully understood without taking the allocation they opt for into account. Take a  $B$  player who increases the likelihood of interaction structure  $S_2$  in the competitive payoffs setting. She may wish to obtain the chance

<sup>17</sup>Two shares (relative frequencies) are compared via one-sided Fisher’s Exact tests, three and more frequencies, e.g. the share of  $B$  participants paying for the  $S_2$ ,  $S_1$ , or who do not pay anything at all, are compared via Chi-square tests using exact, i.e. simulated,  $p$ -values.

<sup>18</sup>A brief reading example: In treatment LIE with neutral payoffs, there were 47  $B$  participants. 17% of them paid for  $S_1$  and 6% for  $S_2$ . The 17% who paid for  $S_1$  made at the median,  $S_1$  13% more likely than  $S_2$ . The 6% who paid for  $S_2$ , made, at the median,  $S_2$  30% more likely than  $S_1$ . 77% of 47  $B$  participants left the default 50-50 chance of arriving in either  $S_1$  or  $S_2$ . For absolute figures, see appendix E

Which allocation do  $B$ -participants impose when they hold the power to do so?  
 selfish: (payoff B: 100, payoff A: 0); altruistic: (payoff B: 0, payoff A: 100)

treatment payoff regime interaction structure # nr. of $B$ players.	LIE <sup>20</sup>			SPY			SABOTAGE		
	payoff neutral $S_1$	neutral $S_2$	competitive $S_2$	payoff neutral $S_1$	neutral $S_2$	competitive $S_2$	payoff neutral $S_1$	neutral $S_2$	competitive $S_2$
	#25	#22	#25	#20	#33	#40	#22	#30	#28
selfish	80%	77%	32%	90%	94%	100%	82%	87%	29%
equal chance	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
altruistic	20%	23%	68%	10%	6%	0%	18%	13%	71%

**Table 5:**  $B$ 'S CHOICES OF THE PAYOFF ALLOCATION IN THE 'FAIR' ( $S_1$ ) AND THE 'UNFAIR' ( $S_2$ ) INTERACTION STRUCTURE.

to lie, spy, or sabotage to her own material advantage. She may yet also wish to obtain the power of giving all payoff away in an attempt, for instance, to compensate  $A$  for the rules of the game.<sup>19</sup> Table 5 shows which payoff allocations  $B$  players impose when they hold the power to do so: under payoff neutrality,  $B$  can impose her preferred allocation in  $S_1$  and  $S_2$ , whereas under competitive payoffs,  $B$  can only do so in  $S_2$ . The allocation (B, A) : (0, 100) where  $B$  gives all payoff away is labelled 'altruistic', the allocation (B, A) : (100, 0) where  $B$  keeps all payoff is labelled 'selfish'. 'Equal chance' denotes cases where  $B$  chooses to toss a fair coin between the selfish, and the altruistic allocation. The absolute figures underlying table 5 are detailed in table 11, appendix E.

In treatment SPY, 93  $B$  participants could impose their preferred allocation: all 53 under payoff neutrality, and the 40 who arrived in  $S_2$  under competitive payoffs. 89 of them (96%) took all payoff. In treatments LIE and SABOTAGE instead, we observe a substantial amount of altruism. In SABOTAGE, 72  $B$  participants could impose their preferred allocation: all 52 under payoff neutrality and the 28 who arrived in  $S_2$  under competitive payoffs. 28 of those 72  $B$  participants (39%) gave all payoff away. In treatment LIE, 72  $B$  participants could impose their preferred allocation: all 47 under payoff neutrality, and the 25 who arrived in  $S_2$  under competitive payoffs. 27 of those 72  $B$  participants (38%) gave all payoff away. Thus, we observe significantly more altruism in LIE or SABOTAGE than in SPY (Fisher's exact tests,  $p$ -value < 0.01). We observe no such difference between LIE and SABOTAGE (Fisher's exact test,  $p$ -value = 0.87).

*RESULT 2: We observe significantly more selfish allocations in treatment SPY than in*

<sup>19</sup>It might also be that  $B$  participants choose interaction structure  $S_2$  with the intention of taking all payoff, but once arrived in the lying, spying, or sabotaging procedure, feel too guilty to do so. We analyze  $B$  players' moral motivations to be altruistic in  $S_2$ , and the determinants of choosing  $S_1$  rather than  $S_2$  in section 5.

<sup>20</sup>A brief reading example: In treatment LIE with payoff neutrality,  $B$  can impose her preferred allocation in  $S_1$  and  $S_2$ . Out of 47  $B$  participants (see table 4), 25 arrived in  $S_1$ . 80% of them kept all payoff for themselves, 20% gave all payoff away, and nobody tossed a coin. The remaining 22  $B$  participants arrived in  $S_2$ . 77% of them kept all payoff, 23% gave all payoff away, nobody tossed a coin. With competitive payoffs,  $B$  can only impose the allocation in  $S_2$ . Out of 44  $B$  participants (see table 4), 25 arrived in  $S_2$ , 32% of which kept all payoff, and 68% of which gave all payoff away; nobody tossed a fair coin. For absolute figures, see appendix E.

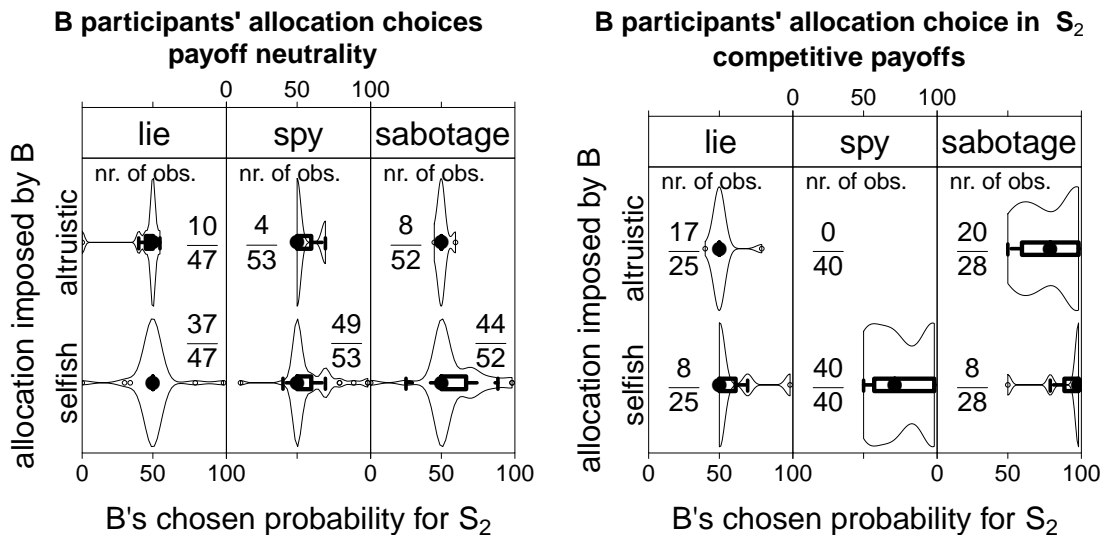
*SABOTAGE or LIE.*

Treatments LIE and SABOTAGE induce significantly more altruism under *competitive payoffs* than under *payoff neutrality* (Fisher’s exact tests,  $p$ -value < 0.01), but not treatment SPY (Fisher’s exact test,  $p$ -value = 0.13). In LIE with competitive payoffs, 17 of those 25  $B$  players (68%) who fabricate information give all payoff away compared to 10 out of 47 (21%) under payoff neutrality. In SABOTAGE with competitive payoffs, 20 of those 28  $B$  participants (71%) who sabotage give all payoff away compared to only 8 out of 52 (15%) under payoff neutrality. In SPY *all*  $B$  participants who spy on  $A$  exploit the information they acquire to *take* all payoff. Section 2 discussed that under competitive payoffs, treatments LIE and SABOTAGE empower  $B$  to impair  $A$ ’s freedom of choice which is not the case under payoff neutrality. Treatment SPY in turn grants  $B$  additional information about  $A$  in both payoff settings. If  $B$  deemed that a procedure should not grant her the power to impair  $A$ ’s freedom of choice, and if this drove  $B$ ’s altruism, then  $B$  participants’ altruism should vary across our  $3 \times 2$  treatments exactly as it does.

*RESULT 3: B participants are significantly more altruistic when the treatment empowers them to impair A’s freedom to choose than when it does not.*

Figure 5 illustrates how much information  $B$  participants’ choices  $\alpha$  of the interaction structure  $S_2$  discloses about their allocation decision. The  $X$  axis shows how likely  $B$  participants made the ‘unfair’ interaction structure  $S_2$  whereas the  $Y$  axis depicts  $B$  participants’ choice of the allocation.

**Figure 5:**  $B$  PARTICIPANTS’ CHOSEN PROBABILITY<sup>21</sup> FOR THE UNFAIR INTERACTION STRUCTURE  $S_2$ , AND THEIR CHOICE OF ALLOCATION WHEN THEY COULD IMPOSE IT (LEFT GRAPH: PAYOFF NEUTRALITY, RIGHT GRAPH: COMPETITIVE PAYOFFS).



<sup>21</sup>A brief reading example: take the right graph (‘competitive payoffs’), and therein the third panel ‘sabotage’. The upper half depicts those 20  $B$  participants of altogether 28 who arrived in  $S_2$  where they could impose the

Under *payoff neutrality* (left graph) *altruistic B* participants in LIE, SPY, and SABOTAGE typically leave the default and arrive with a 50% chance in the 'unfair' interaction structure  $S_2$ . All distributions of probability choices are centered at 50% which is also the median (fat black dot). In treatment LIE, *selfish B participants* also leave the default, and *B* participants' procedural choices do therefore not reveal which allocation they will choose. In SPY and SABOTAGE, however, a visible share of selfish *B* participants increases the probability of  $S_2$  such that the distributions of probability choices (violin in the lower parts of panels SPY and SABOTAGE, left graph) show fat right tails. If a *B* participant's choice falls within such a tail, she hence signals she will impose the selfish allocation.

Turning to *competitive payoffs* (right graph), altruistic *B* participants in LIE typically arrive in  $S_2$  by a 50% chance whereas the distribution of choices by selfish *B* participants shows a fat right tail. Similarly, nearly all selfish *B* participants in SABOTAGE make  $S_2$  certain while most altruists increase the probability of  $S_2$  less pronouncedly. In LIE and SABOTAGE therefore, procedural choices which fall within the right tail of their distribution signal that a selfish allocation will be imposed. In SPY, *every* increase in the probability of  $S_2$  signals *B* will take all payoff.

*Result 4: Only resolute attempts at fabrication and sabotage indicate B will take all payoff. In contrast, every attempt at spying results in B taking all payoff.*

## 5 Moral motivations

In this section, we try to understand the nature of our main results: i) why some *B* participants who opt into the 'unfair' interaction structure give all payoff away while others take all payoff, ii) why this amount of altruism differs across treatments, and iii) why some *B* participants opt into the 'unfair' interaction structure  $S_2$  while others opt into the 'fair' interaction structure  $S_1$ , and why their shares vary significantly across treatments.

It may be that *B* participants who give all payoff away do – in general – simply care more about others' payoffs than they care about their own. Other than out of a natural disposition, however, *B* participants may also choose the altruistic allocation because they care about being taken for a nice person, because they wish to fulfill *A*'s payoff expectations and do not wish to let *A* down, or because they wish to comply with a social norm stipulating equal chances for all parties to obtain the one ex-post nonzero payoff and hence, decided to toss a fair coin between both allocations. Alternatively, *B* participants may care about the distribution of

---

allocation under competitive payoffs and imposed the altruistic allocation. The horizontal boxplot shows that within this group of altruistic *B*s, 25% (the left boundary of the boxplot rectangle equals the 25% quantile of the distribution) chose probabilities for  $S_2$  smaller or equal to roughly 60%; 75% (the right boundary of the boxplot rectangle) chose values smaller or equal to 100%. The black dot shows the median probability for  $S_2$  in this group. The horizontal violin around this boxplot is wide at values which many *B* participants chose, and narrow where few choices are located. Looking at the selfish group in the lower part of panel sabotage with 8/28 observations, the violin shows that most selfish *B*s set the probability of  $S_2$  to values close to or at 100%.

rights in each interaction structure which systematically disadvantages  $A$ .<sup>22</sup> Treatment SPY grants  $B$  privacy of her own choice but empowers  $B$  to monitor  $A$ 's. However, it is not within  $B$ 's power to grant  $A$  more, and hence, equal rights of information since she cannot disclose the interaction structure to  $A$ , see also section 6. Treatments SABOTAGE and LIE in contrast grant  $B$  the freedom to choose between two actions and – in the competitive payoffs setting – empower  $B$  to grant or to deny  $A$  the same freedom of choice.  $B$  participants who hold ethical ideals about the equality of decision rights may wish to choose according to this ideal, and may therefore be willing to forego payoff to do so – either by instating  $A$ 's rights, or by compensating  $A$  through payoff. As outlined above, spying infringes another right than fabrication and sabotage, and does not allow  $B$  to increase  $A$ 's information rights. Hence,  $B$  may deem a different compensation adequate in each treatment. In order to understand whether one, or several of the moral ideals outlined above are at play, we first need a means to describe how a given  $B$  participant arrives at the conclusion that a particular course of action is either right or wrong.

Jean Piaget and Lawrence Kohlberg have studied extensively how individuals in the field make such moral judgements, see e.g. (Kohlberg 1969, 1984; Piaget 1948). They observed individuals who referred to the absence of punishment or the existence of a reward, to others' expectations, or to a social norm – i.e. which action the majority of people in a society or a peer group would adopt – to derive the right course of action. Other judgements invoked the status quo ('it is right to do what we have always done in this situation'), or referred to the law. Individuals would also refer to the social contract and look at whether an action respected the individual's civic rights, or the equality of rights across individuals stipulated therein. Finally, some moral judgements would refer to concepts *beyond* the social contract such as human rights, human dignity, or some other general ethical principle considered to be universally valid. Kohlberg (e.g. 1969, pp. 375) classifies these various moral ideals into six types. Table 6 reviews the two types in particular which invoke individual rights.<sup>23</sup> We characterize  $B$  participants' use of all six ways of argumentation – see appendix F for a complete classification and examples – by subjects' scores from the moral judgement test administered in phase 3 to model  $B$  participants' altruism and procedural choices.

We begin with *competitive payoffs* where we classify  $B$  participants into type i) who pays for interaction structure  $S_1$  and arrives there ( $n=14$ ), type ii) who sets the probability for interaction structure  $S_2$  to  $\alpha \geq 50\%$  and, if arriving in  $S_2$ , uses her lying, spying or sabo-

---

<sup>22</sup>The corresponding formal preference models built upon the moral ideals mentioned here are guilt aversion (ideal: comply with others' expectations) as in (Battigalli and Dufwenberg 2007), preferences for equal expected payoffs (ideal: comply with a social norm that everybody's chances to obtain the payoff should be equal) as in (Bolton, Brandts, and Ockenfels 2005), and purely procedural preferences (ideal: equality of rights across parties) (Chlaß, Güth, and Miettinen 2009). We discuss each of these models in section 6.

<sup>23</sup>To date, only one type of economic preferences builds upon these classes. Chlaß et al. (2009) let individuals choose between largely outcome-invariant allocation procedures which distribute parties' rights of information or participation either equally or unequally. Individuals often choose a procedure where they expect lower payoffs. Procedural choices systematically linked to the degree by which subjects referred to the equality of civic rights, that is, social contract reasoning *postclass 1* in table 6. Chlaß and Moffatt (2012) find that dictator game giving links to *postclass 2*, and hypothetical transfers by recipients in dictator games to *postclass 1*.

Argumentation	Motivation for moral behaviour
postconventional	<i>postclass 1.</i> Social contract orientation, in which duties are defined in terms of the social contract and the respect for others' rights as recorded in that contract. Emphasis is upon equality and mutual obligation within a democratic order.
	<i>postclass 2.</i> 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.

**Table 6:** KOHLBERG'S TWO CLASSES OF POSTCONVENTIONAL (OUTCOME-INVARIANT) MORAL ARGUMENTATION (ISHIDA 2006).

tagging option to give all payoff away<sup>24</sup> (n=81), and a type iii) who prefers not to influence the interaction structures and arrives in  $S_2$  where she takes all payoff (n=14). In a series of simple binary Logit models, we contrast each of these types with the most selfish type iv) that we observe: the one who pays for and arrives in  $S_2$  where she exploits her lying, spying, or sabotaging option to keep all payoff for herself. This classification covers all  $B$  participants which we observe. We regress each pair of types on two dummies for treatments LIE and SABOTAGE, on individuals' moral judgement characteristics, on risk and envy preferences, and where collected, on individuals' materialism-postmaterialism scores.<sup>25</sup>

The left of tables 7 compares the procedural type i) who pays for interaction structure  $S_1$  to the most selfish type iv) who is our reference category. Out of the treatment dummies, only LIE increases the frequency of type i) by 49%,  $p$ -value  $< 0.01$ . The more strongly a given  $B$  participant refers to *postclass 1* arguments – the social contract and the respect for civic rights – the *more* likely she is of procedural type i) whereas the use of *postclass 2* arguments – general ethical principles of conscience – makes individuals less likely to be of procedural type i).  $B$  participants who invoke ethical principles are more likely not to influence the draw of the interaction structure and hence, more likely to be of type iii). They may deem it unethical to exert any of their power over  $A$ 's position of rights at all. Risk attitudes and other control variables do not show a significant impact, see also appendix I.

PROCEDURAL TYPE (I)			ALTRUISTIC TYPE (II)		
<i>Argument</i>	<i>Effect</i>	<i>std.err.</i>	<i>Argument</i>	<i>Effect</i>	<i>std.err.</i>
<i>lie</i>	0.49	0.17 <sup>a</sup>	<i>lie</i>	0.50	0.04 <sup>a</sup>
<i>postclass 1</i>	0.16	0.08 <sup>b</sup>	<i>sabotage</i>	0.49	0.05 <sup>a</sup>
<i>postclass 2</i>	-0.15	0.07 <sup>b</sup>	<i>postclass 1</i>	0.10	0.03 <sup>a</sup>
<i>[risk aversion</i>	-0.05	0.03]	<i>[risk aversion</i>	0.01	0.02]

**Table 7:** WHICH DETERMINANTS MAKE THE PROCEDURAL TYPE (I) (N=56), AND THE ALTRUISTIC TYPE (II) (N=123) MORE LIKELY THAN THE MOST SELFISH TYPE (IV)?

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

<sup>24</sup>If she arrives in  $S_1$  where she cannot determine the allocation, no restriction is imposed on the allocation. We also classified two  $B$  participants as type ii) who make  $S_1$  slightly more likely, but arrive in  $S_2$  and give all payoff away.

Treatment dummy LIE turns insignificant if one introduces an interaction effect  $lie \times postclass\ 1$ : ethical concerns about the opponent’s equality in civic rights hence also explains the excess occurrence of type i) in treatment LIE (21%  $p\text{-value} < 0.01$ ). Turning to the right of tables 7, the altruistic type is 50%,  $p\text{-value} < 0.01$ , more prevalent in treatment LIE and 49%,  $p\text{-value} < 0.01$ , more prevalent in SABOTAGE than in SPY. The more strongly  $B$  participants refer to *postclass 1* arguments, the more likely they are of type ii) who gives all payoff away in the ‘unfair’ interaction structure. Surprisingly, types i) and ii) share a common ethical motivation (which they also share with the fair-coin type iii), see appendix I) – they all care more strongly about the equality of individuals’ civic rights and the social contract than the most selfish type iv) who spies, lies, and sabotages to her own advantage.<sup>26</sup> A one-unit increase in individuals’ use of these arguments increases the likelihood of type ii) by 10%  $p\text{-value} < 0.01$ . The distributions of the explanatory variables in tables 7 do not differ across treatments LIE, SABOTAGE, and SPY. As before, risk attitudes or further control variables do not show a significant effect, see appendix I.

Their motivations being the same, types i) and ii) may differ in their view how to rectify the infringement of  $A$ ’s rights: type ii) might seek the ‘unfair’ interaction structure to exploit her power for  $A$ ’s good and give her all payoff whereas type i) may prefer to directly reinstate  $A$ ’s rights by opting for the ‘fair’ interaction structure. If true, both types should differ in their attitudes toward *power*. We use individuals’ materialism and postmaterialism scores from part 3 to test this idea. Materialists value the existence of hierarchy, order, duty, power and should consequently classify more often as type ii). Specifically Gensicke’s and Klages’s value category II – which is part of Inglehart’s materialism dimension, see appendix H – captures attitudes toward power. Postmaterialists value individual autonomy, dislike power and should more often classify as type i). Indeed, a one-unit increase on Klages’s and Gensicke’s (2006) postmaterialism scale<sup>27</sup> makes the procedural type i) who avoids power, by 12%,  $p\text{-value} < 0.01$ , more likely. A one-unit increase on hedonistic materialism (value category II) increases the likelihood of being type ii) who opts into  $S_2$  and gives all payoff away by 12%,  $p\text{-value} = 0.01$ . Accounting for these variables, the effect of *postclass 1* arguments for

---

<sup>25</sup>All estimated Logit models were tested downwards (reduced) from large specifications which included all two way interactions to those determinants which were significant. We report marginal effects of explanatory variables, i.e. by how many per cent the response dummy is more likely to take on a value of One, if the respective explanatory variable increases by one unit. The moral preferences are computed as in previous studies (Chlaß, Güth, and Miettinen 2009; Chlaß and Moffatt 2012): we take the mean rating over all four arguments referring to the same type (class) of moral argumentation for all six types of moral argumentation (class 1 to 6 in table F) and adjust each mean rating for the difference between the largest and smallest value a subject ever ticks in the entire test. These averages are then normalized on the entire sample of  $B$  participants, subtracting the sample mean and dividing by the sample standard deviation. Initial model specifications always include the complete set of six moral preferences elicited in the test for each participant.

<sup>26</sup>Since the effect of *postclass 1* arguments is less significant for type i), she might have a weaker concern than ii) and might wish to give away less payoff. Note, however, that also in the ‘unfair’ interaction structure  $S_2$ ,  $B$  could have avoided giving away all payoff for sure and tossed a fair coin between both allocations if she wanted to allocate some payoff to  $A$  on expectation.

<sup>27</sup>We take the mean rating over all questionnaire items belonging to the same dimension, see Appendix F.

type i) from tables 7 increases in size and significance to 45%,  $p\text{-value} < 0.01$ , see table 16 in appendix I. Both types do therefore seem to care equally strongly for the infringement of  $A$ 's rights but – due to their attitudes toward power – choose different strategies to compensate their opponent. In summary, we find that the moral ideal underlying  $B$  participants' willingness to forego payoff always springs from the same source – an ethical concern that  $B$  holds the power to infringe  $A$ 's equal position of rights.

A similar logic can explain the remaining type iii) who arrives in the 'unfair' interaction structure by the toss of a fair coin where she takes all payoff. Amongst types i), ii), and iii), she foregoes least expected payoff. These selfish individuals who avoid influencing the interaction structure to their own advantage score 14%,  $p\text{-value} < 0.03$  stronger on postmaterialist values and 15%,  $p\text{-value} < 0.04$  lower on materialist values than the most selfish reference type iv). At the same time, they also share a concern for  $A$ 's position of rights; the strength of this concern makes the fair-coin type 15%,  $p\text{-value} < 0.04$  more likely.

In the *payoff neutral treatment* where  $B$  does not impair  $A$ 's freedom of choice through opting for the unfair interaction structure  $S_2$ , this ethical concern crowds out, and no ethical ideal can be confirmed to underlie  $B$ 's behaviour. Ethical concerns for  $A$ 's position of rights also crowd out if we make parties' position of rights more equal and give also  $A$  some power to influence  $B$ 's freedom of choice through a symbolic punishment and reward option which  $B$  cannot avoid in any interaction structure<sup>28</sup>.

*Result 5: Ethical ideals about the equality of rights explain  $B$ 's willingness to forego payoff and the variation in this willingness across treatments. Attitudes toward materialist and postmaterialist values explain how  $B$  prefers to rectify the infringement of her ethical ideal.*

## 6 Underlying Preferences & Discussion

In this section, we discuss which preferences might underlie  $B$ -participants' behaviour and whether or not our results confirm or contradict their being at play. We restrict our attention to the non-trivial, i.e. the competitive payoff setting where  $B$  can only take all payoff for sure if she opts for  $S_2$ , that is, if she lies, spies, or sabotages.

*Self-interested opportunism.* If  $B$  only cares about her own material payoff, she opts into the 'unfair' interaction structure  $S_2$  for sure. She does so by paying 5 ECU to set  $\text{Prob}(S_2) = \alpha = 1$ . In interaction structure  $S_2$ , she chooses allocation ( $B$ : 100,  $A$ : 0) either by opting for a strategy combination  $\{B : RL^A, A : \{\cdot\}\}$ , or  $\{B : LR^A, A : \{\cdot\}\}$ . Hence,  $B$  receives  $100 - 5 = 95$  ECU and  $A$  receives 0 ECU in treatments LIE, SPY, and SABOTAGE<sup>29</sup>.

<sup>28</sup>Appendix B shows the normal form for  $S_1$  and  $S_2$  with punishment or reward:  $A$  can now reduce, or increase the extent to which  $B$  prefers each strategy over the other in  $S_1$ , and in  $S_2$ . Table 17 shows the results from our companion paper (Chlaß and Riener 2015): neither types i), ii) or iii) are motivated by postclass 1 arguments anymore, if contrasted with the most selfish type iv). Other moral ideals crowd in.

<sup>29</sup>That 95 ECU is the largest possible payout can be seen from comparing the payout of 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



Clearly, self-interested opportunism can neither explain the differences in altruism, nor the variation in  $B$  participants' procedural choices across treatments LIE, SPY, and SABOTAGE, nor the link with individuals' moral judgement from section 5.

*Pure altruism.* If  $B$  only cares about her opponent's material payoff, she pays 5 ECU for setting  $\text{Prob}(S_2) = \alpha = 1$  to opt into interaction structure  $S_2$ . Therein, she chooses 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  $A$  receives 100 ECU in LIE, SPY, and SABOTAGE. Altruistic preferences should therefore be unlikely to explain any differences in allocation choices or procedural choices between treatments LIE, SPY, and SABOTAGE.

*Preferences for equal expected payoffs.*  $B$  may be willing to forego some of her maximal payoff to grant  $A$  more equal chances on the one ex-post nonzero payoff (Bolton, Brandts, and Ockenfels 2005). Put differently,  $B$  may be inequity-averse over expected payoffs and e.g. have utility  $u_B = a_B \cdot E(y_B) - 0.5b_B (E(y_B) \cdot 100^{-1} - 0.5)^2$  with  $y_B$  her own expected payoff,  $a_B \geq 0$   $B$ 's inequity aversion against disadvantageous inequality, and  $b_B \geq 0$   $B$ 's inequity aversion against advantageous inequality. In  $S_1$ , two perfectly selfish players would each choose to toss the fair coin between  $L$  and  $R$  which at the same time, guarantees ex-ante equality in payoffs. In  $S_1$ ,  $B$ 's corresponding utility is hence  $a_i \cdot 50$  with no disutility from advantageous inequality. In  $S_2$ ,  $B$  can also toss a fair coin which equalizes expected payoffs irrespective of  $A$ 's choice and moreover,  $B$  can mix over her strategies such as to generate any distribution of chances on the one ex-post nonzero payoff she prefers. If  $B$  has  $a_B, b_B$  such that she cannot reach her preferred distribution of chances in  $S_1$ , she prefers  $S_2$ . Since payoffs are the same in LIE, SPY, and SABOTAGE, this decision is always identical. Unless participants differ systematically in their degrees of inequity aversion across treatments, preferences for equal expected payoffs are unlikely to explain any of the differences we observe between LIE, SPY, and SABOTAGE. Moreover, preferences for equal expected payoffs stipulate that individuals refer to social norms to judge which action is right.<sup>30</sup> In our setting,  $B$  participants' preferences to do so did not explain their choices of  $S_1$ , or their altruism in  $S_2$ . Both linked to a different moral ideal suggesting other preferences.<sup>31</sup>

*Preferences for kind procedures (Sebold 2010).*  $A$  and  $B$  may care for the *kindness* of a procedural choice (the kindness of a person who chooses a procedure is equal to the kindness of

---

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.

<sup>30</sup>Preferences for equal expected payoffs are built around a social norm that parties' outcomes should ex-ante be equal. The moral judgement test which we use elicits individuals' preferences over these ideals, and hence, test whether the 'necessary conditions' for inequity aversion, reciprocity, guilt aversion etc. hold.

<sup>31</sup>Theoretically, social norms may stipulate that carrying out activities such as lying and sabotaging, is per se morally more severely wrong than spying. Two conflicting norms in each treatment – stipulating expected payoff equality versus avoiding the unfair procedure  $S_2$  – with the second having a different power of attraction in LIE, SPY, and SABOTAGE might therefore have explained some of the treatment differences which we report. Empirically, however, we do not find any evidence that  $B$  participants' preference to invoke social norms guides their willingness to forego payoff in our setting.

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 can only choose between efficient strategies: 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  $q_L \cdot 100 + (1 - q_L) \cdot 0 - (q_L \cdot 100 + (1 - q_L) \cdot 0 + q_L \cdot 0 + (1 - q_L) \cdot 100) / 2$ .<sup>32</sup>, 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 Nash-equilibrium in  $S_1$ , then  $B$ 's choice of  $L$  and  $R$  is exactly neutral toward  $A$ . Since  $B$  is not unkind in equilibrium,  $A$  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. Even  $B$  participants who prefer kind over unkind procedures therefore opt into  $S_2$  which earns them  $100 - 5$  ECU. Even if  $B$  held off-equilibrium beliefs in  $S_1$ , then whatever reciprocation she expects in  $S_1$  would be identical across SPY, LIE, and SABOTAGE. Moreover, preferences for kind procedures stipulate that players derive utility from procedural choices which intend to induce kind outcomes whereby an outcome is kind if it satisfies some norm of payoff equality. We could not confirm that individuals' tendency to invoke social norms or intentions when judging the right and wrong of an action statistically explained any departures from rational self-interest in LIE, SPY, or SABOTAGE.

*Guilt aversion.* If  $B$  is guilt-averse, she wishes to avoid disappointing  $A$ 's payoff expectations, or wishes to avoid being blamed by  $A$  for doing so (Battigalli and Dufwenberg 2007). In phase two, we elicited  $B$ 's expectations about  $A$ 's symbolic punishment or reward plan for a broad range of procedural choices<sup>33</sup> – symbolic in the sense that punishment and reward are too small to induce reciprocal motives. These symbolic punishment and reward plans contain compound information how much  $A$  disapproves of a given procedural choice, and of the corresponding allocation choice she expects.  $B$  in turn could expect symbolic punishment when she believes  $A$  expects to be let down, and a symbolic reward otherwise. However,  $B$ 's

---

<sup>32</sup>  $q_L \cdot 100 + (1 - q_L) \cdot 0$  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).

<sup>33</sup>  $A$ '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.

expectations about  $A$ 's punishment and reward plans are inconsistent with this idea.  $B$  participants expect more symbolic punishment for choosing the unfair set of rules  $S_2$  in SPY than for choosing it 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  $\alpha = 1$ ). Expectations between LIE and SABOTAGE or SPY and SABOTAGE do not differ. Hence,  $B$ 's frequent choices of  $S_2$  in SPY as compared to the rare choices of  $S_2$  in LIE cannot be explained by a desire to avoid what  $A$  would *not* like  $B$  to do, or explain why we observe no altruism in SPY. Also, the normative ideal underlying guilt-aversion – invoking others' expectations to derive the right action – neither explained  $B$  participants' procedural nor their allocation choices. Guilt aversion is therefore unlikely to explain any differences between LIE, SPY, and SABOTAGE.

*Purely Procedural Preferences.*  $B$  participants may have ethical concerns against distributing decision rights, or rights in general unequally across parties (Chlaß, Güth, and Miettinen 2009). Suppose that  $B$ 's linear utility function includes the following element:  $-\beta_B \max\{\#S_B - \#S_A, 0\} - \alpha_B \max\{\#S_A - \#S_B, 0\}$  where  $\#S_B - \#S_A$  counts the difference in cardinalities between parties' pure strategy sets, counting only such strategies which induce genuinely different outcomes (only 'diverse options' expand a party's freedom of choice, see section 2). In LIE and SABOTAGE,  $B$  has two pure strategies which expand her freedom of choice in  $S_1$ , and two in  $S_2$ .  $A$  in turn has two pure strategies in  $S_1$ , and one (or zero) in  $S_2$ .  $B$  can therefore grant  $A$  equality in decision rights by competing fairly, or advantage herself in terms of decision rights by opting for unfair competition. This is not true in SPY where  $B$  cannot reduce  $A$ 's freedom of choice: both parties have two pure strategies which expand their freedom of choice by the same extent for each potential choice by the opponent. Through opting for unfair competition in SPY,  $B$  merely changes the distribution of information. SPY therefore affects a different type of right.  $B$  might yet also hold concerns about the distribution of *information rights*.  $B$ 's utility function might include element  $-\beta_B \max\{\#\mathcal{I}_B^z - \#\mathcal{I}_A^z, 0\} - \alpha_B \max\{\#\mathcal{I}_A^z - \#\mathcal{I}_B^z, 0\}$  where  $\#\mathcal{I}_A^z - \#\mathcal{I}_B^z$  measures the difference between the cardinalities of party  $A$ 's and  $B$ 's information partitions over the terminal histories  $z \in Z$  of a game, and  $\alpha_B$  and  $\beta_B$  express  $B$ 's aversion against advantageous, or disadvantageous inequality in information rights, respectively. Starting with SPY,  $B$  knows her own, but not  $A$ 's choice in  $S_1$  and hence,  $B$ 's information partition over the four terminal nodes of  $S_1$  has cardinality two. In  $S_2$ ,  $B$ 's information partition over the four terminal nodes has cardinality four: she knows the terminal node for sure. Since  $A$  does not know the interaction structure, her information partition has cardinality one always.  $B$ 's choice of  $S_2$  does therefore increase her own information rights, but not by taking information rights away from  $A$ . In LIE and SABOTAGE, the distribution of information rights is identical in  $S_1$  and  $S_2$ :  $B$ 's information partition over the terminal nodes has cardinality two in  $S_1$  and  $S_2$ ;  $A$ 's cardinality is always one. SPY therefore differs in three aspects from LIE and SABOTAGE: information, and not decision rights are at stake,  $B$  does not obtain additional rights in  $S_2$  by taking them away from  $A$ , and  $B$  has no power to grant  $A$ 's equality of rights. Inequity

aversion over the distribution of decision rights could therefore explain the altruism in LIE and SABOTAGE and its absence in SPY; it could also explain the decline of altruism in the payoff neutral-setting where  $B$  cannot reduce or increase  $A$ 's decision rights. Indeed, the moral ideal underneath  $B$ 's altruism in this paper is identical to the moral ideal underlying Chlaß, Güth, and Miettinen's (2009) *purely procedural preferences*. However, these preferences cannot explain why some individuals who value the equality of rights prefer  $S_1$  while some prefer  $S_2$  where they give all payoff away.

*Preferences for power & control.* If  $B$  prefers to maintain power and control (Bartling, Fehr, and Herz 2014), she maximizes her utility by opting for interaction structure  $S_2$  where she exerts full power over the allocation. In  $S_2$ , she holds the exclusive right to decide and implements whatever allocation she prefers. Preferences for power and control can therefore not explain the differences in procedural choices and altruism in  $s_2$  across LIE, SPY, and SABOTAGE. Similarly, the finding that procedural choices and altruism in  $S_2$  should link to ethical ideals about the equality of individual rights suggests a simple preference for power is not at play<sup>34</sup>. Preferences for power can, however, explain why the exact same ethical ideal about the equality of rights underlies  $B$  participants' choices of  $S_1$ , and their altruism in  $S_2$ .  $B$  participants who prefer power and control prefer to opt into  $S_2$  and give payoff away to compensate  $A$  for her unequal rights; those who dislike exerting power would opt into  $S_1$  and actually grant  $A$  equal decision rights. Indeed, we find that  $B$  participants who likely value power – who score high on Klages's materialism values – rather opt into  $S_2$  whereas those who value the autonomy of the individual – Klages's postmaterialists – opt into  $S_1$ . This holds equally for treatments LIE and SABOTAGE where we elicit these values. The same logic applies if  $B$  participants' preferences for power would ultimately stem from a dislike of having others interfere with their own decisions (Neri and Rommelsperger 2014): in  $S_2$ , nobody can interfere with  $B$ 's decision and she can impose whatever allocation she prefers.

*Risk attitudes.* In both interaction structures  $S_1$  and  $S_2$ ,  $B$  chooses between the same ex-post payoffs – 100 ECU, or 0 ECU. Only in  $S_2$ , however, she can obtain 100 ECU for sure. Risk averse  $B$  participants would therefore always prefer  $S_2$ . Since  $B$  cannot obtain a higher ex-post payoff than these 100 ECU through incurring additional risk, also risk-loving or risk-neutral  $B$  participants prefer  $S_2$  where they take all payoff for sure. Risk attitudes can therefore not explain the variations of altruism across our LIE, SPY, or SABOTAGE treatments. Indeed, we could not confirm that risk attitudes explained  $B$  participants' choices of the interaction structures, or their altruism in LIE, SPY, or SABOTAGE.

*Experimenter demand effects.* Other than having addressed any of these preferences, we might— despite a strictly neutral framing — have induced a social experimenter demand effect (Zizzo 2010) in that the existence of an experimenter, or the awareness of participating

---

<sup>34</sup>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, Güth, and Miettinen 2009) [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.

in an experiment affected  $B$  participants' behaviour. If so, a significant share of them should be motivated by a desire to satisfy our expectations and to behave in a way which pleases us. If so,  $B$ s' behaviour should link to the extent by which they refer to others' (our own) expectations about their behaviour. We do not find that  $B$ s' preferences to refer to i) others' expectations, or ii) or to be taken as a nice person when deciding about the right and wrong of an action explain any part of our findings.

## 7 Conclusion

This paper studies by which degree, how, and why, individuals prefer to compete either fairly, or unfairly with an opponent for one ex-post nonzero payoff. In an experimental setting, one party chooses the rules of a constant sum game: she can opt into a set of rules where neither she, nor her opponent has information about the other's choice, and both parties have equal decision rights. She can also opt into a set of rules where she manipulates the consequences of her opponent's action (SABOTAGE), or spies the opponent's choice (SPY), or fabricates and reports this choice to a third party who makes this report payoff-relevant (LIE). A party may sabotage, spy, or fabricate to take all payoff, or to give all payoff away. The material incentive to do so is identical across SABOTAGE, SPY, and LIE.

Our results are first, that individuals resort more often to sabotage and spying than they resort to fabrication. Specifically when the game cannot be won for sure through fair competition, sabotage and spying attempts nearly double from 35% to 70%. Attempts to actively fabricate are comparatively rare and hardly respond to material incentives.

Second, the amount of altruism varies substantially depending on whether the set of rules allows for fabrication, spying, or sabotage. Specifically when only unfair competition allows individuals to obtain the non-zero payoff for sure, 68% of all individuals who fabricate information end up giving all payoff to their opponent, 71% of those who sabotage give all payoff away but *everybody* who spies does so to *take* all payoff.

Individuals who opt into fair competition and those who opt into unfair competition but end up giving all payoff away forego substantial amounts of payoff. To understand the motives underlying these departures from rational self-interest, we elicit the moral ideals which individuals invoke to judge whether an action is right or wrong in a moral judgement test developed by Lind (1978, 2008). The test includes the large set of ideals which Piaget (1948) and Kohlberg (1969, 1984) reported in their extensive field research on how individuals make moral judgements. We use this entire set of moral preferences for each individual to model her i) choice of the fair set of rules, or her ii) choice of the unfair set if she gives all payoff away and contrast each behaviour with those participants who compete unfairly to win the game. Surprisingly, both departures from rational self-interest link to the same moral ideal. The more an individual invokes the equality of civic rights and the social contract when judging

about the right or wrong of an action, the more likely she opts into fair competition, and the more likely she fabricates or sabotages to benefit the opponent. We conclude that fabrication and sabotage induce a psychological cost through infringing the opponent's position of rights and that individuals forego material payoff to rectify this infringement.

The key to understanding why these two types adopt different strategies to rectify the opponent's equal position of rights are their scores along the well-known materialism - post-materialism value scales. The more an individual values power and hierarchy (materialism), the more often they lie or sabotage to give all payoff to the opponent. The more individuals value individual autonomy and dislike power (postmaterialism), the more they prefer to grant their opponent the same rights and to compete fairly with her. Both types therefore seem to adopt different strategies to rectify the violation of the same moral ideal.

The only preference type to date which predicts the variation of altruism which we observe are Chlaß, Güth, and Miettinen's (2009) purely procedural preferences which describe inequity aversion over the distribution of decision and information rights: if only unfair competition wins the game for sure, fabrication and sabotage deplete the opponent's relative position in terms of decision rights; refraining from these activities reinstates the opponent's equal position of rights. Spying, however, does not take information rights away from the opponent since all activities are clandestine anyway, i.e. in our setting, the opponent does not know she is being spied. Therefore, an individual who refrains from spying does not rectify her opponent's equal position of information rights. If fair, and unfair competition can win the game for sure, an individual's decision to compete unfairly merely takes payoff-irrelevant decision rights from the opponent: in this case, fabrication and sabotage do not deteriorate the opponent's equal position of decision rights and no payoff need be foregone to compensate her. This is exactly what we observe.

The heterogeneity in individuals' attitudes toward unfair competition is so substantial that one may well entertain doubts whether competition selects the highest quality if such activities are possible at all: if a highly talented individual has strong reservations against sabotaging others while a less talented competitor has not and manages to successfully sabotage the former, competition will not correctly sort qualities, and have very different welfare effects than we have come to rely upon in economics.

## References

- Abbink, K. and H. Hennig-Schmidt (2006). “Neutral versus Loaded Instructions in a Bribery Experiment”. In: *Experimental Economics* 9 (2), pp. 103–121. ISSN: 1386–4157.
- 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.
- 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), pp. 269–280. ISSN: 0167–4544.
- 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., E. Fehr, M. A. Maréchal, et al. (2009). “Egalitarianism and Competitiveness”. In: *The American Economic Review* 99.2, pp. 93–98.
- Battigalli, P., G. Charness, and M. Dufwenberg (2013). “Deception: The Role of Guilt”. In: *Journal of Economic Behavior & Organization* 93, pp. 227–232.
- Battigalli, P. and M. Dufwenberg (2007). “Guilt in Games”. In: *The American Economic Review* 2.97, pp. 170–176.
- 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.
- 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, A. W., E. Ø. Sørensen, and B. 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.

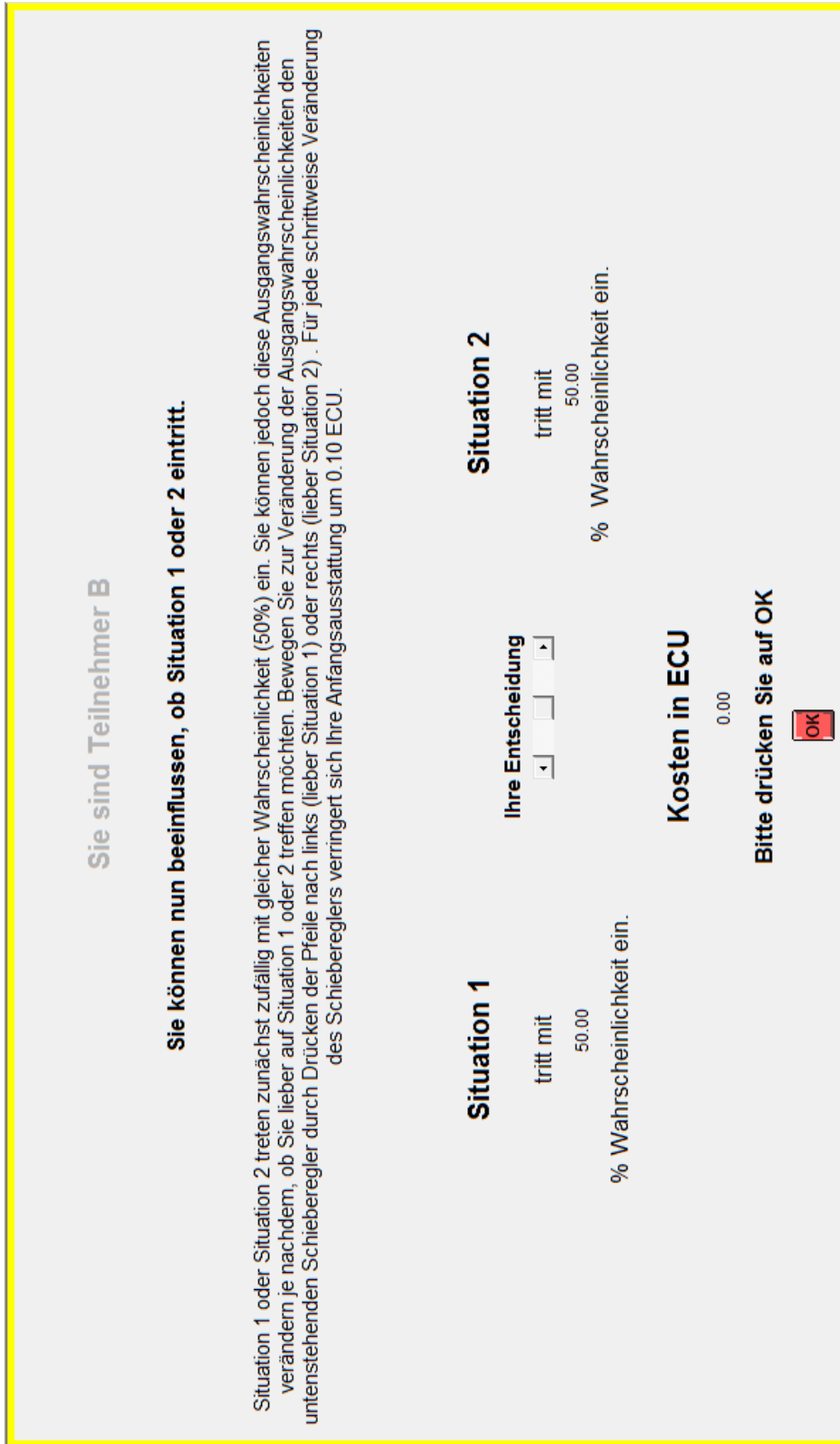
- Chlaß, N., W. Güth, and T. Miettinen (2009). *Purely Procedural Preferences – Beyond Procedural Equity and Reciprocity*. Tech. rep. 2014-03. Stockholm Institute of Transition Economics.
- Chlaß, N. and P. Moffatt (2012). *Giving in Dictator Games, Experimenter Demand Effect, or Preference over the Rules of the Game?* Tech. rep. 2012-44. Jena Economic Research Papers.
- Chlaß, N. and G. Riener (2015). *Participation Crowds out Altruism*. Tech. rep. mimeo.
- Conrads, J. et al. “Honesty in Tournaments”. In: *Economics Letters* 123, pp. 90–93.
- Cooper and Roberts (2011). “After 40 Years, the Complete Pentagon Papers”. In: *The New York Times* 2011-06-07.
- 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.
- Falk, A., E. Fehr, and D. Huffman (2008). *The Power and Limits of Tournament Incentives*. Tech. rep. 2008. University of Zurich.
- Fischbacher, U. (2007). “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments”. In: *Experimental Economics* 10.2, pp. 171–178.
- 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.
- Greenwald, G., E. MacAskill, and L. Poitras (June 2013). “Edward Snowden: the Whistleblower Behind the NSA Surveillance Revelations”. In: *The Guardian* 11.
- Greiner, B. (2004). “An online recruitment system for economic experiments”. In: *Forschung und wissenschaftliches Rechnen* 63. Ed. by K. Kremer and V. Macho.
- Guiso, L., P. Sapienza, and L. Zingales (2006). “Does Culture Affect Economic Outcomes?” In: *The Journal of Economic Perspectives* 20, pp. 23–48.
- Harbring, C. and B. Irlenbusch (2011). “Sabotage in Tournaments: Evidence from a Laboratory Experiment.” In: *Management Science* 57.4, pp. 611–627.
- Harbring, C., B. Irlenbusch, et al. (2007). “Sabotage in Asymmetric 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.



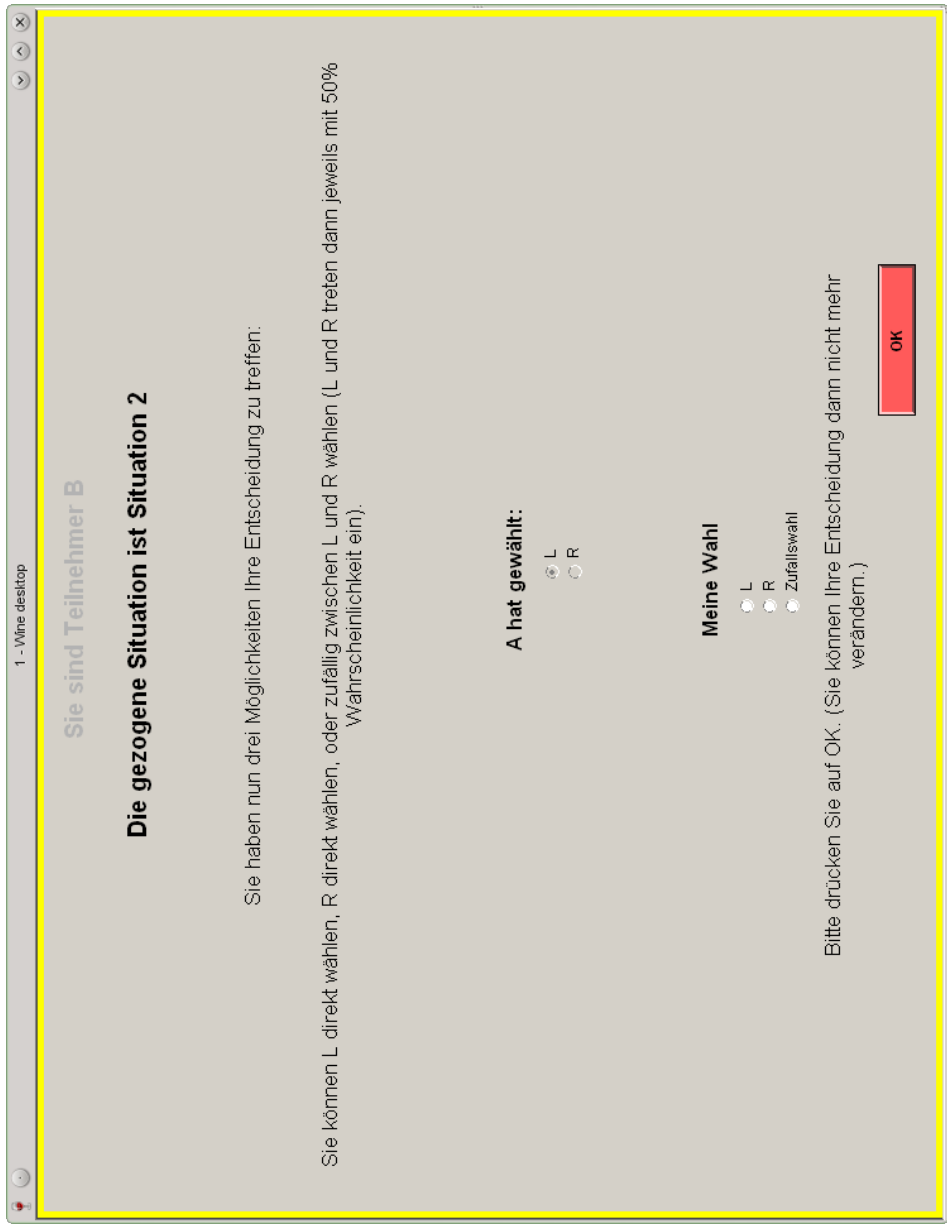
- Inglehart, R. (1977). *The Silent Revolution: Changing Values and Political Styles Among Western Publics*. Princeton University Press.
- Ishida (2006). “How do Scores of the DIT and the MJT differ? A Critical Assessment of the Use of Alternative Moral Development Scales in Studies of Business Ethics.” In: *Journal of Business Ethics* 67, pp. 63–74.
- 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.
- 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.
- Korsgaard, M. A., D. M Schweiger, and H. J. Sapienza (1995). “Building Commitment, Attachment, and Trust in Strategic Decision–Making Teams: The Role of Procedural Justice”. In: *Academy of Management Journal* 38.1, pp. 60–84.
- Larson, J., N. Perloth, and S. Shane (Sept. 2013). “Revealed: The NSA’s Secret Campaign to Crack, Undermine Internet Security”. In: *New York Times* 5.
- Lightle, J. P. (2014). “The Paternalistic Bias of Expert Advice”. In: *Journal of Economics & Management Strategy* 23.4, pp. 876–898.
- 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.
- (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.
- Lind, Georg (2002). *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung*. 2nd ed. Logos Verlag.
- 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.
- Miettinen, T. (2013). “Promises and Conventions – An Approach to Pre-Play Agreements”. In: *Games and Economic Behaviour* 80.80, pp. 68–84.
- Milinski, M. and B. Rockenbach (July 2007). “Spying on Others Evolves”. English. In: *Science*. New Series 317.5837, pp. 464–465. ISSN: 00368075.

- 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.
- 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.
- Solan, E. and L. Yariv (2004). "Games with Espionage". In: *Games and Economic Behavior* 47.1, pp. 172–199.
- Sutter, M. (2009). "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams". In: *The Economic Journal* 119.534, pp. 47–60.
- Tabellini, G. (2010). "Culture and Institutions: Economic Development in the Regions of Europe". In: *Journal of the European Economic Association* 8.4, pp. 677–716.
- Thaler, R. and C. Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
- Whitfield, J. (2002). "Nosy Neighbors". In: *Nature* 419, pp. 242–243.
- Zizzo, D. J. (2010). "Experimenter Demand Effects in Economic Experiments". In: *Experimental Economics* 13.1, pp. 75–98.

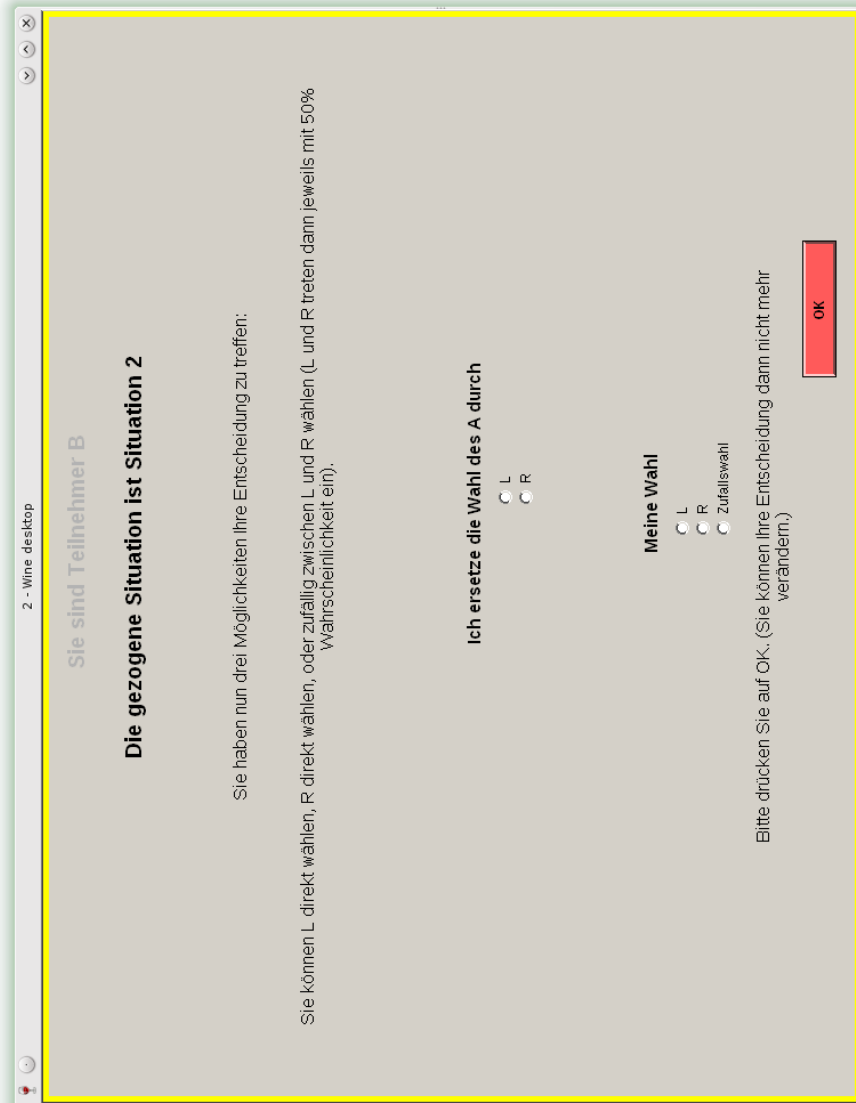
## A Screenshots



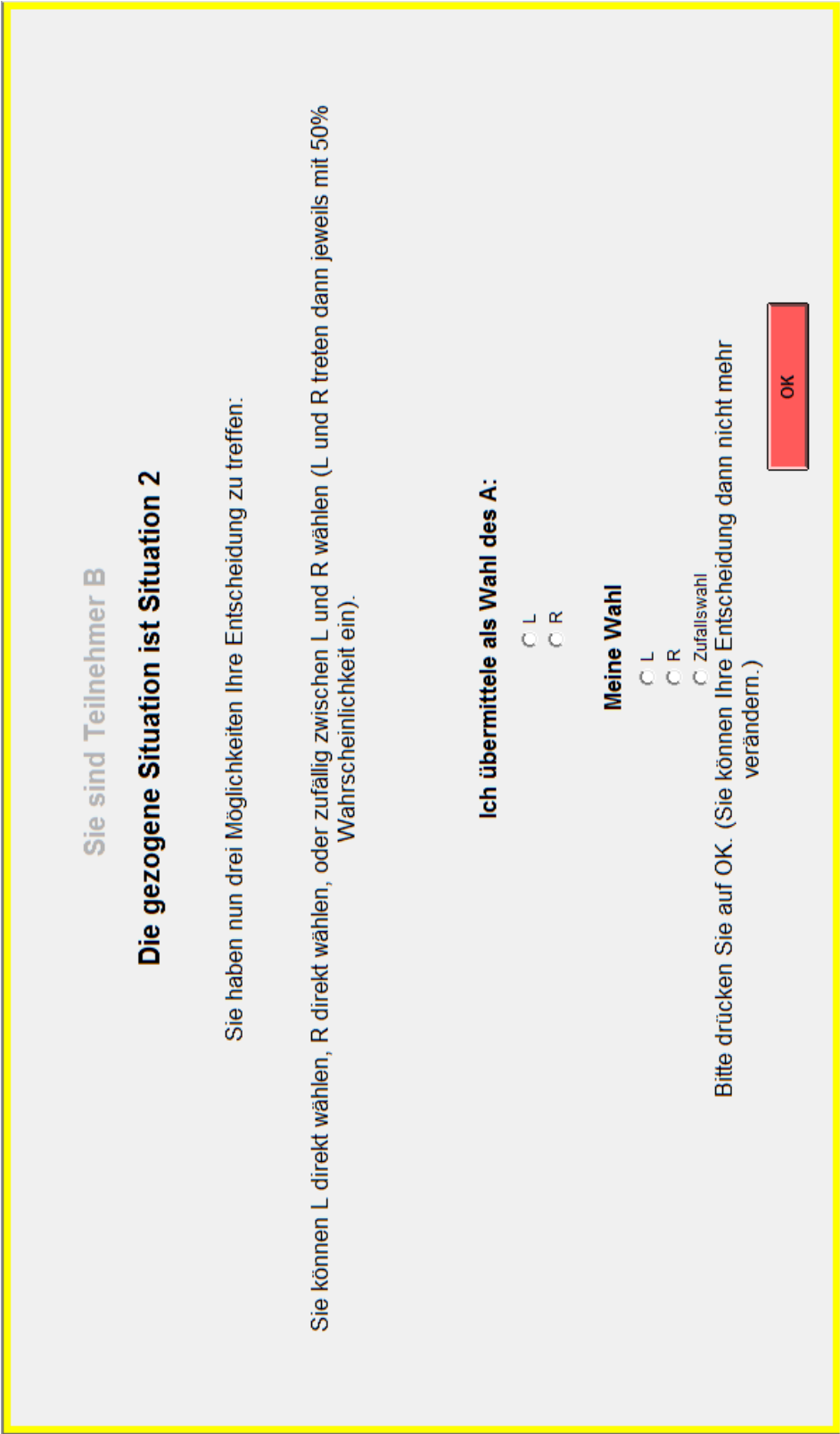
**Figure 6:** *B*'S PROBABILITY CHOICE OF THE SITUATION



**Figure 7:** *B*'S DECISION SCREEN IN THE UNFAIR SET OF RULES, TREATMENT SPY.



**Figure 8:** *B*'S DECISION SCREEN IN THE UNFAIR SET OF RULES, TREATMENT SABOTAGE.



**Figure 9:** B's DECISION SCREEN IN THE UNFAIR SET OF RULES, TREATMENT LIE.

## B Normal form representation of the payoff neutral regime.

**Table 8:** 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.

8a) the 'fair' set of rules

		party A	
		$L$	$R$
party B	$L$	100 0	100 0
	$R$	100 0	100 0

8b) the 'unfair' set of rules

		party A	
		$L$	$R$
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

## C Normal form representation of the competitive payoffs regime with symbolic reward and punishment (Chlaß and Riener 2015).

**Table 9:**  $A$ 'S SYMBOLIC PUNISHMENT AND REWARD OPTION MAKES HER RELATIVE POSITION OF DECISION RIGHTS MORE EQUAL TO  $B$ 'S:  $A$  CAN REDUCE (OR INCREASE) THE EXTENT TO WHICH  $B$  PREFERS  $L$  OVER  $R$  BY 30 ECU, AND REDUCE/INCREASE THE EXTENT TO WHICH  $B$  PREFERS  $RL^A$  OR  $LR^A$  OVER  $LL^A$  AND  $RR^A$  BY 30 ECU IN  $S_2$ . (IBID.)

9a) the 'fair' set of rules

		party A	
		$L$	$R$
party B	$L$	100 - [0, 30] 0 + [-30, 30]	100 - [0, 30] 0 + [-30, 30]
	$R$	0 - [0, 30] 100 + [-30, 30]	0 - [0, 30] 100 + [-30, 30]

9b) the 'unfair' set of rules

		party A	
		$L$	$R$
party B	$LL^A$	100 - [0, 30] 0 + [-30, 30]	100 - [0, 30] 0 + [-30, 30]
	$RL^A$	0 - [0, 30] 100 + [-30, 30]	0 - [0, 30] 100 + [-30, 30]
	$LR^A$	100 - [0, 30] 0 + [-30, 30]	100 - [0, 30] 0 + [-30, 30]
	$RR^A$	0 - [0, 30] 100 + [-30, 30]	0 - [0, 30] 100 + [-30, 30]

## D Defining sabotage: Max and Moritz (Busch 1906).

Figure 10: MAX AND MORITZ FILL THEIR TEACHER'S PIPE WITH BLACK POWDER.

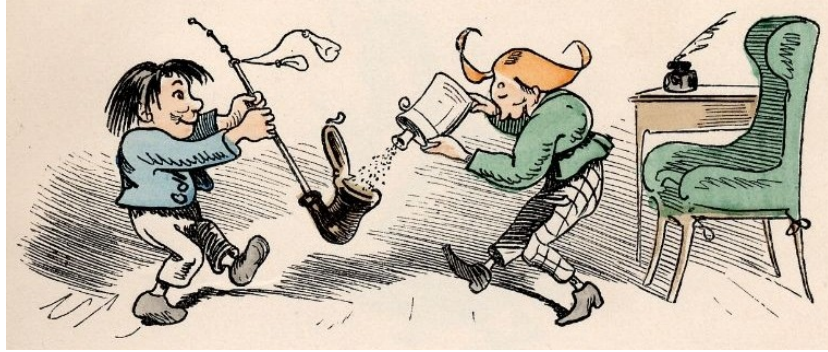
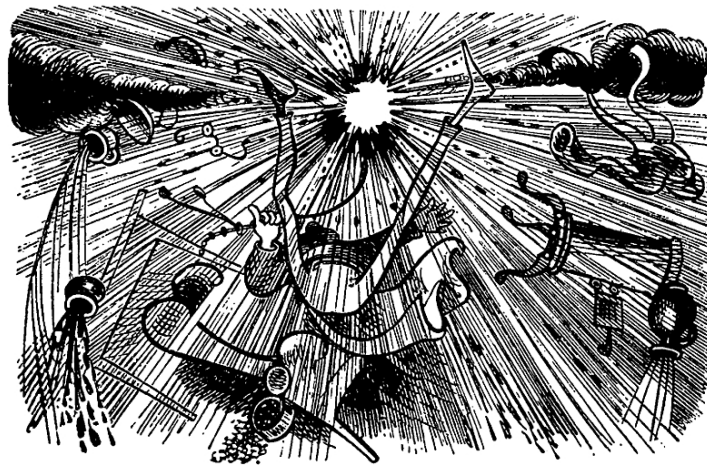


Figure 11: LIGHTING THE PIPE HAS NOW A NEW CONSEQUENCE FOR THE TEACHER.





## E Experimental Results: Absolute figures

Number of  $B$ -participants paying for interaction structure  $S_1$  ('fair') and  $S_2$  ('unfair') per treatment

treatment	LIE <sup>35</sup>				SPY				SABOTAGE			
	payoff neutral		competitive		payoff neutral		competitive		payoff neutral		competitive	
payoff regime	#47		#44		#53		#53		#52		#54	
#nr. of $B$ players												
interaction structure	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$	$S_1$	$S_2$
% who pays	8	3	9	5	2	19	5	36	4	18	2	37
median change of $\alpha$	13%	30%	10%	20%	25%	20%	50%	25%	17.50%	20%	16%	25%
% who does not pay	36		30		32		12		30		15	

**Table 10:** CHOICES OVER PROCEDURES FOR ALL TREATMENTS.

Which allocation do  $B$ -participants impose when they hold the power to do so?  
 selfish: (payoff B: 100, payoff A: 0); altruistic: (payoff B: 0, payoff A: 100)

treatment	LIE <sup>36</sup>			SPY			SABOTAGE		
	payoff neutral		competitive	payoff neutral		competitive	payoff neutral		competitive
payoff regime	$S_1$	$S_2$	$S_2$	$S_1$	$S_2$	$S_2$	$S_1$	$S_2$	$S_2$
interaction structure	$S_1$	$S_2$	$S_2$	$S_1$	$S_2$	$S_2$	$S_1$	$S_2$	$S_2$
# nr. of $B$ players.	#25	#22	#25	#20	#33	#40	#22	#30	#28
selfish	20	17	8	18	31	40	18	26	8
equal chance	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
altruistic	5	5	17	2	2	0	4	4	20

**Table 11:**  $B$ 'S CHOICES OF THE PAYOFF ALLOCATION IN THE 'FAIR' ( $S_1$ ) AND THE 'UNFAIR' ( $S_2$ ) INTERACTION STRUCTURES.

<sup>35</sup>A brief reading example: In treatment LIE with neutral payoffs, there were 47  $B$  participants. Eight of them paid for  $S_1$  and three for  $S_2$ . The eight who paid for  $S_1$  made at the median,  $S_1$  13% more likely than  $S_2$ . The three who paid for  $S_2$ , made, at the median,  $S_2$  30% more likely than  $S_1$ . 36 of 47  $B$  participants left the default 50-50 chance of arriving in either  $S_1$  or  $S_2$ .

<sup>36</sup>A brief reading example: In treatment LIE with payoff neutrality  $B$  can impose her preferred allocation in  $S_1$  and  $S_2$ . Out of 47  $B$  participants, 25 arrived in  $S_1$ . 20 of them kept all payoff for themselves, five gave all payoff away, and nobody tossed a coin. The remaining 22  $B$  participants arrived in  $S_2$ . 17 of them kept all payoff, five gave all payoff away, nobody tossed a coin. Under competitive payoffs,  $B$  can only impose the allocation in  $S_2$ . Out of 44  $B$  participants, 25 arrived in  $S_2$ , eight of which kept all payoff, and seventeen of which gave all payoff away. Nobody tossed a fair coin.

## F Kohlberg's six ways of moral argumentation

**Table 12: 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	<b>Class 1.</b> Orientation to punishment and obedience, physical and material power. Rules are obeyed to avoid punishment. <b>Class 2.</b> 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	<b>Class 1.</b> "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". <b>Class 2.</b> 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	<b>Class 1.</b> 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. <b>Class 2.</b> 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.

## G 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.

## H Klages's and Gensicke's (2006) materialism - postmaterialism scales<sup>37</sup>

**Table 13:** QUESTIONNAIRE ITEMS FOR EACH OF KLAGES'S AND GENSIKKE'S THREE VALUE DIMENSIONS (CATEGORIES) TO IDENTIFY MATERIALISTS, POSTMATERIALISTS, AND MIXED VALUE TYPES IN THE GERMAN POPULATION (KLAGES AND GENSIKKE 2006).

value category I duty and acceptance values	value category II hedonistic and materialistic values	value category III idealistic values and public participation <sup>38</sup>
✓ 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'.

<sup>37</sup>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.

<sup>38</sup>Category III corresponds to Inglehart'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

## I B participant types: do demographics, or other moral preferences play a significant role?<sup>39</sup>

PROCEDURAL TYPE (I)			ALTRUISTIC TYPE (II)		
<i>Argument</i>	<i>Effect</i>	<i>std.err.</i>	<i>Argument</i>	<i>Effect</i>	<i>std.err.</i>
<i>lie</i>	0.49	0.17 <sup>a</sup>	<i>lie</i>	0.50	0.04 <sup>a</sup>
<i>postclass 1</i>	0.16	0.08 <sup>b</sup>	<i>sabotage</i>	0.49	0.05 <sup>a</sup>
<i>postclass 2</i>	-0.15	0.07 <sup>b</sup>	<i>postclass 1</i>	0.10	0.03 <sup>a</sup>
[ <i>risk aversion</i>	-0.04	0.04 ]	[ <i>risk aversion</i>	0.01	0.02 ]
[ <i>Age</i>	0.00	0.02 ]	[ <i>Age</i>	0.00	0.01 ]
[ <i>Gender:male</i>	0.08	0.12 ]	[ <i>Gender:male</i>	0.04	0.06 ]
[ <i>Envy</i>	0.06	0.12 ]	[ <i>Envy</i>	0.02	0.06 ]
[ <i>sabotage treatment</i>	0.05	0.16 ]	[ <i>Kohlberg class 1</i>	-0.10	0.04 <sup>b</sup>
[ <i>Kohlberg class 1</i> <sup>40</sup>	-0.20	0.21 ]	[ <i>Kohlberg class 2</i>	-0.01	0.05 ]
[ <i>Kohlberg class 2</i>	0.09	0.12 ]	[ <i>Kohlberg class 3</i>	0.05	0.04 ]
[ <i>Kohlberg class 3</i>	0.11	0.11 ]	[ <i>Kohlberg class 4</i>	0.05	0.05 ]
[ <i>Kohlberg class 4</i>	0.02	0.12 ]	[ <i>Kohlberg class 5</i>	-0.00	0.05 ]

**Table 14:** WHICH DETERMINANTS MAKE THE PROCEDURAL TYPE (I) (N=56), AND THE ALTRUISTIC TYPE (II) (N=121) MORE LIKELY THAN THE MOST SELFISH TYPE (IV)?

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

PROCEDURAL TYPE (I) WITH (POST)-MATERIALISM SCORES		
<i>Argument</i>	<i>Effect</i>	<i>std.err.</i>
<i>postclass 1</i>	0.48	0.11 <sup>a</sup>
<i>postclass 2</i>	-0.40	0.11 <sup>a</sup>
<i>materialism</i>	-0.10	0.05 <sup>b</sup>
<i>postmaterialism</i>	0.20	0.04 <sup>a</sup>

**Table 15:** MODELING THE PROCEDURAL TYPE I) VS THE MOST SELFISH TYPE IV) ADDING B PARTICIPANTS' MATERIALISM AND POSTMATERIALISM SCORES, WHERE AVAILABLE (N=19)

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

category I does not significantly influence B participants' choices in the experiment.

<sup>39</sup>The core model is a joint estimation of all variables without brackets. In brackets, we see which coefficients and significance levels would result if we jointly added risk attitudes, all demographics, all other Kohlbergian classes, and the sabotage dummy to the core model. Naturally, this extended model has higher variance, i.e. less precision, than the core model and the insignificance of additional controls might be due to this fact. However, none of the additional variables in brackets would have a significant effect if it were added by itself, or in small groups with other controls, to the core model. Hence, the insignificance of all additional controls does not result from the inefficiency of the estimation.

<sup>40</sup>Turns insignificant if we start deleting other insignificant variables and is not significant if added to the core model.

FAIR-COIN TYPE (III) WITH (POST)-MATERIALISM SCORES		
<i>Argument</i>	<i>Effect</i>	<i>std.err.</i>
<i>postclass 1</i>	0.15	0.07 <sup>b</sup>
<i>materialism</i>	-0.15	0.03 <sup>a</sup>
<i>postmaterialism</i>	0.14	0.06 <sup>b</sup>
<i>risk aversion</i>	0.06	0.04

**Table 16:** WHICH DETERMINANTS MAKE TYPE III) WHO TOSSES A FAIR COIN BETWEEN THE INTERACTION STRUCTURES MORE LIKELY THAN THE MOST SELFISH TYPE IV) WITH *B* PARTICIPANTS' MATERIALISM AND POSTMATERIALISM SCORES WHERE AVAILABLE (N=16)

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

## J Purely Procedural Concerns crowd out under punishment/reward<sup>41</sup> (Chlaß and Riener 2015).

<i>Argument</i>	PROCEDURAL TYPE (I)		ALTRUISTIC TYPE (I)		FAIR COIN TYPE (III)	
	<i>Effect</i>	<i>std.err.</i>	<i>Effect</i>	<i>std.err.</i>	<i>Effect</i>	<i>std.err.</i>
<i>Kohlberg 1</i>	-0.16	0.04 <sup>a</sup>	(-)	(-)	-0.10	0.04 <sup>b</sup>
<i>Kohlberg 3</i>	0.20	0.09 <sup>b</sup>	(-)	(-)	(-)	(-)
<i>Kohlberg 4</i>	0.14	0.06 <sup>b</sup>	0.11	0.05 <sup>b</sup>	(-)	(-)
<i>postclass 1</i>	-0.17	0.11 <sup>b</sup>	-0.15	0.05 <sup>a</sup>	0.03	0.05
<i>expected punishment</i>	0.08	0.04 <sup>c</sup>	0.17	0.04 <sup>a</sup>	0.35	0.10 <sup>a</sup>
<i>expected reward</i>	(-)	(-)	-0.07	0.04 <sup>b</sup>	-0.14	0.05 <sup>a</sup>
<i>lie</i>	(-)	(-)	0.56	0.05 <sup>a</sup>	(-)	(-)
<i>sabotage</i>	(-)	(-)	0.25	0.08 <sup>a</sup>	(-)	(-)

**Table 17:** CONTRASTING THE PROCEDURAL TYPE I), THE ALTRUISTIC TYPE II), AND THE FAIR COIN TYPE III) WITH THE MOST SELFISH TYPE IV) WHEN *A* CAN PUNISH OR REWARD *B*'S PROCEDURAL CHOICE.

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

<sup>41</sup>Binary logit models where the dependent variable is a pair of types: either type (I) vs the most selfish type (IV), or type (II) vs type IV) or type (III) vs type (IV). Kohlberg 1,3, and 4 correspond to the Kohlbergian ways of argumentation in classes 1, 3, or 4 from table 12 in section F. Variables which are insignificant and not of interest have been deleted from the specification, variables which have an effect on some, but not all types are marked with (-) when they are insignificant.