

This is the final peer-reviewed accepted manuscript published in JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION (Elsevier): *When Foul Play Seems Fair: Exploring the Link between Just Deserts and honesty*, Fabio Galeotti, Reuben Kline, Raimondello Orsini

The final published version is available online at:

<https://doi.org/10.1016/j.jebo.2017.08.007>

© [2017]. This manuscript version is made available under the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) 4.0 International License

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

(This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>))

***When citing, please refer to the published version.***

# When Foul Play Seems Fair: Exploring the Link between Just Deserts and Honesty\*

Fabio Galeotti<sup>†</sup> Reuben Kline<sup>‡</sup> Raimondello Orsini<sup>§</sup>

August 2, 2017

## Abstract

The distributive justice norm of “just deserts”—i.e. the notion that one gets what one deserves—is an essential norm in a market society, and honesty is an important factor in economic and social exchange. We experimentally investigate the effect of violations of the distributive justice norm of “just deserts” on honesty in a setting where behaving dishonestly entails income redistribution. We find that the violation of the just deserts norm results in a greater propensity toward dishonesty. We then test a more general proposition that violations of just deserts induce dishonesty, even in cases where dishonesty does not have redistributive consequences. Our results confirm this proposition but only for cases in which the violation of just deserts also entails income inequality.

JEL-Classification: C91, D03, D31, D63

Keywords: Meritocracy, Equity, Dishonesty, Just Deserts, Experiment

---

\*We wish to thank Joachim Winter, an anonymous referee, and Stefania Bortolotti, Marco Casari, Caterina Giannetti, Daniel Houser, Kevin McCabe, and Valeria Maggian for their comments and suggestions on earlier versions of this paper. We also thank seminar participants at George Mason, the Beelab 2013 Workshop (Florence), the 8th Conference of the French Association of Experimental Economics (Rennes), the 12th Workshop on Social Economy for Young Economists (Forlì), the NorMAS International Workshop (Leiden), the 2012 ESA International Meeting (New York), the 2012 Antigua Experimental Economics conference and the 2013 NYU-CESS Experimental Political Science Conference for useful feedback and advice. Funding: This work was supported by the Einaudi Institute for Economics and Finance (Grants 2013); the Russell Sage Foundation’s Small Grants in Behavioral Economics; MIUR—FIRB Project SONIC—A behavioral approach to social norms and intertemporal choices (n. RBFR084L83001); University of Bologna—FARB research project: “Mortality Saliency, Conformità a Norme Giuridiche e Sociali, Comportamenti Economici: Modelli Teorici e Metodologie Sperimentali”; and CFICEI. This research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d’Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). The usual disclaimer applies.

<sup>†</sup>Univ Lyon, Université Claude Bernard Lyon 1, Centre Léon Bérard, GATE L-SE UMR 5824, F-69008 Lyon, France; E-mail: galeotti@gate.cnrs.fr

<sup>‡</sup>Stony Brook University, Department of Political Science Social and Behavioral Sciences, N735 Stony Brook, NY 11794, E-mail: reuben.kline@stonybrook.edu

<sup>§</sup>Corresponding author. University of Bologna, Dipartimento di Scienze Economiche, Strada Maggiore 45, 40125 Bologna, Italy; E-mail: raimondello.orsini@unibo.it

“If people feel that they are taken advantage of, why should they not rip off the system in return?”

---

— Elster (1989)

## 1 Introduction: Honesty and Just Deserts

Norms play a crucial role in the functioning of any socio-economic system. In particular the norms of trust and honesty are important norms for economic, political and legal development (Guiso et al., 2008; Uslaner, 2008; Rose-Ackerman, 2001; Rothstein and Uslaner, 2005; Zak and Knack, 2001; Mazar and Ariely, 2006) as they reduce the riskiness of market transactions in a world of uncertainty and incomplete contracts (Robert and Arnab, 2013). While trust and honesty help facilitate market exchange, a market economy paradigmatically relies on a meritocratic equity norm or *just deserts* —i.e. the notion that one gets what one deserves—to give legitimacy to the inequality of the distribution of income in society (Miller, 1979; Arnold, 1987; Mankiw, 2010). The presumption of this norm is that one deserves the income one has, as a function of merit. We define merit as productivity (the result of skill plus effort), adopting the economic approach of rewarding merit on the basis of consequences: actions are to be rewarded for the good they do, so that incentives are well defined to produce a more efficient society in the long run. While the belief in the extent to which just deserts *should* play a role in the distribution of income within a society varies across individuals in large part as a function of political ideology (Lewin-Epstein et al., 2003; d’Anjou et al., 1995; Mitchell et al., 2003), and context (Scott and Bornstein, 2009), most individuals in advanced industrial societies support the principle of merit as one (perhaps among several) criterion in decisions regarding distributive justice (Scott et al., 2001; Kunovich and Slomczynski, 2007).

In this study, we investigate the relationship between these norms: honesty on the one hand, and just deserts on the other. Our principal hypothesis is that violations of the norm of just deserts encourage and justify dishonest behavior as a corrective mechanism against meritocratic inequity, inducing those who feel that the system is inequitable, to “rip it off” in turn. To test our hypothesis, we conduct two earned-income laboratory experiments in which we manipulate the just deserts and inequality of the payoffs from a real effort task as a function of performance relative to the median. In Experiment 1, we focus on a specific setting where dishonesty is instrumental to redistribute income between a person who benefited from inequity and another who did not. The aim is to study the interplay

between inequity, dishonesty and distributional concerns.<sup>1</sup> In the experiment, we pair high-performing subjects with low-performing subjects, allowing the high performers to record the outcomes of a series of private signals in which a misrepresentation of the signals can materially benefit the high performer at the expense of the low performer. High-performing but rule-disadvantaged subjects in the inequitable conditions are expected to behave more dishonestly in order to restore distributional justice. In Experiment 2, we test a more general proposition that the mere experience of inequity induces more dishonesty quite independently from re-distributional concerns. The experimental setting is analogous to the one used in Experiment 1 except that subjects are not matched in pairs. Hence, in Experiment 2, the misrepresentation of the signals does not affect others' earnings.

Our experimental design brings together two growing trends in experimental and behavioral economics: experiments on dishonesty and lying behavior, the majority of which follow the paradigms laid out in Gneezy (2005) and Fischbacher and Föllmi-Heusi (2013), and real-effort experiments in which income (or a privileged role) is “earned” by performance on a quiz or task embedded in the experiment (Hoffman et al., 1994; Ruffle, 1998; Rutström and Williams, 2000; Jakiela, 2011; Esarey et al., 2012; Balafoutas et al., 2013; Durante et al., 2014).

The rest of the paper is organized as follows. In section 2, we briefly survey the related literature. In section 3, we describe the experimental design and the behavioral predictions of Experiment 1. In section 4 we show and discuss the results. Section 5 reports the experimental design, predictions and results of Experiment 2. Section 6 concludes.

## 2 Related Literature

There is extensive evidence that individuals care not only about the nature of the final distribution of income but also about the procedures that brought it about (e.g. Rabin, 1993; Hoffman et al., 1994; Konow, 1996, 2000; Frey et al., 2004; Bolton et al., 2005; Almås et al., 2010; Dal Bó et al., 2010; Jakiela, 2011; Grimalda et al., 2016). There are numerous theoretical and empirical studies of meritocracy in the behavioral economic tradition, with a number of theoretical conceptions of merit-based equity having been put forth, often combined with experimental tests.

Konow (1996, 2000) develops the “accountability principle” in which individuals are to be held responsible for outcomes under their control but not for factors outside of their control.

---

<sup>1</sup>Previous studies focused on the relationship between procedural fairness and distributional preferences (see next section). The main novelty of our experiment is to test whether this relationship holds in a context where subjects must forgo the honesty norm in order to restore equity.

Schokkaert and Devooght (2003) and Ooghe et al. (2007) develop a similar concept they call “responsibility-sensitive egalitarianism”. Cappelen et al. (2007, 2010) manipulate the sources of rewards (i.e., whether they are earned or a result of luck) to test meta-preferences for equity, such as strict egalitarianism, libertarianism and liberal egalitarianism, the latter of which is functionally equivalent to Konow’s accountability principle. Though they find heterogeneous preferences among their experimental subjects, there is a preponderance of “liberal egalitarians.” Almås et al. (2010) demonstrate, through the use of a modified dictator game, that acceptance of equitable inequality (i.e. inequality generated by differential returns to effort and achievement) tends to develop during early adolescence. Taken together, these results demonstrate that just deserts is an important criterion for establishing the equitability of income distributions. Our study builds on these to investigate the link between (un)just deserts and (dis)honesty.

In the laboratory, meritocracy is typically operationalized by assigning initial endowments and/or experimental roles based on performance on a task or quiz embedded within the experiment. In these designs high performance is rewarded with either a larger endowment or a privileged experimental role. This condition is then typically contrasted with a condition in which the endowments or roles are randomly assigned. Among the first to use quizzes and/or tasks—as opposed to chance—to assign roles and/or initial endowments in an experimental setting, Hoffman et al. (1994) find that assigning the role of proposer in the Ultimatum game based on high performance on a quiz, causes responders to accept more unequal income allocations. Burrows and Loomes (1994) find that whether group income was earned or randomly assigned affects behavior in bargaining over the final allocation of payments. Ruffle (1998) finds that in ultimatum and dictator games proposers reward the effort that recipients put forth in an experimental task. This large, interdisciplinary literature offers considerable evidence that just deserts is an important, though not exclusive, determinant of distributive preferences, and that it can be meaningfully operationalized and manipulated in the lab.

There is also a rapidly growing interest in the behavioral economics of cheating and dishonesty. A large literature has grown out of a paradigm developed in Fischbacher and Föllmi-Heusi (2013). In this paradigm, subjects roll a die in private, the outcome of which determines their earnings (and/or the earnings of another subject) in one way or another. The subjects, however, are free to misreport the outcome of the die roll. Researchers are then able to determine whether certain outcomes of the die roll are observed with greater frequency than others. What is often observed is that while subjects, in the aggregate, do misrepresent the outcome, they do not do so in a way which maximizes their earnings. This paradigm has, in a short time, generated a large amount of literature (Buccioli and Piovesan, 2011; Shalvi et al., 2011; Suri et al., 2011; Shalvi et al., 2012; Dai et al., 2017, among many

others).

Several papers have made apparent that lies have a moral and psychological cost. Gneezy (2005) reports experimental evidence indicating that people do not like to lie. Erat and Gneezy (2012) show that many people may dislike lying even in situations where such deception brings material gains for everyone. Cappelen et al. (2013) argue that people face different costs of lying, while Gneezy et al. (2013) propose an experimental design explicitly aimed at measuring subjects' degree of lying aversion. The moral cost of lying seems to be context dependent, being influenced by different factors, such as the distribution of information, the material consequences of lying, the belief to be believed or not by partners, etc. We propose to add the perception of just deserts to the list of factors affecting the moral and psychological costs of breaking the "honesty social norm." We operationalize this factor in our experiment by manipulating the equity of the remuneration system used to allocate resources among the experimental subjects as a function of performance on a real effort task.

In our design, the experimenter has complete information regarding the true outcomes, and therefore rather than rely on deviations from expected frequencies, we know whether dishonesty has occurred in each specific instance. This allows us to conduct an individual-level analysis of the data. The potential disadvantage is that subjects might be aware of the fact that cheating can be observed, and this could affect behavior, by perhaps increasing the cost of lying. This should however equally apply to all our treatments, and only influence the absolute level of dishonest behavior and not the relative comparison between our treatments. In addition, previous studies find that the effect of having the experimenter or an anonymous observer to verify behavior is only minor (e.g. Bäker and Mechtel, 2015; van de Ven and Villeval, 2015; Houser et al., 2016; Gneezy et al., 2016). There are few recent studies that employ variants of the Fischbacher and Föllmi-Heusi (2013)'s task with a "full information" set-up similar to ours. Maggian and Villeval (2016) investigate altruistic and selfish lies in children. Like us, they ask subjects to truthfully or untruthfully report the outcome of a randomly generated binary signal. Kocher et al. (2016) study dishonesty in groups by asking subjects to watch one of six videos of a die roll on a computer and report the outcome (which is known by the experimenter). To study the structure of the intrinsic cost of lying, Gneezy et al. (2016) ask participants to report the number that appears behind one of ten boxes in their computer screen. A similar approach is used by Abeler et al. (2016) who ask participants to take coins from an envelope based on a randomly drawn chip in the computer. These studies, though using a similar dishonesty task to ours, do not investigate the relationship between dishonesty and just deserts.

Within the literature on cheating and lying behavior, Gino and Pierce (2009, 2010) and Houser et al. (2012) are closely related to our study. Gino and Pierce (2009, 2010)

investigate whether a subject is more or less likely to dishonestly help or hurt another individual depending on the relative wealth of the two subjects. In Gino and Pierce (2009), the wealth levels are manipulated through an initial lottery. Hence, unlike our study, the income distribution is not a function of merit. In Experiment 2 of Gino and Pierce (2010), the initial wealth of the subjects is determined by a third-person subjective evaluation of their creativity in a writing task, and, therefore, the link with merit is questionable. In addition, high-evaluated subjects always receive a higher payment than low-evaluated subjects. Hence, equity and equality are confounded. Finally, the subject who behaves dishonestly affects the other subject's payoff but he/she receives a flat fee irrespectively of his/her behavior.

Houser et al. (2012) study whether people who are treated unfairly in a dictator game are more likely to cheat in a subsequent task. They find that subjects who are given a small allocation in a dictator game are more likely to (self-servingly) cheat when reporting the outcome of a subsequent coin flip, thereby increasing their total payoff. Their study does not involve earned income manipulations, thus they posit that the increase in dishonest behavior is a result of a perceived violation of an *egalitarian* norm. Our study employs a similar design to investigate whether dishonesty can result from perceived violations of other equity norms, in particular the norm of just deserts.

Finally, there is a literature that looks at the effects of unfair treatment and bad intentions on cheating in the workplace. Greenberg (1990) finds that employees who experience a cut in their wage display higher theft rates, especially when the pay cuts are not adequately explained. In a laboratory setting, Greenberg (1993) finds that subjects who are told that they would be given a lower wage than a promised one take more money when they are asked to pay themselves. In both cases, the focus is not on just deserts but on whether a pay lower than a prior or promised one induces (more) stealing. Gill et al. (2013) investigate experimentally the effects of random bonuses on cheating and productivity. The authors find that random bonuses increase cheating behavior possibly because they are perceived as unfair by the subjects. John et al. (2014) find that people who get lower pay-rates than others, and are aware of this, cheat more. In both studies, the payment is determined randomly and there is no rationale based on merit that justifies the difference in payments between subjects. This is not the case in our experiment where differences in payment are determined by prior performance.

### 3 Experiment 1

In this section, we describe the experimental design and then we state the main hypotheses of our first experiment.

### 3.1 Participants and Procedures

The experimental sessions were conducted at Stony Brook University (SBU) between April 2012 and February 2013. We recruited 164 subjects using the on-line recruitment system ORSEE (Greiner, 2015). At the end of each session subjects were paid in cash and in private at the subjects' stations. Average payment was \$17.06. Each session lasted around 40-60 minutes, including instructions and final payments. The experiment was computerized and programmed using the software z-Tree (Fischbacher, 2007). Computer desks were divided by partitions in order to avoid eye contact between subjects. Subjects received neutral instructions both on their computer screen and on paper. A copy of the instructions and demographic information can be found in the on-line appendix.

### 3.2 Experimental Design

The experiment was divided into two main stages. First we used a real-effort task to measure performance and, on that basis, assign initial endowments. Subjects were classified as high or low-performing based on their relative performance in the real-effort task. This binary categorization of the participants followed the method used in Gino and Pierce (2009, 2010); then we employed a *custodial stage* where, for each pair of subjects, we measured the level of dishonesty of one subject who was asked to record the transfers of money from a second subject's "escrow" account to her own account. After the two main stages, subjects performed a second real-effort task, analogous to the one in the first stage but with a piece-rate pay, in order to obfuscate the real objective of the experiment and gauge individual motivation and learning during the experiment. The results of this second real-effort task are not discussed here and they are available upon request.<sup>2</sup> Subjects received separate instructions at the beginning of each stage. Before paying the subjects, we also collected demographic information in a post-experiment questionnaire.

#### 3.2.1 Stage 1: The First Real-Effort Task

The real effort task involved counting the occurrences of the letters "e" and "c" for each line of a 15-line text in the German language (see on-line appendix for more details regarding the text). We chose this particular task because it provides a high variation in performance, and it is easy for the subjects to understand. At the same time, it is tedious and mentally costly for the subjects.

Before starting the task, subjects were informed that, after its completion, they will be divided into two equally sized groups based on their relative performance. The performance

---

<sup>2</sup>In a nutshell, we find a very strong correlation between the performance in the first and second real-effort task (Spearman's  $\rho = 0.677$ ,  $p < 0.001$ ). Subjects performed slightly better in the second task (Wilcoxon signed-rank test,  $p < 0.001$ ), probably due to a learning-by-doing effect.



of each subject was measured by taking the deviation in absolute value between the true number of c's and e's in each line of the text and the number recorded by the subject—the lower the score the higher the performance. Subjects whose performance was above the median were designated “high performers”, whereas those below the median were designated “low performers”.<sup>3</sup> Ties in performance at the median were randomly broken by the computer (we had only two cases). Subjects were also told that they will be assigned to one of four possible scenarios which are defined by how the high and low performers are remunerated for the task. Each scenario corresponded to a different treatment of the experiment. The instructions informed the subjects that “in three of the four scenarios the payment to the high performers will be greater than or equal to the payment to the low performers. In one out of the four scenarios the payment to the high performers will be less than the payment to the low performers.” At the end of the task, we also solicited self-reported assessments of effort on a 0-4 scale. The majority of people (84.15%) reported an effort of 3 or 4, and only 3.66% of the subjects (6 out of 164 subjects) reported an effort of zero or one. The performance categorization and the corresponding initial endowment in the first stage was communicated to each subject immediately before starting the second stage.

### 3.2.2 Stage 2: The Custodial Task

In the second stage, each subject was matched with another participant. One subject was assigned the role of participant A, while the other the role of participant B. A was the Owner of an “escrow” account which was credited with \$15. B did not receive any additional endowment and played (anonymously) the role of the Custodian of the partner’s account. After the assignment of the custodial relationships, the decision-making element of stage 2 began. Each subject designated as Custodian (B) received a series of 25 consecutive binary signals generated at random from a symmetric binomial distribution (each signal was displayed in a new screen). The signal generating process was carefully explained to the subjects, and it was described as being equivalent to a coin flip. Each signal was either red or green and known only privately by each Custodian. It was common knowledge that the revelation of the signal itself was private; that is, only the Custodian, but not the Owner of the escrow account knew the actual realization of the signal. The experimenters also knew with certainty the realization of each draw, but, of course, the subjects were anonymous to the experimenters, and thus the decisions of each subject could not be attributed to them personally. The color of each signal was meant to instruct the Custodian whether to transfer

---

<sup>3</sup>This binary conception of performance allowed us to keep the design and the analysis simple. It also reflects natural settings in which signals about relative performance are likely to be noisy and imprecise but nonetheless correlated with the true levels of performance. For a similar approach, see Gino and Pierce (2009) and Gino and Pierce (2010).

an amount equal to  $1/25$  of the escrow account (\$0.60) to herself (in case of a green signal) or leave it in the Owner’s account (in case of a red signal). Each Custodian was asked to record the color of each signal received. Subjects were instructed that it was only the recording of the signal that affected payoffs, not the signal itself. Since each signal was private information of the Custodian, the design allowed the Custodians to misrepresent the signal enabling them to increase or decrease their own payoff at the expense or benefit of their partner.<sup>4</sup>

This particular task allows for possible self-serving redistribution, without any demand effect towards it. Subjects who choose to misrepresent any signal have to afford a moral and psychological cost of not doing the right thing. Moreover, the multiple choice design (having to make 25 consecutive choices) is more reliable than a single choice task, allowing for a more precise definition of the allocation, since it is much less likely that the subject’s choice be determined by impulsivity or randomness.

To make sure that subjects understood the procedures, we asked them to complete some practice rounds with forced input before starting the task. In these rounds, subjects were encouraged to experiment with the recording of the signals and to observe the effect it had on payoffs, both theirs and their partner’s. Figure 1 shows the computer screen that it was seen by the Custodians. During the task, while the Custodians reported the color of the signals, the (unincentivized) exercise for the Owners was to predict the number of red signals that the Custodian would report.

Since our aim was to analyze the effect of perceived inequity on honesty, in this first experiment we confined our analysis on the degree of honesty of those subjects who were expected to be more sensitive to inequity (i.e. those subjects for which inequity takes place to their disadvantage). Therefore, the role of Custodian was played by high-performing subjects.<sup>5</sup>

To characterize the qualitative nature of the Custodians’ dishonesty, we borrow the typology developed in Erat and Gneezy (2012). Their typology of lying is based on its effects on the payoffs of the liar and another party. They differentiate between “altruistic” and “Pareto” white lies, in which the liar benefits another party, either at his own expense (altruistic) or in such a way as to be beneficial to both (Pareto). They define a black lie as one which is detrimental to the other party. There are “spiteful”—hurting both parties—and “selfish”—hurting only the other party—black lies. In our experiment, because the reporting

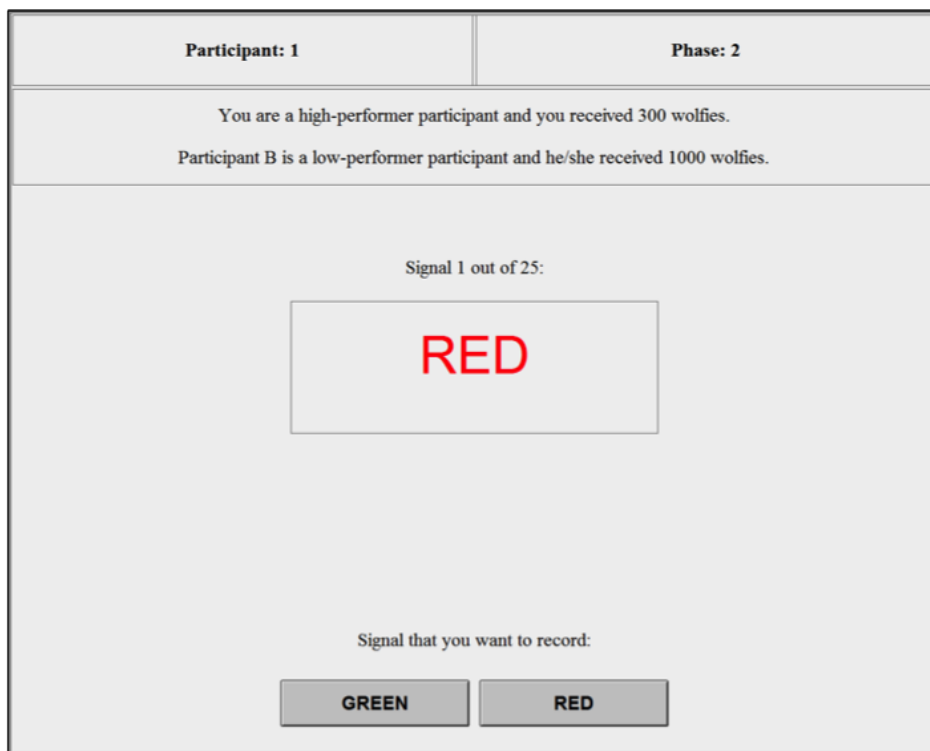
---

<sup>4</sup>Maggian and Villeval (2016) adopted a similar task to measure the lying behavior of children. Differently from us, their game is one-shot, they use shapes—a sun or a star—instead of colors, and subjects can click several times to make their favorite shape appear.

<sup>5</sup>Experiment 2 extends the analysis also to the behavior of low performers, that is subjects who benefited from inequity.

of the signal has a zero-sum effect on payoffs, if the Custodian is dishonest, it is either a “selfish black lie” (selfish dishonesty)—when the Custodian reports a red signal as green—or an “altruistic white lie” (altruistic dishonesty)—when he reports a green signal as red.

Figure 1: Screenshot of Signal Reporting Stage



### 3.2.3 Treatments

In the experiment, we manipulate the remuneration of the first stage as a function of the relative performance of the participants in the real-effort task. In particular, we vary the extent to which the payment mechanism is meritocratic under two alternative conditions: one where everyone receives the same remuneration, and one where some subjects receive a higher remuneration than others. This results in four treatments, summarized in Table 1. The treatments were randomized across sessions.

The most important differences are expected from the comparisons of the inequitable treatments – EI and UI – with the equitable treatments – EE and UE. To estimate the sample size necessary to uncover these hypothesized effects, we conduct an a priori power analysis. This type of analysis would be of limited value without a reasonable estimate of the treatment differences we would expect in our experiment. Such estimates are typically based on the results of previous similar studies. The studies which are most similar to ours are Gino and Pierce (2009, 2010). Each employs “positive inequity” and “negative inequity”

treatments and pairs up two subjects. In particular, Experiment 2 in Gino and Pierce (2010) is very similar to our design, because it is the only to employ an earned income allocation of initial resources. In terms of what they call “dishonest reporting by hurting” (and we call selfish dishonesty) the average difference between these two treatments across all of the experiments in the two studies was approximately 0.3. Thus, we follow van de Ven and Villeval (2015), and estimate the power of our sample sizes to uncover hypothesized differences of 0.35, 0.3 and 0.25. With these estimates of the differences and assuming a type-I error rate of  $\alpha = 0.10$ , our tests are powered at 0.97, 0.93 and 0.82 respectively. Generally, a power level of 0.8 is considered acceptable. Thus, we can be relatively confident that we can uncover our most important hypothesized effects.

Table 1: Treatments of Experiment 1

		Equity	
Equality	Equal/Equitable (48)	Equal/Inequitable (48)	
	Unequal/Equitable (34)	Unequal/Inequitable (34)	

*Notes:* Number of subjects in parenthesis. In EE (EI), 24 (28) subjects receive a remuneration of \$10, while 24 (20) a remuneration of \$3. Variation in the number of participants reflects different attendance rates across sessions.

#### Treatment UI: Unequal and Inequitable

In this treatment, the high performers receive a compensation of \$3 after the first stage, while the low performers receive \$10. Subjects are then randomly paired high with low performers. Hence, the income distribution between partners can be viewed as both unequal and inequitable. This treatment is the utmost inequitable since those who perform better in the real effort task are paid *less* than their partners who perform *worse*.

#### Treatment UE: Unequal and Equitable

In this treatment, the distribution of income is unequal: the high performers receive a remuneration of \$10 after the first stage of the experiment, while low performers receive a remuneration of \$3. Subjects are randomly paired high with low performers. Hence, the income distribution between partners can be viewed as equitable.<sup>6</sup> This treatment captures the typical meritocratic setting: people earn different incomes as a function of their merit.

<sup>6</sup>Depending on the relative valuations placed on high and low performance, a high performer could perceive this scenario as inequitable if she thinks that she was not remunerated enough. While we cannot exclude this possibility, we believe it is unlikely since we only give subjects one salient reference compensation (i.e. the one of the partner). In addition, it should have no impact on our results: if anything, it makes more difficult to detect a difference with the corresponding inequitable treatment (UI).

Treatment EI: Equal and Inequitable

In this treatment, the income distribution is equal: both high and low performers receive the same payment after the first stage. However, it is inequitable: again, high-performing subjects are (randomly) paired with low performers in the second stage of the experiment. If one has even the slimmest belief in just deserts (i.e., that higher performance deserves a higher reward) then this is an inequitable scenario for a high-performing subject, because she receives compensation equivalent to her low-performing partner, despite the common knowledge of her superior performance. This treatment captures the other possible non-meritocratic situation where people are compensated equally despite differential performance. It also enables us to control for distributional preferences: subjects are all paid the same and, therefore, monetary inequality concerns are muted. To control for income effects we conduct two sub-treatments, in which both high and low performers receive a remuneration of \$10 in the *high income* sub-treatment and \$3 in the *low income* sub-treatment.

Treatment EE: Equal and Equitable

This treatment is a special case that serves as a base of comparison for the EI treatment. The distribution of income is equal as in treatment EI. Moreover, it can be seen as equitable: subjects are (randomly) paired high performers with high performers and low performers with low performers in the second stage of the experiment.<sup>7</sup> Thus, the income distribution among partners is both equal and equitable because both subjects within each pair are compensated with the same amount for the same type of performance.<sup>8</sup>

Because in the EE treatment (and only in the EE) the pairings are homogeneous with respect to performance, the relative valuation attached to the levels of performance should not impact the decision at the level of the dyad. There is no information provided within the context of the experiment that would allow the subjects to differentially assign just deserts, as the members of each pair are objectively indistinguishable.<sup>9</sup> Thus, this treatment allows us to determine a level of dishonesty that should—by the design of the experiment—be divorced from concerns of equity. This level can be then compared with the amount of

---

<sup>7</sup>This means that in some cases, and in Experiment 1 unique to this treatment, the low performers are in the decision making role, whereas in each of the others, the high performers are so positioned.

<sup>8</sup>In EE, subjects are matched within the same type (high-high and low-low performers). This could produce in-group biases (i.e. subjects behaving nicer towards the other player). We can exclude this possibility as we do not find any statistically significant difference in behavior between treatments EE and UE (where, based on the same logic, we should expect an out-group bias).

<sup>9</sup>High-performing subjects might indirectly perceive some kind of injustice if they compare themselves not with their matched partner but with the low performing subjects. This should however strengthen our results if we find a difference with respect to the EI treatment. In addition, we can verify whether the behavior of high-performing subjects is different from that of low-performing subjects. Finally, we can compare the EE with the other equitable treatment (UE) testing for any difference in behavior.

dishonesty observed in the corresponding equal and inequitable treatment (EI). Just as in the EI treatment, to control for income effects, we conduct two sub-treatments, where both high and low performers receive a remuneration of \$10 (the *high income* sub-treatment) and \$3 (the *low income* sub-treatment).

The role assignment for each of the treatments described above is summarized in Table 2.

Table 2: Role-assignment in the four treatments

Treatment	Owner	Custodian
EE	High (Low) performer	High (Low) performer
EI	Low Performer	High Performer
UE	Low Performer	High Performer
UI	Low Performer	High Performer

### 3.3 Behavioral Predictions

Our experimental design enables us to measure the extent to which unjust deserts induces dishonesty on the part of the Custodians. In our set-up, unjust deserts stems from two main situations. In one case, the more performing subject is compensated less than the less performing partner. In another case, she is compensated as much as the less performing partner.<sup>10</sup> We focus on negative or disadvantageous inequity by considering scenarios where the performance of the decision maker (the Custodian) is higher. This is because we expect negative or disadvantageous inequity to have a much stronger effect on behavior than positive or advantageous inequity (see e.g. Bloom, 1999; Gino and Pierce, 2009, 2010).<sup>11</sup> We hypothesize the following:

**Hypothesis 1.** *Meritocratic inequity will induce greater selfish dishonesty ( $D_s$ ) in those disadvantaged by the performance-payoff mapping:  $D_s^{EI} > D_s^{EE}$  and  $D_s^{UI} > D_s^{UE}$ .*

This hypothesis can be tested by measuring the extent to which the Custodians selfishly misrepresent the signals in a situation where there is a violation of just deserts compared to a similar scenario where just deserts is not violated. To control for distributional preferences, we compare scenarios where the income distribution is equal and the only thing that varies

<sup>10</sup>A third possible case is when the more performing subject is compensated more than the less performing partner but not enough. We believe that this case is not particularly relevant for our setting since it would require an additional reference point (e.g. a prior, or promised earnings) which we do not provide in our experiment. In our set-up, the only salient reference compensation that we provide is the money that the partner receives. In addition, this third case should, if anything, induce less dishonesty compared to the other two cases.

<sup>11</sup>Scenarios in which the performance of the decision maker is lower than the performance of the other subject are left for future research.

is just deserts (EI vs. EE).<sup>12</sup> We also compare scenarios where both just and unjust deserts result in an unequal income distribution (UI vs. UE). If we find no differences in our comparisons, then we may conclude that just deserts does not affect honesty. If instead we find any significant difference, this would support our hypothesis that dishonesty increases as a result of violations of the just deserts principle.

The effect of unjust deserts may depend on whether inequality is present. Indeed, previous studies indicate that people may be more or less dishonest depending on the consequences that their lies have on others' well-being. In particular, people may be (more) less prone to act dishonestly if it contributes to attenuate (exacerbate) inequality (see e.g. Gneezy, 2005; Hurkens and Kartik, 2009; Erat and Gneezy, 2012; Maggian and Villeval, 2016). In the context of our experiment, subjects may lie more, to their advantage, in response to violation of just deserts if this violation also contributes to increase inequality. We thus state the following hypothesis.

**Hypothesis 2.** *Meritocratic inequity will induce greater selfish dishonesty if it is associated to an unequal income distribution:  $D_s^{UI} > D_s^{EI}$ .*

We can test this hypothesis by comparing the level of selfish dishonesty in the UI treatment and EI treatment. If we find no difference, we may conclude that, in our setting, inequality does not add an effect to that of inequity.

In the experimental literature, there is also evidence of “altruistic white lies”, i.e. costly lies that benefit another person. Erat and Gneezy (2012) find that, in sender-receiver games, 33% of senders are willing to lie in order to increase the earnings of the partner at a personal cost. Though such behavior is possible in our experiment, we did not expect to observe a significant amount of it. Indeed, evidence from experiments with a setting similar to ours indicate that altruistic white lies may be much less frequent than in sender-receiver games. For example, Maggian and Villeval (2016) find that only 4.81% of their subjects were willing to misreport a signal to favor their partner. That said, we can still devise some hypotheses regarding white lies which are analogous to those presented for black lies. Based on previous results on white lies, we may expect that subjects who care particularly about others' well-

---

<sup>12</sup>Dishonesty may be sensitive to social preferences (see e.g. Gneezy, 2005; Hurkens and Kartik, 2009; Erat and Gneezy, 2012; Maggian and Villeval, 2016). In the context of our experiment, being dishonest may be confounded with some of these distributional preferences. In particular, inequality averse subjects may behave more dishonestly to reduce the possible inequality of stage 1. We control for this by comparing EI and EE. Subjects with competitive or spiteful preferences may also behave more dishonestly to reduce the income of their counterpart. Any across-treatment differences cannot however be explained by these preferences since the behavior of competitive or spiteful subjects should be the same in all treatments due to the zero sum nature of our game, and equivalent to the behavior of self-interested subjects. Finally, reciprocity does not apply to our case since the owner of the account is not responsible for the income distribution and did not hurt or help the custodian.

being (i.e. altruistic individuals or inequality-averse subjects) may be willing to altruistically misreport a signal to favor the other person. This drive may however be attenuated if—other things being equal—the other person has already benefited from an unjust allocation of the initial resources. This effect may be larger if the income distribution is not only unjust but also unequal. We can state the two following hypotheses.

**Hypothesis 3.** *Meritocratic inequity will induce less altruistic dishonesty ( $D_a$ ) on the part of those disadvantaged by the performance-payoff mapping:  $D_a^{UI} < D_a^{UE}$  and  $D_a^{EI} < D_a^{EE}$ .*

**Hypothesis 4.** *Meritocratic inequity will induce even less altruistic dishonesty if it is combined with an unequal income distribution:  $D_a^{UI} < D_a^{EI}$ .*

## 4 Results of Experiment 1

In this section, we present the main results of our experiment. We first focus on selfish dishonesty, and then on altruistic dishonesty. Afterwards, we consider an alternative measure of dishonesty that takes into account both selfish and altruistic dishonesty. Next, we explore the dynamics of dishonesty by looking at whether dishonesty changes over time. Finally, we test the robustness of our results in a regression analysis.

### 4.1 Selfish dishonesty

We define the *rate of selfish dishonesty* as the number of red signals that a subject reports as green divided by the total number of red signals that he or she received, yielding one observation per subject. This variable provides a normalized measure of the rate of (selfish) black lies. Figure 2 displays the rates of selfish dishonesty for each experimental treatment. Since we did not detect any statistically significant difference between the high and low income sub-treatments of the EE and EI treatment respectively ( $p = 0.513$  and  $0.678$ ), we pooled the data of the sub-treatments together to increase power. Nor did we find any statistically significant difference between the rates of selfish dishonesty of high and low performers in the EE treatment ( $p = 0.470$ ), so we pool these data as well.

In order to justify our pooling across these sub-treatments, we conduct a power analysis which estimates, given our observed data, what sample size would be required to find a statistically significant effect at  $\alpha = 0.10$ . These estimates are 372, 275 and 195 subjects per condition for EE high vs. low income; EI high vs. low income; and EE high vs. low performers, respectively. Given that even the smallest of these estimates is much larger than most such laboratory studies, this pooling appears to be appropriate.

The rate of selfish dishonesty lies between the 14% of the EE treatment and the 34% of the UI treatment. While this, at first sight, might seem quite low if one considers that



a rational and profit-maximizing agent should always misreport a red signal, it is actually in line with the results of other studies on lying behavior. For instance, the proportion of income maximizers estimated by Fischbacher and Föllmi-Heusi (2013) is 22%, while Dai et al. (2017) estimates a 28%. More in general, the meta-study of Abeler et al. (2016) indicates that people appropriate only 21.6% of the gains they could make by lying. Similar percentages of lying behavior are observed in cheating games where the experimenter has complete information regarding the individual lying behavior. For instance, Gneezy et al. (2016) find that between 26% and 33% of subjects lie in these observed games.

It seems that the *inequitable* treatments (EI and UI) were characterized by a much greater degree of selfish dishonesty. With non-parametric tests we can investigate more formally whether there is any significant difference in the rates of selfish dishonesty across treatments, to test Hypotheses 1 and 2.<sup>13</sup> If we compare the rate of selfish dishonesty across treatments, in the aggregate, we observe significantly higher rates of selfish dishonesty in the UI treatment compared to the UE and EE treatment respectively ( $p = 0.061$  and  $0.019$ ), and in the EI treatment compared to the UE and EE treatment respectively ( $p = 0.067$  and  $p = 0.011$ ). There is no significant difference between UI and EI ( $p = 0.989$ ), and between EE and UE ( $p = 0.567$ ). Based on these findings, we present our first result which is consistent with Hypothesis 1 but rejects Hypothesis 2:<sup>14</sup>

**Result 1.** *There is more selfish dishonesty in the inequitable treatments than in their corresponding equitable treatments. Inequality does not amplify the effect of inequity.*

## 4.2 Altruistic dishonesty

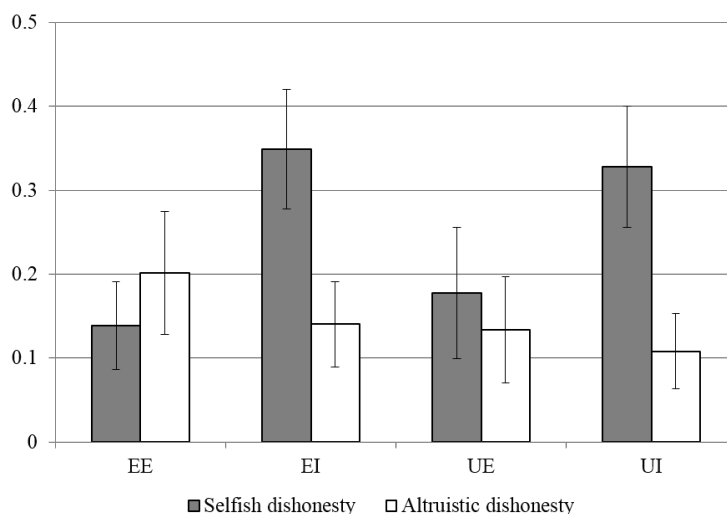
We now turn to altruistic white lies. We define the *rate of altruistic dishonesty* as the number of green signals reported as red divided by the total number of green signals received. This variable provides a normalized measure of the rate of white lies. Figure 2 displays the average rates of altruistic dishonesty for each experimental treatment.<sup>15</sup> The rates of altruistic dishonesty are much lower than the rates of selfish dishonesty (Wilcoxon signed-

<sup>13</sup>Throughout the paper, we employ Mann-Whitney U tests unless otherwise specified. All tests are two-sided. For one-sided hypotheses, we also conduct one-sided statistical tests (available upon request). The results are qualitatively the same.

<sup>14</sup>We did not detect any statistically significant difference across treatments in the number of red signals received by the Custodians ( $p > 0.1$  for all comparisons), so this result, as well as the following findings, cannot be explained by differences in the number of red signals—randomly generated by the computer for each Custodian—received by different Custodians across treatments.

<sup>15</sup>As for the analysis of the black lies, we checked whether there is any statistically significant difference between the high and low income sub-treatments of the EE and EI treatment respectively. We also checked for any statistically significant difference between the rates of altruistic dishonesty of high and low performers in the EE treatment. In all these comparisons, we did not find any difference ( $p > 0.1$ ), and, therefore, we pooled the data together.

Figure 2: Selfish and Altruistic Dishonesty by Treatment



Note: Error bars identify standard errors of the mean.

rank test,  $p = 0.002$  pooling all treatments together). This result is driven by the UI and EI treatments where the rate of selfish dishonesty is more than twice as big as the rate of altruistic dishonesty ( $p = 0.003$  and  $0.01$  respectively). In the two equitable treatments (EE and UE), there is no statistically significant difference between the rate of altruistic and selfish dishonesty ( $p = 0.760$  and  $p = 0.734$  respectively). To test hypotheses 3 and 4, we can check whether the rate of altruistic dishonesty differs across treatments. On average, the rate of altruistic dishonesty is larger in the equitable treatments (EE and UE) and in EI compared to UI. However, the differences are not large enough to achieve statistical significance ( $p > 0.1$  for all pairwise comparisons). We can thus present our next result.

**Result 2.** *Altruistic dishonesty is generally low (below 20%) and does not statistically differ across treatments.*

A possible explanation of this result is that only a few intrinsically-motivated subjects are willing to misreport the signals to favor the counterpart (see Maggian and Villeval, 2016). Therefore, even if we hypothesize that unjust deserts has an effect on the behavior of these subjects, it affects too small a proportion of the population to observe an effect at the aggregate level.

Interestingly, if we classify subjects based on whether they truthfully or untruthfully report the 25 signals (see Table 3), we find that a large proportion of those subjects who misreported the signals to their disadvantage also, and to a larger extent, engage in selfish dishonesty. We call these individuals *randomizing selfish dishonest subjects*. A possible ex-

planation of their behavior is that these subjects want to balance out their selfish dishonesty with some altruistic dishonesty in order to make their black lies appear as random mistakes (recall that in our experiment the experimenter knows if a subject has misreported the signal or not and, therefore, reputational costs may be very high). If the aversion to lie for personal benefits is not (or not only) due to an intrinsic cost of lying but also to reputational concerns as recent models suggest (e.g. Gneezy et al., 2016; Abeler et al., 2016; Dufwenberg and Dufwenberg, 2016), some people could, instead of completely refraining from lying, alternate some selfish lies with some (fewer) altruistic lies in order to signal that they are not really liars but only “distracted” participants and, thus, reduce the reputational costs associated with dishonesty.<sup>16</sup> We also identify a handful of *randomizing altruistic dishonest subjects* who balance out their altruistic dishonesty with some selfish dishonesty. Other categories of subjects reported in Table 3 are *fully selfish (altruistic) dishonest subjects* who only report green (red) signals, irrespective of the actual signal received; *partially selfish (altruistic) dishonest subjects* who never misreport a green (red) signal and sometimes misreport a red (green) signal; *honest subjects* who report truthfully all the signals; and *noisy dishonest subjects* who misreport the same proportion of green and red signals.

Table 3: Percentages of honest and (selfish and altruistic) dishonest subjects

Types	EE	EI	UE	UI	All	Selfish dishonesty	Altruistic dishonesty
Selfish dishonest	25% (6)	58.33% (14)	35.29% (6)	64.7% (11)	45.12% (37)	0.45	0.08
– Fully	0% (0)	12.5% (3)	11.76% (2)	5.88% (1)	7.32% (6)	1	0
– Partially	12.5% (3)	25% (6)	23.53% (4)	29.41% (5)	21.95% (18)	0.27	0
– Randomizing	12.5% (3)	20.83% (5)	0% (0)	29.41% (5)	15.85% (13)	0.44	0.22
Altruistic dishonest	16.66% (4)	16.67% (4)	29.41% (5)	5.88% (1)	17.08% (14)	0.16	0.57
– Fully	8.33% (2)	0% (0)	0% (0)	0% (0)	2.44% (2)	0	1
– Partially	8.33% (2)	4.17% (1)	17.65% (3)	0% (0)	7.32% (6)	0	0.39
– Randomizing	0% (0)	12.5% (3)	11.76% (2)	5.88% (1)	7.32% (6)	0.38	0.61
Honest	50% (12)	25% (6)	35.29% (6)	29.41% (5)	35.37% (29)	0	0
Noisy dishonest <sup>a</sup>	8.33% (2)	0% (0)	0% (0)	0% (0)	2.44% (2)	0.70	0.70

*Notes:* The frequencies are reported in parenthesis. The table also reports the mean rates of selfish and altruistic dishonesty (last two columns). <sup>a</sup>One noisy dishonest subject misreported all the signals. This explains why, for this category, the rates of selfish and altruistic dishonesty are very high.

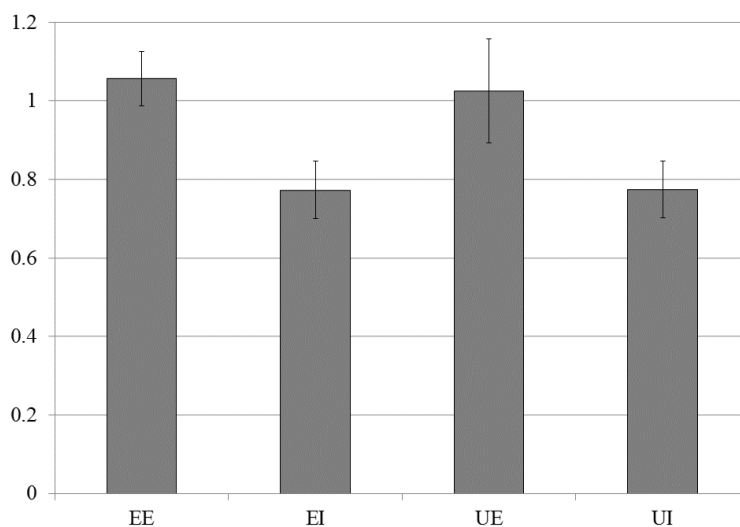
### 4.3 Overall dishonesty

Our measures of dishonesty used so far are appropriate if we think of the decision to be dishonest as playing out independently for each signal received by the Custodian. Subjects, however, may have a supergame in mind and try to realize an acceptable final distribution over the course of the 25 signals irrespectively of which signals need to be misreported. This

<sup>16</sup>Of course, our experiment does not provide a precise test of this, and it is likewise possible that some lies are genuine mistakes, even if it is difficult to believe that mistakes are systematic given the very simplistic nature of the task. Nonetheless, we think it is an interesting finding to report which could be investigated more in future research.

is particularly relevant for randomizing dishonest subjects who misreport both green and red signals. To analyze such a possibility, we measure honesty as the ratio of the total number of red signals reported to the total number of red signals received (*rate of red signals reported*). This measure takes into account both types of dishonesty, selfish and altruistic, and thus could be greater than one if a subject engaged in altruistic white lies (i.e. reported the signal as red when it was in fact green). Figure 3 displays the average rate of red signals reported for each experimental treatment.<sup>17</sup> If we conduct pairwise comparisons of the treatments, honesty was statistically significantly lower in the UI treatment compared to the UE and EE treatment respectively ( $p = 0.072$  and  $0.011$ ), and in the EI treatment compared to the EE treatment ( $p = 0.023$ ).<sup>18</sup> There is no statistically significant difference between EI and UI ( $p = 0.706$ ). This is in line with Result 1: subjects react to just deserts by reporting a larger proportion of red signals when the income distribution is equitable and a smaller proportion when the income distribution is inequitable.

Figure 3: Rate of Red Signals Reported



Notes: Error bars identify standard errors of the mean.

#### 4.4 Dynamics

Since we have multiple observations for each custodian (subjects reported 25 signals), we can also look at the dynamics of dishonesty. Indeed, it is possible that subjects lie only up to a certain point at which just deserts is restored, and then no longer. If this is the case,

<sup>17</sup>Again, we find no differences between the sub treatments in EE and EI, so therefore pool the data.

<sup>18</sup>In comparing EI vs. UE, the  $p$ -value is just above the 10% level (i.e.  $p = 0.102$ ). A one-sided test provides a weakly significant result ( $p = 0.051$ ).

the hypothesized treatment effects would last only until a certain threshold of experienced or misreported signals. To explore this possibility, we compare, for each treatment, the rate of selfish dishonesty, altruistic dishonesty, and red signals reported respectively across periods. We find no differences (Friedman tests,  $p \geq 0.1$  for all comparisons).<sup>19</sup>

There are two possible interpretations of this result. First, the majority of people may find it easier to justify small lies spread over time rather than repeated lies in a short or given period of time. This is consistent with the idea that people want to maintain a positive self-image when they are lying (Shalvi et al., 2011; Mazar et al., 2008). A second possible interpretation is that inequity has a fixed long-lasting effect on dishonest behavior that persists irrespectively of whether just deserts is restored or not. Unfortunately, our experiment does not enable us to disentangle between these two alternative explanations, and we leave this for future research.

We can also look at how the difference in payoffs between custodian and owner evolves over time as a result of violations of the just deserts principle. As already mentioned, one behaves more (selfishly) dishonestly in response to violation of meritocracy in order to modify one's outcomes such that the resulting distribution is more equitable. If this is the case, the difference in payoffs between custodian and owner should increase over time. This is indeed what we find in Figure 4. As the graph shows, the relative payoff of the custodian increases over time in the two inequitable treatments (Page's trend tests,  $p = 0.004$  in UI and  $p < 0.001$  in EI) but it is flat in the equitable treatments ( $p > 0.1$  for both).<sup>20</sup> We should stress that this effect cannot be driven by equality concerns. Note in fact that inequality increases in EI as a result of dishonesty. The figure also shows that the difference in payoffs between custodian and owner increases at a steady rate. This confirms our previous finding that dishonesty remains constant over time.

#### 4.5 Regression analysis

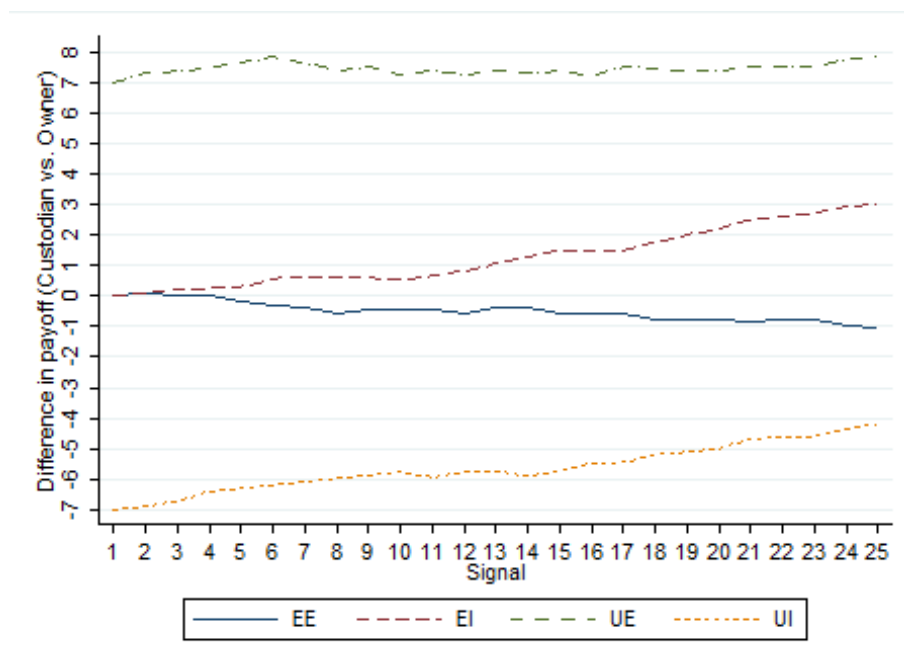
To test the robustness of our results we conduct a regression analysis. In particular, we run a logistic regression pooling all the data together. The dependent variable is a dichotomous variable which takes value 1 if a signal is recorded as red and 0 if it is recorded as green. We employ random effects to control for the non-independence of the observations at the subject level (recall that we have 25 signals per subject).<sup>21</sup> The independent variables are the current private signal (1 if red, 0 if green), the number of consecutive periods without a green signal (this controls for the idea that a long drought of green signals would make participants

<sup>19</sup>To minimize the impact of potential differences in the number of red signals received across periods, we grouped the 25 signals in five-period blocks.

<sup>20</sup>See footnote 19.

<sup>21</sup>We have also run regressions with error clustering instead of random effects. The results are qualitatively the same and are available upon request.

Figure 4: Evolution of the Difference in Payoffs Between Custodian and Owner



likely to lie more), the treatment dummies (the baseline category is the EE treatment),<sup>22</sup> the signal period (2 to 25), the effort reported by the subject in the first real-effort task, and interaction terms of the treatment dummies with the current private signal and the number of consecutive periods without a green signal. Table 4 reports the average marginal effects.

The results of the regression analysis indicate that both self-reported effort and time have no effect on the probability of reporting a red signal. Also, there is no effect in terms of dynamics (i.e. number of consecutive periods without a private green signal). We can contrast the marginal effects of receiving a private red or green signal across our different treatments. Subjects were more likely to record a red signal when they received a private red signal in EE and UE compared to both EI and UI (87% and 82.1% vs. 64.7% and 66.7% respectively,  $p < 0.01$  for all comparisons). The probability of reporting a red signal when the private signal received was green (altruistic dishonesty) was in general low and similar across treatments except in the UE treatment where it was significantly lower than the EE treatment (10.7% vs. 23.3%,  $p = 0.047$ ).

These results confirm Hypothesis 1 that increased dishonesty is associated with a decrease in perceived just deserts quite independently of the equality of the incomes. They also corroborate the insights gained from the previous analysis conducted using non-parametric

<sup>22</sup>In the on-line appendix, we also report the same regressions but with the two inequity treatments pooled together (i.e. we use only one dummy to identify the treatments with inequity).

Table 4: Regressions Results

	Logit regression		
	dy/dx	Std. Err.	P>z
Effort	-0.021	0.029	0.461
Round	0	0.001	0.972
<i>Private signal = green</i>			
× EE	0.233	0.053	<0.001
× EI	0.156	0.04	<0.001
× UE	0.107	0.038	0.005
× UI	0.142	0.046	0.002
<i>Private signal = red</i>			
× EE	0.858	0.038	<0.001
× EI	0.6	0.061	<0.001
× UE	0.853	0.048	<0.001
× UI	0.617	0.075	<0.001
<i>N. periods without a green</i>			
× EE	0.006	0.01	0.545
× EI	-0.014	0.01	0.185
× UE	-0.018	0.013	0.158
× UI	-0.004	0.015	0.784
Obs	2050		
Df	13		
Prob > F	0		

tests.

## 5 Experiment 2

The focus of Experiment 1 is on a specific setting where dishonesty is instrumental to re-establishing a just distribution between a person who benefited from inequity and another who did not. It is however possible that the perception of unjust distributive rules decreases the moral cost of dishonesty quite independently from re-distributional concerns. In other words, the mere experience of inequity could induce a higher propensity to violate the general norm of honesty even if dishonesty does not contribute to restore equity. To test this possibility, we conducted Experiment 2. As in Experiment 1, we manipulate the remuneration of the first stage as a function of the relative performance of the participants in a real-effort task. We then test whether subjects behave more or less dishonestly in a subsequent unrelated task which, unlike Experiment 1, does not allow for any income redistribution.

In what follows of this section, we discuss the main differences of Experiment 2 with respect to Experiment 1 and we present the main hypotheses and results.

### 5.1 Participants and Procedures

We ran Experiment 2 at SBU, between December 2014 and March 2015. We recruited 232 subjects (excluding those who had participated in Experiment 1). Average earnings were equal to \$20.75. All the other procedures were the same as in Experiment 1.

### 5.2 Experimental Design

Experiment 2 is divided into three treatments: *UI-no-match*, *EI-no-match*, and *UE-no-match*. The design is identical to Experiment 1 except for one key aspect: in the Custodial Stage, subjects were not anymore matched in pairs.<sup>23</sup> In particular, they were put in charge of an account (which was *not* owned by another subject) and asked to record the color of 25 signals. Every time they recorded a green signal, they were entitled to take \$0.60 from the account, while every time they recorded a red signal, they had to leave the account untouched. As in Experiment 1, subjects were instructed that it was only the recording of the signal that affected their payoffs, not the signal itself. This time, however, no other participant benefits or loses money from a misrepresentation of the signal. Custodians when lying do not take resources from another subject. Actually, they subtract resources from the experimenter's funds. Hence, a misrepresentation of the signal could not be used to redistribute income from the non-deserving to the deserving subject. In addition, we got rid of all possible influences (e.g. in-group or out-group biases) arising from inter-personal comparisons with the matched subject.

In the UI-no-match treatment, the distribution of the income from the real-effort task was unequal and inequitable (like in UI); in UE-no-match, it was unequal but equitable (like in UE); and in EI-no-match, it was equal but inequitable (like in EI). The treatments of Experiment 2 are summarized in Table 5. Using the observed differences between the equitable and inequitable treatments in Experiment 1 as our baseline, a power analysis reveals that we would need 35 subjects per conditions to detect such an effect at the 0.10 significance level. Thus, our design for Experiment 2 appears adequately powered.

---

<sup>23</sup>In Experiment 2, we also added two short incentivized tasks at the end of the experiment not related to the current research question. They are briefly described in the on-line appendix. Note that these two tasks could not distort subjects' decisions since they were just added at the end of the session. Compared to Experiment 1, we also collected more data to increase the statistical power of our tests since we expected less dishonesty overall (recall that, in Experiment 2, we removed income redistribution as a potential motive for behaving dishonestly).



Table 5: Treatments of Experiment 2

Treatment	Income distribution	Matching
UE-no-match (74)	Unequal and Equitable	No matching
EI-no-match (88)	Equal and Inequitable	No matching
UI-no-match (70)	Unequal and Inequitable	No matching

*Notes:* Number of subjects in parenthesis. In EI-no-match, we conduct a high-income (40 subjects) and low-income (48 subjects) sub-treatments to control for income effects. In the high-income (low-income) condition, participants receive a remuneration of \$10 (\$3) after the real-effort task. Variation in the number of participants reflects different attendance rates across sessions.

### 5.3 Behavioral Predictions

By comparing the different treatments of Experiment 2, we can investigate whether the mere experience of inequity induces rule-disadvantaged subjects to violate the honesty norm to their advantage in a context where dishonesty is not instrumental to restore equity. We can study this for different levels of inequity: a “weak” level where equality is preserved (EI-same-match) and a “strong” level where inequity goes hand in hand with inequality (UI-same-match). More formally, the hypotheses that we test are the following:

**Hypothesis 5.** *A strong (weak) form of meritocratic inequity induces rule-disadvantaged subjects to behave more dishonestly even if equity cannot be restored:  $D_s^{UI-no-match} > D_s^{UE-no-match}$  ( $D_s^{EI-no-match} > D_s^{UE-no-match}$ ).*

**Hypothesis 6.** *The impact of meritocratic inequity on the behavior of rule-disadvantaged subjects is greater when it entails inequality:  $D_s^{UI-no-match} > D_s^{EI-no-match}$ .*

We can then bridge the two experiments together (Experiments 1 and 2) to test whether re-distributional concerns matter in explaining the relationship between meritocratic inequity and dishonesty. This can be done by comparing the behavior of high-performing subjects in the inequitable treatments across experiments. The main difference in the inequitable treatments between the two experiments is whether a misreport affects or not the payoff of another participant. If dishonesty is higher in Experiment 1 than 2, it means that re-distributional concerns are an important explanation of the impact of meritocratic inequity on honesty.

**Hypothesis 7.** *Dishonesty is higher after a strong (weak) violation of the just deserts principle if it helps to redistribute income from the undeserved to the deserved subjects:  $D_s^{UI} > D_s^{UI-no-match}$  ( $D_s^{EI} > D_s^{EI-no-match}$ ).*

The null hypothesis is that there is no difference in dishonesty between the two experiments ( $D_s^{UI-no-match} = D_s^{UI}$  and/or  $D_s^{EI-no-match} = D_s^{EI}$ ), meaning that the impact of

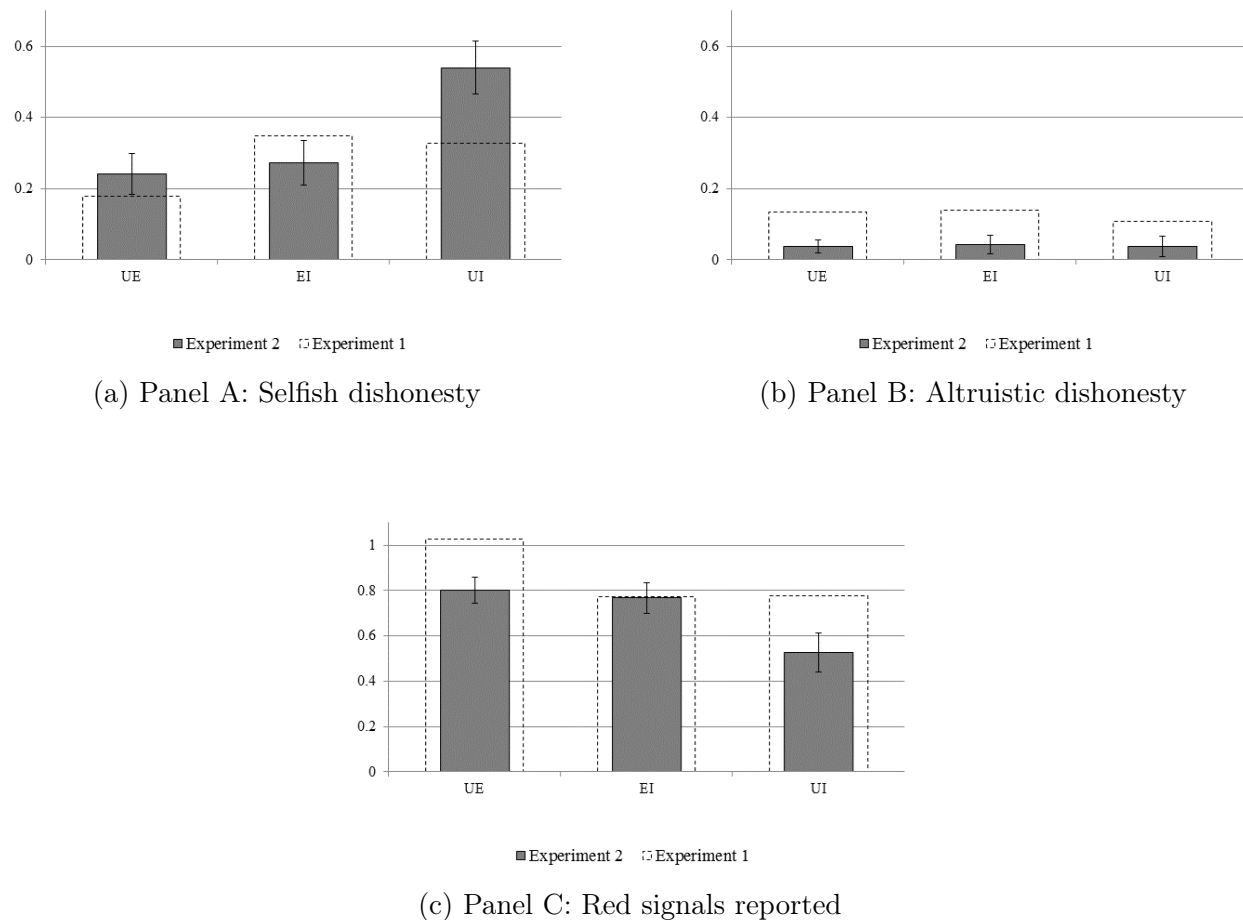
meritocratic inequity on honesty is not driven by re-distributional concerns.

In outlining the hypotheses above, our focus has been on the behavior of those subjects who are expected to be more sensitive to inequity (i.e. subjects who were disadvantaged by the performance-payoff mapping). In Experiment 2, we, by design, also collected data on the behavior of rule-advantaged subjects. We report an analysis of their behavior in the on-line appendix.

In the context of Experiment 1, we devised some hypotheses regarding altruistic dishonesty. In particular, we hypothesized that people could behave more or less dishonestly to their disadvantage to obtain a more equitable final distribution of the income. These re-distributional motives do not apply to Experiment 2. For this reason, we do not expect to observe an effect of meritocratic inequity *per se* on altruistic dishonesty.

#### 5.4 Results of Experiment 2

Figure 5: Selfish dishonesty, altruistic dishonesty and red signals reported in Experiment 2



Note: Error bars identify standard errors of the mean.

Figure 5 displays the rates of selfish dishonesty, altruistic dishonesty, and red signals reported for high-performing subjects in all treatments of Experiment 2, and the corresponding treatments of Experiment 1 (dotted bars).

To test Hypotheses 5 and 6, we compare the level of selfish dishonesty of high-performing subjects across the treatments of Experiment 2. We find that high-performing subjects lied much more in UI-no-match than UE-no-match and EI-no-match respectively ( $p = 0.003$  and  $0.006$ ), while there is no significant difference between UE-no-match and EI-no-match ( $p = 0.983$ ). A similar (but opposite) pattern is obtained if we compare the rate of red signals reported (UI-no-match vs. UE-no-match,  $p = 0.003$ ; EI-no-match vs. UE-no-match,  $p = 0.811$ ; UI-no-match vs. EI-no-match,  $p = 0.013$ ).<sup>24</sup> We can thus report the following result which supports Hypothesis 5 for the strong form of meritocratic inequity, and Hypothesis 6.

**Result 3.** *A violation of the just deserts principle induces rule-disadvantaged subjects to behave more dishonestly to their advantage even if this does not produce any income redistribution. This effect only takes place when the initial (unjust) income distribution is also unequal.*

To investigate whether subjects react more to a violation of the just deserts principle if they have the opportunity to redistribute income (Hypothesis 7), we compare the behavior of high-performing subjects in the *inequitable* treatments of Experiments 1 and 2.<sup>25</sup> We find that subjects misreport fewer red signals in Experiment 2 compared to Experiment 1 only when the initial income distribution is equal (EI vs. EI-same,  $p = 0.075$ ). In the unequal and inequitable condition, we find (although not significantly) the opposite pattern ( $p = 0.108$ ).<sup>26</sup> This gives the following result which confirms Hypothesis 7 for the weak form of inequity, but rejects it for the strong form.

**Result 4.** *The absence of redistributive concerns reduces the amount of selfish dishonesty only when inequity does not entail inequality.*

To test the robustness of our results we conduct a regression analysis on the data of Experiment 2. The specification is the same as in Experiment 1 except that we now employ treatment dummies for Experiment 2 as explanatory variables (UI-no-match, EI-no-match, and UI-no-match). Table 6 reports the marginal effects of the regression.

<sup>24</sup>As expected, we did not find statistically significant differences in altruistic dishonesty ( $p > 0.1$  for all comparisons).

<sup>25</sup>We also compare the behavior of high-performing subjects in the equitable treatments (UE and UE-no-match). We find no difference in the rate of selfish dishonesty ( $p = 0.975$ ), altruistic dishonesty ( $p = 0.255$ ) and red signals reported ( $p = 0.224$ ).

<sup>26</sup>This also results in a significant lower rate of red signals reported ( $p = 0.054$ ). In both inequitable conditions, we also find that subjects in Experiment 2 misreport fewer signals to their disadvantage compared

Table 6: Regression Results (Experiment 2)

	Logit regression		
	dy/dx	Std. Err.	P>z
Effort	-0.021	0.031	0.494
Round	0	0.001	0.673
<i>Private signal = green</i>			
× UE-same	0.107	0.034	0.002
× EI-same	0.055	0.02	0.005
× UI-same	0.056	0.021	0.009
<i>Private signal = red</i>			
× UE-same	0.682	0.058	< 0.001
× EI-same	0.691	0.054	< 0.001
× UI-same	0.432	0.06	< 0.001
<i>N. periods without green</i>			
× UE-same	-0.018	0.009	0.036
× EI-same	-0.012	0.008	0.119
× UI-same	-0.018	0.006	0.005
Obs	2900		
Df	10		
Prob > F	< 0.001		

*Notes:* The regression is conducted on the sample of high-performing subjects. In the on-line appendix, we report the results of the regression with low-performing subjects.

In line with the results of Experiment 1, we find that self-reported effort and time have no effect on the probability of reporting a red signal. Differently from Experiment 1, the probability of reporting a red signal decreases if the number of consecutive periods without a private green signal increases. This is statistically significant in the UE-same and UI-same treatments ( $p = 0.036$  and  $p = 0.005$ ). If we contrast the marginal effects across treatments, we find that high-performing subjects were less likely to record a red signal when they received a private red signal in UI-same compared to both UE-same and EI-same (43.2% vs. 68.2% and 69.1% respectively,  $p < 0.01$  for both comparisons). These results corroborate the insights gained from the previous analysis conducted using non-parametric tests.

As we did for Experiment 1, we also analyzed the dynamics of dishonest behavior, and classified subjects based on whether they truthfully or untruthfully report the 25 signals. The results are similar to the findings of Experiment 1 and are presented in the on-line appendix. As we mentioned earlier, we also look at the behavior of low-performing subjects who benefited from the inequity condition. The results (reported in the on-line appendix)

---

to Experiment 1 ( $p = 0.003$  for EI vs. EI-no-match, and  $p = 0.019$  for UI vs. UI-no-match). This confirms our previous claim that altruistic dishonesty is less observed when there are no re-distributional motives.

suggest that the violation of the equity norm may also have some (weaker) spillover effects on the behavior of rule-advantaged subjects.

## 6 Conclusion

The results of our experiments confirm our main prediction that the violation of the just deserts principle induces greater selfish dishonesty on the part of those disadvantaged by this violation. In Experiment 1, we find that the effect is the same irrespective of whether the initial income distribution is equal or unequal. Once we remove the possibility to redistribute income between rule-disadvantaged and advantaged subjects (Experiment 2), the effect of (un)just deserts on dishonesty persists only when the initial payoffs are unequal. These findings suggest that the violation of the equity norm has a spillover effect on dishonest behavior—quite independently of redistributive motives—only when inequity is associated with inequality. In other words, when the system is inequitable in producing an unequal income distribution, people are willing to “rip it off” in turn, irrespective of whether dishonesty has a redistributive effect. In contrast, a weaker form of inequity which preserves equality induces more dishonesty only when the latter is instrumental to re-establishing a just income distribution between rule-disadvantaged and rule-advantaged individuals.

In both experiments, we also investigated whether a violation of just deserts induces people to tell less altruistic white lies. We do not find evidence of this. We however find that the majority of those who tell altruistic white lies also engage in self-serving dishonesty, and to a larger extent than altruistic dishonesty, as if they wish to disguise their black lies as simple mistakes. To the best of our knowledge, we are the first to show this kind of behavior.

Our experiments were conducted with American subjects, who are quite sensitive to meritocratic concerns. Kunovich and Slomczynski (2007) provide evidence that the levels of both actual and perceived meritocracy are quite high in the United States with respect to other countries. Americans tend to have higher expectations for meritocracy, so they may be more sensitive to its absence. Hence, if American experimental subjects are “pre-treated” with higher expectations of just deserts, they can be more sensitive to violations of the just deserts principle, reacting more and displaying a stronger connection between inequity and dishonesty. A cross-national experimental comparison to assess the extent to which our findings extend to other countries would therefore prove enlightening in terms of the generality of the behavior we observe with American subjects. Comparative research is particularly needed in this field, since if perceived levels both of honesty and just deserts vary significantly among countries, there might be self-fulfilling beliefs at work to sustain this diversity. We can imagine that countries in which the principle of just deserts is weaker

might induce more dishonesty in their citizens.

Our study showed that individuals may respond to an absence of just deserts with dishonesty. This could result in a societal distribution of resources which is lacking to an even greater degree in just deserts. In future research, it could be interesting to investigate whether all this creates a self-reinforcing spiral into an ever-more dishonest and less virtuous society.

## References

- Abeler, J., Nosenzo, D., and Raymond, C. (2016). Preferences for truth-telling. *CESifo Working Paper Series No. 6087*.
- Almås, I., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, 328(5982):1176–1178.
- Arnold, N. S. (1987). Why profits are deserved. *Ethics*, 97(2):387–402.
- Bäker, A. and Mechtel, M. (2015). Peer settings induce cheating on task performance. *IAAEU Discussion Paper Series in Economics No. 06/2015*.
- Balafoutas, L., Kocher, M. G., Putterman, L., and Sutter, M. (2013). Equality, equity and incentives: An experiment. *European Economic Review*, 60:32–51.
- Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal*, 42(1):25–40.
- Bolton, G. E., Brandts, J., and Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, 115(506):1054–1076.
- Buccioli, A. and Piovesan, M. (2011). Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78.
- Burrows, P. and Loomes, G. (1994). *The Impact of Fairness on Bargaining Behaviour*, pages 21–41. Physica-Verlag HD, Heidelberg.
- Cappelen, A., Hole, A., Sorensen, E., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.
- Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Responsibility for what? fairness and individual responsibility. *European Economic Review*, 54(3):429–441.

- Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2013). When do we lie? *Journal of Economic Behavior & Organization*, 93:258–265.
- Dai, Z., Galeotti, F., and Villeval, M.-C. (2017). Cheating in the lab predicts fraud in the field : An experiment in public transportations. *Management Science*, forthcoming.
- Dal Bó, P., Foster, A., and Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *The American Economic Review*, 100(5):2205–29.
- d’Anjou, L., Steijn, A., and Van Aarsen, D. (1995). Social position, ideology, and distributive justice. *Social Justice Research*, 8(4):351–384.
- Dufwenberg, M. and Dufwenberg, M. (2016). Lies in disguise - a theoretical analysis of cheating. *CESifo Working Paper Series No. 6208*.
- Durante, R., Putterman, L., and van der Weele, J. (2014). Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association*, 12(4):1059–1086.
- Elster, J. (1989). *Solomonic Judgements: Studies in the Limitations of Rationality*. Cambridge paperback library. Cambridge University Press, New York, NY.
- Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58(4):723–733.
- Esarey, J., Salmon, T., and Barrilleaux, C. (2012). Social insurance and income redistribution in a laboratory experiment. *Political Research Quarterly*, 65(3):685–698.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.
- Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.
- Frey, B. S., Benz, M., and Stutzer, A. (2004). Introducing procedural utility: Not only what, but also how matters. *Journal of Institutional and Theoretical Economics (JITE)*, 160(3):377–401.
- Gill, D., Prowse, V., and Vlassopoulos, M. (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior & Organization*, 96:120 – 134.
- Gino, F. and Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*, 20(9):1153–1160.

- Gino, F. and Pierce, L. (2010). Lying to level the playing field: Why people may dishonestly help or hurt others to create equity. *Journal of Business Ethics*, 95(1):89–103.
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.
- Gneezy, U., Kajackaite, A., and Sobel, J. (2016). Lying aversion and the size of the lie. Available at SSRN: <https://ssrn.com/abstract=2852055>.
- Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293–300.
- Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, 75(5):561–568.
- Greenberg, J. (1993). Stealing in the name of justice: Informational and interpersonal moderators of theft reactions to underpayment inequity. *Organizational Behavior and Human Decision Processes*, 54(1):81–103.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.
- Grimalda, G., Kar, A., and Proto, E. (2016). Procedural fairness in lotteries assigning initial roles in a dynamic setting. *Experimental Economics*, 19(4):819–841.
- Guiso, L., Sapienza, P., and Zingales, L. (2008). Alfred marshall lecture: social capital as good culture. *Journal of the European Economic Association*, 6(2-3):295–320.
- Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7:346–380.
- Houser, D., List, J. A., Piovesan, M., Samek, A., and Winter, J. (2016). Dishonesty: From parents to children. *European Economic Review*, 82:242–254.
- Houser, D., Vetter, S., and Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56(8):1645–1655.
- Hurkens, S. and Kartik, N. (2009). Would i lie to you? on social preferences and lying aversion. *Experimental Economics*, 12(2):180–192.
- Jakiela, P. (2011). Social preferences and fairness norms as informal institutions: experimental evidence. *American Economic Review*, 101(3):509–513.



- John, L. K., Loewenstein, G., and Rick, S. I. (2014). Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat. *Organizational Behavior and Human Decision Processes*, 123(2):101–109.
- Kocher, M. G., Schudy, S., and Spantig, L. (2016). I lie? we lie! why? experimental evidence on a dishonesty shift in groups. *CESifo Working Paper Series No. 6008*.
- Konow, J. (1996). A positive theory of economic fairness. *Journal of Economic Behavior & Organization*, 31(1):13–35.
- Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *The American Economic Review*, 90(4):1072–1091.
- Kunovich, S. and Slomczynski, K. M. (2007). Systems of distribution and a sense of equity: A multilevel analysis of meritocratic attitudes in post-industrial societies. *European sociological review*, 23(5):649–663.
- Lewin-Epstein, N., Kaplan, A., and Levanon, A. (2003). Distributive justice and attitudes toward the welfare state. *Social Justice Research*, 16(1):1–27.
- Maggian, V. and Villeval, M. C. (2016). Social preferences and lying aversion in children. *Experimental Economics*, 19(3):663–685.
- Mankiw, N. G. (2010). Spreading the wealth around: reflections inspired by joe the plumber. *Eastern Economic Journal*, 36(3):285–298.
- Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6):633–644.
- Mazar, N. and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, 25(1):117–126.
- Miller, D. (1979). *Social justice*. Oxford University Press, Oxford, New York.
- Mitchell, G., Tetlock, P. E., Newman, D. G., and Lerner, J. S. (2003). Experiments behind the veil: Structural influences on judgments of social justice. *Political Psychology*, 24(3):519–547.
- Ooghe, E., Schokkaert, E., et al. (2007). Equality of opportunity versus equality of opportunity sets. *Social Choice and Welfare*, 28(2):209–230.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 83(5):1281–1302.

- Robert, I. and Arnab, M. (2013). Is dishonesty contagious? *Economic Inquiry*, 51(1):722–734.
- Rose-Ackerman, S. (2001). Trust and honesty in post-socialist societies. *Kyklos*, 54(2-3):415–443.
- Rothstein, B. and Uslaner, E. M. (2005). All for all: Equality, corruption, and social trust. *World politics*, 58(01):41–72.
- Ruffle, B. (1998). More Is Better, But Fair Is Fair: Tipping in Dictator and Ultimatum Games. *Games and Economic Behavior*, 23(2):247–265.
- Rutström, E. and Williams, M. B. (2000). Entitlements and fairness: an experimental study of distributive preferences. *Journal of Economic Behavior & Organization*, 43(1):75–89.
- Schokkaert, E. and Devooght, K. (2003). Responsibility-sensitive fair compensation in different cultures. *Social Choice and Welfare*, 21(2):207–242.
- Scott, J. T. and Bornstein, B. H. (2009). What’s fair in foul weather and fair? distributive justice across different allocation contexts and goods. *Journal of Politics*, 71(3):831–846.
- Scott, J. T., Matland, R. E., Michelbach, P. A., and Bornstein, B. H. (2001). Just deserts: An experimental study of distributive justice norms. *American Journal of Political Science*, 45(4):749–767.
- Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological science*, 23(10):1264–1270.
- Shalvi, S., Handgraaf, M. J., and De Dreu, C. K. (2011). Ethical manoeuvring: why people avoid both major and minor lies. *British Journal of Management*, 22(s1):S16–S27.
- Suri, S., Goldstein, D. G., and Mason, W. A. (2011). Honesty in an online labor market. In von Ahn, L. and Ipeirotis, P. G., editors, *Papers from the 2011 AAAI Workshop*. AAAI Press, Menlo Park, CA.
- Uslaner, E. M. (2008). *Corruption, inequality, and the rule of law*. Cambridge University Press Cambridge, New York, NY.
- van de Ven, J. and Villeval, M. C. (2015). Dishonesty under scrutiny. *Journal of the Economic Science Association*, 1(1):86–99.
- Zak, P. J. and Knack, S. (2001). Trust and growth. *The Economic Journal*, 111(470):295–321.