



HAL
open science

Three Essays in behavioral Ethics on Honesty and Fairness

Julien Benistant

► **To cite this version:**

Julien Benistant. Three Essays in behavioral Ethics on Honesty and Fairness. Economics and Finance. Université de Lyon, 2019. English. NNT : 2019LYSE2085 . tel-02494245

HAL Id: tel-02494245

<https://theses.hal.science/tel-02494245>

Submitted on 28 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre NNT : 2019LYSE2085

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 486 Sciences Économique et de Gestion

Discipline : Sciences économiques

Soutenue publiquement le 21 mai 2019, par :

Julien BENISTANT

Three essays in behavioral ethics on honesty and fairness.

Devant le jury composé de :

Benoît TARROUX, Professeur des universités, Gate Lyon Saint-Etienne, Président

Mathieu LEFEBVRE, Professeur des universités, Université de Starsbourg, Rapporteur

Astrid HOPFENSITZ, Maître de conférences HDR, Toulouse School of Economics, Rapporteur

Agne KAJACKAITE, Chercheuse, Berlin Social Sciences Center (WZB), Examinatrice

Marie-Claire VILLEVAL, Directrice de Recherche, C.N.R.S., Directrice de thèse

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas de modification](#) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer ni l'adapter.

UNIVERSITÉ DE LYON - ECOLE DOCTORALE SCIENCES ECONOMIQUES ET GESTION

UNIVERSITÉ LUMIÈRE LYON 2

Groupe d'Analyse et de Théorie Economique

Thèse de Doctorat (NR) de Sciences Economiques

Présentée par

Julien Benistant

en vue de l'obtention du grade de docteur de l'Université de Lyon

délivré par l'Université Lumière Lyon 2

THREE ESSAYS IN BEHAVIORAL ETHICS ON HONESTY
AND FAIRNESS

Directrice de Thèse :

Marie Claire Villeval - Directrice de recherche CNRS, Université de Lyon

Jury :

Astrid Hopfensitz - Maître de conférence, Toulouse School of Economics - Rapportrice

Agne Kajackaite - Chercheuse, WZB Berlin Social Science Center

Mathieu Lefebvre - Professeur, Université de Strasbourg - Rapporteur

Benoît Tarrow - Professeur, Université de Lyon

University Lumière Lyon 2 is not going to give any approbation or disapprobation about the thoughts expressed in this dissertation. They are only the author's ones and need to be considered such as.

Remerciements

La thèse est souvent vue comme un exercice solitaire, mais il se trouve que c'est avant tout une expérience sociale. En effet, cette thèse n'aurait jamais pu voir le jour sans un grand nombre de personnes que je me dois de remercier.

En tout premier lieu, je remercie chaleureusement Marie Claire Villeval, ma directrice de thèse qui a pris le risque d'accueillir un doctorant n'étant pas de la discipline et qui a toujours fait preuve d'une grande disponibilité malgré son emploi du temps de ministre. Votre rigueur et votre vision de la recherche continuera à m'influencer, j'en suis certain.

Je tiens aussi à remercier Emmanuel Procyk qui m'a permis de ne pas complètement abandonner les neurosciences au cours de ma thèse et qui a toujours pris le temps de répondre à mes questions. De même pour Nadège Bault, qui m'a en plus permis de passer quelques mois à Trento. Ce séjour a été une excellente expérience tant professionnelle que personnelle. Merci aussi à Fabio Galeotti pour sa contribution en tant que coauteur d'un des chapitres de cette thèse mais aussi pour son soutien au cours de la thèse.

Je remercie aussi Astrid Hopfensitz et Mathieu Lefebvre d'avoir accepté d'être les rapporteurs de cette thèse ainsi qu'Agne Kajackaite et Benoit Tarroux pour avoir accepté de faire partie de mon jury de thèse.

Je tiens aussi à remercier tous les doctorants du GATE, Alice, Benjamin, Charlotte, Claire, Clément, Jocelyn, Liza, Maxime, Marius, Morgan, Tatiana, Thomas, Tidiane, Tijuo, Valentin, Vincent, Wilfried, Yohann. Sans vous, cette aventure n'aurait pas eu la même saveur. Merci pour les soirées, les barbecues, les parties de ping-pong endiablées

et les pauses café qui m'ont permis de tenir dans les moments difficiles. Je tiens aussi à remercier particulièrement Rémi Suchon avec qui j'ai écrit une partie de cette thèse et qui a très largement facilité mon introduction à l'Economie, en plus de participer aux activités susmentionnées.

Je n'oublie pas non plus de remercier le reste des membres du GATE-LAB qui ont toujours partagé leurs connaissances et leurs conseils. Plus largement, merci à tous les membres du GATE pour leur aide précieuse et pour avoir fait de ce laboratoire un endroit parfait pour y faire une thèse.

En dehors du monde académique, je tiens à remercier les « gars » et tous mes amis qui ont toujours été là et qui ont su quand m'aider à garder les pieds sur terre. Merci au docteur Rossi de m'avoir ouvert la voie et de m'avoir fait bénéficier de ton expérience.

Enfin, merci à tous les membres de ma famille qui, même s'ils ne sont pas du milieu, ont toujours suivi attentivement les péripéties de cette thèse.

Contents

| | |
|--|-----------|
| General Introduction | 1 |
| 1 Unethical Behavior and Group Identity in Contests | 21 |
| 1.1 Introduction | 21 |
| 1.2 Experimental Design and Procedures | 26 |
| 1.2.1 Experimental Design | 26 |
| 1.2.1.1 No-Identity Treatment | 27 |
| 1.2.1.2 Identity Treatment | 29 |
| 1.2.1.3 Additional Measures | 30 |
| 1.2.2 Procedures | 32 |
| 1.3 Behavioral Conjectures | 33 |
| 1.4 Results | 35 |
| 1.4.1 Misreporting behavior across conditions | 35 |
| 1.4.2 Group Identity and Misreporting Behavior | 43 |
| 1.4.3 Beliefs | 45 |
| 1.5 Robustness tests | 47 |
| 1.6 Discussion and Conclusion | 52 |
| A Instructions | 61 |
| B Tables | 77 |
| C Figures | 83 |

| | | |
|----------|---|------------|
| 2 | The impact of feedback and competitive payment schemes on dishonesty | 87 |
| 2.1 | Introduction | 87 |
| 2.2 | Related Literature | 90 |
| 2.3 | Experimental Design and Procedures | 92 |
| 2.3.1 | Design | 92 |
| 2.3.1.1 | Payment schemes | 93 |
| 2.3.1.2 | Information | 93 |
| 2.3.1.3 | Reporting | 94 |
| 2.3.1.4 | Additional measures | 95 |
| 2.3.2 | Procedures | 96 |
| 2.4 | Behavioral Conjectures | 96 |
| 2.5 | Results | 98 |
| 2.5.1 | Lying Behavior Across Treatments | 98 |
| 2.5.2 | Dynamics of Lying Behavior Over Time | 105 |
| 2.5.3 | Gender Heterogeneity | 107 |
| 2.6 | Discussion and Conclusion | 109 |
| A | Instructions | 117 |
| A.1 | Two-Sided treatments | 117 |
| A.2 | One-Sided treatments | 121 |
| B | Tables | 127 |
| C | Figures | 130 |
| 3 | It does (not) get better: the effect of relative gains and losses on subsequent giving | 137 |
| 3.1 | Introduction | 137 |
| 3.2 | Experimental design and procedures | 141 |
| 3.2.1 | Experimental game | 141 |
| 3.2.2 | Experimental conditions | 143 |
| 3.2.3 | Additional measures | 144 |

| | | |
|-------|--|------------|
| 3.2.4 | Procedures | 146 |
| 3.3 | Theoretical framework and hypotheses | 147 |
| 3.3.1 | The model | 147 |
| 3.3.2 | Hypotheses | 152 |
| 3.4 | Results | 153 |
| 3.4.1 | Behavioral results | 153 |
| 3.4.2 | Emotions | 156 |
| 3.5 | Discussion and Conclusion | 160 |
| A | Instructions | 167 |
| B | Tables | 170 |
| C | Figures | 172 |
| D | Proofs | 173 |
| D.1 | Wage and profit | 173 |
| D.2 | Continuity and concavity of the utility function | 173 |
| D.3 | Details of the calculation for proof 2 & 3 | 174 |
| | General Conclusion | 177 |

List of Tables

| | | |
|-------|--|-----|
| 1.1 | Performance, frequency and intensity of selfish misreporting, by treatment and condition | 38 |
| 1.2 | Determinants of the decision to make a selfish misreport | 41 |
| 1.3 | Determinants of the relative intensity of selfish misreporting | 42 |
| 1.4 | Determinants of beliefs on misreporting decisions in past sessions | 47 |
| 1.5 | Average reported random number by treatment and condition | 51 |
| B.1 | Characteristics of participants, by treatment | 77 |
| B.2 | Determinants of performance | 78 |
| B.3 | Frequency and intensity of non-selfish misreporting, by treatment and condition | 80 |
| B.4 | Determinants of selfish misreporting in the Identity-No-Info treatment . | 81 |
| B.5 | Determinants of the reported random number in the Identity-No-Info and the Identity-No-Scrutiny treatments | 82 |
| 2.3.1 | Treatment dimensions and conditions | 95 |
| 2.5.1 | Reports by treatment | 99 |
| 2.5.2 | Dunn test: Pairwise comparisons of the value of the reports between treatments | 101 |
| 2.5.3 | Pairwise comparisons of treatments for all outcomes | 104 |
| B.1 | Demographic characteristics of participants, by treatment | 127 |
| B.2 | Determinants of the roll outcome reported | 128 |
| B.3 | Determinants of beliefs in the TR-FF treatment | 129 |

| | | |
|-------|---|-----|
| 3.2.1 | Number of participants by experimental condition | 147 |
| 3.4.1 | Average transfers by experimental condition | 153 |
| 3.4.2 | Treatment effects over the different levels of wage. | 154 |
| 3.4.3 | Treatment effect on loss averse <i>vs.</i> not loss averse participants, at wage=€5. | 155 |
| 3.4.4 | Marginal effects of physiological arousal on transfers over the different level of wage. | 160 |
| B.1 | Demographic characteristics. | 170 |
| B.2 | Pairwise correlations between responses in the emotion questionnaire. . | 171 |

List of Figures

| | | |
|-------|--|-----|
| 1.1 | Timeline of the experiment | 32 |
| 1.2a | Distribution of numbers reported in the Self-condition, by treatment . . | 36 |
| 1.2b | Distribution of numbers reported in the Opp-condition, by treatment . . | 36 |
| 1.3 | Average allocation in ECU as a third-party in scenario 3, by period (stage 3 of the group identity induction procedure) | 44 |
| C.1 | First decision in the SVO test | 83 |
| C.2 | Average allocation in ECUs as a third party in scenario 1 (left panel) and scenario 2 (right panel), by period (stage 3 of the group identity induction procedure) | 83 |
| C.3 | Distribution of the reported numbers in the Opp-condition, by treatment and identity matching | 84 |
| C.4 | Distribution of the reported numbers in the Self-condition, by treatment and identity matching | 85 |
| 2.5.1 | Marginal effects of time on the probability to report each outcome, by treatment | 106 |
| 2.5.2 | Probability of reporting a red outcome, by gender and treatment | 108 |
| C.1 | Average reported value by treatments and across blocks of four periods | 130 |
| C.2 | Standard deviation of the total value of reports within true pairs and within hypothetical pairs in the TR and PR treatments | 131 |
| C.3 | Standard deviation of the total value of reports between true pairs and between hypothetical pairs in the TR and PR treatments | 132 |

| | | |
|-------|---|-----|
| C.4 | between minus within standard deviation for true pairs and hypothetical pairs in the PR and TR treatments | 133 |
| C.5a | Probability to report a blue outcome for male and female participants, by treatment | 134 |
| C.5b | Probability to report a yellow outcome for male and female participants, by treatment | 134 |
| C.6 | Average beliefs on others' reports, by treatment and period. | 135 |
| 3.4.1 | Electrodermal activity across conditions | 157 |
| 3.4.2 | Indexes of positive (top) and negative emotions across conditions | 158 |
| C.1 | Average responses for each items of the emotion questionnaire, by condition | 172 |

General Introduction

Morality is the basis of things and truth is
the substance of all morality

Mahatma Gandhi

Moral or ethical principles are guidelines for people's social conduct.⁵ They play a central role in human behavior: people are willing to forgo material resources or even risk their physical integrity to be in line with them. Some people give money to charities or homeless individuals, some abstain to lie to others while it would be profitable, and some even sacrifice their life for morally-guided causes or principles. These examples explain why ethical principles have attracted the attention of many scholars from various disciplines.

Philosophers were the first to study ethics, using a normative or prescriptive approach. Their goal is to determine how people should behave to be ethical or moral. Among the multiple theories proposed by the moral philosophers, two of them have been particularly influential. The first one is the deontological ethics that states that the morality of an action shall be judged on the action's adherence to rules or principles. The second one, utilitarianism, examines the consequences of the action rather than the action itself to judge about its morality. These two theories have been at the center of the debate between philosophers. To clarify the distinction between the two approaches and determine whether people are more likely to follow one or another, scholars often

⁵The terms moral and ethic are used interchangeably in this thesis. For a discussion of the difference between ethics and moral see for example Gino and Shalvi (2015).

employ the so-called trolley and footbridge dilemmas. In these hypothetical ethical dilemmas, people are asked to choose between two options with one leading to the death of five individuals while the other involves the death of only one individual. Although simplistic in appearance, the action associated with the latter option can be morally reprehensible. In the trolley dilemma, people have to imagine the following scenario:

A runaway trolley is heading for five people who will be killed if it proceeds on its present course. The only way to save them is to pull a lever that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one?

In the footbridge dilemma, a variant of the trolley dilemma, people have to imagine the following scenario:

A runaway trolley threatens to kill five people. You are standing on a footbridge spanning the tracks between the oncoming trolley and the five people. Next to you is a railway worker who is wearing a large backpack. The only way to save the five people is to push this worker off the bridge and onto the tracks below. The man will die as a result, but his body will stop the trolley from reaching the others. Ignoring legal concerns, is it okay to save the five people by pushing this stranger to his death?

In both scenarios, if people decide neither to pull the lever nor to push the worker, their decisions follow the prescription of deontological ethics: one cannot willingly harm others, independently of the consequences of his action. Conversely, if people decide to pull the lever or to push the worker, their choices are in line with the utilitarian vision of ethics. In that case, people favors the least socially harmful option by choosing to kill one person instead letting five die. In experiments, when participants face these dilemmas, they tend to favor the utilitarian option in the trolley dilemma (*i.e.*, pulling the lever) while they favor the deontological option in the footbridge dilemma (*i.e.*, not

pushing the man) (e.g., Kamm, 1998; Greene et al., 2001). This apparent inconsistency in people's moral judgment fueled the debate in moral philosophy but also shows the limitations of the normative approach of ethics. Indeed, it suggests that people's ethical judgments, and by extension decisions, do not follow rules cast in stone but are rather influenced by the context.

Consequently, alongside this prescriptive approach of ethics, other scholars started to use a positive approach of ethics. Instead of determining what people should do to behave ethically, their goal is to determine why people behave ethically in some cases but not in others.

Psychologists have a long tradition of investigating (un)ethical behavior with non-incentivized experiments and have developed several theories to explain people's moral behavior. Some theories propose that people need skills in moral reasoning to behave ethically. A certain level of moral development has to be reached to be able to follow ethical principles (e.g., Kohlberg, 1981; Rest, 1986). More recently, some psychologists challenged these theories and introduced the idea that moral decision-making is mostly intuitive (e.g., Haidt, 2001), moral reasoning being only used to create *post-hoc* justifications about what people have decided intuitively. By working with neuroscientists, they also pointed out the central role of emotions in intuitive moral judgment and decision-making which depend on the individuals' empathic abilities (e.g., Greene et al., 2001; Moll et al., 2005, 2008; Raine and Yang, 2006; Raine, 2008; Decety and Batson, 2009).

In Economics, the history of the investigation of moral behavior is more complicated. In the early development of modern economics, moral and ethical principles were seen as central to explain economic behavior. Adam Smith, who is often referred as the founder of modern economic science, depicted in his first book that virtues such as justice, prudence and benevolence are the main driving forces of human behavior (Smith, 1759).⁶ With Adam Smith's second book, *The Wealth of Nation* (Smith 1723-1790, 2000), and the subsequent work of economists such as David Ricardo or John

⁶For a recent re-investigation of Adam Smith's work see Smith and Wilson (2019)

Stuart Mill, moral became less and less a central topic in the discipline. The weakening of the links between economics and the notion of morality is well illustrated by the concept of the Homo Economicus, born in the 19th century. This theory depicts the economic agent as a self-interested, rational and amoral being.⁷ According to the, more recent, "economic-of-crime" theory, when confronted with an ethical dilemma, the Homo Economicus is supposed to behave ethically only if the material rewards of the ethical outcome outweigh the monetary incentives associated with the unethical act (e.g., Becker, 1968). Thus, the economic agents are supposed to react only to exogenous changes of the monetary incentives, such as the probability of being caught when acting unethically. It was only in the past decades that economists started to rethink the Homo Economicus model based upon experimental evidence showing that people value ethical principles and are concerned about other's welfare giving birth to the field of behavioral economics (e.g., Camerer and Loewenstein, 2004; Thaler, 2016).⁸

In spite of their methodological and conceptual differences, scholars from psychology and economics who study (un)ethical behavior contribute to the field of behavioral ethics.⁹ In this field of research, scholars aim for a similar goal which is: "the study of systematic and predictable ways in which individuals make (un)ethical decisions" (Bazerman and Gino, 2012). Their observations converged toward one main result. Ethical decision-making is not absolute, people who are ethically straight in one context can behave unethically in a different one. Thus, morality is malleable rather than being a stable trait or the results of a simple monetary cost-benefit analysis (e.g., Messick and Tenbrunsel, 1996; Bazerman and Gino, 2012)

⁷The words of Amartya Sen, the winner of the 1998 Nobel Prize for Economics illustrate the detachment of economists from moral and ethics: the nature of economics "has been substantially impoverished by the distance that has grown between economics and ethics" (Sen, 1987).

⁸This paragraph is of course a simplification of the History of economic thought as this is not the main topic of this dissertation. Before the birth of behavioral economics, some authors, such as Gary Becker, had already stressed out the importance of considering individuals as not purely self-interested (e.g., Becker, 1976, 1978).

⁹To illustrate the inter-disciplinarity of this field one can look at the special issue on behavioral ethics published in 2015 in *Current Opinion in Psychology*. The contributions range from biology and psychology to economics and management.

This thesis contributes to this flourishing field of research and focuses on two ethical principles, honesty and fairness. In the following sections I review the main results in the literature on (dis)honesty and fairness with a focus on the contributions of behavioral economics and I present the contributions of this thesis to each of these literatures.

(dis)Honesty

The notion of dishonesty encompasses a large set of behavior, ranging from a small lie to a colleague up to embezzling money and misreporting your income to the tax authorities. Such behavior is, by nature, hard to observe as people usually try to conceal their misdeeds. To solve this issue, behavioral economists developed incentivized laboratory economic games in which players can make profitable lies at the expense of either another participant or the experimenter. Studies on tax compliance were the first to use this type of economic games in experimental settings (*e.g.*, Alm et al., 1992a,b, 1993). Following this early work, behavioral economists developed two main types of games to study lying behavior: the sender-receiver games and the self-reporting games.¹⁰

Sender-receiver games are two-player communication games (*e.g.*, Gneezy, 2005; Hurkens and Kartik, 2009; Sutter, 2009) which follow a similar design and that originated from the literature on cheap-talk games (*e.g.*, Crawford and Sobel, 1982). The first player, the sender, receives private information about the monetary consequences of two options. Then, he chooses to send a message stating which option is the more profitable for the second player, the receiver who is not informed. Based on this message, the receiver has to choose one of the two options. To induce an ethical dilemma, incentives are misaligned, one option usually favors the sender over the receiver. In that case, the sender has a material interest to deceive the receiver by sending him the wrong message.

¹⁰ Alongside lying, which can be defined as telling something untrue, behavioral economists studied other types of dishonest behavior. One of them is deceptive behavior which is close from lying behavior but do not necessarily involve telling a lie. Indeed, someone can deceive another individual by telling him the truth while knowing that the other will not believe him (see Mahon (2016) for a thorough discussion on lying and deception). Another dishonest behavior studied is cheating. Unlike the previously described behavior, cheating is more general and refers to rule-breaking behavior.

Variations of this standard version of the game have been largely used to study deceptive behavior but using this baseline version, Gneezy (2005) and (Sutter, 2009) reported that around 40% of the senders chose to deceive the receiver.

In self-reporting games, players have to report the outcome of a simple task, usually tossing a coin (coin task) or rolling a die (die-under-the-cup task) (*e.g.*, Fischbacher and Föllmi-Heusi, 2013; Bucciol and Piovesan, 2011). These games are mostly conducted under conditions of total anonymity from both the others players and from the experimenter. Players are aware of the payoffs associated with all the possible outcomes and have incentives to lie depending on the random outcome. The lack of observability of lying at the individual level forces the experimenter to compare the distribution of the reported outcome to the theoretical distribution ($1/2$ for heads on a coin, $1/6$ for each side of a die) to estimate lying level at the aggregate level. In another type of self-reporting games, the outcome that players have to report is not based on a random outcome but on their performance in a real effort task (*e.g.*, Mazar et al., 2008). In that case, the experimenters can detect dishonesty at the individual level if they know the true performance of the players. In the earliest studies that used self-reporting games, 86% of the participants reported the most profitable side of the coin (Bucciol and Piovesan, 2011) and 35% participants claimed to have rolled the most profitable outcome of the die in the die-under-the-cup task (Fischbacher and Föllmi-Heusi, 2013). A recent meta-analysis reports that the latter result is robust and found in several others studies using a similar design (Abeler et al., 2019).

Studies that used any of these games always reported that people are neither fully dishonest nor fully honest. To account for this heterogeneity, scholars have proposed the existence of an intrinsic cost of lying (*e.g.*, Erat and Gneezy, 2012; López-Pérez and Spiegelman, 2013; Gibson et al., 2013; Gneezy et al., 2013; Abeler et al., 2014; Kajackaite and Gneezy, 2017). Starting from this point, the literature divided in two strands, some studies investigate the nature of the intrinsic cost of lying while others focus on the exogenous factors that modulate it.

In the former strand of the literature, three propositions have been made about the cost of lying. The self-concept maintenance theory states that people can be dishonest as long as they can label themselves as ethical individuals (*e.g.*, Bénabou and Tirole, 2002; Mazar et al., 2008; Shalvi et al., 2011). Thus, the cost of lying represents how much a lie would hurt one's self-image. More recently, the reputation theory has expanded the self-concept maintenance theory by integrating social-image concern (*e.g.*, Bénabou and Tirole, 2006; Fischbacher and Föllmi-Heusi, 2013; Kajackaite and Gneezy, 2017; Khalmetski and Sliwka, 2017; Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018). In addition to care about their self-image people are expected to care about how others see them. Consequently, the cost of lying reflects the willingness of people to not appear dishonest to others. These theoretical advances are important to the literature as they help to understand some central behavioral results. For example, in Fischbacher and Föllmi-Heusi (2013), authors report that a large fraction of the participants are only partially lying: participants report to have rolled a 4 while reporting a 5 would be more profitable. This can be explained by the fact that participants care both about their self and social-image. They lie partially to avoid hurting their moral image and to hide their misdeeds to the experimenter. Alongside the self and social-image concerns, some scholars have proposed that the cost of lying represents a pure-aversion to lying (*e.g.*, Erat and Gneezy, 2012; López-Pérez and Spiegelman, 2013). In that case people emotionally suffer from lying due to its unethical nature. This pure aversion is in line with the deontological view of ethics which states that the morality of a decision is solely based on whether an action follows some set of rules, which in that case is truthfulness.

Alongside the aforementioned studies, another strand of the literature focuses on the exogenous factors that affect people tendency to lie.

The social context has been shown to affect dishonesty through two channels. First, the act of lying has often social consequences: others can be harmed or can benefit from it. Previous studies have shown that the more harmful the consequences of a lie are, the less likely people behave dishonestly (*e.g.*, Gneezy, 2005). Conversely, when others can

benefit from their lie, people are more likely to be dishonest (Erat and Gneezy, 2012). Consequently, people are willing to use opportunities to blind themselves about the repercussions of their dishonesty (*e.g.*, Shalvi et al., 2011; Pittarello et al., 2015).

Second, people have opportunities to observe others' behavior as they live in a social world and others' behavior can be dishonest. Evidence from the laboratory reports that dishonesty is contagious (*e.g.*, Gino et al., 2009; Fosgaard et al., 2013; Innes and Mitra, 2013) although not systematically (see *e.g.*, Fortin et al., 2007). When one observes peers dishonesty, it signals that it is socially acceptable to lie. Thus, individuals who lie in this context suffer from a lower cost of lying leading them to be more likely to act dishonestly. Such distortion of the social norm on truth-telling has been observed as endogenous to some professions (Cohn et al., 2014) or in culture where institutions are weaker (Gächter and Schulz, 2016).

On a more familiar note for economists, honesty is also affected by the type of incentives at stake in the decision environment. Intuitively, one may think that dishonesty is more likely as the incentives get larger and larger, however a recent meta-analysis shows that this is not the case (Abeler et al., 2019). Some evidence suggests that this is explained by the reputation cost of lying (Kajackaite and Gneezy, 2017). People are more likely to lie to avoid a loss than to increase their gain (Grolleau et al., 2016; Garbarino et al., 2018) which is in line with the literature on loss aversion (*e.g.*, Kahneman et al., 1986). Similarly, the nature of the incentive schemes affects the likelihood of dishonesty. Competitive incentives, widely used to improve workers performance, have been shown to increase dishonest behavior (*e.g.*, Schwieren and Weichselbaumer, 2010; Belot and Schröder, 2013; Feltovich, 2018), even when competition is about social status rather than money (Charness et al., 2014). A simple explanation of this result is that lying is usually more profitable when competitive incentives are at stake. In that case the marginal gain of lying is increased in competition. Additionally, competition could prime a competitive mindset which would dehumanize the opponents. Then, dishonest actions which would harm them seem less blameworthy (*e.g.*, Rigdon and D'Esterre, 2017). Finally, the social norm in competitive context could prescribe that dishonesty is

more acceptable than in other contexts. All these explanations are in line with the idea that the incentive to lie, or to “play dirty” is actually built into the competitive reward structure (Kohn, 1992).

In this thesis we contribute to the literature on dishonesty by investigating how it is modulated by contextual and social factors in competition. Previous research has already been done on competition but have let aside factors that are often encountered in real-life settings, such as social identity or the dynamic aspect of competition.

Chapter 1 tests whether cheating in competition is sensitive to one’s opponent social proximity and to the nature of the lie. Competition often occurs in social context in which competitors may have social ties with some of their opponents creating a feeling of belonging to a same social group. Such social identity have been shown to bias people who tend to favor in-group member (*e.g.*, Akerlof and Kranton, 2000; Chen et al., 2009). Thus, we can expect that social identity would potentially mitigate the negative relationship between competition and dishonesty.

We implemented a repeated two-player Tullock contest in which the relative final scores of the participants determine their probability to win the competition. The participants’ final score is the sum of their performance in a real effort task and a random number that they have to report. In a 2 by 2 design, we vary within-subject whether participants have to report their own random number or their opponent’s one. Participants are able to lie to increase their likelihood to win by either overreporting their own random number or underreporting their opponent’s one. We also manipulate, between-subject, the existence of social identity by inducing it, or not, following the procedure proposed by Chen et al. (2009). It allows us to vary whether participants are paired with an in-group or an out-group opponent. We hypothesized that participants would be less likely to lie about their opponent’s number than about their own, as lying about others’ performance is morally more reprehensible than lying about your own performance (Rigdon and D’Esterre, 2015). We also expect that participants would be less likely to lie when paired with an in-group opponent than when paired with an

out-group opponent, as people tend to favor members of their own social group over others (*e.g.*, Tajfel and Turner, 1979; Akerlof and Kranton, 2000; Chen et al., 2009; Goette et al., 2006, 2012).

We show that participants are as likely to misreport in any condition, neither the social group of their opponent nor the nature of the number that they report affect their dishonest behavior. Thus, dishonest behavior in competition is insensitive to social distance, unlike what have been found in non-competitive contexts (*e.g.*, Jiang, 2015; Cadsby et al., 2016; Meub et al., 2016). To ensure that our results are robust, we implement a new treatment in which there is no possible scrutiny from the experimenter as scrutiny has been shown to affect dishonesty (*e.g.*, Gneezy et al., 2018). Indeed, participants who are sensitive to social image are less likely to misreport when there is scrutiny and these participants could also be sensitive to their opponent's social identity. However, unlike expected, participants' dishonesty is not affected by their opponent's identity in this new treatment. Yet, participants are now less likely to misreport their opponent's number than their own number. This last result is in line with the idea that reputation is an important component of the cost of lying.

Chapter 2 explores whether both a dynamic competition and social information on others' (dis)honest behavior affects individuals' dishonesty. These two aspects have received only little attention in the literature while this is an important component of competition in real-life settings. In dynamic settings people can gather information about their opponents throughout the competition which can modulate dishonesty.

We implemented a repeated variant of the die-under-the-cup task (Fischbacher and Föllmi-Heusi, 2013) with participants being matched in pairs. Participants have to report the outcome of a die roll and the sum of their reports in the successive periods determine their performance. Using a 2 by 2 between-subject design, we vary whether participants are paid based on a piece-rate or a tournament. We also vary whether participants receive continuous feedback about their counterpart's performance or only a final feedback. Additionally, in tournament settings, we vary between-subject whether only one

participant could lie within the pair. This last variation allows us to test whether the ability to lie of one's opponent is at the core motive for higher likelihood of dishonesty observed in competitive settings compared to non-competitive ones. We expect to replicate previous findings showing that under tournament incentives participants would lie more than under piece-rate incentives (*e.g.*, Schwierien and Weichselbaumer, 2010; Belot and Schröder, 2013). We also argue that the effect of tournament incentives on dishonesty depends on one's opponent ability to lie: if he is not, we expect dishonest behavior to be less likely than when he is able to lie. Finally, continuous feedback on participants' counterpart's reports is expected to increase cheating behavior, compared to the final feedback treatments. We argue that this effect would be larger under competition as the incentives to follow one's counterpart dishonesty are larger than in the piece-rate payment scheme.

Our results support only partially our assumptions. We show that continuous feedback increases dishonesty in the piece-rate, compared to the final feedback treatment, but not in the tournament. However, participants mimic their counterpart's dishonest behavior in both treatments leading to a significant increase of dishonesty over time. We also find that participants lie more under tournament incentives than under piece-rate incentives, as expected. Also, participants are less likely to lie when their opponent cannot cheat than when he can, confirming the the apprehension of the opponent's dishonesty is a driver of lying behavior in competition. Finally, in tournaments in which the opponent cannot lie participants cheat as much as in the piece-rate payment scheme. This suggests that the tournament incentives themselves are not sufficient to induce more dishonesty than non-competitive incentives. Unlike expected, our data also show that the lack of effect, at the aggregate level, of continuous feedback in the tournament is driven by males' reaction to the uncertainty about their opponent's dishonesty. They are more dishonest than in any other conditions when they are not informed about their opponent's reports while females are not.

Fairness

Alongside the principle of honesty, social behavior is also largely influenced by another ethical principle, fairness. In behavioral economics, fairness refers to people's social preference for equitable outcomes and encompasses concepts such as altruism and reciprocity (*e.g.*, Fehr and Schmidt, 1999; Dana et al., 2011). Altruistic behavior has been reported in experiments using dictator games (Forsythe et al., 1994). In this game, a player in the role of a "dictator" has to split an endowment of money between himself and another anonymous player (the recipient). The dictator decides freely how much he is willing to share without any possibility of being punished. Multiple experiments report that dictators share, on average, 30% of their endowment (see the meta-analysis of Engel, 2011), suggesting that people have a sense of fairness which drives them to consider others' earnings when making decisions. To show whether people reciprocate to unfair behavior, behavioral economists used the ultimatum game. In this game, unlike in the dictator game, the recipient can reject the offer of the proposer (a.k.a. the dictator). If he does, both players receive no earnings. Experimental results show that recipients are willing to reject unfair offers although it leads them to earn nothing (*e.g.*, Güth et al., 1982; Camerer, 2003). Overall, the results of these two games show that people value fairness as they exhibit altruism and reciprocate negatively when being treated unfairly.

To account for such behavior, several theories of social preferences have been proposed by behavioral economists. In particular, inequality aversion theory states that individuals dislike inequalities, even the advantageous ones (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). They also proposed that people are looking to maximize social welfare and incorporate others' intention (Charness and Rabin, 2002). Finally, some theories focused on the role of reciprocity as they supposed that individuals reciprocate to benevolent behavior (Rabin, 1993; Falk and Ichino, 2006). Yet, like honesty, fairness is also affected by the social context. Individuals' concern for social image (Bénabou and Tirole, 2002) modulates their generosity in the dictator game: when individuals are exposed to others' scrutiny, they tend to give more (*e.g.*, Andreoni and Bernheim, 2009). Furthermore, the social norm at stake in a given context can affect people sharing behavior in the dictator game (Bicchieri and Xiao, 2009; Krupka and Weber, 2013). Other

factors influencing fairness include the possibility to hide the social consequences of acting selfishly (*e.g.*, Dana et al., 2006), the social distance with the recipient (*e.g.*, Chen et al., 2009), the legitimacy of the endowment to redistribute (*e.g.*, Cherry et al., 2002) and the framing of the decision (*e.g.*, List, 2007; Bardsley, 2008).

Chapter 3 contributes to this literature by identifying a new factor that affects individuals' level of generosity. There are some evidences that people form reference point from which they compare their actual outcome. A lower outcome than the reference point is experienced as a loss while a higher outcome is experienced as a gain (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Köszegi and Rabin, 2006). While thoroughly studied, we don't know how such losses or gains spillover on social preferences and more precisely, on altruism. To determine whether such an effect is prevalent or not we designed an experiment in which participants start by performing a real effort task for a fixed wage, followed by a dictator game in which they have to share a €5 endowment. We implement a 3x2 design varying the level of wage earned for the real effort task and the of the revelation of the wage. Participants can receive three levels of wage, low (€5), medium (€10) or high (€15). In a first set of conditions, the level of wage is assigned at the beginning of the experiment and participants are not aware of the other wage levels. We pay participants the corresponding wage in an envelope before they perform the real effort task. In the other set of conditions, the level of wage is only indicated after individuals performed the real effort task. Each levels are equally likely to be selected and participants are aware of the potential three levels of wage. Participants receive, in an envelope, the medium wage before the real effort task from which we withdraw money if the selected wage is low or add money if the selected wage is high. In this set of conditions, we expect that participants evaluate the selected wage against the potentials levels while they cannot do so in the first set of conditions. We also collect both self-reported and physiological measures of emotions. Thanks to these measurements, we are able to determine whether the evaluation of

others potential wages levels induces an emotional response that affects subsequent transfers in the dictator game.

We derive our assumptions from a model in which social preferences are dependent of a reference point (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Köszegi and Rabin, 2006). In this model, participants have an expectation about how much they will earn. That depends on the reference wage and the hypothetical transfer they would choose if they actually received their reference wage. We assume that participants' reference wage is the medium wage, as both the status-quo based and the expectation-based formation of the reference point would predict. This means that if the high wage is selected, it will be encoded as a gain and if the low wage is selected, it will be encoded as a loss. Our model predicts that, for the low wage, participants who receive information about the other levels of wage will transfer less than the participants who are not informed. Conversely, for the high wage, informed participants will transfer more than non-informed participants. The first prediction is supported by our results but only for loss averse participants. The second prediction is also supported by our results for all participants regardless of their attitude toward losses. Regarding emotions, in spite of participants experiencing emotions of the expected valence, when the actual wage is revealed they have no explanatory power on transfers. It suggests that emotions are not responsible for the spillover effects of losses and gains on transfers in the dictator game.

By manipulating the salience of the counter factual wages, and showing that this impacts generosity in a subsequent dictator game, we show the importance of the context on altruism. These results are in line with the aforementioned studies showing that subtle manipulations of the context in which the decision is taken impact fairness related-decisions. A premature conclusion could be that social preferences are very contextual by nature, which would cast doubt on their predictive power. An alternate conclusion is that social preferences are rather stable within individuals but the understanding of behavior driven fairness concern requires a thorough examination of the circumstances

of the decisions. This examination in turn can provide interesting insights on the social and psychological motivations of individuals (Falk and Heckman, 2009).

Bibliography

- Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. *Journal of Public Economics* 113, 96–104.
- Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica*. Forthcoming.
- Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics* 115(3), 715–753.
- Alm, J., B. Jackson, and M. McKee (1992a). Institutional uncertainty and taxpayer compliance. *American Economic Review* 82(4), 1018–1026.
- Alm, J., B. R. Jackson, and M. McKee (1992b). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal* 45(1), 107–114.
- Alm, J., B. R. Jackson, and M. McKee (1993). Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior Organization* 22(3), 285 – 303.
- Andreoni, J. and B. D. Bernheim (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica* 77(5), 1607–1636.
- Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics* 11(2), 122–133.
- Bazerman, M. H. and F. Gino (2012). Behavioral Ethics: Toward a Deeper Understanding of Moral Judgment and Dishonesty. *Annual Review of Law and Social Science* 8, 85–104.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy* 76(2), 169–217.
- Becker, G. S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature* 14(3), 817–826.
- Becker, G. S. (1978). *The Economic Approach to Human Behavior*. Number 9780226041124 in University of Chicago Press Economics Books. University of Chicago Press.
- Belot, M. and M. Schröder (2013). Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour. *Journal of Economic Behavior and Organization* 93(C), 233–238.

- Bénabou, R. and J. Tirole (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics* 117(3), 871–915.
- Bénabou, R. and J. Tirole (2006). Incentives and Prosocial Behavior. *American Economic Review* 96(5), 1652–1678.
- Bicchieri, C. and E. Xiao (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making* 22(2), 191–208.
- Bolton, G. E. and A. Ockenfels (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review* 90(1), 166–193.
- Buccioli, A. and M. Piovesan (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology* 32(1), 73–78.
- Cadsby, C. B., N. Du, and F. Song (2016). In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization* 128, 59–71.
- Camerer, C. and G. Loewenstein (2004). Behavioral Economics: Past, Present, Future. In *Advances in Behavioural Economics*, pp. 1–52.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. New York, NY, US: Russell Sage Foundation.
- Charness, G., D. Masclet, and M. C. Villeval (2014). The Dark Side of Competition for Status. *Management Science* 60(1), 38–55.
- Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics* 117(3), 817–869.
- Chen, Y., S. X. Li, and Y. Chen (2009). Group identity and social preferences. *American Economic Review* 97(4), 431–457.
- Cherry, T. L., P. Frykblom, and J. F. Shogren (2002). Hardnose the Dictator. *American Economic Review* 92(4), 1218–1222.
- Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature* 516(7529), 86–89.
- Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica* 50(6), 1431–1451.
- Dana, J., G. Loewenstein, and R. Weber (2011). Ethical Immunity: How people violate their own moral standards without feeling they are doing so. In D. De Cremer and A. E. Tenbrunsel (Eds.), *Behavioral business ethics: Shaping an emerging field*, pp. 197–216. New York: Psychology Press.
- Dana, J., R. A. Weber, and J. X. Kuang (2006, sep). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1), 67–80.

- Decety, J. and D. C. Batson (2009, 08). *Empathy and Morality: Integrating Social and Neuroscience Approaches*, pp. 109–127.
- Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory* 175(175), 248–264.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics* 14(4), 583–610.
- Erat, S. and U. Gneezy (2012). White Lies. *Management science* 58(4), 723–733.
- Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science* 326, 535–538.
- Falk, A. and A. Ichino (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics* 24(1), 39–57.
- Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114(3), 817–868.
- Feltovich, N. (2018). The interaction between competition and unethical behaviour : experimental evidence. *Experimental Economics*. Forthcoming.
- Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise – An Experimental Study on Cheating. *Journal of the European Economic Association* 11(40), 525–547.
- Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior* 6, 347–369.
- Fortin, B., G. Lacroix, and M. C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics* 91(11-12), 2089–2112.
- Fosgaard, T. R., L. G. Hansen, and M. Piovesan (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior and Organization* 93, 279–284.
- Gächter, S. and J. F. Schulz (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531(7595), 496–499.
- Garbarino, E., R. Slonim, and M. C. Villeval (2018). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization*. In press.
- Gibson, R., C. Tanner, and A. F. Wagner (2013). Preferences for Truthfulness : Heterogeneity among and within Individuals. *American Economic Review* 103(1), 532–548.
- Gino, F., S. Ayal, and D. Ariely (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological Science* 20(3), 393–8.
- Gino, F. and S. Shalvi (2015). Editorial overview: Morality and ethics: New directions in the study of morality and ethics. *Current Opinion in Psychology* 6, v–viii.

- Gneezy, U. (2005). Deception: The Role of Consequences. *American economic review* 95(1), 384–394.
- Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review* 108(2).
- Gneezy, U., B. Rockenbach, and M. Serra-Garcia (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization* 93, 293–300.
- Goette, L., D. Huffman, S. Meier, and B. L. Goette (2006). The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups. *American Economic Review* 96(2), 212–216.
- Goette, L., D. Huffman, S. Meier, and M. Sutter (2012). Competition between organizational groups : Its impact on altruistic and anti-social motivations. *Management Science* 58(5), 948–960.
- Greene, J. D., R. B. Sommerville, and L. E. Nystrom (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science (New York, N.Y.)* 293, 2105–2108.
- Grolleau, G., M. G. Kocher, and A. Sutan (2016). Cheating and loss aversion: Do people cheat more to avoid a loss? *Management Science* 62(12), 3428–3438.
- Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior Organization* 3(4), 367 – 388.
- Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108(4), 814–834.
- Hurkens, S. and N. Kartik (2009). Would i lie to you? on social preferences and lying aversion. *Experimental Economics* 12(2), 180–192.
- Innes, R. and A. Mitra (2013). Is dishonesty contagious? *Economic Inquiry* 51(1), 722–734.
- Jiang, T. (2015). Other-regarding Preferences and Other-regarding Cheating - Experimental Evidence from China, Italy and the Netherlands. Mimeo.
- Kahneman, D., J. L. Knetsch, and R. Thaler (1986, September). Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *American Economic Review* 76(4), 728–741.
- Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–291.
- Kajackaite, A. and U. Gneezy (2017). Incentives and cheating. *Games and Economic Behavior* 102, 433–444.
- Kamm, F. M. (1998). *Morality, Mortality Volume I: Death and Whom to Save From It*. New York: Oxford University Press.

- Khalmetski, K. and D. Sliwka (2017). Disguising Lies – Image Concerns and Partial Lying in Cheating Games. CESifo Working Paper no. 6347.
- Kohlberg, L. (1981). *Essays on Moral Development, Vol. I: The Philosophy of Moral Development*. Harper Row.
- Kohn, A. (1992). *No contest : the case against competition*. Revised edition. Boston : Houghton Mifflin.
- Köszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121(4), 1133–1165.
- Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association* 11(3), 495–524.
- List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy* 115(3), 482–493.
- López-Pérez, R. and E. Spiegelman (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics* 16(3), 233–247.
- Mahon, J. E. (2016). The definition of lying and deception. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.
- Mazar, N., O. Amir, and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45(6), 633–644.
- Messick, D. M. and A. E. Tenbrunsel (Eds.) (1996). *Codes of Conduct: Behavioral Research into Business Ethics*. Russell Sage Foundation.
- Meub, L., T. Proeger, T. Schneider, and K. Bizer (2016). The victim matters – experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters* 23(16), 1162–1167.
- Moll, J., R. De Oliveira-Souza, and R. Zahn (2008). The neural basis of moral cognition: sentiments, concepts, and values. *Annals of the New York Academy of Sciences* 1124, 161–80.
- Moll, J., R. Zahn, R. D. Oliveira-souza, and F. Krueger (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience* 6, 799–809.
- Pittarello, A., M. Leib, T. Gordon-Hecker, and S. Shalvi (2015). Justifications Shape Ethical Blind Spots. *Psychological Science*, 1–15.
- Rabin, M. (1993, December). Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83(5), 1281–1302.

- Raine, A. (2008, oct). From Genes to Brain to Antisocial Behavior. *Current Directions in Psychological Science* 17(5), 323–328.
- Raine, A. and Y. Yang (2006). Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience* 1(3), 203–13.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.
- Rigdon, M. L. and A. D’Esterre (2015). The Effects of Competition on the Nature of Cheating Behavior. *Southern Economic Journal* 81.
- Rigdon, M. L. and A. D’Esterre (2017). Sabotaging Another: Priming Competition Increases Cheating Behavior in Tournaments. *Southern Economic Journal* 84(2), 456–473.
- Schwieren, C. and D. Weichselbaumer (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology* 31(3), 241–253.
- Sen, A. (1987). *On Ethics and Economics*. Oxford and New York: Basil Blackwell.
- Shalvi, S., M. J. J. Handgraaf, and C. K. W. De Dreu (2011). Ethical Manoeuvring: Why People Avoid Both Major and Minor Lies. *British Journal of Management* 22, 16–27.
- Smith, A. (1759). *The Theory of Moral Sentiments*. McMaster University Archive for the History of Economic Thought.
- Smith, V. L. and B. J. Wilson (2019). *Humanomics*, pp. i–i. Cambridge Studies in Economics, Choice, and Society. Cambridge University Press.
- Smith 1723-1790, A. (2000). *The wealth of nations*. New York : Modern Library, 2000.
- Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal* 119, 47–60.
- Tajfel, H. and J. Turner (1979). *An Integrative Theory of Intergroup Conflict*. Monterey, CA: Brooks-Cole.
- Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American Economic Review* 106(7), 1577–1600.
- Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics* 106(4), 1039–1061.

Chapter 1

Unethical Behavior and Group Identity in Contests¹

1.1 Introduction

Based on 2,410 cases of occupational fraud collected between January 2014 and October 2015 the Association of Certified Fraud Examiners (ACFE, 2016) estimates that unethical behavior in organizations leads to a loss of about US \$6.3 billion. Occupational fraud is a scourge for any organization, as it may discourage honest effort and ruin the reputation of a company if uncovered. Dishonesty in organizations may depend on the presence of “bad apples” (e.g., Gino et al., 2009), but also on the institutional environment, including weak societal norms (e.g., Gächter and Schulz, 2016), corporate cultures encouraging materialistic values (e.g., Cohn et al., 2014), and competitive schemes (e.g., Shleifer, 2004; Falk and Szech, 2013). Competitive incentives aim at motivating employees (Tullock, 1980; Lazear and Rosen, 1981) but may generate sabotage, as shown both theoretically (Lazear, 1989; Konrad, 2000) and empirically (e.g., Harbring et al., 2007; Carpenter et al., 2010; Harbring and Irlenbusch, 2011; Kilduff et al., 2016; Feltovich, 2018); cheating opportunities may lead less honest people to self-select into more competitive payment

¹This chapter is a joint work with Marie Claire Villeval. Published in the Journal of Economic Psychology.

schemes (Faravelli et al., 2015).

The prevalence of unethical behavior in competitive settings varies with the intensity of competition (e.g., Schwierien and Weichselbaumer, 2010; Harbring and Irlenbusch, 2011; Cartwright and Menezes, 2014; Conrads et al., 2014) or the outcome of a previous competition (Pittarello et al., 2013; Banerjee et al., 2018; Schurr and Ritov, 2016). Among these environmental factors, one can suspect that the strength of preexisting links between potential competitors in a group may affect the prevalence of unethical behavior. For example, an employee who competes for a promotion against one of his team members may be less likely to behave unethically than when faced with an employee from another, more distant, group. Our main research objective in this paper is to identify the influence of group identity and social distance between competitors on unethical behavior in competitive settings.

Specifically, we study whether the social identity of their opponent in a competition affects individuals' willingness to misreport an outcome that contributes to determining their payoff. Are people less willing to misreport in order to increase their chance of winning a competition when doing so harms an in-group? On the other hand, are they more willing to misreport when cheating harms an opponent from another group? We also examine whether group identity has a different impact, if any, when individuals have to report their own outcome and when they have to report their opponent's outcome to determine who is the winner.

By connecting the literature on group identity, on competitive behavior and on lying, our contribution is threefold. First, we complement the economic analysis of the role of group identity (Tajfel and Turner, 1979; Akerlof and Kranton, 2000) in social interactions by examining the influence of social identification on unethical behavior. Studies of group identity and cooperation have shown the importance of in-group favoritism, *i.e.*, people treat more generously someone who shares the same group identity than

someone who belongs to another social group (e.g., Charness et al., 2007; Chen et al., 2009; Goette et al., 2012).² Much less research has been conducted on the importance of group identity in competitive settings (e.g., Datta Gupta et al., 2013; Chen et al., 2015; Chowdhury et al., 2016; Kato et al., 2016).³ Our contribution to this literature is exploring the impact of group identity on unethical behavior.

Second, we contribute to the understanding of lying behavior by studying how group identity affects individuals' unethical behavior. The previous literature has shown that priming group identity increases the prevalence of norm violation when identity is associated with weaker morality (Cohn et al., 2015) or stronger materialistic values (Cohn et al., 2014). Studying fairness-restorative dishonesty across games, Della Valle and Ploner (2017) show that being treated unfairly by an out-group peer in a dictator game increases dishonesty in a subsequent mind game compared to a baseline treatment without social identity, whereas dishonesty is unaffected by the unfair decision of an in-group dictator. In contrast, we investigate a competitive setting where dishonest reporting is not at the cost of the experimenter but at the expense of the opponent.⁴ Some studies have found evidence of in-group favoritism in deceptive behavior (e.g., Jiang, 2015; Cadsby et al., 2016; Chakravarty and Maximiano, 2016), but results are

²In-group favoritism relaxes punishment toward out-groups (Bernhard et al., 2006; Goette et al., 2012); it positively affects coordination (Efferson et al., 2008), trust and reciprocity (Buchan and Croson, 2004; Hargreaves Heap and Zizzo, 2009), generosity in dictator games (Güth et al., 2009), cooperation in ultimatum bargaining games (McLeish and Oxoby, 2011) and in dilemma games such as public goods games (Eckel and Grossman, 2005), common-pool resource games (Ruffle and Sosis, 2006), and prisoner dilemma games (Goette et al., 2006; Charness et al., 2007; Guala et al., 2013; Li and Liu, 2017), as well as the willingness to preserve other's image at a cost (Eriksson et al., 2017). When they interact with in-groups, individuals trust more, reciprocate more, contribute more to public goods, and cooperate more. The other side of the medal is that when they interact with out-groups, they are usually less cooperative (e.g., McLeish and Oxoby, 2011; Chen et al., 2014).

³Datta Gupta et al. (2013) report that choosing the gender of a potential competitor influences competitiveness. Chen et al. (2015) find that a salient group identity stiffens the competition between different universities. Chowdhury et al. (2016) show that a more salient group identity (natural vs. minimal) escalates conflicts in a contest by motivating people to increase their effort. Kato et al. (2016) report that co-workers sharing a common group identity do not react to competitive monetary incentives whereas co-workers with different group affiliation compete more aggressively.

⁴Moreover, in our experiment the ex ante inequality of chances in the contest results from differences in ability levels and in luck, not from the selfishness of another player, as explained below; fairness-restorative dishonesty should only emerge when the subject anticipates that his opponent will cheat. Finally, in our main treatments lying can be measured at the individual level and we use a mind game as a robustness test.

contrasted.⁵ Hruschka et al. (2014) showed that behaving dishonestly to favor in-groups instead of following an impartial rule is more likely in societies with lower institutional effectiveness. Our contribution is considering how competitive incentives affect the impact of group identity on unethical behavior in a setting where both competitors can cheat.

Finally, we contribute to the analysis of ethics in competition. As mentioned above, several studies have shown, theoretically and empirically, that some people bend their moral rules to increase their chance of winning a competition, for example through sabotage (e.g., Harbring and Irlenbusch, 2011; Harbring et al., 2007; Chowdhury and Gürtler, 2015). But we know little about whether sabotaging a competitor is morally equivalent to cheating on one's own performance, when both actions similarly increase one's chance of winning. Rigdon and D'Esterre (2015) found that individuals are more likely to cheat by overreporting their own performance than by underreporting their opponent's performance under both competitive and noncompetitive incentives. Lying about an opponent's outcome increases the moral cost of lying compared to lying about one's own outcome. We try to replicate this finding and test whether inducing group identity amplifies this effect when the opponent is an in-group and cancels it out when the opponent is an out-group.⁶

⁵Jiang (2015) and Cadsby et al. (2016) report that in a noncompetitive environment, subjects are more willing to lie when lying benefits a member of their group compared to when it benefits a member of another group. Similarly, cheating is lower when it harms another student compared to when cheating is at the expense of the experimenter (Meub et al., 2016). Using a sender-receiver game with a natural group identity based on friendship, Chakravarty and Maximiano (2016) showed that deception is less likely when the receiver is a sender's friend. In contrast, using the same game Feldhaus and Mans (2014) found that the sender's decision to deceive the receiver is independent of whether they share the same group identity or not.

⁶Note that people motivated by social status may behave competitively even without competitive monetary incentives. Charness et al. (2014) showed that individuals cheat more by artificially increasing their own performance than by sabotaging group members to improve their performance rank under a flat-payment scheme. In contrast, we introduce competitive incentives and no monetary cost for lying.

To sum up, the novelty of our approach is studying the impact of group identity on dishonesty in competitive settings.⁷ In a repeated two-player Tullock contest, the relative final scores of the players determine their probability of winning the contest. The final score of a player is the sum of his performance in a real-effort task and a random number that could represent his idiosyncratic environmental working conditions. After performing the task, subjects are informed of their performance and their number, and of the performance and the number of their opponent. We vary within-subjects which information players have to report. In one condition, they have to report their own random number (“Self-condition”), while in the other they have to report their opponent’s number (“Opp-condition”).⁸ When reporting, subjects were able to inflate their own number or to deflate their opponent’s number,⁹ depending on the condition, at no risk of detection by the opponent due to the use of a probabilistic contest. We manipulated between-subjects the existence of group identity in two main treatments. In the No-Identity treatment, no identity is induced. In the Identity treatment, a minimal identity is induced following the procedure developed in Chen et al. (2009). According to the case subjects are matched with a player who shares the same identity or with a subject with the other identity.

Our results show that 53.7% of the subjects misreport at least once and when they lie, they do it to the full extent. Contrary to our conjecture, in both treatments they

⁷In a lab-in-the-field experiment in rural India, (Banerjee et al., 2018) showed that high-caste individuals are more likely to lie to increase their payoffs while harming a member from the backward caste when they learn that they have lost a tournament under a quota policy protecting backward caste members. However, in this study the opportunity to lie follows a distinct tournament game. In contrast, we introduce an opportunity to lie in a contest.

⁸Reporting on behalf of another employee is typical of the 360-degree feedback in which each employee rates other employees (team members, subordinates, and supervisors). Over one-third of U.S. companies use this type of feedback (see Bracken et al., 2001). An advantage of multi-rater feedback is that managers can collect information that they would not be able to obtain by looking directly at quantitative outputs (see Lepsinger and Lucia, 2009, e.g.). This can include the idiosyncratic shocks that we intend to capture with our random numbers. The reliability of such feedback has been questioned because of possible manipulations and a lack of transparency (on the U.S. Department of Foreign Affairs, see Bent (2018); on Amazon, see Kantor and Streitfeld (2015)). Note, however, that these schemes are used for development and indirectly for hiring and promotion.

⁹Lying in the Opp-condition is similar to sabotage. Compared to other papers, however, sabotage is not modelled as an effort with a monetary cost; here, misreporting involves only a moral cost.

misreport their opponent's outcome as often and by as much as when they misreport their own outcome. We find no effect of group identity on misreporting: there is no effect from social distance between the individual and his opponent on the frequency, the intensity, and the nature of lies. Although they express in-group favoritism in distributive choices and although they believe that group identity affects others' misreporting behavior, people do not lie less against an in-group than an out-group opponent, or in the absence of group identity. The absence of difference could be driven by the fact that the experimenter knows the truth: those who misreport in this environment may be less sensitive to social image and less sensitive to group identity. Therefore, we conducted an additional treatment in which scrutiny is eliminated by using a mind game. In this new treatment subjects report higher outcomes in the Self-condition than in the Opp-condition in absolute value (as in Rigdon and D'Esterre, 2015) but group identity remains insignificant. These results suggest that in competitive settings, unethical behavior is mainly driven by an unconditional desire to win.

The next section introduces our experimental design and the procedures. Section 3 outlines our behavioral conjectures. Sections 4 and 5 report our results and those from robustness tests, respectively. Finally, section 6 discusses these results and concludes.

1.2 Experimental Design and Procedures

We first introduce our experimental design for the main treatments and then, the procedures.

1.2.1 Experimental Design

Our experiment is based on a two-player contest played for 16 periods with a reshuffling of pairs after each period. The probability of each pair member winning the contest depends on a comparison of the two pair members' scores. A score is defined as the sum of the performance in a real-effort task and of an idiosyncratic random number

that subjects have to report, as explained below. We implemented a 2x2 mixed factorial design. One dimension varies within-subjects whether the individual has to report his own random number (“Self-condition”) or his opponent’s number (“Opp-condition”). The Self- and Opp-conditions were implemented over eight periods each, and were presented in a totally randomized order across periods. The other dimension varies between-subjects the presence of group identity in two treatments, namely the Identity treatment and the No-Identity treatment. We describe each treatment first and then, the additional measures collected in the experiment.

1.2.1.1 No-Identity Treatment

The competition game consists of a two-player Tullock-lottery contest in which two subjects are paired and compete against each other to win a prize. The outcome of the contest depends on the relative final scores. In each period t , the probability to win of a given subject i matched with a subject j is given by the following formula:

$$p_{it}(\text{win}) = \frac{\text{Finalscore}_{it}}{\text{Finalscore}_{it} + \text{Finalscore}_{jt}} \quad (1.1)$$

The higher the subject’s final score relative to that of his opponent, the higher is his probability of winning. The winner receives 100 ECUs (Experimental Currency Units) and the loser 50 ECUs. The final score of a subject is the sum of his performance (the “subscore”) in a real-effort task and an idiosyncratic random number. The task is an encryption task, as in Erkal et al. (2011) and Charness et al. (2014). It consists of decoding letters into numbers within 30 seconds with the help of a correspondence table displayed on the computer screen. In our design, each problem is composed of one letter only. Subjects have to enter the corresponding number and to validate their answer. They receive immediate feedback on whether their answer was correct or not. If the answer was not correct, they have to enter a new number. If the answer was correct, another letter is displayed on the screen. The subscore for the period is the total number of

correct answers provided within 30 seconds.

Once the 30 seconds have elapsed, subjects are informed both of their own subscore and of their opponent's subscore. Then, the program draws two random numbers independently from a uniform distribution in the interval $[-5, 5]$, one for the subject and one for his opponent, with rebate between the two draws. These numbers can represent the environmental conditions in which the task has to be performed (for example, the environmental conditions that influence the success of a project, the quality of the managers in charge of different units, ...). Subjects are informed of these two numbers. The rest of the period depends on the condition and subjects are informed on the condition prevailing in the period after they have performed the task.

In the Self-condition, subjects have to report their own random number in the computer. In the Opp-condition, they have to report instead their opponent's number. Subjects who are willing to increase their chance of winning the prize have the opportunity to misreport their own or their opponent's random number, depending on the condition. In the instructions, nothing was mentioned about this possibility. The reported number is added to the subscore to determine the final score.

Finally, the outcome of the contest is computed according to equation 1.1 and subjects are informed whether or not they won the prize. They are only reminded of their final score in the Self-condition or their opponent's final score in the Opp-condition. Indeed, to avoid the contamination of lying and blind revenge, players are not informed whether their own number was reported truthfully or not in the Opp-condition, or whether the opponent misreported his own number or not in the Self-condition. In no condition do they learn the distance between their own final score and their opponent's final score. To determine the subjects' earnings in this part, two periods (one in each condition) were randomly selected at the end of the session, and the payoffs in these periods were added up.

1.2.1.2 Identity Treatment

The Identity treatment is similar to the No-Identity treatment, except that group identity is induced prior to the contest game. The identity induction part is composed of three stages, as described by Chen et al. (2009). The first two stages aim at inducing identity, while the third one serves to assess the validity of identity manipulation.

In the first stage, subjects have to review five pairs of paintings by Klee and Kandinsky. Without receiving any clue about the painters, they report their preferred painting in each pair. According to their preferences toward one or another painter, they are split into two groups, the “Klee group” and the “Kandinsky group”. Unlike Chen et al. (2009), we form groups of equal size to get a balanced matching in the contests. Subjects with the weakest preference toward one artist or another are switched into the other group, if needed. The whole procedure is made common knowledge. Subjects are privately informed of their group affiliation.

In the second stage, subjects have to review two additional paintings and they are asked to identify which artist painted each of them. Before entering their answers individually, they have the opportunity to participate in a chat discussion over eight minutes, exclusively with their group members. The aim is to reinforce the feeling of belonging to a group, since the chat allows group members to help each other. Each correct answer pays 50 ECUs.

The third stage consists of five periods of a third-party allocation task. Subjects have to share a number of ECUs between two anonymous participants. They are not allowed to allocate ECUs to themselves. The objective is to identify in-group favoritism in the allocation decisions. From the first period up to the fifth, the number of ECUs to be allocated increases from 50 to 250 by increments of 50 ECUs. In each period, subjects are presented three scenarios. In the first scenario, the two participants share the same group identity as the decision-maker. In the second scenario, the two participants belong

to the other group. In the third scenario, one participant belongs to the same group and the other one belongs to the other group.¹⁰

After completion of the third stage, subjects play the contests. The rules are the same as in the No-Identity treatment, except that subjects are informed of the group identity of their opponent in each period after performing the task. In half of the 16 periods, subjects face an opponent from their group (Same Group, or “SG pairs” hereafter). In the other half, they face an opponent from the other group (Other Group, or “OG pairs” hereafter). Thus, the Identity treatment includes four sets of four periods: SG pairs in the Self-condition, SG pairs in the Opp-condition, OG pairs in the Self-condition and OG pairs in the Opp-condition. The order between these four conditions was totally random across periods.

1.2.1.3 Additional Measures

Belief Elicitation

Some subjects may misreport because of their preference or because they believe that their opponent will not report truthfully (e.g., López-Pérez and Spiegelman, 2013). At the end of the contest game, we elicited the subjects’ beliefs about the frequency of misreports in previous sessions. In the No-Identity treatment, subjects are presented two scenarios based on the same contest game. In the first scenario, two participants, A and B, have to report their own random number, with participant A having a higher subscore than participant B. Subjects have to indicate how many A participants among 10 and how many B participants among 10 randomly selected in previous sessions they think misreported their random number. In the second scenario, participants A and B have to report their opponent’s number. Similarly, subjects have to indicate how many A participants among 10 and how many participants B among 10 they think misreported

¹⁰The procedure for payment followed Chen et al. (2009). We generated a random sequence of ID numbers. Subjects allocated ECUs between the two participants whose ID directly followed their own ID in the sequence. Accordingly, a subject’s payoff was the sum of the ECUs allocated to him by the two people whose ID preceded him in the sequence.

the random number of their opponent. In the Identity treatment, subjects have to guess how many A and B participants out of 10 drawn from past sessions misreported in four scenarios: in two scenarios A and B share the same group identity, and in the two other scenarios they belong to different groups. One of these guesses was randomly selected for payment. A correct guess paid 100 ECUs.¹¹

Online Questionnaire on Social Preferences

Since the outcome of the contest leads to very unequal payoffs, the players' social preferences may have affected their behavior in our experiment. Therefore, about a week prior to the session in the laboratory, subjects had to fill out an incentivized online questionnaire. The main part of this questionnaire consists of the Social Value Orientation (SVO) test of Murphy et al. (2011). In this test subjects face six decisions successively. Each decision consists of choosing an allocation of ECUs between himself and another participant among nine options. Figure C.1, in the Appendix C, shows the first decision in the test. For this decision, the first option offers an equal split (85 ECUs for oneself and for the other), and the degree of inequality increases between the second and the ninth option (up to 85 ECUs for oneself and 15 ECUs for the other). The other five decisions have the same structure. The subject's six decisions allow us to compute an index, called the SVO angle, for each subject.¹² The lower this index is, the less pro-social the subject is.

At the beginning of the questionnaire, subjects are informed that one of their six decisions will be randomly selected and that they will be matched twice with another subject participating in the same laboratory session. In the first matching, the subject is paid the amount he allocated to himself and the matched partner is paid the amount the

¹¹We acknowledge that some liars may exaggerate their estimate of the others' willingness to misreport to self-excuse their own misbehavior. We have tried to minimize this bias by incentivizing the belief elicitation but we cannot exclude it. Alternatively, eliciting beliefs prior to the competition game would have attracted the attention of the subjects on the lying opportunity and probably introduced an experimenter demand effect.

¹²We compute the mean allocation that the subject chose for himself and the mean allocation he chose for the other. Then, we subtract 50 from each of these means. Finally, the inverse tangent of the ratio between these means is computed, resulting in the SVO index.

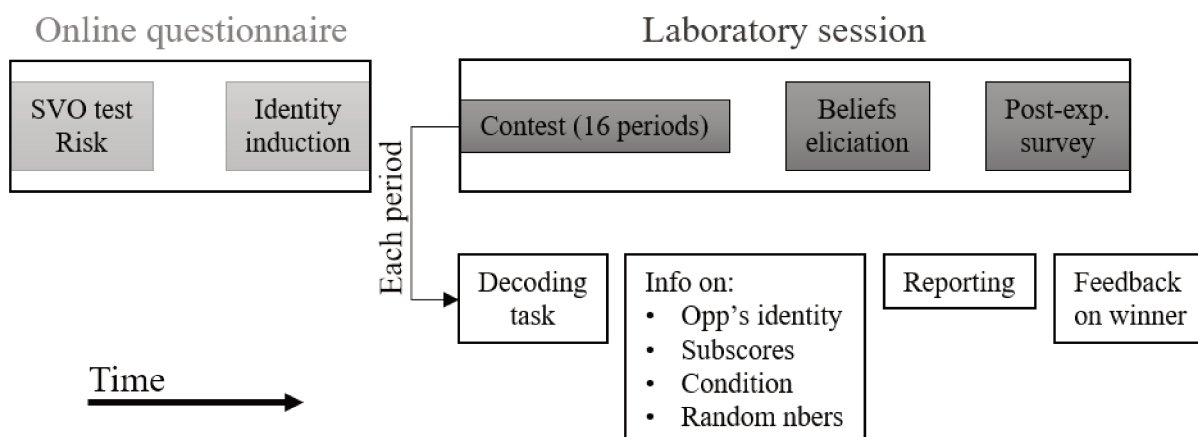


Figure 1.1 – Timeline of the experiment

subject allocated to the other. In the second matching, the subject is paid according to the choice made by the matched partner for the other and this matched partner receives the amount he allocated to himself. Feedback on the earnings in this task is given only at the end of the lab session.

In addition to the SVO test, subjects have to answer ten questions taken from the risk-elicitation questionnaire by Weber et al. (2002) with no other objective than diverting the subjects' attention from the elicitation of social preferences in the experiment. Figure 1.1 summarizes the timeline of the experiment.

1.2.2 Procedures

The experiment was conducted at GATE-Lab, Lyon (France). 192 subjects were recruited using HROOT (Bock et al., 2014). 77.60% of the subjects were students from local engineering, business, and medical schools. We conducted four sessions of the No-Identity treatment with 64 subjects and eight sessions of the Identity treatment with 128 subjects. Table B.1 in the Appendix B summarizes the participants' characteristics, with no significant difference between treatments. The online questionnaire was programmed with LimeSurvey (Limesurvey GmbH) and the laboratory experiment

with Z-tree (Fischbacher, 2007). In the lab sessions, instructions were distributed to the subjects and read aloud (see Appendix A). In the Identity treatment we first distributed the instructions describing the first two stages of the group identity induction procedure. The instructions for the third stage were distributed only after these two stages were completed. The instructions used neutral language. Those for the belief elicitation were displayed directly on the screens at the end of the contest game.

Sessions lasted on average 70 minutes for the No-Identity treatment and 90 minutes for the Identity treatment. Someone who was not aware of the content of the experiment paid participants in private in a separate room. This was made clear in the instructions. ECUs were converted into Euros, at the rate of 100 ECUs to €2.50. Mean earnings were €19.29 (SD=1.78), including a €5 show-up fee and an average €1.57 (SD=0.52) for the online task.

1.3 Behavioral Conjectures

Since there is no monetary cost for lying, individuals who are willing to maximize their earnings and have no moral cost for lying should misreport to the full extent to maximize their probability of winning the contest, regardless of their performance at the task and of their random number. In contrast, individuals who suffer a finite moral cost from lying should be less likely to misreport or should misreport only partially, and those who endure an infinite cost of lying should never misreport. This conjecture relies on the recent literature on lying (e.g., Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013; Gibson et al., 2013; Kajackaite and Gneezy, 2017; Abeler et al., 2019) showing that dishonesty is usually partial.

In contexts in which they have to report performance or outcomes, individuals who are willing to maximize their payoffs should misreport to the same extent (albeit not in the same direction) their outcome and the outcome of their opponent. However, Rigdon

and D'Esterre (2015) have found evidence that individuals misreport more their own performance than the performance of another person, in both competitive and noncompetitive settings. Similarly, we conjecture that in our experiment subjects are more likely to overreport their own number than underreport their opponent's number, because harming directly another person may be perceived as aggressive. We summarize our first conjecture as follows:

Conjecture 1 (Misreporting): (a) Subjects misreport both their own number and their opponent's number; and (b) they are more likely to overreport their number than underreport the opponent's number.

Previous literature has shown that people who identify with a social group tend to favor their in-group members in terms of cooperation, trust and reciprocity compared to out-group (Goette et al., 2006; Charness et al., 2007; Chen et al., 2009). In-group favoritism can come with out-group aggressiveness even without any strategic consideration (parochial altruism) (Abbink et al., 2012; Goette et al., 2012; Kolmar and Wagener, 2012). Accordingly, we anticipate that the moral cost of lying is higher when the victim of the lie is an in-group, and lower when it is an out-group. Thus, we expect subjects to be less likely to misreport when they are paired with an in-group (or more with an out-group) in the Identity treatment, especially in the Opp-condition.

Finally, in-group favoritism can originate from two main channels: preferences and beliefs about how others will behave. Previous studies on group identity and distributive choices (e.g., Yamagishi and Mifune, 2008; Güth et al., 2009; Ockenfels and Werner, 2014), or cooperation (Guala et al., 2013) favor the role of beliefs. Thus, we conjecture that subjects believe that in-group opponents are less likely to lie than out-group opponents. As a result and regardless of preferences, they may deceive less when matched with an in-group and more when matched with an out-group. Thus, we conjecture that pairs composed of in-groups are less likely to misreport (or more if composed of out-groups). Our second conjecture is the following:

Conjecture 2 (Social identity and misreporting): (a) When facing an in-group opponent, subjects misreport less than when facing an out-group opponent, especially in the Opp-condition; and (b) differentiated beliefs about the opponent's behavior according to his group identity are correlated with reporting behavior.

1.4 Results

In this section we focus exclusively on reporting behavior. Indeed, a separate analysis of the subject's performance reveals no significant difference in the subscores between conditions.¹³ We start by presenting general results on reporting behavior under the different conditions. Next, we analyze the impact of group identity on misreporting. Finally, we examine the relationship between the subjects' beliefs and their reporting behavior.

1.4.1 Misreporting behavior across conditions

We introduce our first result:

Result 1: (a) More than half of the subjects misreport self-interestedly at some point, and those who misreported did it to the full extent most of the time. *(b)* They misreported to the same extent and with the same frequency when they reported for themselves (Self-condition) or for their opponent (Opp-condition).

Result 1 gives some support to Conjecture 1a but rejects Conjecture 1b.

Support for Result 1. Figures 1.2a and 1.2b display the distribution of numbers reported

¹³For non-parametric statistics, we consider the averages at the session level. we consider only the first period. Two-sided Wilcoxon signed-rank tests indicate no significant difference in the subscores across conditions (No-Identity treatment: Self- vs. Opp-condition: $p=0.400$; Identity treatment, same group identity: Self- vs. Opp-condition: $p=0.273$; for the Identity treatment with mixed group identity Self- vs. Opp-condition: $p=0.780$). Considering instead the data from the first period leads to the same conclusion. A regression analysis including all periods except the first one is reported in Table B.2 in the Appendix B.

by the subjects in the Self-condition and in the Opp-condition, respectively, by treatment.

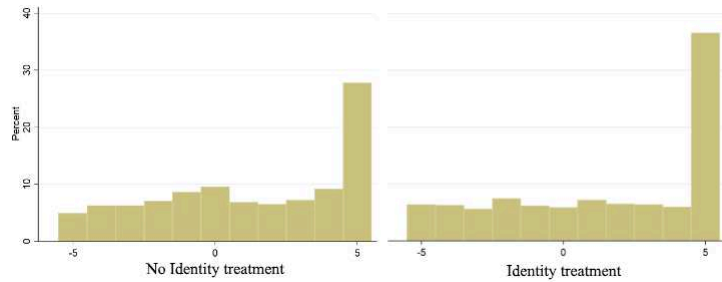


Figure 1.2a – Distribution of numbers reported in the Self-condition, by treatment

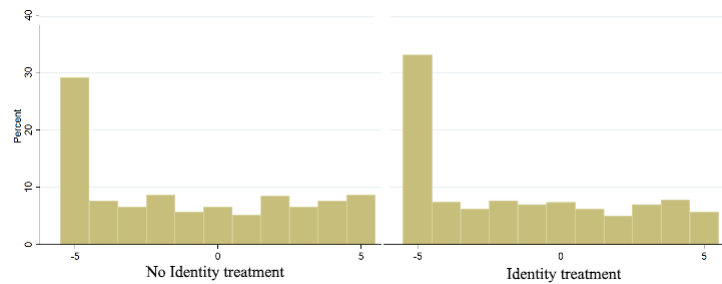


Figure 1.2b – Distribution of numbers reported in the Opp-condition, by treatment

In each condition, each number is expected to appear uniformly 9.09% of the time if reports are truthful. Figures 1.2a and 1.2b suggest that the distribution of reported numbers is different from a uniform distribution in both conditions and treatments.¹⁴ The frequency of 5s reported in the Self-condition was 27.73% in the No-Identity treatment and 36.52% in the Identity treatment; conversely, the frequency of -5s reported in the Opp-condition was 29.10% and 33.01%, respectively.¹⁵ This provides evidence of

¹⁴For the following non-parametric tests, we consider only the first of the 16 periods. Each subject is taken as one independent observation. Each subject was taken as one independent observation. Tests were two-sided. χ^2 goodness-of-fit tests rejected the uniform distribution in the Identity treatment ($p < 0.001$), but not in the No-Identity treatment ($p = 0.172$). To increase the number of observations for these tests we multiply the reported numbers by -1 in the Opp-condition allowing us to merge the data from the two conditions. A similar strategy is used for the binomial tests reported in the footnote 14.

¹⁵Binomial tests on the first period indicate that in the Identity treatment these numbers are significantly different from 9.09% ($p < 0.001$). Conversely, in the No-Identity treatment these numbers do not reach standard levels of significance ($p = 0.079$).

self-interested misreporting. Our protocol allows us to identify misreports precisely. A (small) majority of subjects told selfish lies. 103 subjects out of 192 (53.65%) misreported self-interestedly at least once during the 16 periods. Among these subjects, 77 misreported at least once in each of the two conditions (40.10%), 12 subjects misreported only in the Self-condition (6.25%), and 14 subjects only in the Opp-condition (7.29%).

Table 2.3.1 displays the absolute and relative frequencies of misreports, their average absolute intensity, and the absolute and relative frequencies of full misreports, by treatment and condition.¹⁶ Absolute intensity is defined as the difference between the reported random number and the actual number. Full misreport is equal to 1 if the subject lied and reported the highest possible number for himself or the lowest possible number for his opponent, and 0 otherwise. This Table includes only selfish misreports (those increasing the subject's probability of winning). Table B.3 in Appendix B displays the same information on the cases where a subject underreported his own number or overreported his opponent's number.

Table 2.3.1 shows that in the No-Identity treatment 27% of the subjects reported a higher number for themselves (131/475) and the same percentage reported a lower number for their opponent (120/438) when they had an opportunity to lie. In the Identity treatment, 32% (296/912) and 31% (285/908) of the subjects misreported in the Self- and in the Opp-condition, respectively. Subjects who lied selfishly misreported to the full-extent in the vast majority of cases (from 84% to 91% of the observations, depending on the treatment and condition). This is consistent with Kajackaite and Gneezy (2017): since the experimenter is able to observe misreporting (*ex post facto*), those with a higher moral cost of lying may misreport to a lesser extent than if there were no scrutiny at all, but those who are willing to misreport have no reluctance to lie in full.

¹⁶In Table 2.3.1 and the following tables and tests about selfish lies, we exclude the cases in which subjects could not misreport selfishly (their number was equal to 5 in the Self-condition and their opponent's number was equal to -5 in the Opp-condition).

Table 1.1 – Performance, frequency and intensity of selfish misreporting, by treatment and condition

| Treatment | Condition | | Average subscore | Relative frequency of misreports | Relative frequency of full misreports | Average absolute intensity of misreports | N |
|-------------|-----------|--------------------|------------------|----------------------------------|---------------------------------------|--|------|
| No-Identity | Self | All | 6.93 (0.07) | 27.52% | 83.97% | 5.32 (0.25) | 512 |
| | | Opp | 6.82 (0.08) | 27.40% | 85.00% | 4.92 (0.26) | 512 |
| | | All | 6.64 (0.06) | 32.46% | 90.88% | 5.29 (0.17) | 1024 |
| Identity | Self | In-Group opponent | 6.40 (0.09) | 30.57% | 90.71% | 5.21 (0.24) | 512 |
| | | Out-Group opponent | 6.89 (0.07) | 34.36% | 91.02% | 5.36 (0.24) | 512 |
| | | All | 6.77 (0.05) | 31.39% | 87.02% | 5.45 (0.17) | 1024 |
| | Opp | In-Group opponent | 6.73 (0.08) | 31.42% | 86.62% | 5.86 (0.23) | 512 |
| | | Out-Group opponent | 6.82 (0.07) | 31.36% | 87.41% | 5.05 (0.25) | 512 |

Notes: Standard deviations are in parentheses. The average subscore is the average performance in each period of the decoding task. The relative frequency of misreports excludes the observations in which the subject's own random number is 5 in the Self-condition and those in which the opponent's random number is -5 in the Opp-condition. Full misreports correspond to the cases in which a subject misreports the highest possible number for himself (+5) or the lowest possible number for his opponent(-5). N is the total number of observations.

Table B.3 in Appendix B shows that non-self-interested misreports were very rare. In 11 cases (emanating from 10 different subjects and representing 0.79% of the relevant observations) subjects underreported their own number; in 53 cases (emanating from 33 different subjects and representing 3.81% of the cases), they overreported their opponent's number. These cases are either errors or altruistic white lies motivated by the willingness to increase the opponent's chance to succeed. We favor the first interpretation because most subjects made these types of reports only once or twice.¹⁷ The rest of our analysis is based only on self-interested misreporting.

To compare selfish misreporting behavior across conditions, we first report two-sided Wilcoxon signed-rank tests based on averages at the session level. These tests show no significant difference between the decisions to misreport in the Self- and the Opp-conditions in the No-Identity treatment ($N=8$, $p = 0.715$) and in the Identity treatment ($N=16$, $p = 0.779$). The differences are not significant either if we consider the absolute intensity of misreporting ($p = 0.715$ and $p = 0.780$, respectively). We report next

¹⁷One subject overreported the opponent's number in seven out of eight cases and underreported his own number in two out of eight cases. But he also overreported his own number in four cases. His comments left in the post-experimental questionnaire suggest that the subject was confused.

a regression analysis that investigates the determinants of misreporting, considering first the decision to misreport, and second the relative intensity of misreporting. The results are shown in Tables 1.2 and 1.3. Table 1.2 displays the marginal effects of seven random-effects logit models in which the dependent variable is the decision to misreport, with robust standard errors. These models vary the conditions and treatments included. Models (1) to (3) pool the decisions made in the two conditions. Models (4) and (5) consider the decisions made in the Self-condition and models (6) and (7) those made in the Opp-condition. Models (2), (4), and (6) include only the No-Identity treatment, and models (3), (5), and (7) only the Identity treatment. In model (1), the independent variables include dummy variables for the Identity treatment and for the Opp-condition, taking the Self-condition and the No-Identity treatment as the reference categories. Since the decision to misreport may be affected by the relative performance at the task, they also include the subject's subscore, the difference between subscores when the subject outperforms his opponent ($\max\{0, \text{subscore}_{i,t} - \text{subscore}_{j,t}\}$), the difference between subscores when the subject is outperformed by his opponent ($\max\{0, \text{subscore}_{j,t} - \text{subscore}_{i,t}\}$), and a dummy in case of a tie in performance. Since the decision to misreport may also be affected by relative luck, the independent variables include the subject's random number, the difference between the subject's and his opponent's numbers when the subject is luckier than his opponent ($\max\{0, \text{number}_{i,t} - \text{number}_{j,t}\}$), the difference between numbers when the subject is less lucky than his opponent ($\max\{0, \text{number}_{j,t} - \text{number}_{i,t}\}$), and a dummy in case of a tie. To control for a possible non-linear evolution of behavior over time, we added a time trend and its squared value. We included the subject's gender and the SVO angle (a higher angle indicates that the subject is more pro-social in the SVO test). The other characteristics include the subject's age and his number of past participations in a laboratory experiment. Finally, session fixed effects control for unobserved heterogeneity across sessions. The other models include the same independent variables, except the condition and treatment variables. In addition, models (3), (5), and (7) include a dummy variable ("In-group

opponent") equal to 1 if the pair members share the same group identity and 0 otherwise.

Table 1.3 reports the marginal effects from seven random-effects Tobit regressions on the relative intensity of selfish misreporting. The dependent variable is the ratio between the actual and the maximum possible intensity of misreporting.¹⁸ We use Tobit models since data are censored on the right when subjects misreport to the full extent (*i.e.*, subjects could not cheat more).¹⁹ These models have the same other characteristics as those reported in Table 1.2.

Contrary to Conjecture 1b, models (1) to (3) in Table 1.2 show that the probability of misreporting does not differ significantly between the Self- and the Opp-conditions in any treatment.²⁰ Table 1.3 indicates that the relative intensity of misreporting is higher in the Self than in the Opp-condition. However, the effect is small and only significant in the Identity treatment.

The two tables also indicate that the determinants of behavior differ across conditions, suggesting that subjects did not focus exactly on the same information. In the Self-condition, the likelihood of misreporting (only in the No-Identity treatment) and the relative intensity of misreporting react mainly to bad luck in terms of one's random number (models (6)). In contrast, in the Opp-condition the value of these coefficients increases when the performance of the subject increases. Additionally, the relative intensity of misreporting is significantly higher the more the opponent is ahead in terms of random number in both treatments (models (4) and (5)). The probability to misreport

¹⁸For example, if a participant gets a random number of 2 and reports a 4 the ratio is equal to 0.66; if he reports a 5, the ratio is equal to 1. The dependent variable is equal to 1 if a participant misreported a 5 in the Self-condition or a -5 in the Opp-condition. It is equal to 0 if the subject did not misreport. Note that non-parametric tests based on the average relative intensity of misreporting at the session level give the same results than the tests based on the absolute intensity of misreporting.

¹⁹We do not include in these regressions the cases of non-selfish misreporting as they are probably mostly mistakes. Note that if we include them, the results of Table 3 remain qualitatively similar.

²⁰To assess the significance of this null result we calculated the achieved power of our design for three levels of size effect. We use a simulation-based approach on model (1) in Table 1.2 using STATA with a p-value sets at 5%. Our achieved power is 30%, 82% and 99% for a low, medium and large effect size (Cohen's *d* is equal to 0.3, 0.6 and 0.9, respectively).

Table 1.2 – Determinants of the decision to make a selfish misreport

| <i>Dependent variable:</i> <i>Selfish misreport</i> | <i>All conditions</i> | | | <i>Opp-condition</i> | | <i>Self-condition</i> | |
|--|-----------------------|----------------------|---------------------|----------------------|---------------------|-----------------------|--------------------|
| | All (1) | No-Identity (2) | Identity (3) | No-Identity (4) | Identity (5) | No-Identity (6) | Identity (7) |
| Treatments | | | | | | | |
| Identity treatment | -0.042 (0.128) | - | - | - | - | - | - |
| Opp-condition | -0.017 (0.014) | 0.007 (0.027) | -0.023 (0.016) | - | - | - | - |
| In-group opponent | - | - | -0.023 (0.022) | - | -0.027 (0.017) | - | -0.039 (0.087) |
| Subscore _{<i>i,t</i>} | 0.017** (0.008) | 0.026* (0.015) | 0.004 (0.011) | 0.043** (0.020) | 0.011 (0.008) | 0.030 (0.021) | 0.012 (0.048) |
| Tie: subscore _{<i>i,t</i>} = subscore _{<i>j,t</i>} | 0.026 (0.021) | -0.028 (0.037) | 0.058* (0.032) | -0.065 (0.049) | 0.056** (0.027) | 0.008 (0.055) | 0.047 (0.120) |
| Max{0, subscore _{<i>i,t</i>} - subscore _{<i>j,t</i>} } | -0.005 (0.008) | -0.002 (0.015) | -0.007 (0.011) | -0.016 (0.018) | -0.005 (0.008) | <-0.001 (0.020) | -0.003 (0.017) |
| Max{0, subscore _{<i>j,t</i>} - subscore _{<i>i,t</i>} } | 0.024*** (0.008) | 0.040** (0.016) | 0.010 (0.013) | 0.036** (0.018) | 0.018* (0.010) | 0.045 (0.028) | 0.015 (0.046) |
| Random number _{<i>i,t</i>} | <0.001 (0.003) | -0.005 (0.006) | 0.011* (0.006) | 0.017* (0.010) | 0.004 (0.003) | -0.040*** (0.010) | -0.004 (0.007) |
| Tie: number _{<i>i,t</i>} = number _{<i>j,t</i>} | -0.004 (0.025) | 0.002 (0.056) | 0.024 (0.032) | -0.007 (0.088) | -0.004 (0.023) | 0.081 (0.063) | -0.012 (0.086) |
| Max{0, number _{<i>i,t</i>} - number _{<i>j,t</i>} } | -0.011** (0.005) | -0.004 (0.010) | -0.019** (0.008) | -0.006 (0.013) | -0.013** (0.005) | 0.011 (0.014) | -0.004 (0.020) |
| Max{0, number _{<i>j,t</i>} - number _{<i>i,t</i>} } | 0.010*** (0.003) | 0.011* (0.006) | 0.017** (0.008) | 0.025*** (0.008) | 0.009** (0.004) | -0.001 (0.009) | 0.002 (0.005) |
| Period | 0.025*** (0.008) | 0.039** (0.017) | 0.023** (0.011) | 0.042* (0.024) | 0.019** (0.009) | 0.020 (0.020) | 0.014 (0.011) |
| Period squared | -0.001** (0.000) | -0.002** (0.001) | -0.001 (0.001) | -0.002 (0.001) | -0.001 (0.001) | -0.001 (0.001) | <-0.001 (0.001) |
| SVO angle | -0.006*** (0.002) | -0.008*** (0.003) | -0.005** (0.003) | -0.007** (0.003) | -0.005** (0.002) | -0.007* (0.004) | -0.007 (0.031) |
| Male | 0.165*** (0.054) | 0.191** (0.087) | 0.151 (0.134) | 0.177** (0.087) | 0.146 (0.093) | 0.151 (0.108) | 0.161 (0.235) |
| Other individual characteristics | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Session fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| [1pt] Number of observations | 2734 | 914 | 1820 | 438 | 908 | 476 | 912 |
| Number of clusters | 192 | 64 | 128 | 64 | 128 | 64 | 128 |
| Pseudo-loglikelihood | -690.760 | -278.859 | -232.498 | -147.249 | -399.179 | -156.057 | -223.060 |
| Wald Chi2 | 123.544 | 54.300 | 91.508 | 49.371 | 135.267 | 38.682 | 64.569 |
| p>chi2 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.002 | <0.001 |

Notes: This Table reports marginal effects from random-effects logit regressions with robust standard errors clustered at the subject level in parentheses. The regressions exclude the observations in which the subject's own random number is 5 in the Self-condition and those in which the opponent's random number is -5 in the Opp-condition. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 1.3 – Determinants of the relative intensity of selfish misreporting

| <i>Dependent variable: Relative intensity of selfish mis- reporting</i> | <i>All conditions</i> | | | <i>Opp-condition</i> | | <i>Self-condition</i> | |
|---|-----------------------|--------------------|----------------------|----------------------|---------------------|-----------------------|---------------------|
| | All (1) | No-Identity (2) | Identity (3) | No-Identity (4) | Identity (5) | No-Identity (6) | Identity (7) |
| Treatments | | | | | | | |
| Identity treatment | -0.019 (0.041) | - | - | - | - | - | - |
| Opp-condition | -0.006** (0.003) | 0.003 (0.005) | -0.009*** (0.003) | - | - | - | - |
| In-group opponent | - | - | -0.004 (0.003) | - | -0.003 (0.004) | - | -0.007* (0.004) |
| Subscore _{<i>i,t</i>} | 0.005*** (0.001) | 0.007** (0.003) | 0.004** (0.002) | 0.010** (0.004) | 0.005** (0.002) | 0.009** (0.005) | 0.002 (0.002) |
| Tie: subscore _{<i>i,t</i>} = subscore _{<i>j,t</i>} | 0.008* (0.004) | -0.004 (0.008) | 0.014*** (0.005) | -0.001 (0.013) | 0.011 (0.007) | -0.005 (0.012) | 0.013** (0.007) |
| Max{0, subscore _{<i>i,t</i>} - subscore _{<i>j,t</i>} } | -0.003* (0.001) | -0.003 (0.003) | -0.003 (0.002) | -0.004 (0.004) | -0.004 (0.003) | -0.003 (0.005) | -0.002 (0.002) |
| Max{0, subscore _{<i>j,t</i>} - subscore _{<i>i,t</i>} } | 0.006*** (0.002) | 0.008** (0.003) | 0.005*** (0.002) | 0.010* (0.005) | 0.005* (0.003) | 0.008 (0.005) | 0.003 (0.003) |
| Random number _{<i>i,t</i>} | -0.000 (0.001) | -0.001 (0.001) | 0.000 (0.001) | 0.001 (0.002) | 0.001 (0.001) | -0.005** (0.002) | -0.000 (0.001) |
| Tie: number _{<i>i,t</i>} = number _{<i>j,t</i>} | 0.006 (0.005) | 0.013 (0.011) | 0.002 (0.006) | 0.018 (0.016) | 0.011 (0.009) | 0.020 (0.016) | -0.002 (0.008) |
| Max{0, number _{<i>i,t</i>} - number _{<i>j,t</i>} } | -0.000 (0.001) | 0.001 (0.002) | -0.001 (0.001) | 0.001 (0.002) | -0.002 (0.001) | 0.003 (0.002) | -0.000 (0.001) |
| Max{0, number _{<i>j,t</i>} - number _{<i>i,t</i>} } | 0.003*** (0.001) | 0.003** (0.001) | 0.002*** (0.001) | 0.006*** (0.002) | 0.004*** (0.001) | 0.002 (0.002) | 0.001 (0.001) |
| Period | 0.004*** (0.001) | 0.006** (0.003) | 0.003** (0.002) | 0.011** (0.004) | 0.006** (0.002) | 0.001 (0.004) | 0.001 (0.002) |
| Period squared | -0.000 (0.000) | -0.000* (0.000) | -0.000 (0.000) | -0.001** (0.000) | -0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| SVO angle | -0.002*** (0.001) | -0.002* (0.001) | -0.002** (0.001) | -0.002* (0.001) | -0.002* (0.001) | -0.002 (0.001) | -0.002** (0.001) |
| Male | 0.043** (0.018) | 0.039 (0.030) | 0.047** (0.022) | 0.041 (0.032) | 0.045* (0.023) | 0.036 (0.031) | 0.048** (0.022) |
| Other individual characteristics | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Session fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 2734 | 914 | 1820 | 438 | 908 | 476 | 912 |
| Number of right censored observat. | 729 | 212 | 517 | 102 | 248 | 110 | 269 |
| Number of clusters | 192 | 64 | 128 | 64 | 128 | 64 | 128 |
| Pseudo-loglikelihood | -937.044 | -371.219 | -546.565 | -204.396 | -317.826 | -222.372 | -341.095 |
| Wald Chi2 | 233.733 | 89.845 | 164.208 | 66.988 | 110.332 | 51.543 | 79.043 |
| p>chi2 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |

Notes: This Table reports marginal effects from random-effects Tobit regressions with standard errors in parentheses. The dependent variable is the ratio between actual and maximum possible misreporting. The regressions exclude the observations in which the subject's own random number is 5 in the Self-condition and those in which the opponent's random number is -5 in the Opp-condition. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

the opponent's number is also significantly higher the more the opponent is ahead either in terms of performance (in the No-Identity treatment) or in terms of luck. Moreover, subjects are significantly more likely to misreport their opponent's number over time, whereas no clear time pattern can be discerned in the Self-condition, possibly because people have originally more scruples sabotaging others. The increase in the relative intensity of misreporting over time observed in the Opp-condition is consistent with the escalation of self-serving dishonesty reported by Garrett et al. (2016).

Finally, selfishness, as measured by a lower SVO angle in the SVO test, and being a male tend to increase the likelihood and the relative intensity of misreporting, which is consistent with the previous literature (e.g., Dreber and Johannesson, 2008; Nieken and Dato, 2016; Muehlheusser et al., 2015, on gender; and Grosch and Rau, 2017, on pro-sociality and gender). However, the effect of gender varies across treatments and conditions.

1.4.2 Group Identity and Misreporting Behavior

We next present our main finding regarding the impact of group identity:

Result 2: Group identity has no significant impact on the willingness to misreport and on the relative intensity of misreporting, regardless of the condition.

This rejects Conjecture 2a.

Support for Result 2: First, to validate our group identity induction, we test whether subjects exhibit in-group favoritism in the other-other allocation task (scenario 3) in the third stage of the procedure. In this scenario, they had to decide as a third party how to share a certain amount of ECUs between person A (in their in-group) and person B (in their out-group). The amount to share increased by increments of 50 ECUs across periods. Figure 1.3 displays the average allocation of ECUs in each of the five periods.

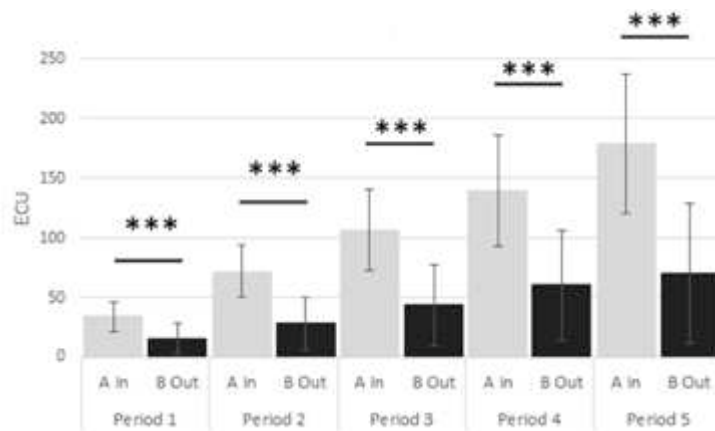


Figure 1.3 – Average allocation in ECU as a third-party in scenario 3, by period (stage 3 of the group identity induction procedure)

Notes: "In" stands for in-group and "Out" for out-group. In scenario 3, person A is an in-group member and B is an out-group member. *** indicate significance at the 1% level in two-sided pair-wise t-tests.

In this task, subjects favored in-groups and discriminated against out-groups. When pooling all the data, we find that they allocated 2.68 times more to an in-group member (106.33 ECUs) than to an out-group member (39.67 ECUs) (two-sided pairwise t-test, $p < 0.001$). As seen in Figure 3, the difference is significant in each period. Importantly, in the two other scenarios the allocations differed significantly neither when both persons A and B were from the same group as the subject (scenario 1), and when both were from the other group (scenario 2) (see Figure C.2 in the Appendix C). This shows that the induction of group identity was successful.

Although subjects expressed in-group favoritism in this task, we found no significant difference in the decisions to misreport between the No-Identity and the Identity treatments, regardless of the condition. This is supported by the regressions reported in Tables 1.2 and 1.3.²¹ In model (1), the Identity treatment variable is not significant.

²¹To assess the significance of this null result we calculated the achieve power of our design for three levels of size effect. We use a simulation-based approach on model (3) in Table 1.2 using STATA with a p-value sets at 5%. Our achieve power is 21%, 56% and 88% considering a low, a medium and a large effect size (Cohen's d equals to 0.3, 0.6 and 0.9, respectively).

Comparing models (2) and (3), (4) and (5), or (6) and (7) shows similar patterns in the extent of misreporting in the Identity and the No-Identity treatments (Table 1.3), but less significant variables in the Identity treatment compared to the No-Identity treatment regarding the probability to misreport (Table 1.2).²² Moreover, two-sided Wilcoxon signed-rank tests based on averages at the session level show that there is no significant difference in misreporting between subjects matched with an in-group and those matched with an out-group ($N=16$, $p = 0.161$). This is validated by models (3), (5), and (7) relative to the Identity treatment.

Overall, this analysis suggests that in such a competitive environment, a competitor is first seen as an opponent, and the social distance with him in terms of minimal group identity is not perceived as a relevant characteristic leading one to change one's norm of behavior.

1.4.3 Beliefs

We introduce our last result:

Result 3: Individuals believe that group identity affects others' misreporting behavior, although they are not themselves influenced by the group identity of their opponent.

This does not support Conjecture 2b.

Support for Result 3: We did not elicit the subjects' beliefs about their opponent's misreporting behavior in each period to avoid biasing their behavior. Instead, beliefs were elicited at the end of the session by asking subjects to guess how many people among 10 participants randomly selected in past sessions misreported in scenarios corresponding to the Self- and to the Opp-conditions, depending on whether the participant's subscore was above or below that of his opponent. Thus, subjects reported four beliefs in the

²²The fact that in Table 1.2 almost none of the independent variables are significant in the Identity treatment is not driven by significant effects of opposite signs between observations corresponding to in-group matching and observations corresponding to out-group matching that would cancel out. Re-estimating models (3), (5), and (7) on the subsamples of pairs sharing the same group identity and pairs with different group identities shows qualitatively the same results as in the models pooling both types of matches.

No-Identity treatment and eight beliefs in the Identity treatment, since in the latter treatment we distinguished beliefs about pairs of in-groups and beliefs about pairs of out-groups. Note that at the end of the session, subjects were not informed whether their opponents lied or not during the contests.

Table 1.4 reports OLS regressions in which the dependent variable is the individual's reported belief in the different scenarios. Robust standard errors are clustered at the individual level. Model (1) pools the data from both treatments and includes a control for the Identity treatment, while models (2) to (4) consider only the No-Identity treatment and models (5) to (7) only the Identity treatment. To correlate beliefs with behavior in the contests, models (3) and (6) are restricted to the sub-sample of subjects who never misreported and models (4) and (7) to the sub-sample of subjects who misreported at least once. In all models, the independent variables include a dummy equal to 1 if the belief is about the Self-condition scenario ("Reporting for the Self scenario"), and 0 if it is about the Opp-condition scenario. "Higher subscore player" is equal to 1 if the belief is about the player with a higher subscore, and 0 otherwise. Finally, we control for the same individual characteristics as in the previous tables. In the models relative to the Identity treatment, the independent variables include also an "In-group pair" dummy equal to 1 if the belief is about two participants who shared the same group identity, and 0 otherwise.

Table 1.4 shows that subjects believe that people lied less when ahead of their opponent in terms of performance (models (1), (5), and (7)) and that they lied more about their own number than about their opponent's number (model (1)). This is mainly driven by subjects who themselves misreported in the Identity treatment (model (7)). Finally, and regardless of whether they lied in the game or not, subjects believe that fewer people misreported when paired with someone from the same group (models (5) to (7)).

Note that in the No-Identity treatment, none of these variables are significant, regardless of the actual behavior. In the Identity treatment, these beliefs are consistent with our conjectures but not with the subjects' actual behavior. Even if they believe that individuals are less likely to cheat against an in-group, they do not misreport less when they are themselves in this situation: neither preferences nor beliefs lead them to compete more fairly against an opponent.

Table 1.4 – Determinants of beliefs on misreporting decisions in past sessions

| <i>Dependent variable: Belief about the number of players who misreported</i> | All subjects | No-Identity treatment | | | Identity treatment | | |
|---|---------------------|-----------------------|--------------------------------|--------------------------|----------------------|--------------------------------|--------------------------|
| | | All subjects | Subjects who never misreported | Subjects who misreported | All subjects | Subjects who never misreported | Subjects who misreported |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Identity treatment | -0.583 (0.975) | - | - | - | - | - | - |
| Reporting for Self scenario | 0.333** (0.137) | 0.133 (0.225) | 0.220 (0.459) | 0.077 (0.243) | 0.383** (0.162) | 0.211 (0.217) | 0.555** (0.245) |
| Higher subscore player | -0.380** (0.152) | -0.352 (0.364) | 0.100 (0.701) | -0.641 (0.413) | -0.387** (0.168) | 0.211 (0.243) | -0.984*** (0.213) |
| In-group pair | - | - | - | - | -0.426*** (0.096) | -0.398*** (0.143) | -0.453*** (0.132) |
| SVO angle | -0.025* (0.014) | -0.008 (0.019) | 0.035 (0.027) | -0.012 (0.023) | -0.031* (0.017) | 0.011 (0.025) | -0.047** (0.022) |
| Male | 0.509 (0.458) | -0.585 (0.669) | -0.161 (0.800) | -1.532* (0.775) | 0.866 (0.544) | 0.204 (0.621) | 1.012 (0.621) |
| Constant | 5.230*** | 7.020*** (1.049) | 2.997** (1.229) | 9.429*** (1.169) | 4.085*** (1.047) | 1.042 (1.189) | 5.891** (2.345) |
| Other individual characteristics | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Session fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Nb of obs. | 1280 | 256 | 100 | 156 | 1024 | 512 | 512 |
| Nb of subjects | 192 | 64 | 25 | 39 | 128 | 64 | 64 |
| R2 | 0.054 | 0.050 | 0.117 | 0.183 | 0.076 | 0.115 | 0.204 |
| $p > F$ | 0.066 | 0.300 | 0.021 | 0.002 | <0.001 | <0.001 | <0.001 |

Notes: The Table reports OLS regressions with robust standard errors clustered at the individual level in parentheses. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

1.5 Robustness tests

In our experiment the extent to which people lie to increase their chance of winning the competition is not affected by whether the opponent is an in-group or an out-group and they tend to lie to the same extent when they report about themselves or about

their opponent. However, the fact that lying can be detected by the experimenter at the individual level may generate a selection bias: individuals who dare to cheat in this environment are probably those who have the least moral cost from misbehaving (Kajackaite and Gneezy, 2017) and perhaps also care less about group identity. In contrast, people who are more sensitive to the environment may be more willing to refrain from cheating when the experimenter can detect their lies, may care more about their opponent's identity, and if so may be less willing to misreport their opponent's number than their own. To test the robustness of our finding regarding the null impact of group identity and condition on cheating, we designed a novel Identity-No-Scrutiny treatment and an Identity-No-Info treatment used as a control.

Design

The Identity-No-Scrutiny treatment (involving 104 new subjects in 5 sessions) is similar to the initial Identity treatment except that we use a mind game in which a lie cannot be detected by the experimenter at the individual level.²³ Precisely, in each condition, after performing the task and after being informed of his subscore and of both the subscore and the group identity of his opponent, the subject has to think in his mind about a letter between A and K, inclusive. Then, the subject's screen displays a table that matches randomly each letter with a number between -5 to +5. The subject is requested to report the number that corresponds to the letter he has selected in his mind before seeing the table. For example, if the subject thought about letter B, he has to report the second value in the set; if the random set is "+2, -5, 0, -1, +4, +1, -2, -4, +5, +3, -3", then he has to report -5. In both conditions misreporting is possible and undetectable.

However, this treatment introduces two changes compared to the Identity treatment. First, the experimenter is no longer able to detect a lie at the individual level, which may increase lying by subjects who have a higher moral cost of lying and who are more

²³Mind games have been used to study lying behavior without scrutiny by the experimenter in die-rolling tasks (Jiang, 2015; Shalvi and De Dreu, 2014; Barfort et al., 2015; Potters and Stoop, 2016; Kajackaite and Gneezy, 2017), and in coin toss tasks (Garbarino et al., 2018).

sensitive to the environment. Second, when deciding on whether lying or not, there is now uncertainty about the opponent's random number in the Self-condition and about one's own random number in the Opp-condition. This difference could be seen as unimportant, as the main uncertainty lies in the fact that one's opponent can lie about the actual random number. However, the results reported in Tables 1.2 and 1.3 show that subjects are sensitive to the difference between their own and their opponent's random numbers. Therefore, we also conducted the Identity-No-Info treatment as a control treatment (involving 100 new subjects in 5 sessions). This treatment is similar to the Identity treatment except that when they have to report, subjects are not informed about the number of their opponent in the Self-condition and about their own number in the Opp-condition. Like in the Identity-No-Scrutiny treatment there is uncertainty about the difference between the two numbers, but the degree of scrutiny is the same as in the original Identity treatment.

A significant difference in reporting behavior depending on whether the subject is matched with an in-group or an out-group in the Identity-No-Scrutiny treatment would suggest that scrutiny turns on or off the effect of group identity on misreporting. A significant difference depending on whether the subject reports his own number or his opponent's number in the Identity-No-Info treatment would suggest that uncertainty about the difference between the two numbers turns on or off the effect of self *vs.* other reporting on cheating. Finally, the comparison between the reporting behavior in the Opp- and the Self-conditions in the Identity-No-Scrutiny and the Identity-No-Info treatments allows us to test whether scrutiny or uncertainty affects how subjects report for themselves *vs.* for their opponent.

Results

We find evidence of misreporting in both new treatments, although in the Identity-No-Scrutiny treatment the evidence is only based on comparisons of the theoretical and empirical distributions of the reported numbers since we do not know the true values of

the random draws. A visual inspection of the reported numbers' distributions in the new treatments (Figures C.3 and C.4 in the Appendix C) clearly shows that these distributions are not uniform, like in the initial treatments (Figures 1.2a and 1.2b). Non-parametric statistics on individual data can only be conducted on the first period since data are no longer independent in the following periods. Two-sided χ^2 goodness-of-fit tests reject a uniform distribution of the reported numbers in the first period in both treatments ($p < 0.001$). Compared to the first period of the initial Identity treatment, there is no significant difference with the Identity-No-Info treatment (two-sided Wilcoxon rank-sum test, $p = 0.272$) but there is a significant difference with the Identity-No-Scrutiny treatment ($p < 0.001$). Subjects report significantly higher (lower) numbers for themselves (for the opponent) when there is no possible scrutiny from the experimenter, which is in line with the literature (Gneezy et al., 2018). The proportion of extreme numbers (5 in the Self-condition and -5 in the Opp-condition) is also significantly different from the expected proportion if reports were honest in both treatments (two-sided binomial tests, $p < 0.001$).²⁴

Table 5 reports the average reported random numbers across conditions in the new treatments. Two-sided Wilcoxon signed-rank tests based on averages at the session level²⁵ reveal no significant difference in reporting when the opponent is an in-group or an out-group in both the Identity-No-Scrutiny treatment (N=10, $p = 0.786$) and the Identity-No-Info treatment (N=10, $p = 0.225$). This is not due to a failure in the induction of group identity, as subjects show a clear in-group bias in the other-other allocation task (they allocate more ECUs to an in-group than to an out-group in Scenario 3, $p < 0.001$ for all endowments, two-sided pairwise t-tests). In contrast, we find a significant difference between the Self- and the Opp-conditions in the Identity-No-Scrutiny treatment (N=10,

²⁴To increase the number of observations for both the χ^2 goodness-of-fit and the binomial tests, we multiply the reported numbers by -1 in the Opp-condition. This allows us to pool the data from the Self and Opp-conditions.

²⁵Here also, for the tests the reported numbers in the Opp-condition are multiplied by -1.

$p = 0.043$), but not in the Identity-No-Info treatment ($N=10$, $p = 0.500$).²⁶ In Appendix B we report a regression analysis controlling for the subjects' socio-demographic characteristics. Table B.4 studies the probability to misreport selfishly in the Identity-No-Info treatment using Logit models, and Table B.5 studies the determinants of the reported number in the Identity-No-Info and the Identity-No-Scrutiny treatments, using Tobit models. These regressions confirm the results based on the non-parametric tests.

Table 1.5 – Average reported random number by treatment and condition

| Treatment | Condition | Opponent's identity | Average reported number | N |
|-------------|-----------|---------------------|-------------------------|-----|
| Identity | Self | All | 1.52 (0.06) | 800 |
| | | In-group opponent | 1.53 (0.09) | 400 |
| | | Out-group opponent | 1.50 (0.09) | 400 |
| No-Scrutiny | Opp | All | -1.00 (0.07) | 800 |
| | | In-group opponent | -1.01 (0.09) | 400 |
| | | Out-group opponent | -0.99 (0.09) | 400 |
| Identity | Self | All | 0.91 (0.06) | 832 |
| | | In-group opponent | 0.85 (0.09) | 416 |
| | | Out-group opponent | 0.98 (0.09) | 416 |
| No-Info | Opp | All | -0.95 (0.06) | 832 |
| | | In-group opponent | -0.95 (0.09) | 416 |
| | | Out-group opponent | -0.95 (0.09) | 416 |

These results show that the experimenter's scrutiny does not turn on or off the effect of group identity on reporting: in no treatment individuals adjust their reports to the group identity of their opponent. Similarly, uncertainty about the difference between the two numbers does not turn on or off the effect of self vs. other reporting, as individuals do not report differently their own numbers and their opponent's numbers either in the Identity-No-Info (see Table B.4 in the Appendix B) or in the initial Identity treatment (see Table 1.2). However, when the experimenter's scrutiny is removed, individuals report

²⁶To assess the significance of these results we calculated the achieved power of our design for three levels of size effect. We used the software G*Power with a p-value sets at 5%. Our achieved power is 13%, 38% and 69% considering low, medium and large effect sizes (Cohen's d is equal to 0.3, 0.6 and 0.9, respectively).

significantly higher absolute numbers in the Self-condition than in the Opp-condition (see Table B.5, in the Appendix B). An interpretation is that when the moral cost of misreporting is reduced, some individuals who would not lie under scrutiny are lying, especially to overreport their own number. This supports our last result:

Result 4: Even when the experimenter's scrutiny is removed, individuals do not adjust their behavior to the group identity of their opponent but in this setting they report higher absolute numbers for themselves than for their opponent.

1.6 Discussion and Conclusion

Studying ethics in competition, we explored whether social distance from an opponent in terms of group identity and the nature of reporting (about oneself or the opponent) affects the willingness to cheat to increase one's chance to succeed. First, we found that less than half of the subjects never misreport, and those who misreport do it to the full extent most of the time. This goes against the idea of a convex cost of lying (e.g., Mazar et al., 2008; Lundquist et al., 2009; Fischbacher and Föllmi-Heusi, 2013). Recent models (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Khalmetski and Sliwka, 2017) have shown that individuals may refrain from reporting the highest outcome because this may alter their reputation *vis-à-vis* others or the experimenter. Behavior in our experiment is more consistent with the notion of a fixed cost of lying under scrutiny (e.g., Kajackaite and Gneezy, 2017). Individuals suffering a sufficiently high moral cost abstain from misreporting, whereas the others lie in full. Moreover, cheaters may be less reluctant to lie in full in our game since the report determines only a fraction of the final score which itself affects the probability of winning but not directly the actual outcome of the competition. This may weaken the feeling of responsibility.

A second finding is that cheaters underreport their opponent's outcome as much as they exaggerate their own outcome. This contrasts with Rigdon and D'Esterre (2015) who found that people lie less when reporting the performance of another subject than

when reporting their own performance. However, our results become consistent with theirs when we remove the experimenter's scrutiny (as in their study), as in this condition subjects report higher numbers for themselves than for their opponent in absolute terms. This could be explained by the existence of conditional liars: individuals who are more sensitive to the moral cost of lying may be more likely to lie when there is no scrutiny and also more likely to misreport their own outcome than another person's outcome perhaps if the latter type of lie is perceived as aggressive.

Finally, the fact that group identity does not affect misreporting in any condition contrasts both with our expectation and those of the subjects that people would be more likely to lie at the detriment of an out-group. It also differs from (Banerjee et al., 2018) who found in the context of castes in India that the lying behavior of people from the dominant caste was discriminatory against previous competitors from the backward caste after they learned that they lost the competition. The difference may result from the fact that in our experiment, group identity was minimal and the opportunity to lie was embedded in the contest game, whereas it followed competition in their case. It also differs from Della Valle and Ploner (2017) who found that dishonesty increases after being treated unfairly by a dictator from another group identity. The difference can be due to our competitive environment but not to the fact that lying can be detected at the individual level since in our treatment using a mind game group identity does not affect either reporting behavior. Note that our results are in line with those of Feldhaus and Mans (2014) who found no effect of social identity on lying in a sender-receiver game, and with those of Chowdhury et al. (2016) who found an effect of group identity on group conflict when group identity was natural but not when it was artificially induced as in our study.

In the context of distributive choices, Guala and Filippin (2017) proposed that group identity has no systematic effect across contexts because it is a heuristic rather than a social preference or a norm. Our interpretation is that in the context of repeated

competition, the willingness to win is stronger than affiliation. A competitor, whether socially close or not, is above all an opponent to beat. This does not mean that ethics is indifferent to the composition of teams but the impact of group identity on unethical behavior may be more sensitive to the environment than previously considered. It would thus be important to investigate in which environments group identity is more, or less, likely to influence unethical practices. For example, we know that misreporting is more important when it is about luck than about performance (Kajackaite, 2018) but we do not know whether the opponent's group identity would matter when misreporting is about performance and not about luck. Another extension could study whether variations in the nature and intensity of competition affect the impact of group identity on misreporting.

Bibliography

- Abbink, K., J. Brandts, B. Herrmann, and H. Orzen (2012). Parochial altruism in inter-group conflicts. *Economics Letters* 117(1), 45–48.
- Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica*. Forthcoming.
- ACFE (2016). Report to the nations on occupational fraud and abuse. Technical report.
- Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics* 115(3), 715–753.
- Banerjee, R., N. D. Gupta, and M. C. Villeval (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review* 101, 567–604.
- Barfort, S., N. A. Harmon, F. G. Hjorth, and A. L. Olsen (2015). Dishonesty and Selection into Public Service in Denmark: Who Runs the World’s Least Corrupt Public Sector? Discussion Papers 15-12, University of Copenhagen. Department of Economics.
- Bent, W. (2018). Speaking Out: The State Department Needs to Reevaluate Its Use of 360-Degree Reviews. *The Foreign Service Journal*. Find at <http://www.afsa.org/state-department-needs-reevaluate-its-use-360-degree-reviews>.
- Bernhard, H., U. Fischbacher, and E. Fehr (2006). Parochial altruism in humans. *Nature* 442(7105), 912–5.
- Bock, O., I. Baetge, and A. Nicklisch (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71, 117–120.
- Bracken, D., C. W Timmreck, and A. Church (2001). *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Buchan, N. and R. Croson (2004). The boundaries of trust: Own and others’ actions in the US and China. *Journal of Economic Behavior and Organization* 55(4), 485–504.
- Cadsby, C. B., N. Du, and F. Song (2016). In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization* 128, 59–71.
- Carpenter, J., P. H. Matthews, and J. Schirm (2010). Tournaments and Office Politics : Evidence from a Real Effort Experiment. *American Economic Review* 100(1), 504–517.
- Cartwright, E. and M. L. C. Menezes (2014). Cheating to win: Dishonesty and the intensity of competition. *Economics Letters* 122(1), 55–58.
- Chakravarty, S. and S. Maximiano (2016). Deception , social preferences and friendship. Working Papers 1008, Purdue University, Department of Consumer Sciences.

- Charness, G., D. Masclet, and M. C. Villeval (2014). The Dark Side of Competition for Status. *Management Science* 60(1), 38–55.
- Charness, G., L. Rigotti, and A. Rustichini (2007). Individual Behavior and Group Membership. *American Economic Review* 97(4), 1340–1352.
- Chen, Y., S. X. Li, and Y. Chen (2009). Group identity and social preferences. *American Economic Review* 97(4), 431–457.
- Chen, Y., S. X. Li, T. X. Liu, and M. Shih (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. *Games and Economic Behavior* 84, 58–86.
- Chen, Z., D. Ong, and R. M. Sheremeta (2015). Competition Between and Within Universities: eoretical and Experimental Investigation of Group Identity and the Desire to Win. MPRA Paper No. 67522.
- Chowdhury, S. M. and O. Gürtler (2015). Sabotage in Contests : A Survey. *Public Choice* 164(1), 135–155.
- Chowdhury, S. M., J. Y. Jeon, and A. Ramalingam (2016). Identity and Group Conflict. *European Economic Review* 90, 107–121.
- Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature* 516(7529), 86–89.
- Cohn, A., M. A. Maréchal, and T. Noll (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies* 82(4), 1289–1308.
- Conrads, J., B. Irlenbusch, R. M. Rilke, A. Schielke, and G. Walkowitz (2014). Honesty in tournaments. *Economics Letters* 123(1), 90–93.
- Datta Gupta, N., A. Poulsen, and M. C. Villeval (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry* 51(1), 816–835.
- Della Valle, N. and M. Ploner (2017). Reacting to Unfairness: Group Identity and Dishonest Behavior. *Games* 8(3), 28.
- Dreber, A. and M. Johannesson (2008). Gender differences in deception. *Economics Letters* 99(1), 197–199.
- Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory* 175(175), 248–264.
- Eckel, C. C. and P. J. Grossman (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization* 58(3), 371–392.
- Efferson, C., R. Lalive, and E. Fehr (2008). The Coevolution of Cultural Groups and Ingroup Favoritism. *Science* 321(5897), 1844–1849.

- Eriksson, T., L. Mao, and M. C. Villeval (2017). Saving face and group identity. *Experimental Economics* 20(3), 622–647.
- Erkal, N., L. Gangadharan, and N. Nikiforakis (2011). Relative Earnings and Giving in a Real-Effort Experiment. *American Economic Review* 101(7), 3330–3348.
- Falk, A. and N. Szech (2013). Morals and Markets. *Science* 340(6133), 707–712.
- Faravelli, M., L. Friesen, and L. Gangadharan (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior and Organization* 110, 160–175.
- Feldhaus, C. and J. Mans (2014). Who do you lie to? Social identity and the costs of lying. Working Paper Series in Economics No. 74 from University of Cologne, Department of Economics.
- Feltovich, N. (2018). The interaction between competition and unethical behaviour : experimental evidence. *Experimental Economics*. Forthcoming.
- Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise – An Experimental Study on Cheating. *Journal of the European Economic Association* 11(40), 525–547.
- Gächter, S. and J. F. Schulz (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531(7595), 496–499.
- Garbarino, E., R. Slonim, and M. C. Villeval (2018). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization*. In press.
- Garrett, N., S. C. Lazzaro, D. Ariely, and T. Sharot (2016). The brain adapts to dishonesty. *Nature Neuroscience* 19, 1727–1732.
- Gibson, R., C. Tanner, and A. F. Wagner (2013). Preferences for Truthfulness : Heterogeneity among and within Individuals. *American Economic Review* 103(1), 532–548.
- Gino, F., S. Ayal, and D. Ariely (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological Science* 20(3), 393–8.
- Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review* 108(2).
- Goette, L., D. Huffman, S. Meier, and B. L. Goette (2006). The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups. *American Economic Review* 96(2), 212–216.
- Goette, L., D. Huffman, S. Meier, and M. Sutter (2012). Competition between organizational groups : Its impact on altruistic and anti-social motivations. *Management Science* 58(5), 948–960.

- Grosch, K. and H. A. Rau (2017). Gender differences in honesty: The role of social value orientation. *Journal of Economic Psychology* 62, 258–267.
- Guala, F. and A. Filippin (2017). The Effect of Group Identity on Distributive Choice: Social Preference or Heuristic? *Economic Journal* 127(602), 1047–1068.
- Guala, F., L. Mittone, and M. Ploner (2013). Group membership, team preferences, and expectations. *Journal of Economic Behavior & Organization* 86, 183–190.
- Güth, W., M. Ploner, and T. Regner (2009). Determinants of in-group bias: Is group affiliation mediated by guilt-aversion? *Journal of Economic Psychology* 30(5), 814–827.
- Harbring, C. and B. Irlenbusch (2011, mar). Sabotage in Tournaments: Evidence from a Laboratory Experiment. *Management Science* 57(4), 611–627.
- Harbring, C., B. Irlenbusch, M. Kräkel, and R. Selten (2007). Sabotage in Corporate Contests – An Experimental Analysis. *International Journal of the Economics of Business* 14(3), 367–392.
- Hargreaves Heap, S. P. and D. J. Zizzo (2009). The Value of Groups. *American Economic Review* 99(1), 295–323.
- Hruschka, D., C. Efferson, T. Jiang, A. Falletta-Cowden, S. Sigurdsson, R. McNamara, M. Sands, S. Munira, E. Slingerland, and J. Henrich (2014). Impartial Institutions, Pathogen Stress and the Expanding Social Network. *Human Nature* 25(4), 567–579.
- Jiang, T. (2015). Other-regarding Preferences and Other-regarding Cheating - Experimental Evidence from China, Italy and the Netherlands. Mimeo.
- Kajackaite, A. (2018). Lying about luck versus lying about performance. *Journal of Economic Behavior and Organization* 153, 194–199.
- Kajackaite, A. and U. Gneezy (2017). Incentives and cheating. *Games and Economic Behavior* 102, 433–444.
- Kantor, J. and D. Streitfeld (2015). Inside Amazon: Wrestling Big Ideas in a Bruising Workplace. *New York Times*.
- Kato, T., S. Pian, and P. Shu (2016). Competition and social identity in the workplace: Evidence from a Chinese textile firm. *Journal of Economic Behavior and Organization* 131(Part A), 37–50.
- Khalmetski, K. and D. Sliwka (2017). Disguising Lies – Image Concerns and Partial Lying in Cheating Games. CESifo Working Paper no. 6347.
- Kilduff, G. J., A. D. Galinsky, E. Gallo, and J. J. Reade (2016). Whatever It Takes To Win: Rivalry Increases Unethical Behavior. *Academy of Management Journal* 59(5), 1508–1534.

- Kolmar, M. and A. Wagener (2012). Group Identities in Conflicts. Working Paper, University of St. Gallen.
- Konrad, K. A. (2000). Sabotage in Rent-seeking Contests. *Journal of Law, Economics & Organization* 16(1), 155–165.
- Lazear, E. P. (1989). Pay equality and Industrial Politics. *Journal of Political Economy* 97(3), 561–580.
- Lazear, E. P. and S. Rosen (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89(5), 841–864.
- Lepsinger, R. and A. D. Lucia (2009). *The Art and Science of 360-Degree Feedback* (2nd editio ed.).
- Li, S. X. and T. X. Liu (2017). Group Identity and Cooperation in Infinitely Repeated Games. Mimeo.
- López-Pérez, R. and E. Spiegelman (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics* 16(3), 233–247.
- Lundquist, T., T. Ellingsen, E. Gribbe, and M. Johannesson (2009). The aversion to lying. *Journal of Economic Behavior and Organization* 70(1-2), 81–92.
- Mazar, N., O. Amir, and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45(6), 633–644.
- Mcleish, K. N. and R. J. Oxoby (2011). Social Interactions and the Salience of Social Identity. *Journal of Economic psychology* 32(1), 172–178.
- Meub, L., T. Proeger, T. Schneider, and K. Bizer (2016). The victim matters – experimental evidence on lying, moral costs and moral cleansing. *Applied Economics Letters* 23(16), 1162–1167.
- Muehlheusser, G., A. Roider, and N. Wallmeier (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters* 128, 25–29.
- Murphy, R. O., K. A. Ackermann, and M. J. J. Handgraaf (2011). Measuring Social Value Orientation. *Judgment and Decision Making* 6(8), 771–781.
- Nieken, P. and S. Dato (2016). Compensation and Honesty: Gender Differences in Lying.
- Ockenfels, A. and P. Werner (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior and Organization* 108, 453–462.
- Pittarello, A., E. Rubaltelli, and R. Rumiati (2013). You can't be better than me: The role of the reference point in modulating people's pursuit of wealth. *Journal of Economic Psychology* 37, 65–76.

- Potters, J. and J. Stoop (2016). Do cheaters in the lab also cheat in the field? *European Economic Review* 87, 26–33.
- Rigdon, M. L. and A. D’Esterre (2015). The Effects of Competition on the Nature of Cheating Behavior. *Southern Economic Journal* 81.
- Ruffle, B. J. and R. Sosis (2006). Cooperation and the in-group-out-group bias: A field test on Israeli kibbutz members and city residents. *Journal of Economic Behavior and Organization* 60(2), 147–163.
- Schurr, A. and I. Ritov (2016). Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences* 113(7), 201515102.
- Schwieren, C. and D. Weichselbaumer (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology* 31(3), 241–253.
- Shalvi, S. and C. K. W. De Dreu (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences* 111(15), 5503–5507.
- Shleifer, A. (2004). Does competition destroy Ethical Behavior ? *American Economic Review* 94(2), 414–418.
- Tajfel, H. and J. Turner (1979). *An Integrative Theory of Intergroup Conflict*. Monterey, CA: Brooks-Cole.
- Tullock, G. (1980). *Efficient Rent-Seeking : Toward a Theory of the Rent Seeking Society*. Texas A&M University Press.
- Weber, E. U., A.-R. A.-R. Blais, and N. E. Betz (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making* 15, 263–290.
- Yamagishi, T. and N. Mifune (2008). Does Shared Group Membership Promote Altruism? *Rationality and Society* 20(1), 5–30.

A Instructions

Instructions for the No Identity treatment (translated from French)

Thank you for participating in this experiment on decision making. Please do not communicate with the others participants until the end of the session. In this session, all your decisions are kept anonymous.

During the experiment, the transactions will be made in ECUs (Experimental Currency Units) and not in Euros. All your earnings will be expressed in ECUs. At the end of the experiment, the total amount of ECUs you have earned will be converted into Euros at the following rate:

$$100 \text{ ECUs} = 2.5 \text{ Euros}$$

You will also receive a 5 Euros show-up fee. Your total earnings will be paid to you in cash at the end of the session. The payments will be made privately in a separate room. The other participants will not be informed of your earnings.

The session consists of three parts. You will receive the instructions for each part after having completed the previous part.

Each of you has completed an online questionnaire before participating in this session. An ID was sent to you by email to keep your answers to this questionnaire anonymous. You will be asked to enter this ID into your computer at the very beginning of the session. Your earnings in this preliminary part have been computed and they will be paid to you at the end of this session.

Part 1

General description

This part consists of 16 periods in which you will have to perform a task. In each period your task consists in solving problems during 30 seconds. Each problem solved increases your "subscore" by one point.

In each period, you are randomly matched with another participant. Thus, it is very unlikely that you are matched with the same participant twice in a row.

At the end of the period, a participant in the pair earns 100 ECUs and the other earns 50 ECUs. Your payoff depends on your "final score" relative to the final score of the other participant. Your final score consists of two elements: your subscore plus a random number that you or the other participant have to report. If your final score is higher than the final score of the other participant, the higher is the difference between your final score and the other participant's final score, the more likely you will get 100 ECUs.

At the end of the session, two periods among the 16 will be randomly selected by the program to determine your earnings in this part.

Description of each period

A period consists of 4 stages.

- Stage 1: the task

You have to perform a task during 30 seconds. This task consists in decoding letters into numbers. A conversion table is displayed permanently on your screen. After having decoded the displayed letter, you have to press the OK button to validate your answer.

You are immediately informed of whether your answer is correct or not. If your answer is incorrect, you have to enter another number. The next problem will be displayed only after you have entered a correct answer. If your answer is correct, a new letter is displayed and this will continue for 30 seconds. The conversion table changes in each period. It is the same for all the participants.

- Stage 2: matching pairs and information

The program forms pairs randomly. You are informed about your subscore and the subscore of the other participant. The other participant receives the same information as you.

- Stage 3: final score of the period

Your final score is determined by the sum of your subscore and the reported random number.

The program selects independently a random number for you and a random number for the other participant (with rebate). The random numbers are integers between -5 and 5. You are informed about your random number and the other participant's random number. You have to memorize these numbers because you will be asked to report one of them on the following screen in order to determine your final scores.

Depending on the period, one of the two following cases occurs:

- **Case 1:** you have to report your own random number. The other participant reports his own random number.
- **Case 2:** you have to report the random number of the other participant. The other participant has to report your random number.

Once the random numbers have been reported, the program computes your final score and the other participant's final score. The final scores are determined differently in the two cases.

In case 1, your final score is equal to your subscore (the number of problems solved) plus your random number as reported by yourself. The computation is the same for the other participant.

Case 1: Your final score = your subscore + your random number as reported by yourself
Other participant's final score = his subscore + his random number as reported by himself

In case 2, your final score is equal to your subscore (the number of problems solved) plus your random number as reported by the other participant. The computation is the same for the other participant.

Case 2: Your final score = your subscore + your random number as reported by the other participant
Other participant's final score = his subscore + his random number as reported by yourself

- Stage 4: Determination of the payoff of the period

The higher is your final score relative to the final score of the other participant, the more likely you will earn 100 ECUs. The lower is your final score relative to the final score of the other participant, the more likely you will earn 50 ECUs.

More precisely, your chance of winning 100 ECUs is determined as follows:

$$\text{Your chance of winning 100 ECUs} = \frac{\text{Your final score}}{\text{Your final score} + \text{other participant's final score}}$$

This formula indicates that for your given final score, if the other participant's final score increases, your chance to win 100 ECUs decreases. In contrast, for a given final score of the other participant, if your final score increases, your chance to win 100 ECUs also increases. Your chance to win depends on your final score but also on the final score of the other participant, and the same logic applies to him.

- End of the period
 - In case 1, you are informed of your final score. You are not informed of the final score of the other participant.
 - In case 2, you are informed of the final score of the other participant. You are not informed of your final score.

In both cases, you are informed of your payoff, either 50 ECUs or 100 ECUs.

The next period starts automatically. You have to perform the same task and new pairs are formed.

Summary

In each period:

- You solve problems during 30 seconds.
- You are randomly matched with another participant.
- You are informed about your subscore and the subscore of this other participant.
- You are informed about your random number and the other participant's random number.
- You report one of the two random numbers, depending on the period.
- You are informed of your payoff and of your final score, or of the final score of the other participant, depending on the period.

Just before the end of the session, there will be a last stage. The instructions will be displayed on your screen. At this time, do not hesitate to ask questions by raising your hand and/or by pressing the red button.

Part 2

The instructions for this part will be directly displayed on your screen. If you have questions at this time, do not hesitate to ask questions by raising your hand and/or by pressing the red button.

End of the session

At the end of the session, your screen will display the following information:

- The two periods selected in part 1 and your earnings for this part.
- Your earnings for the second part.
- Your total earnings for the session.

Your earnings for the online part and the show-up fee will be added to the earnings of the session. The payment will be made by someone who does not know the content of the session and has no access to the program and the data.

Instructions for the Identity treatment (translated from French)

Thank you for participating in this experiment on decision making. Please do not communicate with the others participants until the end of the session. In this session, all your decisions are kept anonymous.

During the experiment, the transactions will be made in ECUs (Experimental Currency Units) and not in Euros. All your earnings will be expressed in ECUs. At the end of the experiment, the total amount of ECUs you have earned will be converted into Euros at the following rate:

$$100 \text{ ECUs} = 2.5 \text{ Euros}$$

You will also receive a 5 Euros show-up fee. Your total earnings will be paid to you in cash at the end of the session. The payments will be made privately in a separate room. The other participants will not be informed of your earnings.

The session consists of three parts. You will receive the instructions for each part after having completed the previous part.

Each of you has completed an online questionnaire before participating in this session. An ID was sent to you by email to keep your answers to this questionnaire anonymous. You will be asked to enter this ID into your computer at the very beginning of the session. Your earnings in this preliminary part have been computed and they will be paid to you at the end of this session.

Part 1

At the beginning of this part and for the rest of the session, participants are divided into two groups of equal size. Each group bears the name of an artist: Klee or Kandinsky. You will be classified into one of the two groups, based on which painter you prefer.

The following procedure is used to determine to which group you will belong to.

Your screen will display five pairs of painting successively. In each pair, one painting was painted by Klee and the other one by Kandinsky. We will ask you to choose which one you prefer in each pair.

Then, you will be matched with people who have the closest preferences from yours. The program will assign each participant to either the Klee group or the Kandinsky group, in order to form two groups of equal size. Having a stronger preference for the paintings of a given artist will increase the likelihood to be assigned to this group. If there are too many participants in one group, those who had the weakest preference for this artist will be assigned to the other group.

Then, you will be informed of which group you belong to. During the rest of the task, you will remain in the same group.

In a second stage, two more paintings will be displayed on your screen. You will have to select the artist who you think made each painting. Each correct answer pays 50 ECUs.

Before entering your answers, you will have access to a chat program to get help from or offer help to other members of your group. You will not be able to communicate with members of the other group. For example, if you are a member of the Klee group you will only be able to chat with the others members of the Klee group, and conversely if you belong to the Kandinsky group. To chat, you have to write your message in the bottom part of the chat box and press enter to validate.

You will be given 8 minutes to access the chat program. Once the 8 minutes have elapsed, a new screen will be displayed in which you will have to enter your answers, individually. You

will be informed of your payoff for this part at the end of the session. If you have questions about these instructions, please raise your hand or press the red button on the side of your desk. We will answer to your questions in private.

Part 2

This part consists of five periods. In each period you will receive a certain endowment in ECUs. The amount varies from period to period. You will be asked to allocate these ECUs between two other participants (called Person A and Person B) under three scenarios:

- Scenario 1: both are from your own group
- Scenario 2: both are from the other group
- Scenario 3: one is from your group, and one is from the other group.

For each scenario you must allocate your entire endowment between the two participants. Allocations have to be integers. Do not allocate any ECU to yourself. Your decisions will be used to determine other participants' payoffs for this part at the end of the session. Similarly, your payoff for this part will be determined by others decisions.

A screenshot is reproduced below.

The screenshot shows a software interface for a decision-making task. At the top left, it says "Period 1 de 4". At the top right, it says "Remaining time [sec]: 25". The main text reads: "You are endowed with 50 ECU that you have to allocate between the participants for each case. Amount allocated have to be integers." Below this, there are three rows of input fields, each representing a scenario. Each row has two columns of input fields, one for Person A and one for Person B, with a plus sign and "= 50" between them. The scenarios are: 1. Both participants from your own group. 2. Both participants from the other group. 3. One participant from your group and one from the other group. An "OK" button is located at the bottom right.

At the end of the session, the program will randomly select a period that is used to calculate the payoffs. Each period of decision will have an equal chance of being selected. For the selected period, a new random draw will determine which scenario will be used to calculate the payoffs.

Then, the program will randomly select two participants. The allocation for Person A made by the first participant and the decision for Person B made by the second participant will be added to determine your payoff.

Depending on the random draw the selected participants can be from your group or not. If the first scenario is selected, the two participants will be from your group. If the second scenario is selected, the two participants will be from the other group. If the third scenario is selected, the first participant will be from your group and the second one from the other group.

You will be informed of your payoff for this part at the end of the session.

After reading again these instructions, if you have any questions please raise your and/or press the red button.

Part 3

General description

This part consists of 16 periods in which you will have to perform a task. In each period your task consists in solving problems during 30 seconds. Each problem solved increases your “subscore” by one point.

In each period, you are randomly matched with another participant. Thus, it is very unlikely that you are matched with the same participant twice in a row.

At the end of the period, a participant in the pair earns 100 ECUs and the other earns 50 ECUs. Your payoff depends on your “final score” relative to the final score of the other participant. Your final score consists of two elements: your subscore plus a random number that you or the other participant have to report. If your final score is higher than the final score of the other participant, the higher is the difference between your final score and the other participant’s final score, the more likely you will get 100 ECUs.

At the end of the session, two periods among the 16 will be randomly selected by the program to determine your earnings in this part.

Description of each period

A period consists of 4 stages.

- Stage 1: the task

You have to perform a task during 30 seconds. This task consists in decoding letters into numbers. A conversion table is displayed permanently on your screen. After having decoded the displayed letter, you have to press the OK button to validate your answer. You are immediately informed of whether your answer is correct or not. If your answer is incorrect, you have to enter another number. The next problem will be displayed only after you have entered a correct answer. If your answer is correct, a new letter is displayed and this will continue for 30 seconds. The conversion table changes in each period. It is the same for all the participants.

- Stage 2: matching pairs and information

The program forms pairs randomly. You are informed of the other participant’s group (Klee or Kandinsky). Then, you are informed about your subscore and the subscore of the other participant. The other participant receives the same information as you.

- Stage 3: final score of the period

Your final score is determined by the sum of your subscore and the reported random number.

The program selects independently a random number for you and a random number for the other participant (with rebate). The random numbers are integers between -5 and 5. You are informed about your random number and the other participant’s random number. You have to memorize these numbers because you will be asked to report one of them on the following screen in order to determine your final score.

Depending on the period, one of the two following cases occurs:

- Case 1: you have to report your own random number. The other participant reports his own random number.

- Case 2: you have to report the random number of the other participant. The other participant has to report your random number.

Once the random numbers have been reported, the program computes your final score and the other participant's final score. The final scores are determined differently in the two cases.

In case 1, your final score is equal to your subscore (the number of problems solved) plus your random number as reported by yourself. The computation is the same for the other participant.

Case 1: Your final score = your subscore + your random number as reported by yourself
Other participant's final score = his subscore + his random number as reported by himself

In case 2, your final score is equal to your subscore (the number of problems solved) plus your random number as reported by the other participant. The computation is the same for the other participant.

Case 2: Your final score = your subscore + your random number as reported by the other participant
Other participant's final score = his subscore + his random number as reported by yourself

- Stage 4: Determination of the payoff of the period

The higher is your final score relative to the final score of the other participant, the more likely you will earn 100 ECUs. The lower is your final score relative to the final score of the other participant, the more likely you will earn 50 ECUs.

More precisely, your chance of winning 100 ECUs is determined as follows:

$$\text{Your chance of winning 100 ECUs} = \frac{\text{Your final score}}{\text{Your final score} + \text{other participant's final score}}$$

This formula indicates that for your given final score, if the other participant's final score increases, your chance to win 100 ECUs decreases. In contrast, for a given final score of the other participant, if your final score increases, your chance to win 100 ECUs also increases. Your chance to win depends on your final score but also on the final score of the other participant, and the same logic applies to him.

- End of the period

- In case 1, you are informed of your final score. You are not informed of the final score of the other participant.

- In case 2, you are informed of the final score of the other participant. You are not informed of your final score.

In both cases, you are informed of your payoff, either 50 ECUs or 100 ECUs.

The next period starts automatically. You have to perform the same task and new pairs are formed.

Summary

In each period:

- You solve problems during 30 seconds.
- You are randomly matched with another participant.
- You are informed about the group of the other participant and about your subscore and the subscore of this other participant.
- You are informed about your random number and the other participant's random number.
- You report one of the two random numbers, depending on the period.
- You are informed of your payoff and of your final score, or of the final score of the other participant, depending on the period.

Just before the end of the session, there will be a last stage. The instructions will be displayed on your screen. At this time, do not hesitate to ask questions by raising your hand and/or by pressing the red button.

End of the session

At the end of the session, your screen will display the following information:

- Your earnings for part 1.
- The period and the case selected for the computation of your earnings in part 2 and your earnings for this part.
- The two periods selected in part 3 and your earnings for this part.
- Your earnings for the last stage.
- Your total earnings for the session.

Your earnings for the online part and the show-up fee will be added to the earnings of the session. The payment will be made by someone who does not know the content of the session and has no access to the program and the data.

Instructions for the Identity-No-Scrutiny treatment (translated from French)

The instructions for part 1, part 2 and for the end of the session are similar to the Identity treatment. They are omitted here.

Part 3

General description

This part consists of 16 periods in which you will have to perform a task. In each period your task consists in solving problems during 30 seconds. Each problem solved increases your “subscore” by one point.

In each period, you are randomly matched with another participant. Thus, it is very unlikely that you are matched with the same participant twice in a row.

At the end of the period, a participant in the pair earns 100 ECUs and the other earns 50 ECUs. Your payoff depends on your “final score” relative to the final score of the other participant. Your final score consists of two elements: your subscore plus a random number that you or the other participant have to report. If your final score is higher than the final score of the other participant, the higher is the difference between your final score and the other participant’s final score, the more likely you will get 100 ECUs.

At the end of the session, two periods among the 16 will be randomly selected by the program to determine your earnings in this part.

Description of each period

A period consists of 4 stages.

- Stage 1: the task

You have to perform a task during 30 seconds. This task consists in decoding letters into numbers. A conversion table is displayed permanently on your screen. After having decoded the letter displayed letter, you have to press the OK button to validate your answer. You are immediately informed of whether your answer is correct or not. If your answer is incorrect, you have to enter another number. The next problem will be displayed only after you have entered a correct answer. If your answer is correct, a new letter is displayed and this will continue for 30 seconds. The conversion table changes in each period. It is the same for all the participants. Each letter correctly encoded increases your subscore by 1 point.

- Stage 2: matching pairs and information

The program forms pairs randomly. You are informed of the other participant’s group (Klee or Kandinsky). Then, you are informed about your subscore and the subscore of the other participant. The other participant receives the same information as you.

- Stage 3: final score of the period

Your final score is determined by the sum of your subscore and the reported random number.

The random number is determined by the following procedure.

- On a first screen you will be asked to choose, in your head, a letter between A and K. You will have to memorize this letter. Then, the program will determine independently a random set of 11 integer numbers. This set contains all integers between -5 and +5, sorted out in a random order. Each letter between A and K will be matched with a different number. The table below displays an example of set.

| | | | | | | | | | | |
|---|----|----|----|---|---|----|---|----|---|---|
| A | B | C | D | E | F | G | H | I | J | K |
| 3 | -3 | -2 | -5 | 2 | 0 | -4 | 4 | -1 | 5 | 1 |

- Then, depending on the period, one of the two following cases occurs
 - * **Case 1:** On a new screen, you will be informed on the set of numbers randomly drawn for you in the period. You will not know the set of numbers of the other participant and the other participant will not know your set of numbers. You will be asked to report the number which corresponds to the letter you chose in your head and had to keep in mind. For example, if you chose the letter E, you have to report the number located in the 5th rank in the set starting from the left. In the example above, it corresponds to the number 2. The other participant will have to report the number corresponding to the letter he has chosen and kept in his mind.
 - * **Case 2:** On a new screen, you will be informed on the set of numbers randomly drawn for the other participant in the period. You will not know the set of numbers drawn for you and the other participant will not know his set of numbers. You will be asked to report the number which corresponds to the letter you chose in your head and had to keep in mind. The other participant will have to report the number corresponding to the letter he has chosen and kept in his mind.

Once the random numbers have been reported, the program computes your final score and the other participant's final score. The final scores are determined differently in the two cases.

In case 1, your final score is equal to your subscore (the number of problems solved) plus the random number reported by yourself. The computation is the same for the other participant.

Case 1: Your final score = your subscore + the random number as reported by yourself
 Other participant's final score = his subscore + the random number as reported by himself

In case 2, your final score is equal to your subscore (the number of problems solved) plus the random number reported by the other participant. The computation is the same for the other participant.

Case 2: Your final score = your subscore + the random number as reported by the other participant
 Other participant's final score = his subscore + the random number as reported by yourself

- Stage 4: Determination of the payoff of the period

The higher is your final score relative to the final score of the other participant, the more likely you will earn 100 ECUs. The lower is your final score relative to the final score of the other participant, the more likely you will earn 50 ECUs.

More precisely, your chance of winning 100 ECUs is determined as follows:

$$\text{Your chance of winning 100 ECUs} = \frac{\text{Your final score}}{\text{Your final score} + \text{other participant's final score}}$$

This formula indicates that for your given final score, if the other participant's final score increases, your chance to win 100 ECUs decreases. In contrast, for a given final score of the other participant, if your final score increases, your chance to win 100 ECUs also increases. Your chance to win depends on your final score but also on the final score of the other participant, and the same logic applies to him.

- End of the period
 - In case 1, you are informed of your final score. You are not informed of the final score of the other participant.
 - In case 2, you are informed of the final score of the other participant. You are not informed of your final score.

In both cases, you are informed of your payoff, either 50 ECUs or 100 ECUs.

The next period starts automatically. You have to perform the same task and new pairs are formed.

Summary

In each period:

- You solve problems during 30 seconds.
- You are randomly matched with another participant.
- You are informed about the group of the other participant and about your subscore and the subscore of this other participant.
- You have to choose in your head a letter between A and K
- You are informed either about your random set of numbers or about the other participant's random set of numbers, depending on the period
- You report the random number which corresponds to the letter you previously chose in your head
- You are informed of your payoff and of your final score, or of the final score of the other participant, depending on the period

Just before the end of the session, there will be a last stage. The instructions will be displayed on your screen. At this time, do not hesitate to ask questions by raising your hand and/or by pressing the red button.

End of the session

At the end of the session, your screen will display the following information:

- Your earnings for part 1.
- The period and the case selected for the computation of your earnings in part 2 and your earnings for this part.
- The two periods selected in part 3 and your earnings for this part.
- Your earnings for the last stage.
- Your total earnings for the session.

Your earnings for the online part and the show-up fee will be added to the earnings of the session. The payment will be made by someone who does not know the content of the session and has no access to the program and the data.

Instructions for the Identity-No-Info treatment (translated from French)

The instructions for part 1, part 2 and for the end of the session are similar to the Identity treatment. They are omitted here.

Part 3

General description

This part consists of 16 periods in which you will have to perform a task. In each period your task consists in solving problems during 30 seconds. Each problem solved increases your “subscore” by one point.

In each period, you are randomly matched with another participant. Thus, it is very unlikely that you are matched with the same participant twice in a row.

At the end of the period, a participant in the pair earns 100 ECUs and the other earns 50 ECUs. Your payoff depends on your “final score” relative to the final score of the other participant. Your final score consists of two elements: your subscore plus a random number that you or the other participant have to report. If your final score is higher than the final score of the other participant, the higher is the difference between your final score and the other participant’s final score, the more likely you will get 100 ECUs.

At the end of the session, two periods among the 16 will be randomly selected by the program to determine your earnings in this part.

Description of each period

A period consists of 4 stages.

- Stage 1: the task

You have to perform a task during 30 seconds. This task consists in decoding letters into numbers. A conversion table is displayed permanently on your screen. After having decoded the letter displayed, you have to press the OK button to validate your answer. You are immediately informed of whether your answer is correct or not. If your answer is incorrect, you have to enter another number. The next problem will be displayed only after you have entered a correct answer. If your answer is correct, a new letter is displayed and this will continue for 30 seconds. The conversion table changes in each period. It is the same for all the participants. Each letter correctly encoded increases your subscore by 1 point.

- Stage 2: matching pairs and information

The program forms pairs randomly. You are informed of the other participant’s group (Klee or Kandinsky). Then, you are informed about your subscore and the subscore of the other participant. The other participant receives the same information as you.

- Stage 3: final score of the period

Your final score is determined by the sum of your subscore and the reported random number.

The program selects independently a random number for you and a random number for the other participant (with rebate). The random numbers are integers between -5 and 5. You are informed either of your random number or of the other participant’s random number. You have to memorize these numbers because you will be asked to report them on the following screen in order to determine the final score.

Depending on the period, one of the two following cases occurs:

- **Case 1:** you have to report your own random number. The other participant reports his own random number. You are not informed of the other participant's random number and he is not informed of your random number.
- **Case 2:** you have to report the random number of the other participant. The other participant has to report your random number. You are not informed of your random number and he is not informed of his random number.

Once the random numbers have been reported, the program computes your final score and the other participant's final score. The final scores are determined differently in the two cases.

In case 1, your final score is equal to your subscore (the number of problems solved) plus your random number as reported by yourself. The computation is the same for the other participant.

Case 1: Your final score = your subscore + your random number as reported by yourself
 Other participant's final score = his subscore + his random number as reported by himself

In case 2, your final score is equal to your subscore (the number of problems solved) plus your random number as reported by the other participant. The computation is the same for the other participant.

Case 2: Your final score = your subscore + your random number as reported by the other participant
 Other participant's final score = his subscore + his random number as reported by yourself

- Stage 4: Determination of the payoff of the period

The higher is your final score relative to the final score of the other participant, the more likely you will earn 100 ECUs. The lower is your final score relative to the final score of the other participant, the more likely you will earn 50 ECUs.

More precisely, your chance of winning 100 ECUs is determined as follows:

$$\text{Your chance of winning 100 ECUs} = \frac{\text{Your final score}}{\text{Your final score} + \text{other participant's final score}}$$

This formula indicates that for your given final score, if the other participant's final score increases, your chance to win 100 ECUs decreases. In contrast, for a given final score of the other participant, if your final score increases, your chance to win 100 ECUs also increases. Your chance to win depends on your final score but also on the final score of the other participant, and the same logic applies to him.

- End of the period

- In case 1, you are informed of your final score. You are not informed of the final score of the other participant.

- In case 2, you are informed of the final score of the other participant. You are not informed of your final score.

In both cases, you are informed of your payoff, either 50 ECUs or 100 ECUs.

The next period starts automatically. You have to perform the same task and new pairs are formed.

Summary

In each period:

- You solve problems during 30 seconds.
- You are randomly matched with another participant.
- You are informed about the group of the other participant and about your subscore and the subscore of this other participant.
- You are informed either of your random number or of the other participant's random number.
- You report one of the two random numbers, depending on the period.
- You are informed of your payoff and of your final score, or of the final score of the other participant, depending on the period.

Just before the end of the session, there will be a last stage. The instructions will be displayed on your screen. At this time, do not hesitate to ask questions by raising your hand and/or by pressing the red button.

End of the session

At the end of the session, your screen will display the following information:

- Your earnings for part 1.
- The period and the case selected for the computation of your earnings in part 2 and your earnings for this part.
- The two periods selected in part 3 and your earnings for this part.
- Your earnings for the last stage.
- Your total earnings for the session.

Your earnings for the online part and the show-up fee will be added to the earnings of the session. The payment will be made by someone who does not know the content of the session and has no access to the program and the data.

B Tables

Table B.1 – Characteristics of participants, by treatment

| Main treatments | | | | | |
|------------------------------|------------------------------|---------------------------|-----------------|------------------------|--|
| | No-Identity treatment (a) | Identity treatment (b) | Test | p-value (a) vs. (b) | |
| Average nb of participations | 4.17 (0.41) | 3.66 (0.27) | Unpaired t-test | 0.288 | |
| Average age | 25.11 (1.07) | 23.42 (0.51) | Mann-Whitney | 0.632 | |
| Pro-self in SVO test (%) | 59.37% | 50.78% | Proportion test | 0.260 | |
| Male (%) | 42.19% | 39.06% | Proportion test | 0.677 | |
| Student (%) | 71.87% | 80.47% | Proportion test | 0.178 | |
| Employed (%) | 14.06% | 12.50% | Proportion test | 0.762 | |
| Unemployed (%) | 9.37% | 4.69% | Proportion test | 0.206 | |
| Other (%) | 4.69% | 2.34% | Proportion test | 0.379 | |
| Number of subjects | 64 | 128 | | | |

Note: Standard deviations are in parentheses.

| Additional treatments | | | | | |
|------------------------------|--|---------------------------------------|------------------------|------------------------|------------------------|
| | Identity-No Scrutiny treatment (c) | Identity-No- Info treatment (d) | p-value (c) vs. (d) | p-value (c) vs. (b) | p-value (d) vs. (b) |
| Average nb of participations | 3.88 (0.35) | 3.2 (0.27) | 0.128 | 0.613 | 0.229 |
| Average age | 21.71 (0.41) | 21.18 (0.17) | 0.706 | <0.001 | <0.001 |
| Pro-self in SVO test (%) | 28.8% | 30.0% | 0.856 | <0.001 | 0.002 |
| Male (%) | 48.1% | 44.0% | 0.559 | 0.168 | 0.452 |
| Student (%) | 93.3% | 100.0 % | 0.008 | 0.005 | <0.001 |
| Employed (%) | 1.0% | . | . | . | . |
| Unemployed (%) | 4.8% | . | . | . | . |
| Other (%) | 1% | . | . | . | . |
| Number of subjects | 104 | 100 | | | |

Note: Standard deviations are in parentheses.

Table B.2 – Determinants of performance

| Dependent variable: Subscore | All (1) | No-Identity (2) | Identity (3) |
|---|----------------------|----------------------|----------------------|
| Identity treatment | 0.101 (0.322) | - | - |
| Opp-condition in t-1 | 0.177** (0.056) | 0.028 (0.089) | 0.230*** (0.070) |
| In-group opponent in t-1 | - | - | 0.023 (0.064) |
| Contest lost in t-1 | 0.117** (0.055) | 0.028 (0.089) | 0.136 (0.085) |
| Contest lost in t-1* Opp-condition in t-1 | -0.126 (0.077) | -0.045 (0.134) | -0.167* (0.094) |
| Contest lost in t-1* In-group opponent in t-1 | - | - | -0.017 (0.064) |
| Period | 0.301*** (0.025) | 0.229*** (0.047) | 0.329** (0.029) |
| Period squared | -0.012*** (0.001) | -0.008*** (0.002) | -0.013*** (0.001) |
| SVO angle | -0.005 (0.005) | -0.010 (0.007) | -0.002 (0.007) |
| Male | 0.256* (0.143) | 0.274 (0.231) | 0.229 (0.184) |
| Constant | 6.781*** (0.383) | 7.099*** (0.551) | 6.844*** (0.392) |
| Other individual characteristics | Yes | Yes | Yes |
| Session fixed effects | Yes | Yes | Yes |
| Number of observations | 2880 | 960 | 1920 |
| Number of clusters | 192 | 64 | 128 |
| Chi2 | 422.435 | 121.290 | 374.252 |
| p>Chi2 | <0.001 | <0.001 | <0.001 |

Notes: This Table reports marginal effects from random-effects GLS regressions with robust standard errors clustered at the subject level in parentheses. The dependent variable is the subscore in t . *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table B.2 reports an OLS regression analysis in which the dependent variable is the subscore in t in the real effort task. The three models pool the Self- and Opp-conditions since subjects were not aware of the condition when they performed the task (they were informed only after performing the task). Model (2) includes only the No-Identity

treatment and model (3) only the Identity treatment.

In model (1), the independent variables include dummy variables for the Identity treatment and for the Opp-condition in $t - 1$, taking the Self-condition and the No-Identity treatment as the reference categories. They also include a dummy variable indicating whether the subject lost or won the contest in $t - 1$ and an interaction term between the outcome of the competition in $t - 1$ and being in the Opp-condition in $t - 1$. We add a time trend and its squared value. We also include the subject's gender, SVO angle, age and number of past participations in a laboratory experiment. We add session fixed effects. In model (3) we also include an interaction term between the outcome of the competition in $t - 1$ and a dummy variable ("In-group opponent in $t - 1$ ") equal to 1 if the pair members shared the same group identity in $t - 1$ and 0 otherwise. We include lagged variables because when they perform the task subjects do not know yet which condition will apply in the period and, in the Identity treatment, whether they will be opposed by an in-group or by an out-group.

This Table shows that the treatment does not impact performance. Surprisingly, being in the Opp-condition in $t - 1$ is associated with a significantly higher subscore in the current period although conditions are assigned randomly in each period. This is driven by the Identity treatment. Winning the contest in $t-1$ increases performance in the current period in model (1), but this is endogenous since the probability of winning increases in the ability level. This variable is no longer significant in models (2) and (3). We observe also a non-linear effect of time in all models: the subscore increases over time but at a decreasing pace.

Table B.3 – Frequency and intensity of non-selfish misreporting, by treatment and condition

| Treatment | Condition | | Absolute relative frequency of misreports | Absolute relative frequency of full misreports | Average absolute intensity of misreports |
|--------------|-----------|--------------------|---|--|--|
| No- Identity | Self | All | 4/468 [0.85%] | 0/4 [-] | 2.75 (0.75) |
| | Opp | All | 27/463 [5.83%] | 6/27 [22.23%] | 3.70 (0.44) |
| Identity | Self | All | 7/926 [0.76%] | 2/7 [28.57%] | 2.57 (0.84) |
| | | In-Group opponent | 2/470 [0.43%] | 0/2 [-] | 2 (0.00) |
| | | Out-Group opponent | 5/456 [1.10%] | 2/5 [40.00%] | 2.8 (1.20) |
| | Opp | All | 26/927 [2.80%] | 7/26 [26.92%] | 3.69 (0.58) |
| | | In-Group opponent | 15/457 [3.28%] | 4/15 [26.67%] | 3.53 (0.80) |
| | | Out-Group opponent | 11/470 [2.34%] | 3/11 [27.27%] | 3.91 (0.88) |

Notes: We consider only the cases in which the subject overreports his opponent's number or underreports his own number; we exclude the cases in which the subject's number was equal to -5 in the Self-condition and the opponent's number was equal to 5 in the Opp-condition. A full misreport consists of reporting the lowest possible number for self or the highest possible number for the opponent. Percentages are into brackets. The average intensity of misreports is defined as the mean absolute value of the difference between the reported number and the actual number (with standard deviations in parentheses).

Table B.4 – Determinants of selfish misreporting in the Identity-No-Info treatment

| Dependent variable: Selfish misreport | All conditions | Opp-condition | Self-condition |
|--|----------------------|-------------------|-------------------|
| <i>Treatment</i> | Identity-No-Info | | |
| | (1) | (2) | (3) |
| In-group opponent | -0.026 (0.024) | -0.069 (1.011) | 0.004 (0.027) |
| Self-condition | 0.023 (0.022) | - | - |
| subscore _{<i>i,t</i>} | 0.007 (0.011) | 0.014 (0.153) | -0.003 (0.011) |
| Tie: subscore _{<i>i,t</i>} = subscore _{<i>j,t</i>} | -0.021 (0.035) | 0.016 (0.202) | 0.013 (0.054) |
| Max{0, subscore _{<i>i,t</i>} - subscore _{<i>j,t</i>} } | -0.015 (0.012) | -0.017 (0.258) | 0.002 (0.020) |
| Max{0, subscore _{<i>j,t</i>} - subscore _{<i>i,t</i>} } | -0.016 (0.015) | -0.006 (0.056) | -0.010 (0.025) |
| Random number _{<i>i,t</i>} | -0.006 (0.004) | - | -0.018 (0.021) |
| Random number Opp _{<i>i,t</i>} | 0.010** (0.004) | 0.012 (0.198) | - |
| Period | 0.065*** (0.023) | 0.058 (1.003) | 0.066 (0.075) |
| Period squared | -0.003*** (0.001) | -0.002 (0.041) | -0.003 (0.004) |
| SVO Angle | -0.003 (0.005) | -0.002 (0.224) | -0.003 (0.014) |
| Male | 0.188 (0.174) | 0.225 (11.148) | 0.134 (0.241) |
| Other individual characteristics | Yes | Yes | Yes |
| Number of observations | 1417 | 703 | 710 |
| Number of clusters | 100 | 99 | 100 |
| Pseudo-loglikelihood | -365.046 | -206.925 | -216.402 |
| Wald Chi2 | 77.232 | 92.399 | 28.211 |
| p>chi2 | <0.001 | <0.001 | 0.005 |

Notes: This Table reports marginal effects from random-effects logit regressions with robust standard errors clustered at the subject level in parentheses. In all models, the independent variables are the same as those included in Table 1.2, except the variables coding the difference between the subject's and his opponent's random numbers since subjects have no information about either their own (Opp-condition) or their opponent's number (Self-condition). The regressions exclude the observations in which the subject's own random number is 5 in the Self-condition and those in which the opponent's random number is -5 in the Opp-condition. In model (2) one subject has been excluded: this subject misreported non-selfishly in three periods and selfish misreporting was impossible in three other periods (the opponent's random number was equal to -5). *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table B.5 – Determinants of the reported random number in the Identity-No-Info and the Identity-No-Scrutiny treatments

| Dependent variable: Reported random number | Both treatments | Identity-No-Scrutiny treatment | | |
|--|---------------------|--------------------------------|---------------------|--------------------|
| | Both conditions | Both conditions | Opp-condition | Self-condition |
| | (1) | (2) | (3) | (4) |
| Identity-No-Scrutiny treatment | 0.356*** (0.138) | - | - | - |
| In-group opponent | -0.053 (0.035) | -0.038 (0.051) | 0.065 (0.063) | -0.039 (0.076) |
| Self-condition | 0.218*** (0.037) | 0.380*** (0.057) | - | - |
| Subscore _{<i>i,t</i>} | 0.027 (0.018) | 0.041 (0.026) | -0.017 (0.032) | 0.076** (0.038) |
| Tie: subscore _{<i>i,t</i>} = subscore _{<i>j,t</i>} | 0.006 (0.058) | 0.014 (0.081) | 0.009 (0.097) | 0.035 (0.128) |
| Max{0, subscore _{<i>i,t</i>} - subscore _{<i>j,t</i>} } | -0.043** (0.018) | -0.055** (0.023) | 0.107*** (0.028) | -0.015 (0.036) |
| Max{0, subscore _{<i>j,t</i>} - subscore _{<i>i,t</i>} } | -0.002 (0.020) | 0.007 (0.027) | 0.018 (0.034) | 0.029 (0.042) |
| Period | 0.027 (0.019) | 0.018 (0.027) | -0.091** (0.044) | -0.001 (0.037) |
| Period squared | -0.000 (0.001) | -0.000 (0.001) | 0.004 (0.002) | 0.000 (0.002) |
| SVO Angle | 0.004 (0.004) | 0.010** (0.005) | -0.011* (0.006) | 0.008 (0.005) |
| Male | 0.273** (0.137) | 0.145 (0.169) | -0.312 (0.205) | 0.011 (0.173) |
| Other individual characteristics | Yes | Yes | Yes | Yes |
| Number of observations | 3106 | 1664 | 832 | 832 |
| Wald Chi2 | 108.987 | 81.113 | 47.258 | 19.403 |
| p>chi2 | <0.001 | <0.001 | <0.001 | 0.054 |

Notes: This Table reports marginal effects from random-effects Tobit regressions with standard errors in parentheses. Model (1) pools the Self- and Opp-conditions and the Identity-No-Info and Identity-No-Scrutiny treatments. Models (2) to (4) include only the Identity-No-Scrutiny treatment with model (2) pooling both conditions, model (3) including only the Opp-condition, and model (5) only the Self-condition. To pool the reported numbers from both conditions in models (1) and (2), the dependent variable takes the value of the reported number in the Self-condition and the value of the reported number in the Opp-condition multiplied by -1. In contrast in model (3), the dependent variable takes the actual value (not multiplied by -1) of the number reported for the opponent. In all models, the independent variables are the same as those in Table 1.2, except for variables relative to the random numbers since in the Identity-No-Scrutiny treatment we ignore the true random numbers. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

C Figures

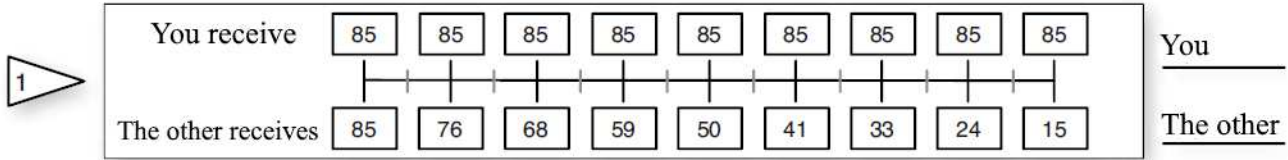


Figure C.1 – First decision in the SVO test

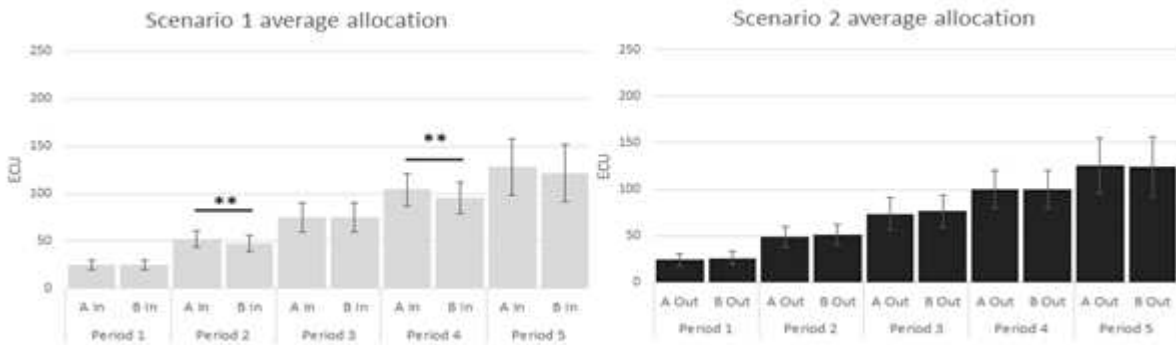


Figure C.2 – Average allocation in ECUs as a third party in scenario 1 (left panel) and scenario 2 (right panel), by period (stage 3 of the group identity induction procedure)

Notes: "In" stands for in-group and "Out" for out-group. In scenario 3, person A is an in-group member and B is an out-group member. *** indicate significance at the 1% level in two-sided pair-wise t-tests.

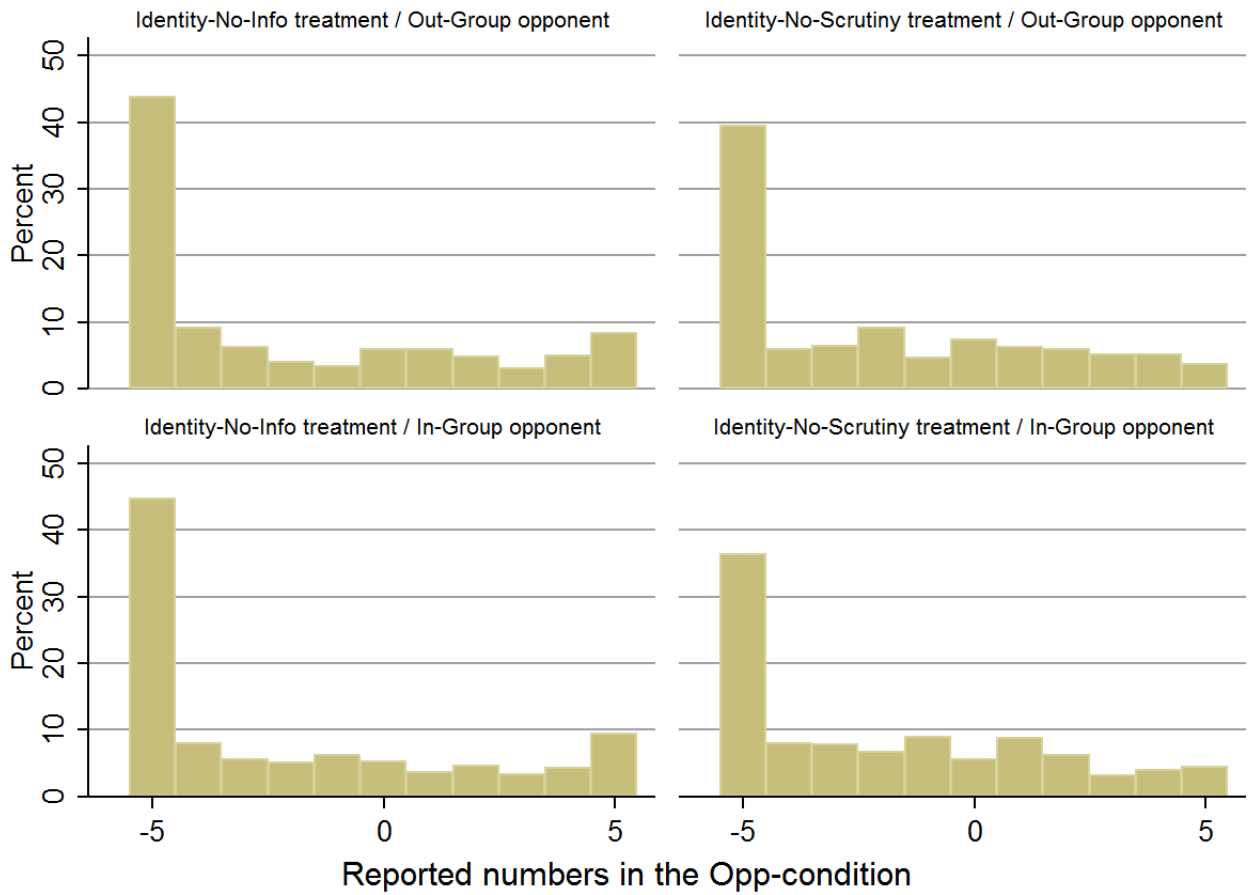


Figure C.3 – Distribution of the reported numbers in the Opp-condition, by treatment and identity matching

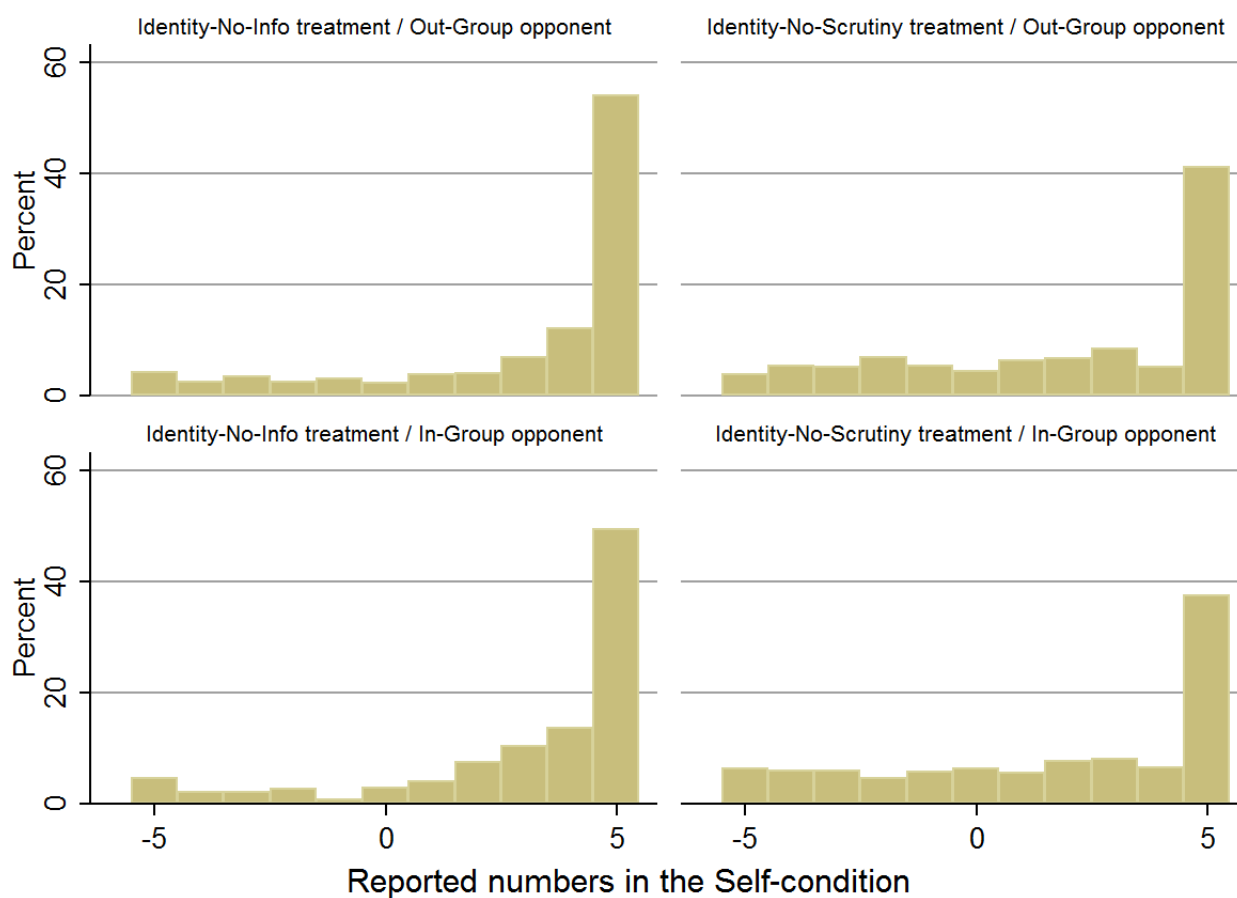


Figure C.4 – Distribution of the reported numbers in the Self-condition, by treatment and identity matching

Chapter 2

The impact of feedback and competitive payment schemes on dishonesty¹

2.1 Introduction

Dishonesty is a major problem in many spheres of the society. Fraudulent practices abound in business and in the financial sector but also in sports, education and everyday life. They are particularly detrimental for companies and the society since they increase transaction costs, reduce trust and undermine efficiency. The rise of competition between and within organizations is frequently evoked to explain the erosion of moral norms, as well as the diffusion of social information about others' immoral behavior which may give a bad example to individuals (e.g., Bowles, 2161).

Financial reward incentives are an important determinant of the level of competition within a firm. A common strategy in companies is to reward employees based on their performance (Prendergast, 1999), either by paying them an individual piece-rate or by adopting tournament incentive schemes. But while tournaments generally induce

¹This chapter is a joint work with Fabio Galeotti and Marie Claire Villeval

higher effort levels compared to piece-rates (e.g., Bull et al., 1987; Van Dijk et al., 2001), they also tend to trigger more cheating behavior (e.g., Cadsby et al., 2010; Schwieren and Weichselbaumer, 2010; Belot and Schröder, 2013; Faravelli et al., 2015; Rigdon and D’Esterre, 2015). Previous studies have also reported that feedback on peer cheating behavior influences individuals’ tendency to cheat in non-competitive settings (e.g., Gino et al., 2009; Innes and Mitra, 2013; Rauhut, 2013; see also Lefebvre et al., 2015, in the context of tax evasion). Information on others’ cheating behavior may lead people to update their beliefs about the existing social norm in the considered context and change their perception of what is socially acceptable. In particular, cheating may become more acceptable in one’s own eyes if one observes peers’ dishonest actions. Yet, and perhaps surprisingly, the impact of social information under competitive pay schemes on cheating behavior has attracted less attention, with the exception of sabotage.² Since individuals have a pecuniary interest to mimic their opponents’ cheating behavior in such competitive settings, this could increase even more cheating behavior compared to non-competitive settings. But feedback on others’ behavior also reduces the uncertainty about the relative performance of their rivals and this may lead to moderate cheating since the success of the competition does not depend on absolute performance levels but on outperforming the competitor. In this study we shed light on this unexplored area of intersection between incentives and social information by investigating how feedback on a counterpart’s performance influences individuals’ cheating behavior under competitive *versus* non-competitive pay schemes.

Previous experimental studies on the effects of incentives and feedback on dishonesty have also mainly focused on static settings where subjects make one-shot decisions or play over very few rounds. Thus, little is known about the dynamic effects of different pay schemes and feedback on cheating behavior. A recent meta-analysis of Abeler et al. (2019) suggests that dishonest behavior does not change over time. Is this also true under competitive pay schemes? How does the provision of social information affect

²Previous studies have shown that individuals informed of their opponents’ sabotaging behavior increase their own sabotage activities. This is motivated by retaliation (e.g., Harbring et al., 2007). However, cheating is different as it does not directly target an other individual.

cheating behavior over time? Our study contributes also to this by considering a more dynamic and realistic setting where subjects can adjust their behavior over time.

To address all the aforementioned research questions, we designed a laboratory experiment where subjects were matched in pairs and played individually, and repeatedly over many periods, a variant of the die-under-the-cup task (Shalvi et al., 2011; Fischbacher and Föllmi-Heusi, 2013). In each period, subjects had to roll a die in private with no scrutiny and to report the outcome in their computer. Their final performance was computed as the sum of their individual reports in all periods.³ We varied between-subjects whether participants were paid according to a piece-rate or a tournament payment scheme, and whether they received a continuous feedback about their counterpart's reports at the end of each period or only at the end of the task. Feedback provides a noisy signal about the honesty of the other player since the reports of several high values of the die outcomes may indicate luck or dishonest reporting.

We also designed a variant of the tournament condition where only one pair member could physically roll the die and cheat by misreporting the outcome. The other player could only roll an electronic die and the outcome was automatically reported by the computer. This allows us to test whether the difference in cheating between the piece-rate and the tournament payment scheme, if any, is driven by the competitive nature of the incentive or by the individuals' beliefs about the dishonest behavior of the rival. If previous differences disappear in the new treatment, this would indicate that dishonest reporting in standard tournaments is driven by the fear of losing an unfair competition and by the fierceness of the competition. If not, this would suggest that the competitive pressure itself generates misbehavior regardless of the counterpart's behavior, for monetary and possibly also non-monetary reasons (e.g., Parco et al., 2005; Sheremeta, 2010; Cason et al., 2018). To test the effect that feedback on relative performance has in this one-sided cheating setting, we also varied whether the feedback on the relative performance is continuous or given only at the end of the competition.

³We purposely chose a task where the actual performance is determined by luck and dishonesty instead of a real effort to ensure that our results are not influenced by participants' abilities.

Our results show that participants over-reported the die outcomes in all treatments. Tournament incentives increased over-reporting compared to the individual piece-rate scheme but the difference is virtually nullified when the opponent cannot cheat. Continuous feedback on the counterpart's reports increased dishonesty compared to a final feedback under the piece-rate pay scheme but not in the tournaments. Despite the latter result, participants reacted to the perceived level of honesty of their counterpart both in tournaments and in the piece-rate pay scheme. Finally, we find, unexpectedly, that males over-reported more than females only in the tournament settings without continuous feedback. This suggests that when there is uncertainty about the opponent's performance (and the related degree of (dis)honesty), males react by increasing their own dishonesty, but not females. We can rule out that this difference is only due to the higher willingness to compete of males.

The next section reviews the related literature. Section 3 describes our experimental design and procedures. Section 4 reports our results and finally, Section 5 discusses these results and concludes.

2.2 Related Literature

Our study contributes to three strands of the literature. The first one concerns the impact of relative performance feedback policies on performance under different pay schemes. An effect of feedback on performance has been found in settings where dishonesty is ruled out by design. For example, Falk and Ichino (2006) and Mas and Moretti (2009) found strong positive peer effects under a flat wage scheme when employees can observe other employees working. In a field experiment conducted in schools, Azmat and Iriberry (2010) found that providing feedback about schoolmates' performance increased pupils' grades. In the context of networks where people are paid a piece-rate Beugnot et al. (2019) identified a strong gender difference in peer effects on productivity, depending on the structure of the network. Eriksson et al. (2009) compared the effects of feedback under a piece rate and a tournament. They found some evidence of negative peer effects

on the quality of work in tournaments, even if low performers rarely quit competition even when feedback indicated that they would lose for sure. Other studies found mixed evidence toward an effect of feedback on performance. Fershtman and Gneezy (2011) found that feedback does not affect effort in tournaments. Kuhnen and Tymula (2009) and Ludwig and Lünser (2012) studied different information structures in tournaments and found heterogeneous effects of feedback. In contrast to these studies, we consider the impact of feedback on honest behavior.

Our study is also related to the literature on the relationship between competition and unethical behavior. A number of experimental studies report that dishonest behavior occurs more frequently in competitive than non-competitive settings, with some variations in the size, stability and distribution of the effect (Cadsby et al., 2010; Schwieren and Weichselbaumer, 2010; Belot and Schröder, 2013; Faravelli et al., 2015; Rigdon and D’Esterre, 2015; Aydogan et al., 2017).⁴ This literature, however, does not explore what is the role played by the information about the opponent’s performance to determine unethical behavior. Our contribution is to identify the impact of feedback about the counterpart’s relative performance on cheating and to investigate whether this effect depends on the incentive scheme and for what reason.

Finally, there is a growing literature on peer effects and dishonesty that is relevant to our study. In a tax evasion experiment, Fortin et al. (2007) found no evidence that people imitate the evading behavior of others but identified strong exogenous peer effects focused on income differences. In a different environment, Lefebvre et al. (2015) showed that tax evasion increases if subjects are informed about the low tax compliance of participants from a previous session, while showing high tax compliance has no effect on behavior. Using a tossing coin task, Fosgaard et al. (2013) observed that men cheat

⁴We focus on the literature on tournaments and contests as this is the main target of our study. Other studies investigated the effect of market competition on unethical behavior (e.g., Falk and Szech, 2013; Rabanal and Rud, 2018; Feltovich, 2018), and the spillover effects of competition on unethical behavior (e.g., Buser and Dreber, 2016; Schurr and Ritov, 2016; Banerjee et al., 2018). It is also worth noting the existence of earlier work showing that competition increases sabotage (e.g., Lazear, 1989; Konrad, 2000; Harbring et al., 2007; Carpenter et al., 2010; Harbring and Irlenbusch, 2011). Finally, some studies looked at the link between the strength of the competition and the occurrence of cheating (Cartwright and Menezes, 2014; Conrads et al., 2013; Feltovich, 2018).

significantly more when they are told that others have cheated. In a die task, Diekmann et al. (2015) found that people lie more if they are informed about the lying behavior of others. When the die task is repeated, Rauhut (2013) showed that subjects who are told about others' reports lie more (less) when they underestimated (overestimated) others' lying behavior compared to uninformed subjects. Similar results are reported in Soraperra et al. (2017) and Lauer and Untertrifaller (2019), and by Innes and Mitra (2013) in the context of a sender-receiver game. Finally, Gino et al. (2009) found that cheating is contagious in a real-effort experiment in which participants could misreport their performance. None of these studies investigated the potential contagion effect of dishonesty in a tournament setting or evaluated the interaction between incentives and feedback, which is the main aim of our study.⁵

2.3 Experimental Design and Procedures

In this section, we first introduce our design before detailing the procedures.

2.3.1 Design

Our experiment is based on a variant of the die-under-cup task (Shalvi et al., 2011; Fischbacher and Föllmi-Heusi, 2013) repeated for 24 periods. We used a six-faced die with three colors rather than numbers to increase the statistical power of our analyzes, like in Dai et al. (2018). The die roll can give three possible outcomes (red, yellow, blue) with equal probability (1/3). The die was inside a sealed cup to increase privacy, as in Shalvi et al. (2011). In each period, participants were requested to roll the die and to report the outcome. The number of points earned for the period was 0, 1 or 2 if they reported a blue, a yellow or a red face, respectively. Participants were allowed to roll the die more than once but they were explicitly told to report only the outcome

⁵Other studies have investigated dishonesty in groups in which individuals are potentially informed about their teammates' reports. Unlike ours, these studies also vary at the same other dimensions such as the possibility to communicate (e.g., Kocher et al., 2017) or the payoffs commonality (e.g., Weisel and Shalvi, 2015; Kocher et al., 2017).

of the first roll. Participants' performance in this task was computed as the sum of points earned across the 24 periods. We implemented a 2x2 between-subject design. In all treatments participants were matched in fixed pair for the whole duration of the experiment. The first dimension manipulated the payment scheme: either participants were paid based on a piece rate scheme or they had to enter a two-player tournament. The second dimension varied whether participants were informed continuously (*i.e.* at the end of each period) or only at the end of the 24 periods about the reports of their counterpart in each period. Finally, we added a variant to the tournament treatment in which one participant in each pair could not misreport. This variant was motivated by the identification of the role of beliefs about the counterpart's dishonesty.

2.3.1.1 Payment schemes

One dimension that we manipulated was how participants' payoffs were determined. In the *Piece Rate* treatment (PR, hereafter), reporting a blue outcome paid 0 point, a yellow outcome paid 1 point and a red outcome paid 2 points. Each point was worth €0.18. Participants' total payoff in this task was determined by the sum of the points accumulated throughout the 24 periods, regardless of the counterpart's report.

In the *Tournament* treatment (TR, hereafter), the pair member with the highest number points accumulated at the end of the 24 periods earned €0.36 per point while the counterpart earned €0. As such, the expected payoffs were the same in the two payment schemes.

2.3.1.2 Information

The second dimension that we manipulated was the nature of the feedback received by the participants about their counterpart's reports.

In the *Final Feedback* treatment (FF, hereafter), participants were only informed at the end of the experiment about the reports made by their counterpart in the 24 periods. Thus, in this treatment there is uncertainty about one's relative performance and about the honesty of the counterpart until the end of the experiment. Providing detailed

information about the counterpart's reports at the end of the experiment allowed us to hold constant the observability of the participants' reports made by their counterpart across treatments.

In the *Continuous Feedback* treatment (CF, hereafter), participants were informed at the end of each period about the report made by their counterpart in that period. A history box with the participant's and the counterpart's reports was available in the computer screen of the participants. Thus, in this treatment, participants were completely informed about the difference in performance between themselves and their counterpart, and its evolution over time. This feedback provided also a noisy information about the honesty of the counterpart. Over time, participants could perceive whether the counterpart was dishonest or not by comparing the distribution of reports with the theoretical uniform distribution, but they could never be certain of whether or not a report of the counterpart in a given period was honest.

2.3.1.3 Reporting

The third dimension that we varied was whether only one or both participants in each pair had to physically roll a die and report the outcome. In the *One-Sided* tournament treatment (TR1, hereafter), only one pair member was requested to roll the die and report the outcome. The other competitor was passive. She or he was shown a video of a die roll and the outcome was automatically reported by the computer. The role of the passive player was randomly assigned to one member in each pair before the first period and roles were kept fixed throughout the 24 periods. Thus, in this condition one competitor's performance in the pair was publicly known to be true. This manipulation was implemented only for the tournament payment scheme. Indeed, if dishonesty is more widespread in the tournament condition than in the piece-rate condition, it might be because of the competitive nature of the tournament scheme *per se* or because people want to restore the fairness of the competition if they anticipate that their counterpart is cheating. This additional treatment manipulation allowed us to mute the second channel. Table 2.3.1 summarizes our treatments.

Table 2.3.1 – Treatment dimensions and conditions

| | Piece-Rate (PR) | Tournament (Two-Sided) (TR) | Tournament (One-Sided) (TR1) |
|--------------------------|--------------------|--------------------------------|---------------------------------|
| Final Feedback (NF) | PR-FF | TR-FF | TR1-FF |
| Continuous Feedback (CF) | PR-CF | TR-CF | TR1-CF |

2.3.1.4 Additional measures

Both at the beginning (before being informed about the main task) and at the end of the 24 periods, we asked participants to rate their level of happiness and emotional arousal.⁶ Before the main task, we also elicited the participants' risk attitude using the procedure of Gneezy and Potters (1997). Participants were endowed with 150 euro cents and asked to choose how much to invest in a lottery. This lottery had two possible outcomes with equal probability of being selected. In case of a failure, participants lost their investment and in case of a success, the amount invested was multiplied by 2.5. Since a risk neutral participant should invest all of this endowment, the amount of the endowment not invested captures the degree of risk aversion.

Finally, at the end of the experiment and after the second round of questions about emotions, we elicited the participants' beliefs about the others' reports. Participants had to state how many participants out of 24 (excluding themselves and their counterpart) reported a red, a yellow and a blue outcome in a given period.⁷ We elicited the participants' beliefs for the first, the thirteenth and the last period. The objective of this task was to understand the participants' perception of the empirical norm in the group and how this evolved over time. One period and one color were randomly selected for payment. Participants received €1.5 for a correct prediction, €1 if their prediction differed within plus or minus 1 from the actual number, €0.5 if it differed within plus or minus 2, and €0 otherwise.

⁶Appendix A reports the instructions and the illustrations used to report happiness and nervousness.

⁷Participants were told that the 24 participants could be from the current session and/or a previous one.

2.3.2 Procedures

We conducted the experiment at the GATE-Lab in Lyon (France). 356 subjects were recruited using HROOT (Bock et al., 2014). 89.9% were students from local engineering, business, and medical schools. We conducted 2 sessions for each treatment except for the TR-CF and PR-CF treatments for which we conducted 3 sessions. More sessions were needed for these treatments, as the independent observation is the pair and not the participant. Table B.1 in Appendix B summarizes the participants' characteristics. The experiment was programmed with Z-tree (Fischbacher, 2007). The instructions were directly displayed on the computer screen of the subjects.⁸ Participants received the instructions for the die task only once the mood and the risk elicitation tasks were completed. The instructions used neutral language and clarification questions were answered in private. Each session lasted approximately 45 minutes. Mean earnings were €12.92 (SD=5.83), including a €5 show-up fee.

2.4 Behavioral Conjectures

A significant amount of evidence shows that individuals suffer from a finite moral cost of lying (e.g., Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013; Kajackaite and Gneezy, 2017; Abeler et al., 2019). Part of this cost can be explained by a potential loss of reputation from being perceived as a cheater by others or even in one's own eyes (Khalmetski and Sliwka, 2017; Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018). When subjects' actions cannot be scrutinized by the experimenter or by other participants, people tend to lie partially instead of fully. In our experiment, scrutiny by the experimenter and by the counterpart (either continuously or at the end) is limited. Indeed, nobody knows what the truth is in each period and, over time, the distribution of reports only gives a probability on whether the individual lied or not. Therefore, we expect a significant fraction of the participants to lie, but mostly partially (*i.e.* since there are only three possible outcomes we anticipate that they will not lie in every period). We

⁸See Appendix A for a translated version of the instructions.

state our first conjecture as follows:

Conjecture 1 (*Lying behavior*): Participants over-report their actual outcomes, but not systematically.

Based on the aforementioned literature and because payoffs depend on the ability to outperform the competitor, we expect that individuals who face competitive incentives are more likely to behave dishonestly than subjects who are paid an individual piece rate. Therefore, we anticipate more dishonesty in the tournament than in the piece rate payment scheme. Furthermore, we expect that when the opponent cannot cheat in the tournament participants are less likely to over-report their own outcome than when their opponent can cheat. If participants suffer from a cost of lying, they have no interest to misreport more than necessary when they know that their opponent cannot lie. We thus expect them to cheat just enough to win the tournament. However, as the marginal benefit of lying is larger in the tournament than in the piece rate scheme, conditionally on winning the tournament, we expect that participants are more likely to over-report in the tournament with a passive opponent compared to participants in the individual piece rate payment scheme. Our second conjecture is as follows:

Conjecture 2 (*Competition*): (a) Participants lie more under a tournament payment scheme than an individual piece rate scheme; (b) They are less likely to lie when their opponent is not allowed to lie compared to when he has such ability; (c) But they are more likely to lie in the one-sided tournament than in the piece rate payment scheme.

Previous studies reported a behavioral contagion effect of dishonesty in settings that are similar to our individual piece rate environment with continuous feedback (*e.g.*, Gino et al., 2009). If there are asymmetric peer effects driven by conformism or rivalry (bad examples have more influence than good ones (*e.g.*, Lefebvre et al., 2015; Dimant, 2018)), we expect to observe that participants who are paid an individual piece rate are more likely to be dishonest when continuously informed about their counterpart's reports. Like in the individual piece rate scheme, feedback on the opponent's reports in the tournament discourages honesty if these reports reveal a likely dishonest behavior.

In addition to a contagion effect, such feedback bears a strategic value in this environment. Unlike in the individual piece rate setting, if one does not mimic the dishonest behavior of the opponent, she or he will lose the tournament and earn nothing. Hence, when people are continuously informed about their opponent's performance they are expected to lie to stick to their opponent's performance. People may lie also if they are not continuously informed about their opponent's performance but they believe that they are matched with a liar and feel they are lagging behind. Whether final feedback in the tournament compared to continuous feedback increases more or less dishonesty depends on the beliefs of the players on the degree of dishonesty of their counterpart. Finally, when it is common knowledge that one's opponent cannot cheat, we expect no difference in the effect of continuous vs. final feedback. This hypothesis holds only for non risk-averse participants as a risk averse one would lie more in the final than in the continuous feedback one-sided tournament. Our third conjecture is as follows:

Conjecture 3 (Social information): (a) In the individual piece rate scheme, on average, participants lie more when they receive continuous feedback than when they do not; in the tournament, the difference depends on the players' beliefs about the dishonesty of their counterpart. (b) In one-sided tournaments, both continuous and final feedback have the same effect on misreporting; (c) In the tournament, participants are more likely to adjust their level of cheating to the reports of their counterpart compared to the individual piece rate scheme.

2.5 Results

We start by presenting our results on reporting behavior in the different treatments. Next, we analyze the dynamics of reporting behavior over time. Finally, we report gender differences in reporting behavior.

2.5.1 Lying Behavior Across Treatments

Our first result is in line with the previous literature:

Result 1 (*Lying behavior*): In all treatments, a significant fraction of the participants over-reported the highest outcome (red) and under-reported the lowest outcome (blue) but not in every period, indicating partial lying. They also under-reported the intermediate outcome (yellow) in the TR and TR1-FF treatments.

Table 2.5.1 – Reports by treatment

| Treatment | Blue outcome reported | Yellow outcome reported | Red outcome reported | Average value of reports | Full extent lying |
|-----------|-----------------------|-------------------------|----------------------|--------------------------|-------------------|
| PR-FF | 23.7% | 30.9% | 45.4% | 1.22 (0.05) | 4.8% |
| PR-CF | 17.4% | 31.3% | 51.4% | 1.34 (0.05) | 1.4% |
| TR-FF | 11.6% | 21.8% | 66.7% | 1.55 (0.06) | 27.3% |
| TR-CF | 12.8% | 23.7% | 63.5% | 1.51 (0.05) | 10.2% |
| TR1-FF | 20.1% | 27.4% | 52.4% | 1.32 (0.06) | 6.9% |
| TR1-CF | 18.5% | 30.5% | 50.9% | 1.32 (0.05) | 9.7% |
| Total | 16.6% | 27.4% | 56.1% | 1.38 (0.02) | 9.4% |

Notes: PR stands for Piece Rate; TR for Tournament, and TR1 for One-sided Tournament. CF stands for Continuous Feedback and FF for Final Feedback. Standard errors are in parentheses. The first three columns report the percentages of blue (low outcome), yellow (medium outcome) and red (high outcome) colors reported over the 24 periods. The fourth column reports the average value of reports over the 24 periods. A reported red outcome pays 2 points, a yellow one pays 1 point and a blue one pays 0 point. In the FF treatments, the average is computed at the participant level. In the CF treatments, it is computed at the pair level. The last column reports the percentage of participants who reported the maximum number of points (48 points) in the task (full extent lying). Passive participants who could not lie in the TR1 treatments are excluded from the statistics reported in this table.

Support to Result 1. Table 2.5.1 reports, for each treatment, the frequency of each reported outcome, the average value of the reported outcomes and the frequency of subjects who lied to the full extent. If participants were reporting the outcome of their rolls honestly, we should observe that each outcome is reported 33% of the time. Table 2.5.1 shows that participants reported a red outcome more than 33% of the time in all treatments (between 45.4% and 66.7%) and they reported a blue or a yellow outcome less than 33% of the time in all treatments (blue: between 11.6% and 23.7%; yellow: between 21% and 30.9%). Two-sided Wilcoxon signed rank tests⁹ comparing the actual and expected reports for each outcome shows that participants significantly over-reported the red outcome and under-reported the blue outcome in all treatments ($p < 0.001$ in all

⁹One could argue that binomial tests would be more suited but in our CF treatments the independent observation is at the pair level, which calls for an analysis at the pair level and prevents us to use binomial tests. In the FF treatments, binomial tests give similar results than Wilcoxon signed-rank tests.

treatments).¹⁰ They also significantly under-reported the yellow outcome but only in the TR and the TR1-FF treatments ($p < 0.001$ in both TR treatments and $p = 0.022$ in the TR1-FF treatment).

In all treatments, the average reported value (between 1.22 and 1.55 points) is higher than the expected value of 1 under the assumption of honest reporting. Two-sided Wilcoxon signed rank tests comparing the average reported performance and the expected performance show that participants over-reported significantly in all treatments ($p < 0.001$ in all treatments). Most participants did not, however, lie to the full extent. The percentage of participants who reported 48 points in total (*i.e.* the maximum number of points that one could report over the 24 periods) is lower than 10%, except in the TR-CF and TR-FF treatments where it reaches 10.2 % and 27.3%, respectively.

To assess the occurrence of lying throughout the experiment, we split the 24 periods in blocks of 4 periods.¹¹ We find that lying occurred from the first block up to the last one in all treatments ($p < 0.05$), with only one exception only marginally significant (the PR-FF treatment in the third block: $p = 0.086$). These findings validate our Conjecture 1.

We now state our second result.

Result 2 (Competition): **a)** Participants over-reported more in the TR than in the PR treatments. **b)** They over-reported more in the TR than in the TR1 treatments where competition was less fierce since it was public knowledge that the opponent could not lie. **c)** They over-reported marginally more in the TR1 than in the PR treatments.

Support to Result 2. Table 2.5.2 reports the p -value of a Dunn test.¹² It shows that participants over-reported significantly more in the two TR treatments than in any other treatment, as predicted by Conjecture 2a. This holds regardless of whether feedback was

¹⁰In this section, all statistical analyses are done at the pair level for the CF treatments and at the participant level for the FF treatment. Participants who were passive in the TR1 treatments are always excluded from the analysis.

¹¹Figure C.1 in Appendix C displays the average value of the reports across blocks of four periods, by treatment. Splitting the data into blocks of four periods avoids the fluctuations of a period-by-period analysis but it is arbitrary. As a robustness test, we conducted a similar analysis with blocks of 2 and 6 periods and the results are similar.

¹²This test is based on a two-sided Kruskal-Wallis test and computes multiple pairwise comparisons on the values of the reports between all treatments.

continuous or final (TR-CF *vs.* PR-FF/PR-CF: $p < 0.001$ and $p = 0.019$, respectively; TR-FF *vs.* PR-FF/PR-CF: $p < 0.001$ and $p = 0.009$, respectively). This is in line with previous research showing that competitive incentives increase dishonesty. The values reported are also significantly higher in the TR than in the TR1 treatments (TR-CF *vs.* TR1-FF/TR1-CF: $p = 0.015$ and $p = 0.014$, respectively; TR-FF *vs.* TR1-FF/TR1-CF: $p = 0.006$ for both comparisons). This shows that individuals are more honest in a tournament when they know that the opponent behaves honestly, as predicted by Conjecture 2b. In a competition, honesty is conditional on one's opponent's honesty since it determines the strength of the competition.

Table 2.5.2 – Dunn test: Pairwise comparisons of the value of the reports between treatments

| | PR-FF | PR-CF | TR-FF | TR-CF | TR-FF |
|--------|------------------|--------------|--------------|--------------|-------|
| PR-CF | 0.035 | - | - | - | - |
| TR-FF | <0.001 | 0.008 | - | - | - |
| TR-CF | <0.001 | 0.019 | 0.383 | - | - |
| TR1-FF | 0.068 | 0.419 | 0.006 | 0.015 | - |
| TR1-CF | 0.062 | 0.424 | 0.006 | 0.014 | 0.494 |

Notes: This Table reports p -values of a Dunn test comparing the average value of reports between all treatments. For the NF treatments the independent observations are at the participant level, while for the CF treatments the independent observations are at the pair level. Participants that could not misreport in the TR1 treatments are excluded from the analysis.

The results also show that compared to piece rate incentives, tournament incentives induce *per se* more lying but only marginally so and not systematically. This supports only moderately Conjecture 2c. Indeed, the values reported in the TR1 treatments are only marginally significantly different from those in PR-FF (TR1-CF *vs.* PR-FF: $p = 0.062$ and TR1-FF *vs.* PR-FF: $p = 0.068$). These values are not significantly higher in the TR1 treatments compared to the PR-CF treatment (TR1-CF *vs.* PR-FF: $p = 0.424$ and TR1-FF *vs.* PR-FF: $p = 0.420$). In competitive settings, believing that the competition may be unfair motivates one's own dishonesty; muting the possibility for the opponent to cheat

reinforces moral conduct.

We now state our third result.

Result 3: Participants lied significantly more when they received continuous feedback compared to final feedback, but only in the PR treatments.

Support to Result 3. In the PR treatments participants were more dishonest when they received continuous rather than final feedback about their counterpart's reports (PR-FF *vs.* PR-CF: $p = 0.035$). This evidence of peer effects – that supports Conjecture 3a – can be driven by the repeated temptation, by some behavioral contagion or by rivalry if people care about their relative status. In contrast, being informed continuously of the opponent's reports affects misreporting neither in the TR nor in the TR1 treatments (TR-FF *vs.* TR-CF: $p = 0.383$; TR1-FF *vs.* TR1-CF: $p = 0.494$). For the TR1 treatments, this result, which is consistent with Conjecture 3b, is not surprising since the informational content of the sequence of reports in the 24 periods does not differ when feedback is provided after each period or at the end of the 24 periods. This finding also suggests that observing repeatedly a counterpart that does not lie does not make people more honest. This is consistent with the asymmetric peer effects found in Lefebvre et al. (2015)) but it could also be due to the fact that truth-reporting is in our case compulsory and does not reveal anything of the intrinsic honesty of the counterpart.

To check the robustness of these results we ran a multinomial logit regression in which the dependent variable is the outcome reported by participant i in period t . The independent variables include treatment dummies with the PR-FF treatment taken as the reference, a time trend, a dummy variable for male participants, the participant's risk attitude and age, a dummy coding whether the participant is a student or not, and a fixed-effect for each group of sessions ran during the same months.¹³ The marginal effects are reported in Table B.2 in Appendix B. Table 2.5.3 reports the results of multiple

¹³This variable controls for the fact that the experimental sessions have been run during three distinct months.

pairwise comparisons of the marginal effects between all treatments computed from the aforementioned multinomial logit model.

Table 2.5.3 shows that participants were more likely to report a red outcome and less likely to report a yellow or a blue outcomes in the TR treatments than in the PR treatments. Furthermore, reports differed between the TR-FF and the TR1 treatment but they were not sensitive to feedback in the TR and TR1 treatments. These results are consistent with the ones obtained from non-parametric tests. However, reports did not differ significantly between the TR-CF and the TR1 treatments, although the sign of the coefficients are in line with the non-parametric tests. Also, only reports of the blue outcome differed between the PR-FF and the PR-CF treatments.

Overall, these results support Conjectures 2a, 2b, and 3b; they only partially support Conjectures 2c and 3a and they reject Conjecture 3c.¹⁴

¹⁴Conjecture 3c predicts that the effect of feedback should be larger in the TR-CF than in the PR-CF treatments. As an additional test, we compared the within-pair standard deviation between the PR-CF and TR-CF treatments. If the conjecture is correct we should see a lower within-pair standard deviation in TR-CF. In contrast to this conjecture, a two-sided Wilcoxon rank sum test rejects that the within-pair average standard deviation was significantly different between the two treatments ($p = 0.217$).

Table 2.5.3 – Pairwise comparisons of treatments for all outcomes

| Blue outcome (0 point) | | | | | |
|--------------------------|-----------------------------|-----------------------------|----------------------------|-------------------|-------------------|
| | PR-FF | PR-CF | TR-FF | TR-CF | TR1-FF |
| PR-CF | -0.048* (0.025) | - | - | - | - |
| TR-FF | -0.113*** (0.026) | -0.065*** (0.024) | - | - | - |
| TR-CF | -0.094*** (0.026) | -0.046* (0.024) | 0.019 (0.025) | - | - |
| TR1-FF | -0.044 (0.032) | 0.005 (0.031) | 0.070** (0.031) | 0.051 (0.031) | - |
| TR1-CF | -0.061** (0.027) | -0.013 (0.026) | 0.052** (0.026) | 0.033 (0.026) | -0.018 (0.029) |
| Yellow outcome (1 point) | | | | | |
| PR-CF | 0.020 (0.027) | - | - | - | - |
| TR-FF | -0.084*** (0.030) | -0.104*** (0.031) | - | - | - |
| TR-CF | -0.059** (0.029) | -0.079*** (0.029) | 0.025 (0.032) | - | - |
| TR1-FF | -0.044 (0.029) | -0.064** (0.031) | 0.040 (0.032) | 0.015 (0.032) | - |
| TR1-CF | -0.018 (0.032) | -0.038 (0.034) | 0.066* (0.034) | 0.041 (0.035) | 0.026 (0.032) |
| Red outcome (2 points) | | | | | |
| PR-CF | 0.028 (0.043) | - | - | - | - |
| TR-FF | 0.198*** (0.048) | 0.169*** (0.050) | - | - | - |
| TR-CF | 0.153*** (0.046) | 0.125*** (0.048) | -0.044 (0.052) | - | - |
| TR1-FF | 0.088* (0.050) | 0.059 (0.055) | -0.110** (0.056) | -0.066 (0.055) | - |
| TR1-CF | 0.079* (0.047) | 0.051 (0.052) | -0.118** (0.053) | -0.074 (0.052) | -0.008 (0.053) |

Notes: The Table reports pairwise comparisons of the marginal effects between all treatments for each outcome. Comparisons are computed from the model reported in Table B.2. Robust standard errors clustered at the pair (CF treatments) or individual level (FF treatments) are in parentheses. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

2.5.2 Dynamics of Lying Behavior Over Time

We now introduce our fourth result.

Result 4: The likelihood to over-report the highest outcome increased over time in the TR-CF and the PR-CF treatments while lying behavior was more stable in the other treatments.

Support to Result 4. Figure 2.5.1 reports the marginal effects of the period variable on the likelihood to report a blue, a yellow or a red outcome in each treatment. These marginal effects are computed based on the multinomial logit regression model reported in Table B.2 in Appendix B. In both the PR-CF and the TR-CF treatments we observe a significant increase of reports of a red outcome over time ($p = 0.012$ and $p = 0.001$, respectively). Reports of the yellow outcome decrease significantly over time only in the TR-FF treatment ($p = 0.001$). This reveals an escalation effect in the presence of social information.¹⁵ Not surprisingly, in the TR1-FF treatment we do not observe any effect of time. The marginal effect of the period is significant for neither outcomes ($p = 0.321$, $p = 0.560$ and $p = 0.682$ for the blue, yellow and red outcome, respectively).

To support the idea that the escalation of lying in the PR-CF and TR-CF treatments results from social interactions,¹⁶ we follow the strategy used in Falk and Ichino (2006) and replicated in Rosaz et al. (2016). We calculate the within-pair and between-pair standard deviations (S.D. hereafter) of the participants' total value of reports at the end of the 24 periods. A comparison of the within-pair and between-pair S.D. between the CF and the FF treatments allows us to determine the impact of social interactions in the PR-CF and TR-CF treatments. If social interactions generate imitation and/or rivalry, the within-pair S.D. should be lower and the between-pair S.D. should be higher in the CF treatments than in the corresponding FF treatments. Finally, the difference between

¹⁵In the PR-FF treatment, participants were less likely to report a yellow outcome over time ($p = 0.040$), but not more likely to report a red outcome over time ($p = 0.239$). Thus, we cannot conclude that they lied more over time in this treatment.

¹⁶One may argue that these results can simply reflect a decrease of the moral cost of lying over repetitions (e.g., Garrett et al., 2016). However, such an effect should be independent of the presence of continuous feedback and we do not find a significant effect of time in the FF treatments, at the exception of PR-FF discussed in footnote 15.

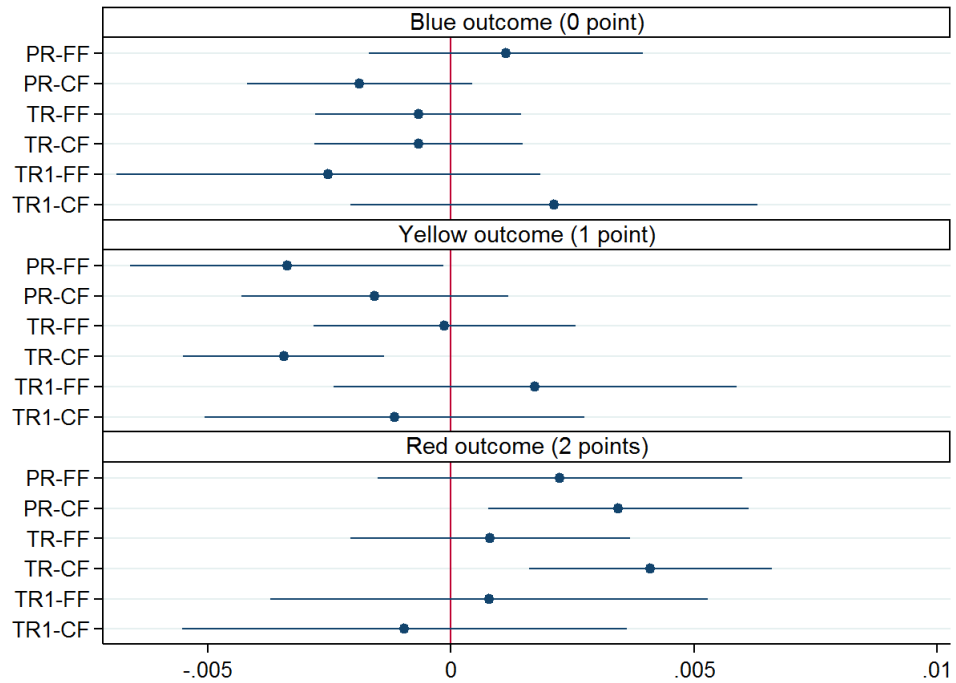


Figure 2.5.1 – Marginal effects of time on the probability to report each outcome, by treatment. *Notes:* The horizontal lines represent 95% confidence intervals.

the between-pair S.D. and the within-pair S.D. should be larger in the CF treatments than in the corresponding FF treatments.

The average within-pair S.D. is equal to 3.692 in the PR-CF treatment and 5.758 in the PR-FF treatment, suggesting a lower within-pair heterogeneity in the PR-CF than in the PR-FF treatments. A similar difference is suggested in the TR treatments, the within-pair S.D. being equal to 2.865 and to 8.035 in the TR-CF and TR-FF treatments, respectively. Two-sided Wilcoxon rank sum tests show that the difference between the PR-CF and PR-FF treatments is significant at the 10% level ($p = 0.053$) and that the difference between the TR-CF and TR-FF treatments is significant at the 1% level ($p < 0.001$). To calculate the standard deviations in the PR-FF and TR-FF treatments we used the actual pairs of participants. As a robustness test, we also ran a simulation-based

approach.¹⁷ The results of this simulation are displayed in Figure C.2 in Appendix C and they confirm the conclusions of the non-parametric tests.¹⁸

The between-pair S.D. is significantly higher in the PR-CF and TR-CF treatments (6.696 and 8.122, respectively) than in the PR-FF and TR-FF treatments (5.103 and 6.269, respectively) (see Figure C.3 in Appendix C). Finally, we compare the difference between the between-pair S.D. and the within-pair S.D. in both CF treatments with the difference computed for the real and simulated pairs in the corresponding FF treatments. Two-sided Wilcoxon rank sum tests show that this difference is significantly larger for the pairs in the FF treatments than for the pairs in the CF treatments (PR treatments: $p < 0.001$; TR treatments: $p < 0.001$). Figure C.4 in Appendix C shows that this difference is larger in the CF treatments than in the simulated pairs of the FF treatments.

Overall, these analyses of within- and between-pair standard deviations support the existence of peer effects due to the dissemination of social information within pairs, which could explain the observed escalation of reports over time in the PR-CF and TR-CF treatments.

2.5.3 Gender Heterogeneity

Ex post we identified some gender differences that can explain some of our general results. We introduce our last result:

Result 5: Males were more likely to over-report their outcome than females in the TR-FF treatment but not in the other treatments, and they were more sensitive to feedback than females in the TR-CF treatment.

Support to Result 5. Figure 2.5.2 displays the probability to report a red outcome, by gender and by treatment. Estimates are computed from the multinomial logit regression model reported in Table B.2 in Appendix B, in which the male variable is interacted with

¹⁷We generated 30,000 configurations of all the hypothetical pairs formed with the participants of the PR-FF and TR-FF treatments that we compared with the average within-pair S.D. in the PR-CF and TR-CF treatments, respectively.

¹⁸Note that, not surprisingly, we found no difference between the within- and between-pair S.D. in the TR1-FF and TR1-CF treatments.

the treatment. Figure 2.5.2 shows that males were significantly more likely to report a red outcome than females in the TR-FF treatment, suggesting a higher dishonesty of males in this environment ($p < 0.001$).¹⁹ In line with this result we also find that 47.8 % of the males in the TR-FF treatment lied to the full extent while only 4.8% of the females did. This gap in the extent of lying is not present in any other treatment.

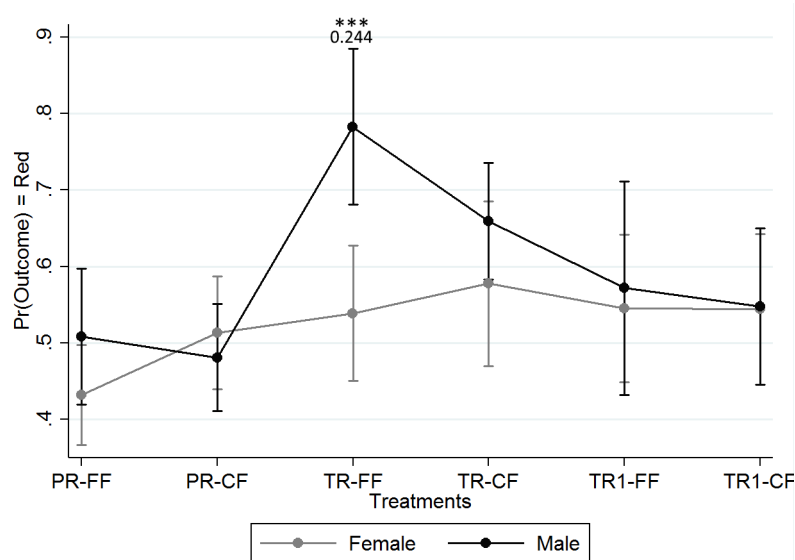


Figure 2.5.2 – Probability of reporting a red outcome, by gender and treatment. *Notes:* Bars represent 95% confidence intervals. The numerical value is the coefficient of the marginal effect of being a male on the probability of reporting a red outcome. Stars represent significance of the marginal effect of the male variable in a given treatment. *** indicate significance at the 1% level.

These results cannot be explained by a higher competitiveness of males because we should observe that males lie more than females in all treatments with competition, which is not the case. An alternative explanation is that males form different beliefs about the honesty of their counterpart. They might expect a higher level of dishonesty of their counterpart in the TR-FF treatment, compared to females.²⁰

¹⁹Males were also less likely to report a blue or a yellow outcome than females in the TR-FF treatment ($p = 0.019$ and $p < 0.001$, respectively). See Figures C.5a and C.5b in Appendix C.

²⁰We cannot test this hypothesis directly because we elicited beliefs about the empirical norm after participants received feedback on their counterpart's reports (to hold this constant across treatments). Thus, participants have been able to update their beliefs as illustrated by the significant correlation between participants' beliefs and their opponent's performance in the TR-FF treatment ($\rho = 0.530$, p -value < 0.001). Table B.3 in Appendix B reports the coefficients of OLS regressions on *ex post* beliefs in TR-FF. It reveals no

Finally, our data suggest that females were less sensitive to feedback than males in the TR-CF treatment. The mean within-pair standard deviation is higher in pairs including at least one female than in pairs formed with two males (two-sided Wilcoxon rank sum test: $p = 0.027$). In contrast, the mean within-pair standard deviation in pairs including at least one male is lower than in pairs formed with two females ($p = 0.056$). This is in line with previous results showing that females react less than males to peer effects when they interact simultaneously with others (Beugnot et al., 2019).

2.6 Discussion and Conclusion

In this study, we investigated how feedback about others' reported outcomes influences individuals' dishonesty in both competitive and non-competitive dynamic settings. We also tested whether the level of dishonesty in competitive settings results mainly from the existence of competition or from the expectation of dishonest reporting by the opponent. As expected, we found evidence of over-reporting under both individual piece rate and tournament incentives from the beginning. Like in the previous literature on one-shot die tasks, we observed, in our dynamic settings, neither full lying behavior nor a uniform distribution of reports (*e.g.*, Gibson et al., 2013). The distribution of reported outcomes was in line with that found in Dai et al. (2018) who used a similar task on a similar sample but in a one-shot setting. This confirms the heterogeneity of lying preferences also in a dynamic setting.

One first important result is that competitive incentives make individuals more likely to behave dishonestly compared to an individual piece rate pay scheme. This replicates, in a new dynamic setting, previous findings obtained in studies using real-effort tasks (*e.g.*, Schwierien and Weichselbaumer, 2010; Rigdon and D'Esterre, 2015; Faravelli et al., 2015; Belot and Schröder, 2013). We also found that individuals are less likely to behave dishonestly in tournaments when it is common knowledge that their counterpart cannot

gender difference in stated beliefs. But again, this cannot reject our interpretation that males were more pessimistic about the dishonesty of others when they receive no social information at all. See also Figure C.6 in Appendix C.

misreport. Dishonesty differs only weakly in this context compared to a piece rate scheme, although the incentives to lie remain higher in a tournament setting. This suggests that the higher level of dishonesty when both competitors can lie is mostly driven by a fear of losing a potentially unfair and more fierce competition.

We also found that individuals, when receiving a continuous feedback on their counterpart's reports, became more dishonest over time in both the tournament and the piece rate pay scheme. This is in line with the idea of a behavioral contagiousness of dishonesty (e.g., Gino et al., 2009; Innes and Mitra, 2013; Lauer and Untertrifaller, 2019). But in contrast to our expectations, the contagiousness of dishonesty was not significantly larger in tournaments than in the piece rate pay scheme. Rivalry over the prize in tournaments did not lead to more imitative behavior than in the piece rate. One explanation could be that under a piece rate pay scheme, feedback generates also rivalry but for social status.

Finally, we found that providing continuous feedback, instead of a final feedback, on the counterpart's reports increases dishonesty at the aggregate level in piece rate settings but not in tournament settings. We argue that this is because when subjects remained uninformed, male competitors lied a lot in the expectation of their opponent's dishonesty. Indeed, males were more dishonest than females in the tournament without continuous feedback, whereas no gender difference held in any other treatment. A possible interpretation is that males have higher expectations about their opponent's dishonesty. Even if our data cannot confirm this, it is in line with a recent study showing that males state higher beliefs than females about the level of sabotage they receive in a competition (Dato and Nieken, 2018).

The previous literature has insisted on the importance of the moral costs of lying. In individual piece rate settings, does feedback on others' reports and the progressive revelation of others' moral weakness mitigate these moral costs? We can proxy indirectly (and imperfectly) these perceived costs through the evolution of mood. Two-sided Wilcoxon signed rank tests show that indeed, in the piece rate treatments participants' happiness and nervousness did not change significantly over time (difference in hap-

piness: -0.088 , $p = 0.539$; difference in nervousness: -0.202 , $p = 0.223$). People lied more over time but knowing that others are also likely liars seems to reduce the moral costs of lying. The previous literature has evidenced not only the moral costs of lying but also the joy of winning in competition. So, when individuals compete with an opponent in the tournament treatments and have the opportunity to over-report their outcomes to increase their chance of winning, what is the resulting effect on mood? In the tournament treatments participants reported less happiness and more nervousness at the end compared to the beginning of the session (difference in happiness: -0.746 , $p = 0.001$; difference in nervousness: 0.738 , $p = 0.002$). Competition increases stress (e.g., Buser et al., 2017) and this may be reinforced by lying. This difference is mostly driven by the losers of the tournament as their nervousness is significantly increased while those of the winners does not (difference in happiness: -0.517 and -0.953 , $p = 0.082$ and $p = 0.002$; difference in nervousness: -0.431 and 1.797 , $p = 0.654$ and $p < 0.001$ for winners and losers, respectively). Interestingly, in the TR1 treatments where competition is much less fierce, happiness did not change over time and nervousness decreased marginally (difference happiness: -0.417 , $p = 0.190$; difference arousal: -0.367 , $p = 0.062$). Being more certain about winning the competition, the participants who were allowed to lie did not seem to suffer a higher moral cost of lying despite their advantageous and unfair situation. However, for passive participants, the TR1 treatments induce less happiness and more nervousness like for participants in the TR treatments, which is certainly driven by the unfairness of their situation (difference in happiness: -1.033 , $p < 0.001$; difference in nervousness: 2.617 , $p < 0.001$).

Our study shows that competitive incentives increase dishonesty mainly because of the uncertainty about one's opponent degree of dishonesty and thus, about the actual fierceness of the competition. Our study stresses out that people's expectations are important to understand dishonest behavior in competition. We suggest that people's beliefs are heterogeneous and that explains the more or less important effect of social information on behavior. Further investigation would be needed to explore the role of beliefs and how they are updated over time depending on the social information

received. Another useful extension of our work would be to explore how precisely social information changes the moral cost of lying, in particular through the perception of what is socially acceptable.

Bibliography

- Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica*. Forthcoming.
- Aydogan, G., A. Jobst, K. D'Ardenne, N. Müller, and M. G. Kocher (2017). The Detrimental Effects of Oxytocin-Induced Conformity on Dishonesty in Competition. *Psychological Science* 28(6), 751–759.
- Azmat, G. and N. Iriberry (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics* 94(7-8), 435–452.
- Banerjee, R., N. D. Gupta, and M. C. Villeval (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review* 101, 567–604.
- Belot, M. and M. Schröder (2013). Sloppy work, lies and theft: A novel experimental design to study counterproductive behaviour. *Journal of Economic Behavior and Organization* 93(C), 233–238.
- Beugnot, J., B. Fortin, G. Lacroix, and M. C. Villeval (2019). Gender and Peer Effects on Performance in Social Networks. *European Economic Review* 113, 207–224.
- Bock, O., I. Baetge, and A. Nicklisch (2014). Hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71, 117–120.
- Bowles, S. (2161). *The Moral Economy: Why Good Incentives are No Substitute for Good Citizens*. Yale University Press.
- Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: An experimental study. *Journal of political Economy* 95(1), 1–33.
- Buser, T. and A. Dreber (2016). The Flipside of Comparative Payment Schemes. *Management Science* 62(9), 2626–2638.
- Buser, T., A. Dreber, and J. Mollerstrom (2017). The impact of stress on tournament entry. *Experimental Economics* 20(2), 506–530.
- Cadsby, C. B., F. Song, and F. Tapon (2010). Are you paying your employees to cheat? an experimental investigation. *The BE Journal of Economic Analysis & Policy* 10(1).
- Carpenter, J., P. H. Matthews, and J. Schirm (2010). Tournaments and Office Politics : Evidence from a Real Effort Experiment. *American Economic Review* 100(1), 504–517.
- Cartwright, E. and M. L. C. Menezes (2014). Cheating to win: Dishonesty and the intensity of competition. *Economics Letters* 122(1), 55–58.

- Cason, T. N., W. A. Masters, and R. M. Sheremeta (2018). Winner-take-all and proportional-prize contests: Theory and experimental results. *Journal of Economic Behavior and Organization* 0, 1–14.
- Conrads, J., B. Irlenbusch, R. M. Rilke, and G. Walkowitz (2013). Lying and team incentives. *Journal of Economic Psychology* 34, 1–7.
- Dai, Z., F. Galeotti, and M. C. Villeval (2018). Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation. *Management Science* 64(3), 983–1476.
- Dato, S. and P. Nieken (2018). Gender Differences in Sabotage : The Role of Uncertainty and Beliefs.
- Diekmann, A., W. Przepiorka, and H. Rauhut (2015). Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. *Rationality and Society* 27(3), 309–333.
- Dimant, E. (2018). Contagion of Pro- and Anti-Social Behavior Among Peers and the Role of Social Proximity.
- Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory* 175(175), 248–264.
- Eriksson, T., A. Poulsen, and M. C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics* 16(6), 679–688.
- Falk, A. and A. Ichino (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics* 24(1), 39–57.
- Falk, A. and N. Szech (2013). Morals and Markets. *Science* 340(6133), 707–712.
- Faravelli, M., L. Friesen, and L. Gangadharan (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior and Organization* 110, 160–175.
- Feltovich, N. (2018). The interaction between competition and unethical behaviour : experimental evidence. *Experimental Economics*. Forthcoming.
- Fershtman, C. and U. Gneezy (2011). The tradeoff between performance and quitting in high power tournaments. *Journal of the European Economic Association* 9(2), 318–336.
- Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise – An Experimental Study on Cheating. *Journal of the European Economic Association* 11(40), 525–547.
- Fortin, B., G. Lacroix, and M. C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics* 91(11-12), 2089–2112.

- Fosgaard, T. R., L. G. Hansen, and M. Piovesan (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior and Organization* 93, 279–284.
- Garrett, N., S. C. Lazzaro, D. Ariely, and T. Sharot (2016). The brain adapts to dishonesty. *Nature Neuroscience* 19, 1727–1732.
- Gibson, R., C. Tanner, and A. F. Wagner (2013). Preferences for Truthfulness : Heterogeneity among and within Individuals. *American Economic Review* 103(1), 532–548.
- Gino, F., S. Ayal, and D. Ariely (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological Science* 20(3), 393–8.
- Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review* 108(2).
- Gneezy, U. and J. Potters (1997). An experiment on risk taking and evaluation periods. *Quarterly journal of economics* (May).
- Harbring, C. and B. Irlenbusch (2011, mar). Sabotage in Tournaments: Evidence from a Laboratory Experiment. *Management Science* 57(4), 611–627.
- Harbring, C., B. Irlenbusch, M. Kräkel, and R. Selten (2007). Sabotage in Corporate Contests – An Experimental Analysis. *International Journal of the Economics of Business* 14(3), 367–392.
- Innes, R. and A. Mitra (2013). Is dishonesty contagious? *Economic Inquiry* 51(1), 722–734.
- Kajackaite, A. and U. Gneezy (2017). Incentives and cheating. *Games and Economic Behavior* 102, 433–444.
- Khalmetski, K. and D. Sliwka (2017). Disguising Lies – Image Concerns and Partial Lying in Cheating Games. CESifo Working Paper no. 6347.
- Kocher, M. G., S. Schudy, and L. Spantig (2017). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science* (August), mns.2017.2800.
- Konrad, K. A. (2000). Sabotage in Rent-seeking Contests. *Journal of Law, Economics & Organization* 16(1), 155–165.
- Kuhnen, C. M. and A. Tymula (2009). Feedback, Self-Esteem and Performance in Organizations. *Ssrn* (December 2015).
- Lauer, T. and A. Untertrifaller (2019). Conditional dishonesty.
- Lazear, E. P. (1989). Pay equality and Industrial Politics. *Journal of Political Economy* 97(3), 561–580.

- Lefebvre, M., P. Pestieau, A. Riedl, and M. C. Villeval (2015). Tax evasion and social information: an experiment in Belgium, France, and the Netherlands. *International Tax and Public Finance* 22(3), 401–425.
- Ludwig, S. and G. K. Lünser (2012). Observing your competitor - The role of effort information in two-stage tournaments. *Journal of Economic Psychology* 33(1), 166–182.
- Mas, A. and E. Moretti (2009). Peers at Work. *American Economic Review* 99(1), 112–145.
- Mazar, N., O. Amir, and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45(6), 633–644.
- Parco, J. E., A. Rapoport, and W. Amaldoss (2005). Two-stage contests with budget constraints: An experimental study. *Journal of Mathematical Psychology* 49(4), 320–338.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of economic literature* 37(1), 7–63.
- Rabanal, J. P. and O. A. Rud (2018). Does competition affect truth telling? An experiment with rating agencies. *Review of Finance* 22(4), 1581–1604.
- Rauhut, H. (2013). Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments. *PLoS ONE* 8(11).
- Rigdon, M. L. and A. D’Esterre (2015). The Effects of Competition on the Nature of Cheating Behavior. *Southern Economic Journal* 81.
- Rosaz, J., R. Slonim, and M. C. Villeval (2016). Quitting and peer effects at work. *Labour Economics* 39, 55–67.
- Schurr, A. and I. Ritov (2016). Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences* 113(7), 201515102.
- Schwieren, C. and D. Weichselbaumer (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology* 31(3), 241–253.
- Shalvi, S., M. J. J. Handgraaf, and C. K. W. De Dreu (2011). Ethical Manoeuvring: Why People Avoid Both Major and Minor Lies. *British Journal of Management* 22, 16–27.
- Sheremeta, R. M. (2010). Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior* 68(2), 731–747.
- Soraperra, I., O. Weisel, R. Zultan, S. Kochavi, M. Leib, H. Shalev, and S. Shalvi (2017). The bad consequences of teamwork. *Economics Letters* 160, 12–15.
- Van Dijk, F., J. Sonnemans, and F. Van Winden (2001). Incentive systems in a real effort experiment. *European Economic Review* 45(2), 187–214.
- Weisel, O. and S. Shalvi (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences* 112(34), 10651–10656.

A Instructions

A.1 Two-Sided treatments

All the instructions were displayed on the computer screen.

Welcome screen

Hello. We thank you for your participant to this experiment on decision making. Please turn off your phone. It is forbidden to talk with the others participants during the duration of the experiment. All your decisions are anonymous.

If you have any questions during the session, please use the red button and an assistant will come to answer privately.

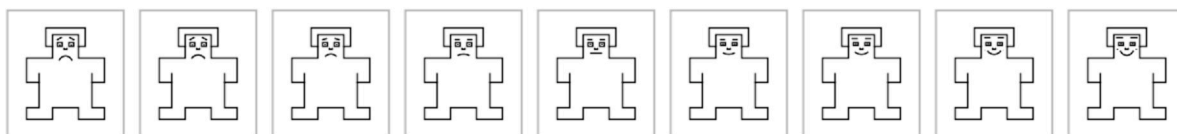
The experiment is divided in two independent parts. In addition, you will be asked some questions at the end of the experiment. At the beginning of each part, you will receive the corresponding instructions.

At the end of the experiment, you will receive the earnings for the first and the part as well as a show-up fee of €5. The payment will be made privately in a separate room by someone who does not know the content of the session in order to respect confidentiality.

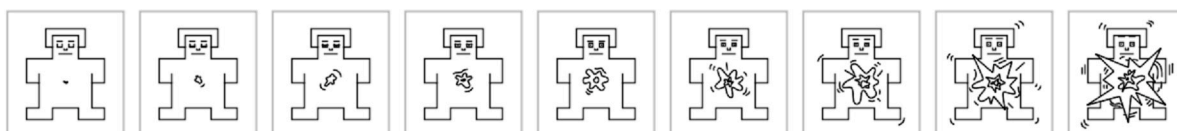
Questionnaire

Before giving you the instructions of the first part, could you please answer sincerely to the two following questions. We remind you that all your answers are anonymous. Please check the picture which best describe your feeling **now**. For instance, if you are feeling deeply happy or satisfy you can indicate it by choosing the picture at the rightmost part of the screen. Conversely, if you are feeling deeply unhappy or dissatisfied, you can also indicate it by choosing the picture at the leftmost part of the screen. Those pictures also allows you to describe a feeling of intermediate happiness by selecting one of them.

Again, check the picture which best describe your feeling **now**. For instance, if you are feeling extremely excited or tensed, you can indicate it by checking the rightmost



picture. Conversely, if you are feeling completely relaxed or calm, you can indicate it by checking the leftmost picture. Those pictures also allows you to describe a feeling of intermediate excitement by selecting one of them.



Part 1

In this part, you will take one decision which will determine your earnings. You are endowed with €150 cents (€1.5) and you have to decide which amount of this endowment (between 0 and €150 cents, included) to invest in a lottery. In this lottery, the amount invested can be either multiply by 2.5 or entirely lost. The probability of each of these events is equal to 50%. Your final gain will be equal to the following sum: amount not invested in the lottery + results of the lottery.

Part 2: Main Task

The task

At the beginning of this part, you will be paired with another participant. You will stay paired with the same participant throughout this part.

This part is composed by 24 periods. In each period, you roll the die which is in the cup on your desk without taking off the lid. Your task is to report the color which appears on the top of the die when looking through the lid. If you wish, you can roll the die more than once in each period. However, you have to report *only the outcome of your first roll*. The points obtained in each period depend on the outcome of the roll as undermentioned:

| Die outcome | Points |
|-------------|----------|
| Red | 2 points |
| Yellow | 1 point |
| Blue | 0 point |

No Feedback condition

At the beginning of each period, you are reminded the number of points obtained in the previous periods as well as the total number of points accumulated so far. At the end of the period 24 (and only at the moment), you will be informed of the number of points obtained by your counterpart in each period as well as the total number of points that she or he has accumulated so far.

Continuous Feedback condition

At the beginning of each period, you are reminded the number of points obtained in the previous periods as well as the total number of points accumulated so far. At the end of each period and at the end of the period 24, you will be informed of the number of points obtained by your counterpart in each period as well as the total number of points that she or he has accumulated so far.

Earnings

At the end of the 24 periods, you will have accumulated a certain number of points.

Piece-rate condition

Your earnings will be determined by this number of points that you have accumulated. Your points will be converted in euros according to the following conversion rate, 1 point = €18 cents, and they will be paid in cash.

Tournament condition

If you have accumulated more points than your counterpart, your earnings will be determined based on this number of points. Your points will be converted in euros according to the following conversion rate, 1 point = €36 cents, and they will be paid in cash.

If you have accumulated less points than your counterpart, your earnings will be equal to €0.

In case of a tie in the number of points accumulated at the end of the period 24 between you and your counterpart, the program will randomly select one of you to receive a gain of € and one of you whose points will be converted into Euros.

Predictions

Before the end of the experiment, you will have the opportunity to earn an additional amount of €1.5. You will have to predict the colors that 24 others participants (excluding your counterpart) have reported in the first, thirteenth and last period. These 24 participants are chosen randomly, by the computer, in the current session or in one of the previous sessions of the same experiment.

Here is the procedure:

For each of these periods (1, 13 and 24), you will indicate how many participants, among these 24, reported the color red (which gives 2 points), the color yellow (which gives 1 point) and the color blue (which gives 0 pint).

At the end of the session, the program will randomly select one of these three periods and one of these three colors. You will be paid for your prediction for the selected period and color.

- If your prediction is correct (meaning that your predicted number of participants for this period and color is equal to the true number), you earn €1.5;
- If your prediction is incorrect to one near (meaning that your predicted number of participants for this period and color is equal to one near to the true number), your earn €1;

- If your prediction is incorrect to two near (meaning that your predicted number of participants for this period and color is equal to two near to the true number), you earn €0.5;
- If your prediction is incorrect to three near or more (meaning that your predicted number of participants for this period and color is equal to three near or more to the true number), you earn €0.

Example: imagine that the program randomly select period 13 and the color yellow. This implies that you will be paid for your prediction on the number of participants who reported the color yellow in period 13. Imagine that your prediction was 12 (on 24 participants) while the correct number is 14. In that case, your prediction is incorrect to two near and your earnings is equal to €0.5.

A.2 One-Sided treatments

All the instructions were displayed on the computer screen.

Welcome screen

Hello. We thank you for your participant to this experiment on decision making. Please turn off your phone. It is forbidden to talk with the others participants during the duration of the experiment. All your decisions are anonymous.

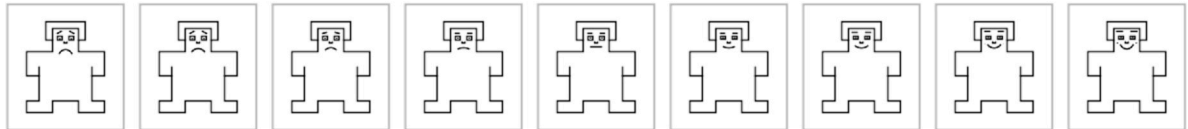
If you have any questions during the session, please use the red button and an assistant will come to answer privately.

The experiment is divided in two independent parts. In addition, you will be asked some questions at the end of the experiment. At the beginning of each part, you will receive the corresponding instructions.

At the end of the experiment, you will receive the earnings for the first and the part as well as a show-up fee of €5. The payment will be made privately in a separate room by someone who does not know the content of the session in order to respect confidentiality.

Questionnaire

Before giving you the instructions of the first part, could you please answer sincerely to the two following questions. We remind you that all your answers are anonymous. Please check the picture which best describe your feeling **now**. For instance, if you are feeling deeply happy or satisfy you can indicate it by choosing the picture at the rightmost part of the screen. Conversely, if you are feeling deeply unhappy or dissatisfied, you can also indicate it by choosing the picture at the leftmost part of the screen. Those pictures also allows you to describe a feeling of intermediate happiness by selecting one of them.



Again, check the picture which best describe your feeling **now**. For instance, if you are feeling extremely excited or tensed, you can indicate it by checking the rightmost picture. Conversely, if you are feeling completely relaxed or calm, you can indicate it by checking the leftmost picture. Those pictures also allows you to describe a feeling of intermediate excitement by selecting one of them.



Part 1

In this part, you will take one decision which will determine your earnings. You are endowed with €150 cents (€1.5) and you have to decide which amount of this endowment (between 0 and €150 cents, included) to invest in a lottery. In this lottery, the amount invested can be either multiply by 2.5 or entirely lost. The probability of each of these events is equal to 50%. Your final gain will be equal to the following sum:

amount not invested in the lottery + results of the lottery.

Part 2: Main Task

The task

At the beginning of this part, you will be paired with another participant. You will stay paired with the same participant throughout this part.

This part is composed by 24 periods. Each one of you has a cup on your desk but only yours or your counterpart's cup has a die in it. In each period, the one who has the cup with the die roll the die which is in the cup on your desk without taking off the lid. The task of this person is to report the color which appears on the top of the die when looking through the lid.

The task of the person without a cup with a die consists to roll an electronic die by clicking of the button "roll the die" on the computer screen. Both for the electronic die or the die in the cup, there are three possible outcomes which have the same chance to be drawn.

If you wish, you can roll the die more than once in each period. However, if you have the die in the cup, you have to report *only the outcome of your first roll*. If you have the electronic die, the computer only automatically records the outcome of the first roll.

The points obtained in each period depend on the outcome of the roll (electronic or not) as undermentioned:

| Die outcome | Points |
|-------------|----------|
| Red | 2 points |
| Yellow | 1 point |
| Blue | 0 point |

No Feedback condition

At the beginning of each period, you are reminded the number of points obtained in the previous periods as well as the total number of points accumulated so far.

At the end of the period 24 (and only at the moment), you will be informed of the number of points obtained by your counterpart in each period as well as the total number of

points that she or he has accumulated so far.

Continuous Feedback condition

At the beginning of each period, you are reminded the number of points obtained in the previous periods as well as the total number of points accumulated so far.

At the end of each period and at the end of the period 24, you will be informed of the number of points obtained by your counterpart in each period as well as the total number of points that she or he has accumulated so far.

Earnings

At the end of the 24 periods, you will have accumulated a certain number of points.

Piece rate condition

Your earnings will be determined by this number of points that you have accumulated. Your points will be converted in euros according to the following conversion rate, 1 point = €18 cents, and they will be paid in cash.

Tournament condition

If you have accumulated more points than your counterpart, your earnings will be determined based on this number of points. Your points will be converted in euros according to the following conversion rate, 1 point = €36 cents, and they will be paid in cash.

If you have accumulated less points than your counterpart, your earnings will be equal to €0.

In case of a tie in the number of points accumulated at the end of the period 24 between you and your counterpart, the program will randomly select one of you to receive a gain of € and one of you whose points will be converted into euros.

Predictions

Before the end of the experiment, you will have the opportunity to earn an additional amount of €1.5. You will have to predict the colors that 24 other participants who received a die (excluding your counterpart or yourself) have reported in the first, thirteenth and last period. These 24 participants are chosen randomly, by the computer, in the current session or in one of the previous sessions of the same experiment.

Here is the procedure:

For each of these periods (1, 13 and 24), you will indicate how many participants, among these 24, reported the color red (which gives 2 points), the color yellow (which gives 1 point) and the color blue (which gives 0 points).

At the end of the session, the program will randomly select one of these three periods and one of these three colors. You will be paid for your prediction for the selected period and color.

- If your prediction is correct (meaning that your predicted number of participants for this period and color is equal to the true number), you earn €1.5;
- If your prediction is incorrect to one near (meaning that your predicted number of participants for this period and color is equal to one near to the true number), you earn €1;
- If your prediction is incorrect to two near (meaning that your predicted number of participants for this period and color is equal to two near to the true number), you earn €0.5;
- If your prediction is incorrect to three near or more (meaning that your predicted number of participants for this period and color is equal to three near or more to the true number), you earn €0.

Example: imagine that the program randomly selects period 13 and the color yellow. This implies that you will be paid for your prediction on the number of participants who reported the color yellow in period 13. Imagine that your prediction was 12 (on 24

participants) while the correct number is 14. In that case, your prediction is incorrect to two near and your earnings is equal to €0.5.

B Tables

Table B.1 – Demographic characteristics of participants, by treatment

| Treatments | Average age (S.D.) | Male (%) | Student (%) | Nb. of part. |
|---|--------------------|----------|-------------|--------------|
| Piece rate - Final Feedback (PR-FF) | 24.78 (1.73) | 50.0% | 78.6% | 42 |
| Piece rate - Continuous Feedback (PR-CF) | 23.07 (0.77) | 52.8% | 90.3% | 72 |
| Tournament - Final Feedback (TR-FF) | 21.41 (0.31) | 52.3% | 97.7% | 44 |
| Tournament - Continuous Feedback (TR-CF) | 21.50 (0.26) | 47.4% | 96.1% | 78 |
| One-Sided Tournament - Final Feedback (TR1-FF) | 22.67 (0.79) | 44.8% | 94.8% | 58 |
| One-Sided Tournament - Continuous Feedback (TR1-CF) | 26.68 (1.58) | 45.2% | 79.0% | 62 |
| Total | 23.29 (0.41) | 48.6% | 89.9% | 356 |

Note: "Nb. of part." means the number of participations in other experiments.

Table B.2 – Determinants of the roll outcome reported

| | Blue (0 point) | Yellow (1 point) | Red (2 points) |
|-----------------------|----------------------|----------------------|---------------------|
| PR-FF | Ref | | |
| PR-CF | -0.048* (0.025) | 0.020 (0.027) | 0.028 (0.043) |
| TR-FF | -0.113*** (0.026) | -0.084*** (0.030) | 0.198*** (0.048) |
| TR-CF | -0.094*** (0.026) | -0.059** (0.029) | 0.153*** (0.046) |
| TR1-FF | -0.044 (0.032) | -0.044 (0.029) | 0.088* (0.050) |
| TR1-CF | -0.061** (0.027) | -0.018 (0.032) | 0.079* (0.047) |
| Period | -0.001 (0.001) | -0.002*** (0.001) | 0.002*** (0.001) |
| Male | -0.019 (0.015) | -0.043*** (0.016) | 0.062** (0.027) |
| Oth. control | Yes | | |
| Observations | 7104 | | |
| Number of clusters | 221 | | |
| Pseudo-R ² | 0.019 | | |
| p>chi2 | <0.001 | | |

Notes: This table reports marginal effects of a multinomial logit regression with robust standard errors in parentheses. Standard errors are clustered at the subject (pair) level in the FF (CF) treatments. The baseline treatment used for the estimations is the PR-FF treatment. Other independent variables include the attitudes toward risk, age, student status, and a control for the month in which data were collected. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table B.3 – Determinants of beliefs in the TR-FF treatment

| Dep. Var Beliefs | Total (1) | First period (2) | 13th period (3) | Last period (4) |
|---------------------|---------------------|------------------------|-----------------------|-----------------------|
| Male | 0.101 (0.158) | -0.008 (0.178) | 0.223 (0.141) | 0.088 (0.198) |
| Age | -0.040* (0.022) | -0.052** (0.026) | -0.040** (0.019) | -0.028 (0.030) |
| Risk | -0.016 (0.027) | 0.014 (0.035) | -0.058* (0.029) | -0.003 (0.029) |
| Constant | 2.319*** (0.583) | 2.444*** (0.699) | 2.775*** (0.527) | 1.738** (0.765) |
| Observations | 44 | 44 | 44 | 44 |
| R ² | 0.081 | 0.079 | 0.143 | 0.091 |

Notes: This table reports coefficients of OLS regressions with robust standard errors in parentheses. The dependent variable is the participant's beliefs on the average value of reports in all periods (model 1), in the first period (model 2), the thirteenth period (model 3), and the last period (model 4). *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.

C Figures

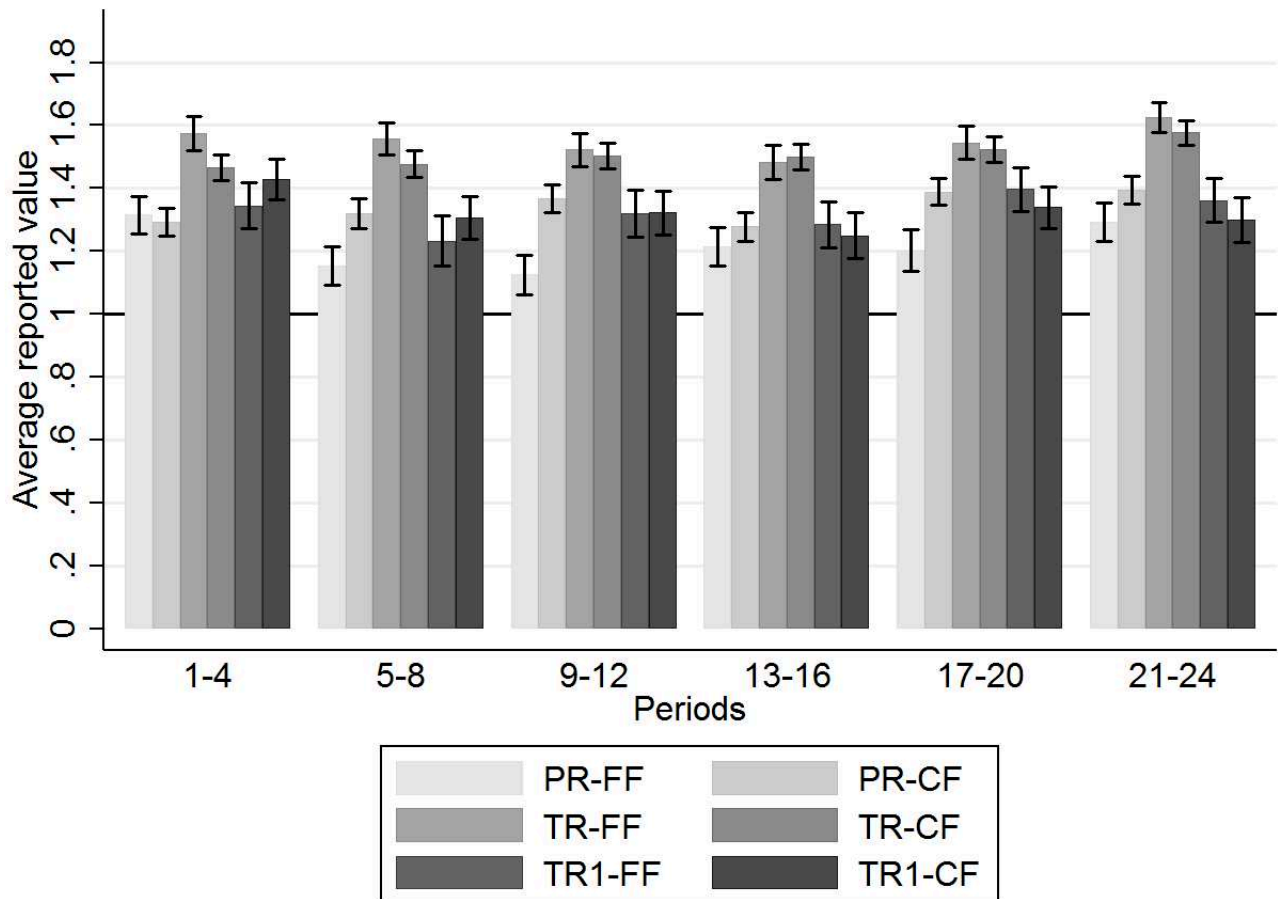


Figure C.1 – Average reported value by treatments and across blocks of four periods. *Notes:* The black bars represent standard errors to the mean. The horizontal block line corresponds to the expected value if participants were reporting honestly.

Within-pair S.D.

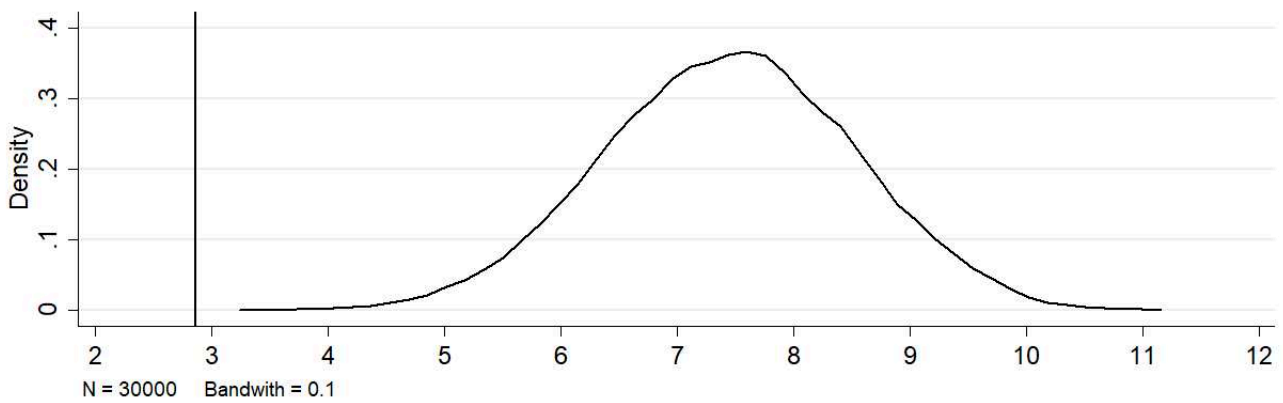
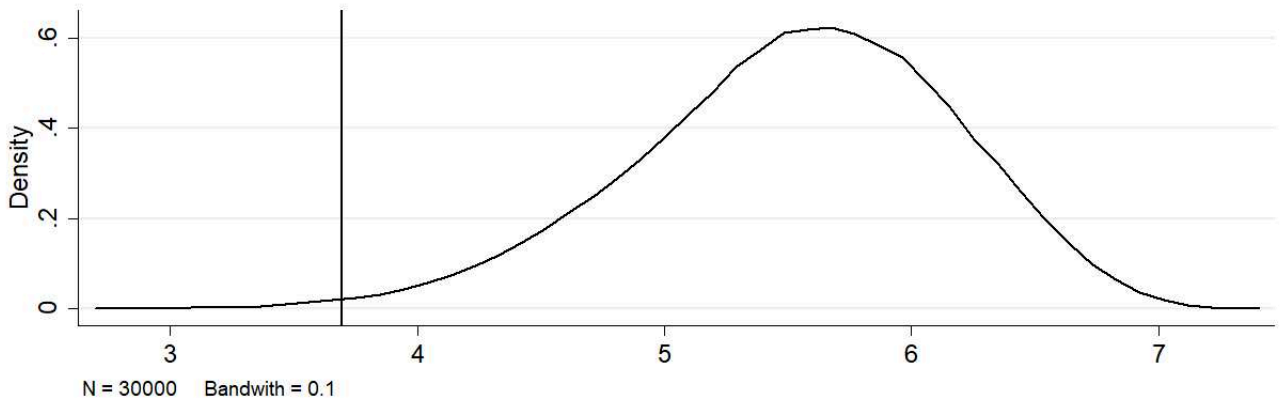


Figure C.2 – Top panel: standard deviation of the total value of reports within true pairs in the PR-CF treatment and within hypothetical pairs in the PR-FF treatment. The vertical line represents the average within-pair S.D. in the PR-CF treatment. **Bottom panel:** standard deviation of the total value of reports within true pairs in the TR-CF treatment and within hypothetical pairs in the TR-FF treatment. The vertical line represents the average within-pair S.D. in the TR-CF treatment.

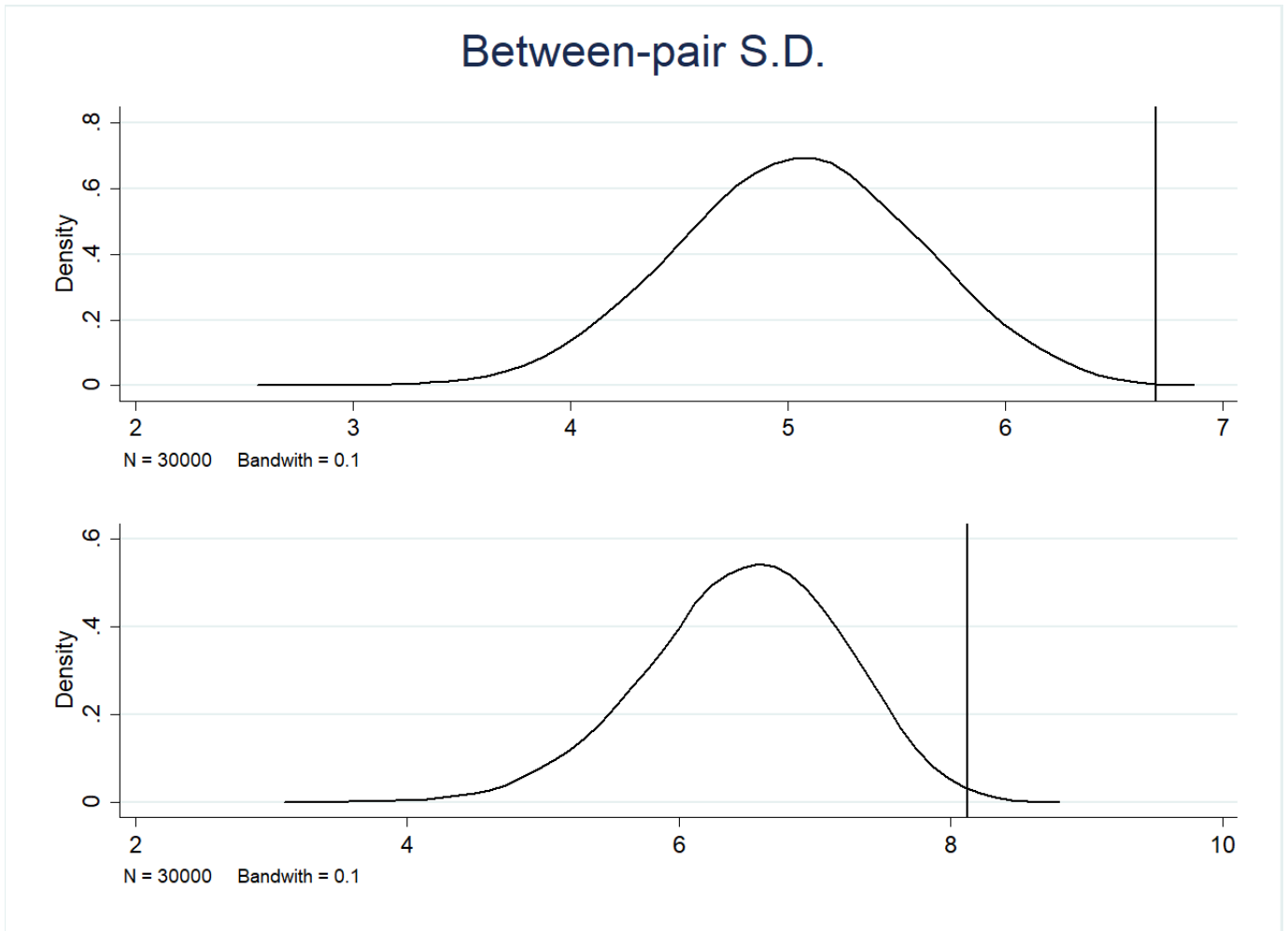


Figure C.3 – Top panel: standard deviation of the total value of reports between true pairs in the PR-CF treatment and between hypothetical pairs in the PR-FF treatment. The vertical line represents the average between-pair S.D. in the PR-CF treatment. **Bottom panel:** standard deviation of the total value of reports between true pairs in the TR-CF treatment and between hypothetical pairs in the TR-FF treatment. The vertical line represents the average between-pair S.D. in the TR-CF treatment.

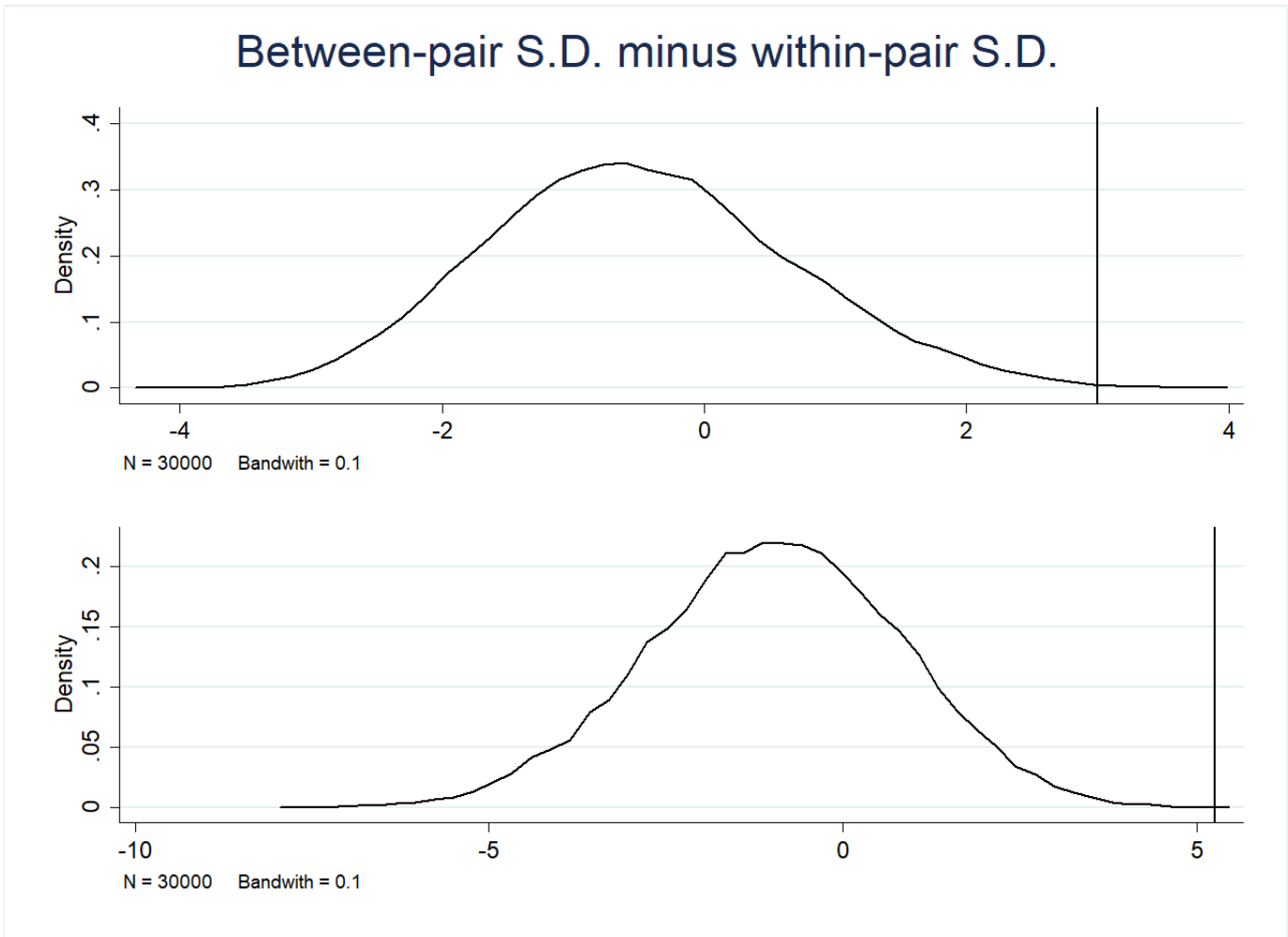


Figure C.4 – Top panel: between minus within standard deviation for true pairs in the PR-CF treatment and hypothetical pairs in the PR-FF treatment. The vertical line represents the average between-pair minus within-pair S.D. in the PR-CF treatment. **Bottom panel:** between minus within standard deviation for true pairs in the TR-CF treatment and hypothetical pairs in the TR-FF treatment. The vertical line represents the average between-pair minus within-pair S.D. in the TR-CF treatment.

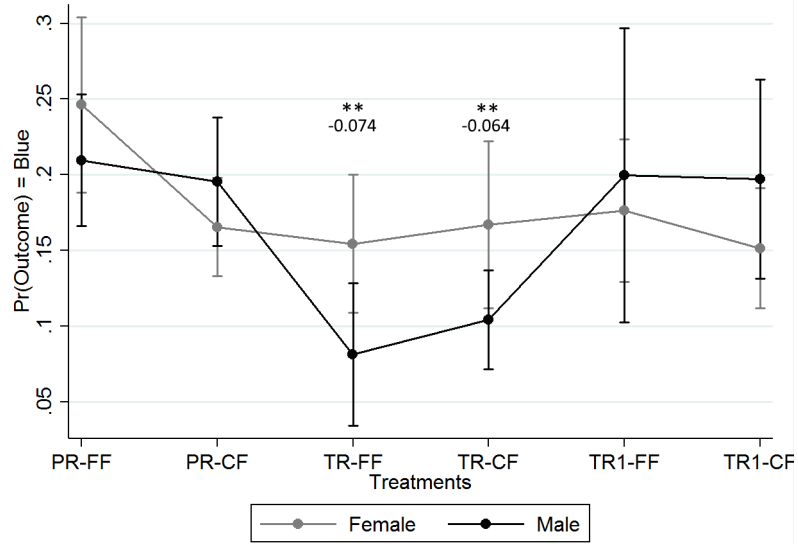


Figure C.5a – Probability to report a blue outcome for male and female participants, by treatment. Bars represent 95% confidence interval. The numerical values are the coefficients of the marginal effect of being a male on the probability of reporting a blue outcome. Stars represent significance of the marginal effect of being a male in a given treatment. ** indicate significance at the 5% level.

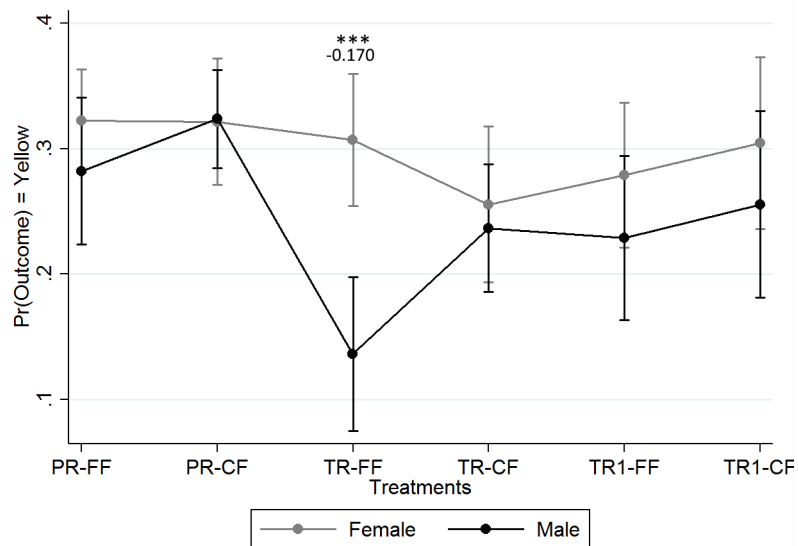


Figure C.5b – Probability to report a yellow outcome for male and female participants, by treatment. Bars represent 95% confidence interval. The numerical value is the coefficient of the marginal effect of being a male on the probability of reporting a yellow outcome. Stars represent significance of the marginal effect of being a male in a given treatment. *** indicate significance at the 1% level.

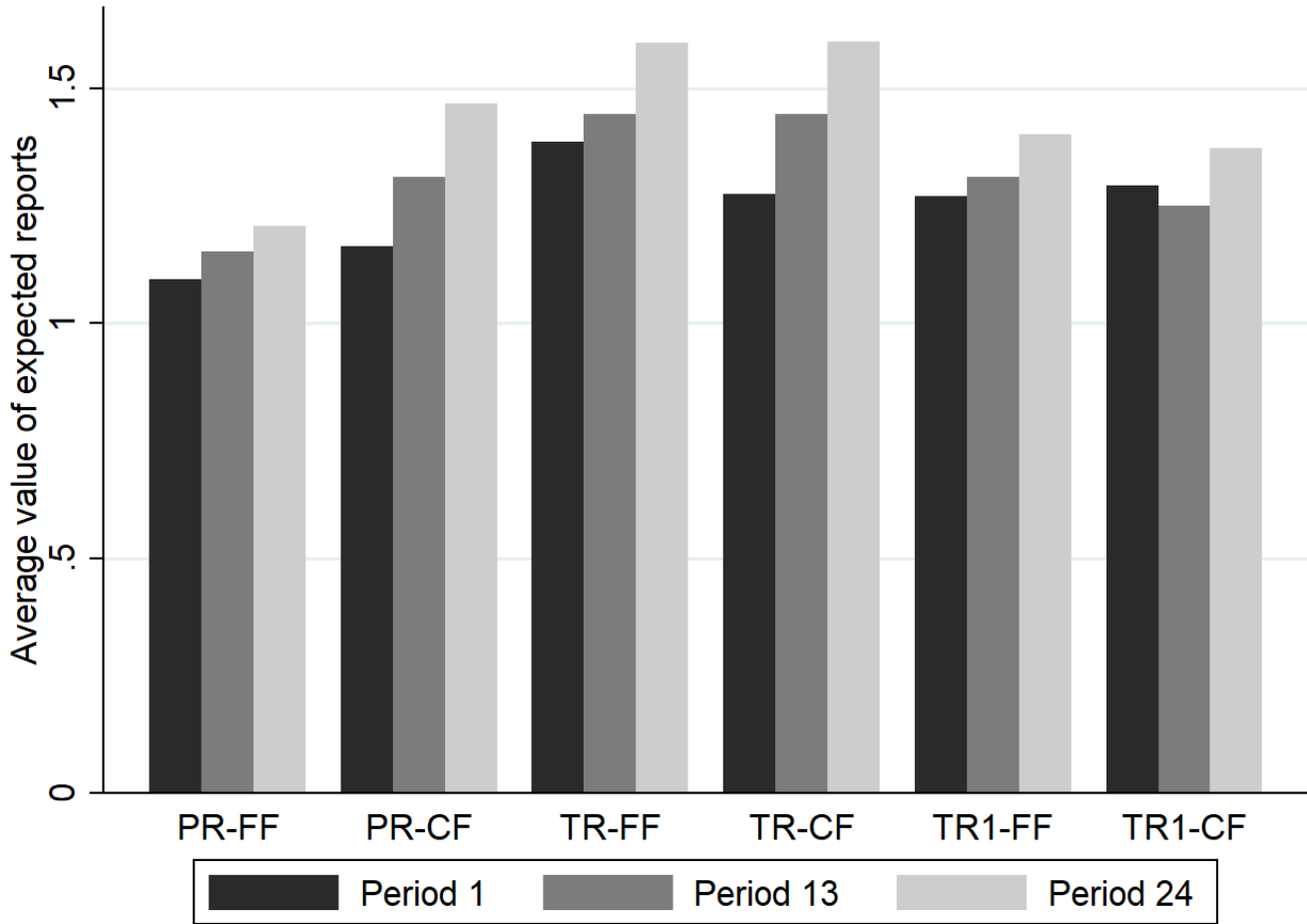


Figure C.6 – Average beliefs on others’ reports, by treatment and period.

Figure C.6 suggests that the beliefs stated by the participants are in line with the reporting behavior we observed in the treatments. Two-sided Wilcoxon rank sum tests on the average beliefs for all three periods only partially confirm it. Participants believe that others overreported more when they received continuous feedback instead of final feedback in the PR treatment, but not in the TR or TR1 treatments (PR: $p = 0.007$, TR: $p = 0.361$ TR1: $p = 0.948$). In the TR1 treatments, beliefs are about other subjects that have to report actively their die outcome. Participants think that others were more dishonest in the TR treatments than in the PR treatments, except between the TR-CF

and PR-CF treatments (TR-FF *vs.* PR-FF/PR-CF: $p < 0.001$ for both comparisons; TR-CF *vs.* PR-FF/PR-CF: $p < 0.001$ and $p = 0.114$, respectively). We find also differences in beliefs between the TR1 and the TR-FF treatments but not between the TR1 and the TR-CF treatments (TR-FF *vs.* TR1-FF/TR1-CF: $p = 0.018$ and $p = 0.012$, respectively; TR-CF *vs.* TR1-FF/TR1-CF: $p = 0.143$ and $p = 0.124$). Finally, participants expect more overreporting in the TR1 treatments than in the PR-FF treatment but not in the PR-CF treatment (TR1-FF/TR1-CF *vs.* PR-FF: $p = 0.014$ and $p = 0.01$, respectively; (TR1-FF/TR1-CF *vs.* PR-CF: $p = 0.960$ for both comparisons). Note that these results may be influenced by the fact that participants were fully informed about their counterpart's reports before being asked to state their beliefs.

Chapter 3

It does (not) get better: the effect of relative gains and losses on subsequent giving¹

3.1 Introduction

Most people are willing to share their own resources with others, and this contributes to the development of more harmonious societies (Gintis et al., 2003). In the economic literature, this behavior can be motivated by inequality aversion (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), the willingness to reward kind intentions (e.g. Charness and Rabin, 2002) or to sustain a good social image (e.g. Andreoni and Bernheim, 2009). These explanations are by nature consequentialist, as they posit that people are willing to share because of the consequences of sharing. Beyond consequences, the context of the decision also conditions the individual willingness to share. This is evidenced by context manipulations in dictator games: for example, adding the possibility to “take” (List, 2007; Bardsley, 2008) or lightly manipulating the legitimacy of dictators (Hoffman et al., 1994; Cherry et al., 2002; Cherry and Shogren, 2008; Carlsson et al., 2013) reduce

¹This chapter is a joint work with Rémi Suchon.

transfers drastically, whereas minimal moral reminders increase transfers (Brañas Garza, 2007).

In this paper, we identify a new potential determinant of one's willingness to give, that pertains to context rather than to consequences: the salience of counter-factual payoffs, that can be more enviable or less enviable. Our premise is that people will compare their actual payoff to the payoff they could have gotten in a different state of the world: if the payoff compares favourably, they encode it as a gain whereas if it compares unfavourably, they encode it as a loss. This type of situations is commonplace: workers compare the wage raise they get to the one they were expecting (Abeler et al., 2011), people feel worse if they miss an opportunity to increase their payoff than if they were not conscious of this opportunity (Filiz-Ozbay and Ozbay, 2007) and customers are more willing to buy a product when they believe the same product is more expensive elsewhere or if they expected to have to pay a higher price (Heidhues and Köszegi, 2008). Does getting a payoff lower than what one could have earned reduce one's willingness to give? Symmetrically, does getting a payoff higher than than what one could have earned increase one's willingness to share?

Our laboratory experiment isolates the pure effect of counter-factual potential payoffs on one's willingness to give. In the first part, participants have to perform a real effort task in exchange of a fixed wage. In the second part, they play a dictator game and they have to decide how much of a fixed endowment they are willing to transfer to an other anonymous participant. In a 3×2 between subjects design, we vary the wage for the real effort task (low, medium, high) and the timing of information on the wage (*Ref* and *NoRef*). In the *NoRef* conditions, participants are informed upfront of the wage in the real effort task and they are given an envelope containing the corresponding wage before performing the real effort task. In the *Ref* conditions, they only know the distribution of potential wages and they are informed that the actual wage will be revealed after the real effort task. In all conditions, participants receive an envelope containing the

medium wage from which we take out money in the Ref_{low} condition or add money in the Ref_{high} condition after the revelation of the actual wage. By informing participants of the distribution of the wage, we let them evaluate their actual wage relative to the other possible levels. To isolate the effects of providing information about counter-factual levels of wage, we compare transfers in the Ref conditions to the transfer in the $NoRef$ conditions, for each levels of wage.

As the comparison of the actual wage to the other potential levels may trigger emotional responses that in turn could translate into other-regarding behaviors (e.g Charness and Grosskopf, 2001; Capra, 2004; Kirchsteiger et al., 2006; Andrade and Ariely, 2009; Persson, 2018), we also collected data on the participants' emotional state throughout the experiment, using both declarative measures (Bosman and van Winden, 2002) and physiological measures (Skin Conductance Response). This data are used to assess the role of emotions in our treatment effects.

A natural conceptual framework to model the treatment effects of the experiment are reference-dependent models of preferences (Kahneman and Tversky, 1979a; Tversky and Kahneman, 1991; Köszegi and Rabin, 2006) because they assume that individuals perform counter-factual evaluations. Building on this, we develop a simple reference-dependent model of giving in which a participant formulates a reference about his payoff upon being informed of the content of the experiment and then evaluates his actual payoff relative to the reference. The reference payoff is determined by the reference wage and the amount he plans to transfer for this reference wage. The reference wage is assumed to be equal to the medium wage (€10) because it is salient (it is both the mean and median value for the wage) and because it corresponds to the status-quo and the expectation-based wage.² The model predicts that participants in the Ref_{low} condition will transfer less than those in the $NoRef_{low}$ because it is a way for them to reduce the

²Baillon et al. (2018) uncover an important heterogeneity in the way people formulate their reference point. However, they find that the "status-quo" is the modal reference point.

gap between their reference payoff and the payoff they actually get. The model also predicts that participants in the Ref_{high} condition will transfer more than those in the $NoRef_{high}$ condition.

The experimental data provide some support to the theory. We find that participants transfer similar amounts in the Ref and $NoRef$ conditions when they earn the medium wage, which suggests that merely being exposed to wage uncertainty has no impact on subsequent transfers. In contrast, loss-averse participants transfer less in the Ref_{low} condition than in the $NoRef_{low}$ while participants transfer more in the Ref_{high} compared to the $NoRef_{high}$ condition, irrespective of whether they are loss-averse. Contrary to the notion that “losses loom larger than gains” (Kahneman and Tversky, 1979b), we find no evidence that losses have a greater impact than gains on subsequent giving decisions. As expected, participants in the Ref_{low} condition experience more negative emotions while participants in the Ref_{high} condition experience more positive emotions. Nevertheless, physiological data have very limited explanatory power: the participants who experience higher emotional arousal do not react more strongly to the treatment than those who exhibit little emotional arousal. This suggests a limited role of visceral drives in our experiment.

We contribute to three streams in the literature. First, our results join a large literature showing that other-regarding behavior such as giving are influenced by the context of the decision (e.g. Dana et al., 2007; List, 2007; Bardsley, 2008, are famous examples.). In our experiment, the context of the decision is manipulated by informing participants of the potential payoffs they could have earned in another state of the world. Second, by identifying a trade-off between the pro-social preference to give and the preference for avoiding private losses relative to a reference, our results contribute to a more limited literature that studies reference-dependent social preferences. For instance, (Charité et al., 2016) provide evidence that dictators are influenced by the reference payoffs of the receivers and Choi (2014) shows that individuals transfer less to charities when their

consumption increase less than expected. On the other hand, Buffat (2016) finds that transfers chosen by participant in his lab experiment are not influenced by the level of an exogenous, randomly implemented transfer. This contradicts the predictions of his expectation-based reference dependent model of social preferences. Very closely related to our study, Ockenfels and Werner (2014) show that workers who are paid less than their reference wage give less in a subsequent dictator game to the principal who determined their wage. In contrast, in our experiment, recipients cannot be held accountable for the earnings in the real effort task. As a consequence, our results offer a new, complementary perspective on the literature studying how worker perceive their wages. It has been shown than workers compare their wage to the wages of their colleagues(e.g. Card et al., 2012; Bracha et al., 2015) or to the wage they use to have before a cut or a raise (e.g Gneezy and List, 2006; Kube et al., 2012; DellaVigna et al., 2016) and this conditions effort provision and job satisfaction. We identify a new type of reference, that is the potential wages one could have received in a different state of the world. In addition, we show that how one's wage compares to a reference also impacts one's social preferences.

Section 2 describes the experimental design and procedures, Section 3 introduces the theoretical framework and the hypotheses tested in the experiment, Section 4 presents the results and Section 5 concludes.

3.2 Experimental design and procedures

3.2.1 Experimental game

Our experimental game consists of three parts. In the first part, participants have to perform a real effort task. In the second part, they play a dictator game. In the third part, we elicit beliefs about the average transfer in the dictator game. Instructions are in Appendix ??.

The real effort task

In the real-effort task, participants have to encrypt a letter into a number 120 times, with the help of a correspondence table displayed on the computer screen. For each new letter, the correspondence table changes, and is randomized at the individual level. If the answer is not correct, they have to enter a new number. Thus, every participants have to solve correctly 120 encryption tasks. We chose this real-effort task because it is both easy to explain and tedious. We fixed the number of letters such that most participants would complete the task in 8 to 10 minutes.

For this task, participants are paid a wage of 5, 10 or 15 euros, depending on the experimental condition. Participants receive an envelop containing either one, two or three €5 banknotes. The envelops were disposed in the cubicles before the session. Our experimental conditions vary the level and timing in which participants are informed of their actual wage for the first part.

The dictator game

In the second part, participants play a dictator game. All participants have to decide how to share a €5 endowment with another anonymous participant. Participants can give any amount between 0 and 5 euros, in increments of 0.1 euro. Participants are informed that they would be randomly matched at the end of the session with another participant. Within each pair, the decision of one participant would be implemented, while the other participant will be the receiver. This procedure enables us to collect the transfer decisions of every participants in an incentive-compatible way. The transfer decisions in this part constitute our main interest variable.

Belief Elicitation

In the third and final part, participants are asked to estimate the average transfer in their session. They were paid €1 if their prediction was equal to the actual average transfer,

plus or minus €0.1.

3.2.2 Experimental conditions

The experimental conditions vary across participants along two dimensions. Both dimensions are related to the wage for the real effort task. The first dimension varies the level of wage itself. Participants earn €5 in the low conditions, €10 in the medium conditions and €15 in the high conditions.

The second dimension varies the timing in which participants are informed of their actual wage. In the *NoRef* condition, participants are informed of their actual wage before the real effort task. The envelopes placed in the cubicles before the beginning of the session contain their actual and definitive wage. Participants are only informed of the wage they get and that every participants in the session get the same wage, but are not informed that other participants in other sessions get different wages.

In contrast, in the *Ref* conditions, participants are told at the beginning of the first part that they could earn a wage of either €5, €10 or €15 with equal probability ($\frac{1}{3}$). They are informed that the actual wage was determined for each sessions before the first session of the experiment, and that they would learn their actual wage at the end of the real effort task. The envelopes placed in the cubicles before the beginning of the session contained two €5 banknotes. In the *Ref_{low}* condition, an experimentalist withdraws one €5 banknote from the envelopes after the actual wage has been revealed to subjects. In the *Ref_{high}* condition, adds one €5 banknote. In the *Ref* conditions, the existence of other potential levels of wage is salient. Participants are able to compare the wage they actually get to the two other potential levels, and presumably encode a low wage as a loss and a high wage as a gain. Participants cannot make such counter-factual evaluation in the *NoRef* conditions.

To measure our treatment effect, we compare transfers in the *Ref* and *NoRef* conditions for each level of wage. This allows us to tightly control for potential wealth effects. By comparing transfers in *Ref_{low}* and in *NoRef_{low}*, we isolate the effect of experiencing a loss. By comparing the transfers in *Ref_{high}* and in *NoRef_{high}*, we isolate the effect of experiencing a gain. The two medium conditions are necessary to insure that the expected outcome in the *Ref* conditions is equal to ten. The expectation-based and status-quo based reference wage are thus equal to ten. The medium conditions also allows us to check whether being exposed to uncertainty about the actual wage for the real effort task has an autonomous impact on transfers.³

3.2.3 Additional measures

Measures of emotions In order to explore how the emotional states of participants are affected by our treatment variations, we use two complementary measures.

We record skin conductance, a measure of electro-dermal activity that proxies participants' emotional arousal (Bradley and Lang, 2000). We focus on the electro-dermal response to a specific event: when participants are informed (or reminded in the *NoRef* conditions) of the actual wage for the real effort task. The response is computed as the average electro-dermal activity in a ten-second window after the revelation (reminder) of the wage from which we subtract an individual baseline computed as the average activity in a thirty second window at the end of the real-effort task.

Recording electro-dermal response allows us to record any, conscious or unconscious, emotional changes. However, it is uninformative about the valence or the type of emotions experienced by participants. Our second measure of emotions is based on a declarative assessment adapted from Bosman and van Winden (2002) that aimed at associating types and valence to the physiological response. After the dictator game

³Individuals generally dislike being exposed to uncertainty (Loewenstein et al., 2001; Gneezy et al., 2006), and this could impact their behavior in the subsequent dictator game.

and before the beliefs elicitation, we ask participants to report their emotional state when they learned (or were reminded) of the actual wage for the real-effort task. On ten successive screens, we present to the participants one emotion and they have to report the intensity of this emotion on a 7-point scale, ranging from “no emotion at all” to “high intensity of emotion”. The 10 emotions were: anger, surprise, disappointment, joy, elation, jealousy, rage, frustration, irritation and sadness.

Online questionnaire on risk and loss aversion The participants reaction to falling below or exceeding the reference point might be affected by loss aversion, because participants who want to avoid losses ex-ante might also react more strongly to losses. Therefore, at least one day before the session in the laboratory, participants had to fill out an incentivized online questionnaire aiming at measuring risk and loss aversion according to the method introduced by Eckel and Grossman (2008). Participants had to choose one lottery in two consecutive sets of six binary lotteries displayed on their screen. The expected payoff of the lotteries ranged from €2.8 to €3.6, and variance increased with the expected payoff. The first set is used to elicit risk attitudes in the gain domain, since expected payoffs are always positive. The second set elicits risk attitudes in the loss domains. In order to keep the expected payoff equal, for the second set of lotteries participants were endowed with €4. Participants were told that their choice in one of the two sets of lottery would be chosen randomly for payment, and that they would be informed of the outcome of the lottery at the end of the lab session, when receiving their payment. We classify participants who choose less risky lotteries in the loss domain than in the gain domain as loss-averse. In addition to this two lottery choices, we collected participants’ demographics.

We separate these measures from the core of the experiment to avoid cross-contamination between behavior in the core of the experiment and the choices in the risk and loss aversion elicitation.

3.2.4 Procedures

The experiment was conducted at the physio-lab of Gate-lab. Participants were recruited using hroot (Bock et al., 2014). The online questionnaire was programmed with LimeSurvey (GmbH., GmbH.) and the laboratory experiment with JAVA. Before the first session, we randomly drew a sequence of conditions thanks to an algorithm that is available upon request. This algorithm ensures that each of the wage levels would be implemented one third of the time. Given this sequence, we planned our sessions by pairs: a *Ref* and a *NoRef* condition for the level of wage determined by the algorithm. The use of the algorithm was explained to the participants in the instructions to avoid deception.

In the invitation mailing, participants were informed that we would use physiological measures. We also told them that they would have to fulfill an online questionnaire at least one day before the session. Upon arrival, participants were given an informed consent form detailing the skin conductance measure. In order to participate, they had to explicitly tick the box indicating consent and sign the form. They were informed that they could freely leave the lab if they did not agree, in this case they would receive the amount corresponding to their choice in the online risk aversion task plus five euros. All participants gave their consent.

After consent, an experimentalist took participants to a bathroom where they were invited to wash their hands thoroughly with a special soap. They then drew a sticker from an opaque bag, which indicated the cubicle in which they were invited to seat. An experimentalist installed the electrodes on the index and middle finger of the left-hand of each participant. The temperature was maintained close to 23 degree Celsius in order to improve the accuracy of the electro-dermal measures. An experimentalist then distributed the instructions for the first two parts, these instructions were read aloud by the same experimenter and participants could ask questions.

Table 3.2.1 – Number of participants by experimental condition

| | low wage | medium wage | high wage |
|--------------|----------|-------------|-----------|
| <i>NoRef</i> | $N = 32$ | $N = 30$ | $N = 28$ |
| <i>Ref</i> | $N = 34$ | $N = 36$ | $N = 30$ |

Overall, 190 participants participated in the experiment. 57 % of the participants are female and the mean age is 23.67. Table 3.2.1 summarizes the number of participants in each experimental condition.⁴

Sessions were run between June and early October 2018. Our procedure insures that, for any given day in which we planned a session, we would have a *NoRef* condition for a given wage and the corresponding *Ref* condition.⁵

Sessions lasted on average 45 minutes. The average payoff is €20.5 (Standard Deviation, SD hereafter: 4.6) including a €5 show-up fee and an average of €3.41 (S.D. 1.9) for the online questionnaire. Table B.1 in Appendix B summarizes the demographic characteristics of participants across conditions.

3.3 Theoretical framework and hypotheses

3.3.1 The model

This section develops a reference-dependent model of giving.

Let us consider an individual i . The payoff of i , π_i , is the sum of two elements: the wage for the real effort task W and of the amount he keeps in the dictator game. This amount is the fraction $0 \leq d \leq 1$ of the endowment E . The profit of the recipient j , π_j , is the sum of the wage for the real effort task W and the amount $(1 - d)E$ that

⁴We invited 12 participants for each sessions, but due to the no-show, we do not have a perfect balance across conditions.

⁵There is one exception: for one session, the software crashed so that we had to trash the data and plan another session some days later to compensate.

the dictator transfers to him. W is determined exogenously and can take three values $W \in \{w_l, w, w_h\}$ with equal probability, with $w_l < w < w_h$ and $w - w_l = w_h - w$. It is common knowledge that i and j receive the same wage.

$$\pi_i = W + dE$$

$$\pi_j = W + (1 - d)E$$

i forms references about his payoff ($\bar{\pi}_i$) and the payoff of the receiver ($\bar{\pi}_j$) when he receives the instructions. These references depend on the wage he expects to get in the real effort task, and the hypothetical transfer he would make for this level of wage. Since w corresponds to both the expectation-based and the status-quo based wage, we assume that participants will take it as reference. We assume that i 's hypothetical transfer corresponds to the transfer i would choose if he would make this choice with no reference in mind, i.e. only with respect to his direct utility function in Equation 3.1.

$$v = m(\pi_i) + \alpha m(\pi_j) \tag{3.1}$$

In this utility function, $m(\cdot)$ represents i 's valuation of his profit with $m'(\cdot) > 0$ and $m''(\cdot) < 0$. $\alpha m(\cdot)$ represents i 's valuation of j 's payoff. $\alpha < 1$ captures the fact that i values his profit more than j 's profit. The value of π_i that maximizes i 's direct utility function is given by Equation 3.2.

$$m'(\pi_i) = \alpha m'(\pi_j) \tag{3.2}$$

i chooses the transfer d that respects the condition in Equation 3.2. Importantly, a higher wage allows i to increase both his and j 's payoff (a more formal proof is presented in Appendix D section D.1). Moreover, Equation 3.2 implies that $\pi_i > \pi_j$ and given that the wage for the real effort part is the same for both participants, participant i transfers less than one half of the endowment in the dictator game.

As explained previously, the hypothetical transfer \bar{d}_w that maximizes the direct utility function for the wage w , determines the references $\bar{\pi}_i$ and $\bar{\pi}_j$.

At the beginning of the dictator game, i is informed of the actual wage for the real effort task, $W \in \{w_l, w, w_h\}$ and chooses his actual transfer d given his utility function. This utility is given by Equation 3.3.

$$\begin{aligned}
 u = & m(\pi_i) + \alpha m(\pi_j) + \mu(m(\pi_i) - m(\bar{\pi}_i)) \\
 & + \mu(\alpha m(\pi_j) - \alpha m(\bar{\pi}_j))
 \end{aligned}
 \tag{3.3}$$

This utility function has the classical features of reference dependent models. $\mu(\cdot)$ is the gain-loss utility function, which captures the impact of exceeding or falling below the reference level of own and others' profit. An important implication of this utility function is that i evaluates his and j 's profit compared to the endogenous references, $\bar{\pi}_i$ and $\bar{\pi}_j$.

Our utility function makes three implicit assumptions common in the literature (see e.g. Köszegi and Rabin, 2006). First, gains and losses are expressed in terms of utility. Second, the gain-loss utility function is additively separable in its arguments. Third, the gain-loss utility function is the same for both components of the direct utility function. We assume a classical shape for the gain-loss utility function:

$$\mu(X) = \begin{cases} X^\gamma & \text{if } X \geq 0 \\ -\lambda(-X)^\beta & \text{if } X < 0 \end{cases}$$

$\lambda \geq 1$ captures the fact that individuals may be more sensitive to losses than to gains. $0 < \gamma < 1$ and $0 < \beta < 1$ capture diminishing sensitivity respectively in the gain and in the loss domains. Note that the utility function u so defined is continuous and concave in π_i for values of β close to 1, a local extremum of this function is thus a global

extremum (proofs are presented in Appendix D section D.2).⁶

Equation 3.4 reports the derivative of the reference-dependent utility function defined in Equation 3.3. The analysis of this derivative at different points allows us to formulate hypotheses on the effect of falling below or exceeding the reference payoff on the subsequent transfer.

$$\begin{aligned}
 u' &= m'(\pi_i) - \alpha m'(\pi_j) \\
 &+ m'(\pi_i)\mu'(m(\pi_i) - m(\bar{\pi}_i)) \\
 &- \alpha m'(\pi_j)\mu'(\alpha m(\pi_j) - \alpha m(\bar{\pi}_j))
 \end{aligned} \tag{3.4}$$

Proposition 1. *If participant i gets his reference wage, he transfers the amount he initially planned to transfer. By doing so, he insures that his and j 's profits are equal to the references. The intuition is that giving more would lead to a loss in terms of i 's own profit which would not be compensated by the gain in j 's profit. Symmetrically, giving less would lead to a gain in terms of i 's profit that is not sufficient to compensate the loss in terms of j 's profit.*

Proof. We study the derivative of the utility function in Equation 3.4 at the point that corresponds to the transfer that maximizes the direct utility function. By definition, at this point $m'(\pi_i) - \alpha m'(\pi_j) = 0$ and $\pi_i = \bar{\pi}_i$ and $\pi_j = \bar{\pi}_j$. Thus, the derivative is null. Since u is concave and continuous, this constitutes the transfers that maximizes i 's utility function. □

Proposition 2. *An individual who receives w_1 as a wage for the real effort part while having $\bar{w} > w_1$ as a reference transfers less than an individual who receives w_1 while having w_1 as a reference. The intuition is the following: because i values his own profit more than j 's profit, i is more sensitive to the loss of some of his own profit than to the loss of some of j 's profit. As a consequence, i reduces his transfers to j in order to reduce the loss in terms of his own profit.*

⁶Both the bounds on the value of α and β and the fact that β is close to 1 reflects empirical calibrations. See Fox and Poldrack (2014) for an overview of the estimates in the literature.

Proof. Consider an individual who planned to have a final profit $\bar{\pi}_i = \bar{w} + \bar{d}_{\bar{w}}$, but actually receives w_l . We show that this individual transfers less than an individual who planned a profit $\bar{\pi}_i = \bar{w}_l + \bar{d}_{\bar{w}_l}$. To do so, we study the sign of the derivative of the utility function for an expected wage of \bar{w} , an actual wage of w_l and a transfer of $\bar{d}_{\bar{w}}$. At this point, Equation 3.4 yields:

$$\begin{aligned} u' &= m'(\pi_i) - \alpha m'(\pi_j) \\ &+ m'(\pi_i)\beta(-\lambda)(-1)(m(\bar{\pi}_i) - m(\pi_i))^{\beta-1} \\ &- m'(\pi_j)\alpha\beta(\lambda)(\alpha m(\bar{\pi}_j) - \alpha m(\pi_j))^{\beta-1} \end{aligned} \quad (3.5)$$

As shown previously, $m'(\pi_i) - \alpha m'(\pi_j) = 0$. The sign of the derivative presented in Equation 3.5 depends only on the sign of $m'(\pi_i)\beta(-\lambda)(-1)(m(\bar{\pi}_i) - m(\pi_i))^{\beta-1} - m'(\pi_j)\alpha\beta(\lambda)(\alpha m(\bar{\pi}_j) - \alpha m(\pi_j))^{\beta-1}$. Using the fact that $m(\cdot)$ is concave and that $\pi_i > \pi_j$, we can show that this expression is positive (Some more details are given in Appendix D section D.3). This means that when i expects a given wage but actually receives a lower wage, i transfers less than an individual who expects *and* gets the low wage. Note that the former expression has a direct interpretation. $m'(\pi_i)\beta(-\lambda)(-1)(m(\bar{\pi}_i) - m(\pi_i))^{\beta-1}$ represents the extra utility of keeping slightly more than $\bar{d}_{\bar{w}_l}E$, which closes the gap between the profit i expects and the profit i actually gets. $-(m'(\pi_j)\alpha\beta(\lambda)(\alpha m(\bar{\pi}_j) - \alpha m(\pi_j))^{\beta-1})$ represents the utility cost of transferring slightly less than $(1 - \bar{d}_{\bar{w}_l})$, which widens the gap between what i expects the recipient would get and the amount the recipient actually gets. Below the reference point, i prefers closing the gap between what he gets and what he was expecting to get, even if it widens the gap between the recipient's profit and what i expected the recipient would get. \square

Proposition 3. *An individual who receives w_h as a wage for the real effort part while having the reference $\bar{W} < w_h$ transfers more than an individual who receives w_h while having the reference w_h . The intuition is that the combination of the concavity of $m(\cdot)$ and diminishing sensitivity makes the extra, unexpected units of profit more valuable when transferred to the recipient rather than kept.*

Proof. Consider an individual who planned to have a profit $\bar{\pi}_i = \bar{w} + \bar{d}_{\bar{w}}$, but actually receives w_h . This individual transfers more than an individual who planned a profit $\bar{\pi}_i = \bar{w}_h + \bar{d}_{\bar{w}_h}$. To do so, we study the sign of the derivative of the utility function for an expected wage of \bar{w} , an actual wage of w_h and a transfer of $\bar{d}_{\bar{w}_h}$. At this point, the Equation 3.4 yields:

$$\begin{aligned} u' &= m'(\pi_i) - \alpha m'(\pi_j) \\ &+ m'(\pi_i)\gamma(m(\pi_i) - m(\bar{\pi}_i))^{\gamma-1} \\ &- m'(\pi_j)\alpha\gamma(\alpha m(\pi_j) - \alpha m(\bar{\pi}_j))^{\gamma-1} \end{aligned} \tag{3.6}$$

The sign of the derivative depicted in Equation 3.6 is determined by the sign of $m'(\pi_i)\gamma(m(\pi_i) - m(\bar{\pi}_i))^{\gamma-1} - m'(\pi_j)\alpha\gamma(\alpha m(\pi_j) - \alpha m(\bar{\pi}_j))^{\gamma-1}$. Using a similar argument than previously, we can show that this expression is negative. This means that when i expects a given wage but actually receives a higher wage, i transfers more than an individual who expects *and* gets the high wage. The intuition for this result is that, because $m(\cdot)$ is concave, and because sensitivity to gains diminishes with gains, and that i 's profit is always greater than j 's, a gain in terms of π_j procures a greater utility to i than a gain in terms of his own payoff. \square

3.3.2 Hypotheses

The previous theoretical framework leads to the following hypotheses:

Hypothesis 1: Participants transfer the same amount in the Ref_{med} and $NoRef_{med}$ condition. This follows from Proposition 1.

Hypothesis 2: Participants in the Ref_{low} condition transfer less than participants in the $NoRef_{low}$. This follows from proposition 2.

Hypothesis 3: Participants in the Ref_{high} condition transfer more than participants in the $NoRef_{high}$ condition. This follows from proposition 3.

3.4 Results

3.4.1 Behavioral results

The average transfer in the dictator game is €1.37 (S.D. 1.40). This represents 27% of the €5 endowment, which is well in line with the literature.⁷ Table 3.4.1 reports the average transfer of participants across conditions.

Table 3.4.1 – Average transfers by experimental condition

| | Wage=5 | Wage=10 | Wage=15 |
|--------------|-------------------------|------------------------|------------------------|
| <i>NoRef</i> | 1.64 (N=32, S.D.= 1.52) | 1.14 (N=30, S.D.=1.14) | 1.32 (N=28, S.D.=1.37) |
| <i>Ref</i> | 1.28 (N=34, S.D.=1.53) | 1.18 (N=36, S.D.=1.51) | 1.69 (N=30, S.D.=1.28) |
| p-value | 0.26 | 0.94 | 0.29 |
| Overall | 1.45 (N=66, S.D.=1.52) | 1.16 (N=66, S.D.=1.34) | 1.51 (N=58, S.D.=1.32) |

Notes: N is the number of observations, S.D. is the standard deviation. *p* – values for two-sided Mann-Whitney tests.

While the tendencies match our predictions, none of the comparisons between the *NoRef* and *Ref* conditions reach statistical significance.

We ran regressions explaining transfers, pooling all the data from the experiment. We interact a dummy variable indicating the *Ref* conditions, with a categorical variable indicating the wage level. In Table 3.4.2, we report the marginal treatment effect for the different levels of wage. Models (1) and (2) are left-censored Tobit models (38% of the transfers are equal to zero). Models (3) and (4) are Logit models, explaining the decision to transfer zero. All models include a categorical variable accounting for month fixed effect.⁸ Models (2) and (4) include individual characteristics. We control for gender, age, the number of previous participations in economic experiments, risk aversion, and

⁷In a meta-analysis including more than 130 papers and 600 treatments, Engel (2011) finds that dictators give on average 28% of their endowment.

⁸We ran sessions in early summer which corresponds to the end of the academic year, and in early fall which corresponds to the beginning of the academic year. The participants differ both on observable and non-observable characteristics, which is why we control for month fixed-effects.

Table 3.4.2 – Treatment effects over the different levels of wage.

| | (1) Transfer | (2) Transfer | (3) Transfer=0 | (4) Transfer=0 |
|-----------------------------|---------------------|----------------------|----------------------|----------------------|
| <i>Treatment effect at:</i> | | | | |
| Wage=5 | -0.558* (0.319) | -0.379 (0.410) | 0.165** (0.078) | 0.146 (0.093) |
| Wage=10 | 0.009 (0.286) | -0.175 (0.292) | 0.028 (0.069) | 0.080 (0.067) |
| Wage=15 | 0.614*** (0.230) | 0.816**** (0.208) | -0.163*** (0.050) | -0.199*** (0.069) |
| Month F.E. | YES | YES | YES | YES |
| Individual Char. | NO | YES | NO | YES |
| N | 190 | 190 | 190 | 190 |

Robust standard errors clustered at the session level in parentheses. (1) and (2) are Tobit models. (3) and (4) are Logit models. Marginal effects reported. Treatment effect corresponds to the marginal effect of the *Ref* condition. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

include dummies indicating whether the participant is loss averse⁹ and whether the participant is a student. Robust standard errors are clustered at the session level.¹⁰

The outcome of these regressions supports the three following behavioral results:

Result 1: For a €10 wage, the *Ref* treatment has no impact on transfers. This result supports Hypothesis 1. None of the parameters capturing treatment effect for the €10 wage is significant in the regressions. This result suggests that being merely exposed to payoff uncertainty does not impact the participants' propensity to share.¹¹ As a consequence, the treatment effects potentially found for the two other levels of wage are not confounded by the effect of being exposed to uncertainty.

⁹Participants who chose a lottery with less variance in the loss domain than in the gain domain in the online questionnaire are categorized as loss averse.

¹⁰There are several reasons to cluster standard errors at the session level even in one-shot experiments (see e.g. Fréchette, 2012). For instance, the experimental procedure to set-up the physiological measures implied interactions between participants and with the experimenter that are hardly measurable and made the room composition in terms of gender salient which can impact behavior (Castillo et al., 2015).

¹¹One could have expected lower transfers in the *Ref* condition than in the *NoRef* condition for the €10 wage because individuals dislike being merely exposed to uncertainty (see the "uncertainty effect" presented in Gneezy et al., 2006).

Table 3.4.3 – Treatment effect on loss averse *vs.* not loss averse participants, at wage=€5.

| | (1) Transfer | (2) Transfer | (3) Transfer=0 | (4) Transfer=0 |
|---|----------------------|----------------------|--------------------|-------------------|
| <i>Treatment effect, at wage=5 for:</i> | | | | |
| Not loss averse participants | 0.013 (0.240) | 0.125 (0.384) | 0.069* (0.036) | 0.048 (0.056) |
| Loss averse participants | -1.185*** (0.452) | -0.983*** (0.299) | 0.237** (0.120) | 0.232 (0.156) |
| <i>p</i> – value diff. | 0.011 | <0.001 | 0.094 | 0.21 |
| Month F.E. | YES | YES | YES | YES |
| Individual Char. | NO | YES | NO | YES |
| <i>N</i> | 190 | 190 | 190 | 190 |

Robust standard errors clustered at the session level in parentheses. (1) and (2) are Tobit models. (3) and (4) are Logit models. Marginal effects reported. Treatment effect corresponds to the marginal effect of the *Ref* condition. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Result 2: For a €5 wage, the “Ref” treatment reduces transfers of loss-averse participants only. The data provide mixed-support to hypothesis 2. The treatment effect is always in the expected direction, but reaches statistical significance only in the models (1) and (3) that do not control for individual characteristics.

As €5 is the lowest potential wage, it is probably encoded as a loss by participants. Loss-averse participants transferred less than non loss-averse participants in the *NoRef_{low}* condition (1.7 *vs.* 0.69, MW: $p = 0.056$), which suggests that loss-averse participants are more sensitive to the treatment. To assess this claim, we run regressions similar to those reported in Table 3.4.2, except that we add a third term to the interactions between treatment and the level of the wage: a dummy variable indicating that the participant is loss averse. The estimates are reported in Table 3.4.3.

We conclude that the effect of experiencing a loss on the transfer decision is more marked for loss-averse participants. This result has an intuitive interpretation: those who try to avoid losses (as evidenced by their choices in the online questionnaire) are also those who want to “chase” losses once they experience them. Here, the only way in

which they can chase losses in terms of their own profit is by transferring less money as dictators.

Result 3: For the €15 wage, participants transfer larger amounts in the *Ref* condition compared to the *NoRef* condition. This result supports Hypothesis 3. The treatment effects for a wage of €15 are large and significant in every specification.

3.4.2 Emotions

We now turn to the analysis of the emotion data. Experiencing losses may trigger negative emotions that could in turn reduce transfers, while experiencing gains may trigger positive emotions that could in turn increase transfers. In what follows, we assess the role of emotions in the treatment effects exposed above.

We first analyze the physiological data.¹² We focus on the electro-dermal activity consecutive to the revelation (or reminder in the *NoRef* conditions) of the actual wage for the real effort task. Figure 3.4.1 plots the normalized activity.

Electro-dermal activity is always greater in the *Ref* than in the *NoRef* condition (MW: p – values are lower than 0.001 for every levels of wage). This shows that the revelation of the wage generated physiological response. However, electro-dermal activity does not differ across level of wage, nor for the *Ref* conditions (KW: $p = 0.713$), neither for the *NoRef* conditions (KW: $p = 0.749$).

In order to associate a valence to the emotional arousal identified in the previous paragraph, we now turn to the analysis of self-reported data. We separate the item of the questionnaire in “positive emotions” (Joy, elation, surprise) and “negative emotions” (Anger, disappointment, rage, irritation, jealousy, frustration, sadness). For each individuals, we average the responses to “positive emotions” items and to “negative emotions”

¹²We drop 22 individual observations from the analysis because of the poor quality of their electro-dermal signal, which represents 11.5% of our sample.

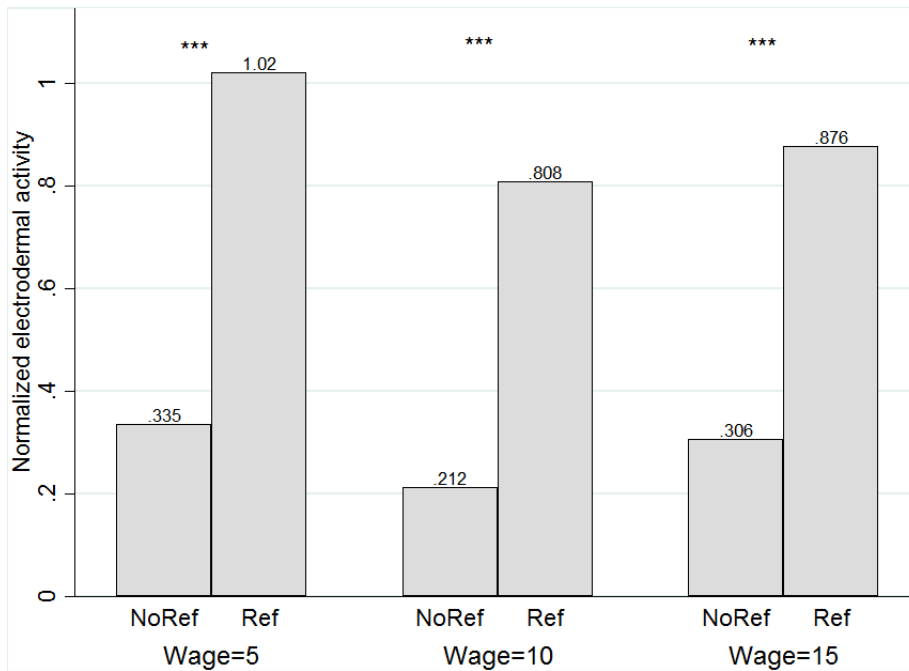


Figure 3.4.1 – Electrodermal activity across conditions

Notes: *** indicates $p < 0.01$ in a Mann-Whitney two-sided test.

items, which gives an individual index of “positive emotions” and an individual index of “negative emotions”.¹³ Figure 3.4.2 plots the positive (top panel) and negative (bottom panel) emotion indexes.

In the *NoRef* conditions, the wage level does not impact positive emotions (Kruskal-Wallis test, KW hereafter: $p = 0.327$) and mildly impacts negative ones (KW: $p = 0.08$). In the *Ref* conditions, both positive and negative emotions strongly depend on the level of wage (KW: $p < 0.001$ for both positive emotions, and negative emotions).

Both positive and negative self-reported emotions differ between the *NoRef* and the *Ref* conditions. Participants express more negative emotions and less positive emotions when they know that they could have earned more than the wage they get (two first

¹³The questionnaire has a good reliability overall (Cronbach’s alpha=0.84). In Appendix B Table B.2, more details are given about the structure of correlation between the items from the questionnaire.

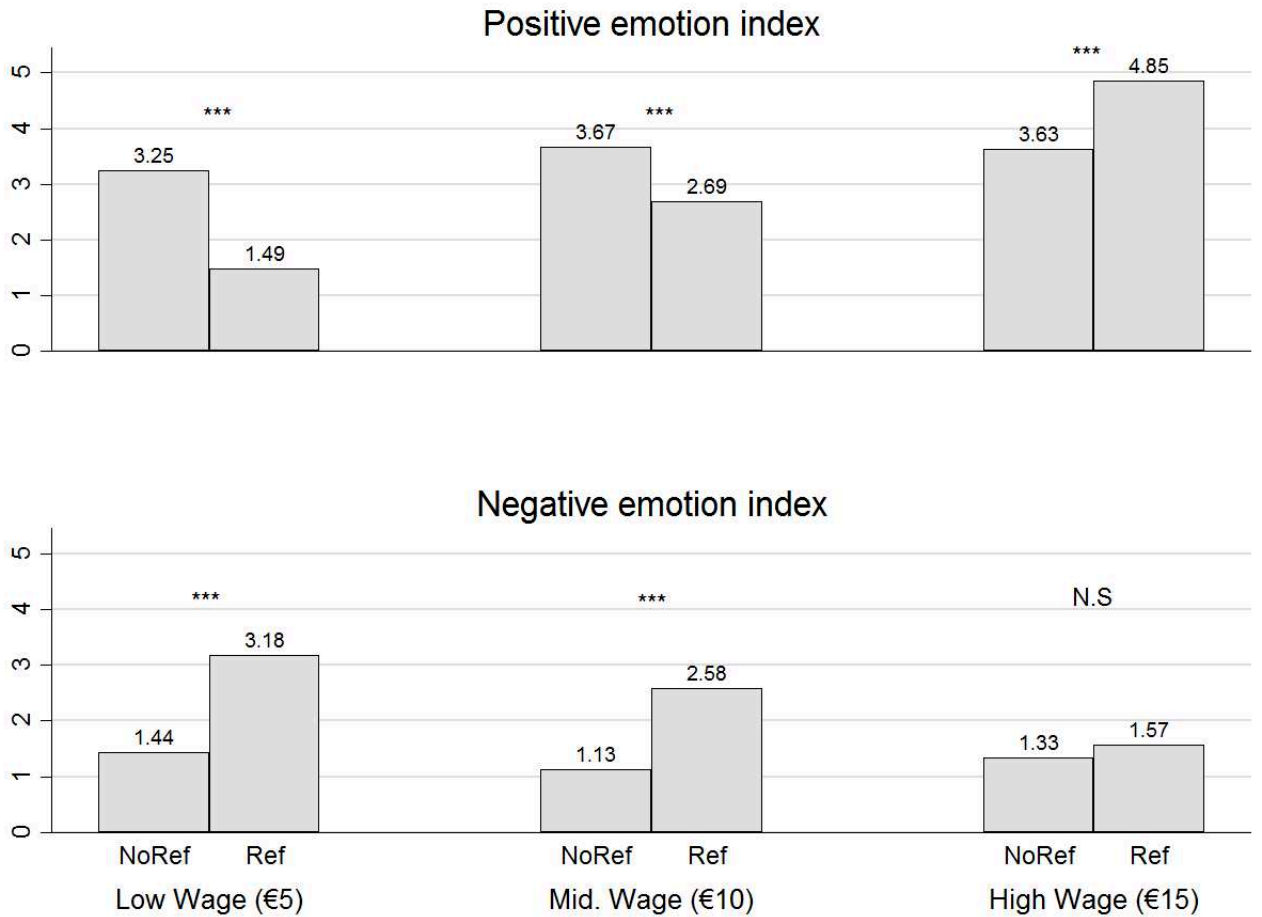


Figure 3.4.2 – Indexes of positive (top) and negative emotions across conditions

Notes: Index of positive emotions: individual average self-report on a 1-7 scale for Joy, Elation and Surprise. Index of negative emotions: individual average self-report on a 1-7 scale for anger, disappointment, rage, irritation, jealousy, frustration and sadness. *** indicates $p < 0.01$ in a Mann-Whitney two-sided test.

comparisons of the top and bottom panels). In contrast, when they get the maximum wage, participants express more positive emotions when they are informed of the other potential wages than when they are not. In both the Ref_{low} and Ref_{mid} , the highest reports were for disappointment, and disappointment was greater in the Ref_{low} than in the Ref_{high} (5.14 vs. 3.58, MW: $p < 0.001$). In the other conditions, the highest reports were always for joy, and the highest joy was reported in the Ref_{high} condition (5.32). Figure C.1 in Appendix C details the average reports for each emotions, across all conditions.

By combining the data from the self-reported and physiological measures of emotion, we can conclude that the salience of the counter-factual wages influenced the emotional state of participants: (i) when participants earn the low wage, informing them of the other wages triggers negative emotions, in particular dissatisfaction ; (ii) when participants earn the high wage, informing them of the other potential wages triggers positive emotions, in particular joy.

In order to explore the role of the emotions experienced at the time of revelation in the treatment effects, we test whether emotional arousal correlates with the decisions in the trust game in the Ref conditions. We add a third term to the interactions between treatment and the level of wage used in Table 3.4.2: the normalized physiological measure of emotion. Table 3.4.4 reports the marginal effects of the normalized physiological measure of emotion for the different levels of wage conditional on being in the treatment. In models (1) and (2), we estimate Tobit models. In models (3) and (4) we estimate Logit models. Models (1) and (3) include month fixed effects. Model (2) and (4) also control for the individual characteristics used in Table 3.4.2.

We find no evidence that emotional arousal impacts transfers. This suggests that the emotions triggered by the treatment do not translate into behaviors.

Table 3.4.4 – Marginal effects of physiological arousal on transfers over the different level of wage.

| | (1) Transfer | (2) Transfer | (3) Transfer=0 | (4) Transfer=0 |
|--|-------------------|-------------------|-------------------|-------------------|
| <i>effect of electro-dermal act. at:</i> | | | | |
| Wage=5 | -0.418 (0.914) | -0.227 (0.840) | 0.015 (0.169) | -0.017 (0.147) |
| Wage=10 | 0.260 (0.928) | 0.081 (0.927) | -0.085 (0.179) | -0.044 (0.195) |
| Wage=15 | -0.233 (0.488) | -0.185 (0.414) | 0.005 (0.093) | -0.004 (0.086) |
| Month F.E. | YES | YES | YES | YES |
| Individual Char. | NO | YES | NO | YES |
| N | 168 | 168 | 168 | 168 |

Robust standard errors clustered at the session level in parentheses. (1) and (2) are Tobit models. (3) and (4) are Logit models. Marginal effects reported. Treatment effect corresponds to the marginal effect of the *Ref* condition. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

3.5 Discussion and Conclusion

In this paper, we tested whether the salience of better or worse counter-factual wages in a real effort task affects giving in a subsequent dictator game. We found that making salient that the wage could have been lower increases the transfer of participants who receive the high wage. Symmetrically, participants who receive the low wage transfer less when they are aware that their were higher potential levels of wage than when they are not, but the effect is concentrated on loss-averse participants. Our results do not support the intuition that losses have greater consequences than gains.

The effects are found by varying the information about the counter-factual levels of wage between participants, holding the actual wage constant. As a consequence, wealth effects do not confound our results. By design, we also mute reciprocity concerns since the recipients in the dictator game cannot be held accountable for the actual wage in the real effort task. In addition, wage uncertainty does not drive our results, since the transfers of participants who get the reference wage do not vary with the timing in which

they are informed of their actual wage. The physiological and self-reported emotional data show that our treatments triggers emotional reactions, but these reactions have little explanatory power. A higher emotional arousal at the time of wage revelation (or reminder) does not affect behavior in the dictator game. This shows that our results are not merely an avatar of the effect of emotions or mood on other-regarding behaviors (On this topics, see e.g Charness and Grosskopf, 2001; Capra, 2004; Kirchsteiger et al., 2006; Pérez-Dueñas et al., 2018). A possible interpretation of our results is that earning a wage that compares favourably to other possible levels of wage is encoded as a gain by participants and dictators tend to share gains. On the other hand, earning a wage that compares unfavorably with other possible levels is encoded as a loss and dictators reduce their transfers in order to compensate for this loss.

From a methodological stand point, some features of our experiment call for discussion. First, narrow bracketing (e.g. Read et al., 1999; Thaler, 1999) might have attenuated the effects because we gave an extra endowment for the dictator game instead of having the gains or losses affect the endowment directly. Possibly, some participants allocate the gain in the real effort task and the gain in the dictator game to two separate mental accounts and separate the decisions in these two parts. Second, participants in our experiment seem to perform counter-factual evaluation of their wage. The results are consistent with the hypothesis that participants compare their wage to the medium wage, which is both salient and the expectation-based and the status-quo-based reference wage. However, our aggregate results might hide individual heterogeneity in how the counter-factual evaluation is performed (see e.g. Baillon et al., 2018). As an illustration, it is possible that some participants encode the medium wage as a gain because their reference is the low wage, while some may encode it as a loss because their reference is the high wage. A possible extension of our research is to vary the number of potential wages within participants in order to refine our understanding of counter-factual evaluation.

Our results contribute to the literature studying social preferences, and more specifically how the context of the decision impacts giving (e.g. List, 2007; Bardsley, 2008). Our manipulation allows to observe the transfers of dictators whose payoffs differ from their reference, and the data point to the existence of a trade-off between social preferences and the natural tendency of loss averse people to chase losses: losses relative to their reference payoff reduce the concern for other while gains increase it. Self-commitment to act altruistically seems to yield in front of the disutility linked to incurring private losses.

Our results also speaks to the literature studying how workers perceive their wages relative to some references. Workers are influenced by how their wage compares to their peers' wages, with consequences for effort provision (Bracha et al., 2015) and job satisfaction (Card et al., 2012). Workers also compare their actual wage to the wage they use to have, as evidenced by the literature on the effects of wage cuts and raises on effort provision (e.g Gneezy and List, 2006; Kube et al., 2012; DellaVigna et al., 2016). The given interpretation is that workers reciprocate to the principal's intentions by adjusting their level of effort. In our experiment, we mute inequality concerns, as all the participants in a given session earn the same wage ; and reciprocity, as the recipient is not accountable for the actual level of wage. As a consequence, we provide evidence that workers may compare their wage to the wages they could have received in a different state of the world, independently of social comparison or wage dynamic. Moreover, losses and gains in terms of wage affect social preferences, rather than effort provision. This means that how workers evaluate their wage relative to a reference might impact the social capital of firms, and promote or impede cooperation between workers.

Bibliography

- Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference Points and Effort Provision. *American Economic Review* 101(April), 470–492.
- Andrade, E. B. and D. Ariely (2009). The enduring impact of transient emotions on decision making. *Organizational Behavior and Human Decision Processes* 109(1), 1–8.
- Andreoni, J. and B. D. Bernheim (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica* 77(5), 1607–1636.
- Baillon, A., H. Bleichrodt, and V. Spin (2018). Searching for the reference point. *Management Science*. Forthcoming.
- Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics* 11(2), 122–133.
- Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71(C), 117–120.
- Bolton, G. E. and A. Ockenfels (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review* 90(1), 166–193.
- Bosman, R. and F. van Winden (2002). Emotional Hazard in a Power-to-Take Experiment. *The Economic Journal* 112(476), 147–169.
- Brañas Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology* 28(4), 477 – 486.
- Bracha, A., U. Gneezy, and G. Loewenstein (2015). Relative pay and labor supply. *Journal of labor economics* 33(2), 297–315.
- Bradley, M. M. and P. J. Lang (2000). Measuring emotion: Behavior, feeling, and physiology. In *Cognitive neuroscience of emotion.*, Series in affective science., pp. 242–276. New York, NY, US: Oxford University Press.
- Buffat, J. (2016). Charitable preferences and expectation-based reference dependence: An experiment on charitable giving. Working Paper.
- Capra, C. M. (2004). Mood-driven behavior in strategic interactions. *The American Economic Review* 94(2), 367–372.
- Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review* 102(6), 2981–3003.
- Carlsson, F., H. He, and P. Martinsson (2013, Jun). Easy come, easy go. *Experimental Economics* 16(2), 190–207.

- Castillo, M., G. Leo, and R. Petrie (2015). Room composition effects on risk taking by gender.
- Charité, J., R. Fisman, and I. Kuziemko (2016). Reference points and redistributive preferences: Experimental evidence. *NBER Working Paper No. 21009*.
- Charness, G. and B. Grosskopf (2001). Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization* 45(3), 301–328.
- Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics* 117(3), 817–869.
- Cherry, T. L., P. Frykblom, and J. F. Shogren (2002). Hardnose the dictator. *The American Economic Review* 92(4), 1218–1221.
- Cherry, T. L. and J. F. Shogren (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters* 101(1), 69 – 72.
- Choi, S. (2014). Reference-dependent preferences and charitable giving.
- Dana, J., R. A. Weber, and J. X. Kuang (2007, Oct). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1), 67–80.
- DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2016). Estimating social preferences and gift exchange at work. Technical report, National Bureau of Economic Research.
- Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization* 68(1), 1 – 17.
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics* 14(4), 583–610.
- Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114(3), 817–868.
- Filiz-Ozbay, E. and E. Y. Ozbay (2007). Auctions with anticipated regret: Theory and experiment. *American Economic Review* 97(4), 1407–1418.
- Fox, C. R. and R. A. Poldrack (2014). Appendix: Prospect theory and the brain. In P. W. Glimcher and E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (Second Edition ed.), pp. 533 – 567. San Diego: Academic Press.
- Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics* 15(3), 485–498.
- Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2003). Explaining altruistic behavior in humans. *Evolution and human Behavior* 24(3), 153–172.

- GmbH., L. Limesurvey: An open source survey tool. LimeSurvey GmbH.
- Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5), 1365–1384.
- Gneezy, U., J. A. List, and G. Wu (2006). The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. *The Quarterly Journal of Economics* 121(4), 1283–1309.
- Heidhues, P. and B. Köszegi (2008). Competition and price variation when consumers are loss averse. *American Economic Review* 98(4), 1245–68.
- Hoffman, E., K. McCabe, K. Shachat, and V. Smith (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7(3), 346 – 380.
- Kahneman, D. and A. Tversky (1979a). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–291.
- Kahneman, D. and A. Tversky (1979b). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–292.
- Kirchsteiger, G., L. Rigotti, and A. Rustichini (2006). Your morals might be your moods. *Journal of Economic Behavior & Organization* 59(2), 155–172.
- Köszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121(4), 1133–1165.
- Kube, S., M. A. Maréchal, and C. Puppe (2012). Do wage cuts damage work morale? evidence from a natural field experiment. *Journal of the European Economic Association* 11(4), 853–870.
- List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy* 115(3), 482–493.
- Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch (2001). Risk as feelings. *Psychological bulletin* 127(2), 267.
- Ockenfels, A. and P. Werner (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior and Organization* 108, 453–462.
- Pérez-Dueñas, C., M. F. Rivas, O. A. Oyediran, and F. García-Torres (2018). Induced negative mood increases dictator game giving. *Frontiers in Psychology* 9, 1542.
- Persson, E. (2018). Testing the impact of frustration and anger when responsibility is low. *Journal of Economic Behavior & Organization* 145, 435 – 448.
- Read, D., G. Loewenstein, and M. Rabin (1999). Choice bracketing. *Journal of Risk and Uncertainty* 19(1/3), 171–197.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making* 12(3), 183–206.

Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics* 106(4), 1039–1061.

A Instructions

Thank you for participation in this experiment on decision making. You are not allowed to communicate with other participants for the length of the experimental session. Your decisions are anonymous. In addition, because we record your skin conductance, please avoid making noise throughout of the session.

The session is composed of three parts. The instructions hereafter describe the content of Parts 1 and 2. The instructions for the third part will be displayed on your screen.

Your decisions can make you earn money. Your payoffs, except for Part 1, will be paid in cash, in private at the end of the session and in a separate room.

We will read the instructions together. Should you have any question, raise your hand. An experimentalist will come to you.

Part 1

For this part, you have to perform 120 encoding tasks. The task consists in encoding a letter from the alphabet into a number. To know the number corresponding to the letter, you must use a correspondence table that is displayed on your screen. After inputting the letter in the dedicated field, you must validate your answer. If your answer is correct, you go on to the next letter; otherwise you have to input a new number. For each letter, the correspondence table changes.

[The next paragraph depends on the conditions.]

[NoRef condition (wage=€5).]

For this task, you will receive a payoff. An envelop containing €5 was placed on your desk for this purpose. The content of the envelop of every participants in the session is

the same. The content of the envelop is your payoff for the first part. The content of the envelop will be reminded to you at the end of Part 1.

[NoRef condition (wage=€10).]

For this task, you will receive a payoff. An envelop containing €10 was placed on your desk for this purpose. The content of the envelop of every participants in the session is the same. The content of the envelop is your payoff for the first part. The content of the envelop will be reminded to you at the end of Part 1.

[NoRef condition (wage=€15).]

For this task, you will receive a payoff. An envelop containing €15 was placed on your desk for this purpose. The content of the envelop of every participants in the session is the same. The content of the envelop is your payoff for the first part. The content of the envelop will be reminded to you at the end of Part 1.

[Ref conditions.]

For this task, you will receive a payoff. An envelop containing €10 was placed on your desk for this purpose. Nevertheless, at the end of the encoding tasks, a random draw will determine your definitive payoff for this part. This random draw was done at the beginning of the set up of the experiment, before the first session, for every sessions. There is 1 chance out of 3 that the content of the envelop remain unchanged, 1 chance out of 3 that we withdraw €5 from the envelop and 1 chance out of 3 that we add €5 in your envelop.

The random draw is at the session level, so that the content of every participant in a

given session is the same. The random draw does not depend on your behavior or on the behavior of other participants. The final content of the envelope is your payoff for Part 1.

Part 2

In this part, the decisions involve a person A and a person B. Person A receives a €5 endowment. She decides how much she wants to transfer to Person B. Every amount between 0 and 5 are possible, rounded to 10 cents. Person A keeps the amount she does not transfer.

Person B does not receive an endowment. She earns the amount that Person A transferred. She has no decision to make.

Every participants in the session take the decision as Person A. At the end of the session, the software will randomly match participants of the session in pairs. For each pair, the software will randomly select a Person A and a Person B. For Person A, her decision will determine her payoff for this part. For Person B, the decision of the Person A to whom she is matched will determine her payoff for this part.

Payoff for Part 2 will be payed at the end of the session.

B Tables

Table B.1 – Demographic characteristics.

| Sessions | Overall | Summer | Fall | <i>p</i> – value |
|------------------|---------|---------------|---------------|------------------|
| Prop. Female | 0.57 | 0.56 | 0.58 | 0.878 |
| Prop. Student | 0.82 | 0.66 | 0.91 | < 0.001 |
| Age | 23.67 | 26.9 | 21.95 | < 0.001 |
| Participation | 3.48 | 6.43 | 1.91 | < 0.001 |
| <i>N</i> | 190 | 66 | 124 | - |
| <i>N</i> Session | 18 | 6 | 12 | - |
| Wage=5 | 6 | NoRef:2 Ref:2 | NoRef:1Ref:1 | - |
| Wage=10 | 6 | 0 | NoRef:3 Ref:3 | - |
| Wage=15 | 6 | NoRef:1 Ref:1 | NoRef:2 Ref:2 | - |

Note: *p* – values for Mann-Withney tests for continuous variables, Fisher-Exacts test for dichotomous variables.

Table B.2 – Pairwise correlations between responses in the emotion questionnaire.

| | Anger | Surprise | Disap. | Joy | Elation | Jealousy | Rage | Frust. | Irrit. | Sad. |
|-------------|-----------|----------|-----------|-----------|-----------|----------|----------|----------|----------|------|
| Anger | 1 | | | | | | | | | |
| Surprise | 0.185** | 1 | | | | | | | | |
| Disap. | 0.594*** | -0.018 | 1 | | | | | | | |
| Joy | -0.207*** | 0.416*** | -0.544*** | 1 | | | | | | |
| Elation | -0.055 | 0.377*** | -0.290*** | 0.650*** | 1 | | | | | |
| Jealousy | 0.547*** | 0.050 | 0.632*** | -0.239*** | -0.106 | 1 | | | | |
| Rage | 0.668*** | 0.071 | 0.548*** | -0.156** | -0.0146 | 0.642*** | 1 | | | |
| Frustration | 0.572*** | -0.002 | 0.762*** | -0.363*** | -0.216*** | 0.668*** | 0.564*** | 1 | | |
| Irritation | 0.779*** | 0.054 | 0.615*** | -0.243*** | -0.144** | 0.672*** | 0.691*** | 0.713*** | 1 | |
| Sadness | 0.679*** | 0.054 | 0.640*** | -0.206*** | -0.096 | 0.677*** | 0.694*** | 0.666*** | 0.772*** | 1 |

$N = 190$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, Bonferroni corrections for multiple comparisons are used.

Table B.2 gives the pairwise correlation between responses to the 10 items of the emotion questionnaire. We ran a factor analysis to check whether the questionnaire can be reduced to a small set of factors. This analysis confirms that two factors capture most of the variation in the responses. The first factor is best explained by the negative emotions: Anger, Disappointment, Jealousy, Rage, Frustration, Irritation and Sadness. The second factor is best explained by the positive emotions: Joy, Elation and Surprise.

C Figures

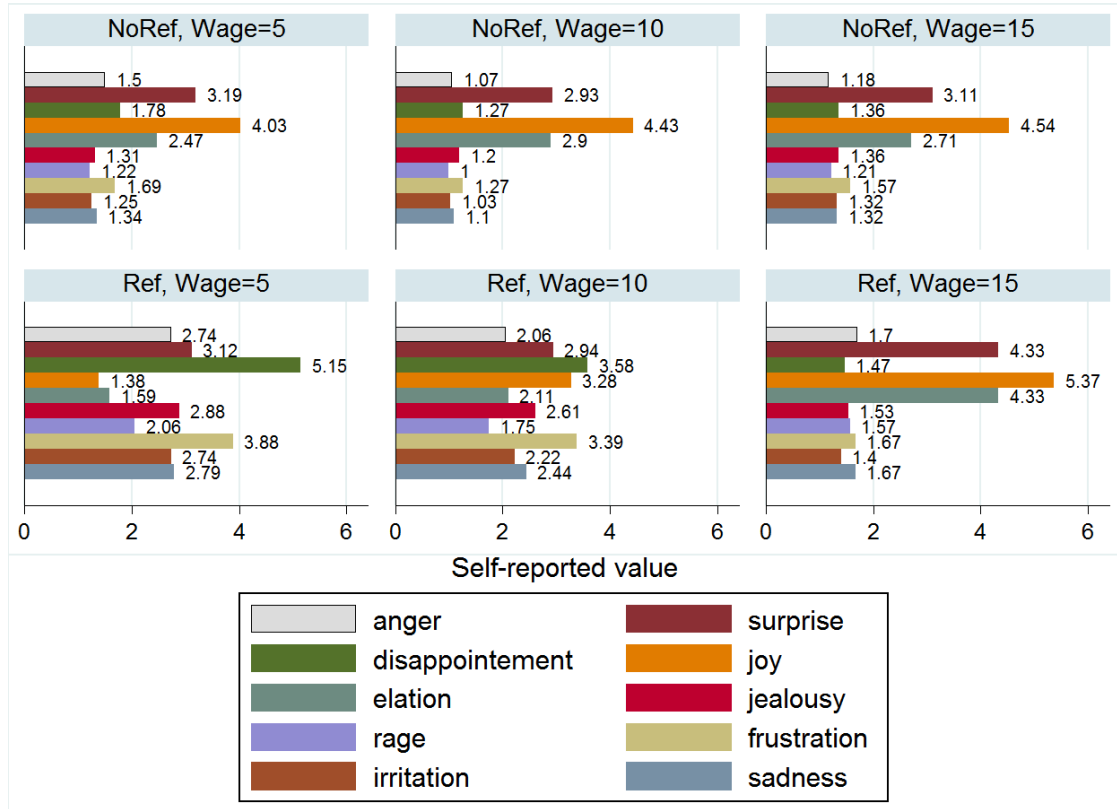


Figure C.1 – Average responses for each items of the emotion questionnaire, by condition

D Proofs

D.1 Wage and profit

The first order condition that defines d_W , the optimal transfer for participant i relative to his direct utility function is given by equation 7.

$$m'(\pi_i) - \alpha m'(\pi_j) = 0 \quad (7)$$

Define $g = m'(\pi_i) - \alpha m'(\pi_j)$. In equation 8, we differentiate g as respect to d and W . We denote differential Δ in order to distinguish it from the transfer d .

$$\begin{aligned} \Delta g = & \frac{\partial \pi_i}{\partial d} \frac{\partial m'(\pi_i)}{\partial \pi_i} \Delta d + \frac{\partial \pi_i}{\partial W} \frac{\partial m'(\pi_i)}{\partial \pi_i} \Delta W \\ & - \alpha \frac{\partial \pi_j}{\partial d} \frac{\partial m'(\pi_j)}{\partial \pi_j} \Delta d - \alpha \frac{\partial \pi_j}{\partial W} \frac{\partial m'(\pi_j)}{\partial \pi_j} \Delta W \end{aligned} \quad (8)$$

We want $\Delta g = 0$ in order for the first order condition to continue holding after changes in d and W . With some omitted steps, $\Delta g = 0$ is found for the relative change of d over W reported in equation 9.

$$\frac{\Delta d}{\Delta W} = - \frac{m''(\pi_i) - \alpha m''(\pi_j)}{E(m''(\pi_i) + \alpha m''(\pi_j))} \quad (9)$$

$0 < \frac{\Delta d}{\Delta W} < 1$, which implies an increase in wage for the dictator will be only partly passed on to the recipient. Thus, an increase in wage will increase both i and j 's profits.

D.2 Continuity and concavity of the utility function

Continuity

The utility function is continuous excepted for transfers $d = \tilde{d}$ where \tilde{d} is such that $\pi_i = \bar{\pi}_i$ and $\pi_j = \bar{\pi}_j$. For this value, it is yet to be proven.

$$\begin{aligned}\lim_{d \rightarrow \tilde{d}^-} u &= m(\bar{\pi}_i) + \alpha m(\bar{\pi}_j) + \lambda(m(\pi_i) - m(\bar{\pi}_i))^\beta + \alpha \lambda(m(\pi_j) - m(\bar{\pi}_j))^\beta \\ \lim_{d \rightarrow \tilde{d}^+} u &= m(\bar{\pi}_i) + \alpha m(\bar{\pi}_j) + (m(\pi_i) - m(\bar{\pi}_i))^\gamma + \alpha(m(\pi_j) - m(\bar{\pi}_j))^\gamma\end{aligned}$$

It is easy to see that $\lim_{d \rightarrow \tilde{d}^-} u = \lim_{d \rightarrow \tilde{d}^+} u$: u is continuous at $d = \tilde{d}$.

Concavity

The concavity of the utility function in the gain domain directly follows from the concavity of $m(\cdot)$ and $\mu(\cdot)$ in the gain domain. In the loss domain, $\mu(\cdot)$ is convex, but for values of β close to one, u is still concave. To show it, we study the sign of the second derivative of the utility function with respect to π_i in the loss domain, depicted in Equation 10:

$$\begin{aligned}u'' &= m''(\pi_i) + \alpha m''(\pi_j) + \lambda \beta m''(\pi_i) (- (m(\pi_i) - m(\bar{\pi}_i)))^{\beta-1} \\ &\quad + \lambda \beta (1 - \beta) m'(\pi_i) (- (m(\pi_i) - m(\bar{\pi}_i)))^{\beta-2} \\ &\quad + \lambda \alpha \beta m''(\pi_j) (- (\alpha m(\pi_j) - \alpha m(\bar{\pi}_j)))^{\beta-1} \\ &\quad + \lambda \alpha \beta (1 - \beta) \beta m'(\pi_j) (- (\alpha m(\pi_j) - \alpha m(\bar{\pi}_j)))^{\beta-2}\end{aligned}\tag{10}$$

For values of β close to 1, this expression is negative, and u is concave. Such values of β denote a low diminishing sensitivity in the loss domain and corresponds to the empirical measures. For an overview of the estimates of β in the literature, see Fox and Poldrack (2014).

D.3 Details of the calculation for proof 2 & 3

Calculation for proof 2. In order to prove proposition 2, we must show that $m'(\pi_i) \beta (-\lambda) (-1) (m(\bar{\pi}_i) - m(\pi_i))^{\beta-1} - m'(\pi_j) \alpha \beta (\lambda) (m(\bar{\pi}_j) - m(\pi_j))^{\beta-1}$ is positive.

Since $m'(\pi_i) - \alpha m'(\pi_j) = 0$, the former expression is positive when $(m(\bar{\pi}_i) - m(\pi_i))^{\beta-1} > (\alpha m(\bar{\pi}_j) - \alpha m(\pi_j))^{\beta-1}$. Given that $\beta - 1 < 0$, it is equivalent to $(m(\bar{\pi}_i) -$

$m(\pi_i) < (\alpha m(\bar{\pi}_j) - \alpha m(\pi_j))$. This can be rewritten $m(\pi_i) - \alpha m(\pi_j) > m(\bar{\pi}_i) - \alpha m(\bar{\pi}_j)$.

Remember that $\pi_i > \pi_j$, $\bar{\pi}_i > \bar{\pi}_j$ and $\bar{\pi}_i > \pi_i$. Moreover, by definition, $\pi_i - \pi_j = \bar{\pi}_i - \bar{\pi}_j$. Since $m(\cdot)$ is concave, the previous implies that the derivative is positive at this point.

Calculation for proof 3. In order to prove proposition 3, we must show that $m'(\pi_i)\gamma(m(\pi_i) - m(\bar{\pi}_i))^{\gamma-1} - m'(\pi_j)\alpha\gamma(\alpha m(\pi_j) - \alpha m(\bar{\pi}_j))^{\gamma-1}$ is negative.

Since $m'(\pi_i) - \alpha m'(\pi_j) = 0$, the former expression is negative when $(m(\pi_i) - m(\bar{\pi}_i))^{\gamma-1} < \alpha(m(\pi_j) - m(\bar{\pi}_j))^{\gamma-1}$. Given that $\gamma - 1 < 0$, it is equivalent to $m(\pi_i) - m(\bar{\pi}_i) > \alpha m(\pi_j) - \alpha m(\bar{\pi}_j)$. This can be rewritten $m(\pi_i) - \alpha m(\pi_j) > m(\bar{\pi}_i) - \alpha m(\bar{\pi}_j)$.

Remember that $\pi_i > \pi_j$, $\bar{\pi}_i > \bar{\pi}_j$ and $\bar{\pi}_i < \pi_i$. Moreover, by definition, $\pi_i - \pi_j = \bar{\pi}_i - \bar{\pi}_j$. Since $m(\cdot)$ is concave, the previous implies that the derivative is negative at this point.

General Conclusion

The chapters of this thesis study the effects of social and contextual factors on decisions related to the ethical principles of honesty and fairness by using controlled and incentivized experiments. In the first two chapters, we investigated dishonesty and how it is affected by social identity (Chapter 1) and by social information and incentives (Chapter 2). In chapter 3, we investigated sharing behavior and how reference-dependent losses and gains in a task can spillover on generosity in a subsequent donation decision.

In Chapter 1, we have used a real-effort experiment to study whether group identity affects unethical behavior in a contest game. We have varied three dimensions: whether minimal group identity was induced or not, whether individuals had to report their own outcome or the outcome of their competitor, and whether pairs of competitors shared the same group identity or not. We found that individuals misreported in the same proportion and to the same extent by inflating their outcome or by decreasing their opponent's outcome, except when any possible scrutiny by the experimenter was removed. Regardless of the possibility of scrutiny by the experimenter, misreporting was affected neither by the competitor's group identity nor by the individual's beliefs about others' misreporting behavior. Our results suggest that in competitive settings, unethical behavior is mainly driven by an unconditional desire to win.

In Chapter 2, we used a variant of the die-under-the-cup task to investigate the dynamics of lies and test whether giving continuous feedback on a counterpart's reported

outcome influences individuals' own dishonesty in a dynamic setting. We explored to which extent this influence depends on the incentive scheme in use by implementing either an individual piece-rate scheme or a tournament scheme. Additionally, we varied whether social information was available or not throughout the game and whether the counterpart was able to cheat or not. Our first main finding confirmed that people lie more when placed in a competitive rather than a non-competitive setting, which is consistent with the former literature (e.g., Schwieren and Weichselbaumer, 2010; Feltovich, 2018). We have also shown that this difference is mostly driven by the counterpart's ability to cheat, suggesting that beliefs about peers' dishonesty is the main driver of dishonesty in competition. Finally, we found that continuous feedback on the counterpart's reports increases individuals' dishonesty under the piece rate scheme but not under the tournament pay scheme. However, participants reacted to their counterpart's potential dishonesty in both treatments creating a vicious *snowball* effect. The fact that, at the aggregate level, continuous feedback did not increase dishonesty in the tournament was due to males who overestimated their opponent's dishonesty when they were uninformed. Our results stress out the importance of considering beliefs about peers' dishonesty to understand the relationships between competition and unethical behavior.

Chapter 3 focused on sharing behavior that is also related to a central ethical principle, fairness. Precisely, we tested whether the salience of counterfactual payoffs impacts generosity. Participants first performed a real-effort task for a fixed wage, and then played a dictator game. Between conditions, we varied the level and the timing of the revelation of the wage. In some conditions, participants knew the wage before the real effort task, and were not informed of the other potential levels. In some other conditions, they were informed of the distribution of the wages before the real effort task, but the actual wage was only revealed afterward. Our hypothesis was that participants in the latter conditions evaluate their actual wage relative to the other potential levels, which in turns impacts their transfers in the subsequent dictator game. The results supported this hypothesis: participants who got the high wage tended to transfer more when they were

informed of the other potential wage levels than when they were not. Symmetrically, participants who got the low wage tended to transfer less when they were informed of the other potential levels than when they were not. We also explored the role of emotions as a channel of our behavioral results by using both skin-conductance and subjective assessment of participants' emotional arousal when the actual wage is revealed. Although the valence and intensity of emotions are in line with our hypotheses (*i.e.*, more negative emotions for the low wage and more positive emotions for the high wage, compared to the medium one), they have no explanatory power on the impact of reference-dependent wages on subsequent sharing behavior.

Limits and extensions

In Chapter 1 we used a minimal social identity paradigm allowing us to control for potential detrimental stereotypes associated with real-world social identity which could affect participants' competitiveness or honesty (*e.g.*, Niederle and Vesterlund, 2007; Cohn et al., 2014, 2015). However, minimal social identity does not represent the full richness of real social identity. For example, in our experiment, ties between participants from the same group are weak and social ties have been shown to be an important aspect of group identity (*e.g.*, Goette et al., 2012). A possible and straightforward extension could be to introduce natural identities. Moreover, our work also investigated the role of consequences of one's lies by showing, conditional on the lack of scrutiny, that people care about whether their lies affect directly or not the opponent's score. An extension would be to investigate why some individuals are concerned about the direct or indirect effect of their lies. A testable hypothesis would be that individuals' endure more emotional distress when an action hurts someone directly rather than indirectly (*e.g.*, Moll et al., 2008).

In Chapter 2 we find that the effect of being continuously informed about the others' reports creates a snowball effect in both piece-rate and tournament settings. Yet, due to males over dishonest behavior when they are not informed in competition, continuous

information does not lead to a higher level of dishonesty in the tournament. We argued that this is driven by males higher expectations of being matched with a dishonest opponent. However, our design cannot specifically identify this channel as we elicited beliefs after that participants were informed about their opponent's reports. Then, one straightforward extension of our work would be to assess participants' initial beliefs about the other competitors' honesty and see whether it explains that providing social information does not lead to a shift in the aggregate level of dishonesty in competition. Another extension would be to examine to which extent the initial beliefs are determined by a difference in the perception of the social norm at stake in competition. Some individuals may think that dishonesty is acceptable in competition while others may not. Finally, in our design we cannot precisely disentangle between the dishonesty due to the mutual influence of each pair member and the dishonesty due to individual peers' characteristics (Manski, 1993). An extension of our work would be to control for this endogeneity issue in the context of competition to better understand the snowball effect that we observe when we give participants information about their counterpart's reports.

In Chapter 3 our model delivers accurate predictions about participants' behavior. However, it assumes that participants form their reference point based on either rational expectations or status-quo and recent studies have highlighted that people are greatly heterogeneous when they form their reference points (*e.g.*, Baillon et al., 2019). Although our design is uninformative about how participants form their reference-point as we only assume that their reference is in-between the high and the low wages. One possible extension would be to multiply the number of wage levels to study more precisely how participants form their reference-point by looking at their impact on subsequent transfers. Using the individuals' sharing behavior we could determine whether they have perceived a previous wage as a loss or a gain, allowing us to determine how they have formed their reference point. Moreover, we found no explanatory power of emotions in our experiment. This does not mean that emotions have no impact in explaining the spillover effects of losses or gains on generosity. Indeed, in our design, all participants

in a given session were given the same wage level, this could hinder the effect of negative emotions, such as disappointment or anger, on generosity, as participants can feel solidarity toward their potential recipient who experienced the same negative feelings. Then, one possible extension would be to get rid of the potential feeling of solidarity by having participants from another experimental session as recipients.

Overall, the results of the three chapters are in line with the behavioral ethics literature showing that ethical principles are highly sensitive to social and contextual clues (e.g., Bicchieri and Xiao, 2009; Gino et al., 2009; Dana et al., 2011; Bazerman and Gino, 2012). The first two chapters of this thesis highlight the sensitivity of people's (dis)honesty in competition. Specifically, we have shown that competition makes people insensitive to their opponent's social identity. We also found that the apprehension of others' dishonesty explained why competition has a negative effect on individuals' honesty. The last chapter shows that giving behavior is influenced by previous experiences of loss or gain.

In this thesis we investigated two ethical principles, fairness and ethics. One line of research arising from this work would be to study the link between these two ethical considerations. For example, one can ask whether a high valuation of fairness (*i.e.*, social preferences) involves a high valuation of honesty and vice versa? This would call for an investigation of the developmental origins of the valuation of ethical principles which has attracted only little attention in behavioral economics (see Maggian and Villeval, 2016, as an exception).

On a final note, our results can also help nourishing the theoretical literature on fairness and honesty. Some attempts have already been made with, for instance, the incorporation of (self-) reputation concern in the moral cost of dishonesty (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018), the addition of normative social norms in the utility function (Krupka and Weber, 2013) or the consideration of moral narratives (Bénabou et al., 2018). In spite of a the lack of effect of emotions in the third chapter of this thesis, a thorough study of the role of emotion in moral decision making in

economics could be a promising line of research. Multiple evidence in psychology and neurosciences stressed out the importance of emotions to understand why ethical behavior is so context-dependent (e.g., Tangney et al., 2007; Decety and Batson, 2009). Several attempts have already been made (e.g., Loewenstein, 2000; Battigalli and Dufwenberg, 2007; Rick and Loewenstein, 2008) and recently, Vernon Smith and Bart Wilson proposed to re-incorporate emotions into economics following the early work of Adam Smith (Smith and Wilson, 2019). In spite of all those promising routes, we are still lacking a simple and elegant way of modeling the richness and diversity of people's moral behavior and this should be the aim of future work in behavioral ethics.

Bibliography

- Baillon, A., H. Bleichrodt, and V. Spinu (2019). Searching for the Reference Point. *Management Science*.
- Battigalli, P. and M. Dufwenberg (2007). Guilt in games. *American Economic Review* 97(2), 170–176.
- Bazerman, M. H. and F. Gino (2012). Behavioral Ethics: Toward a Deeper Understanding of Moral Judgment and Dishonesty. *Annual Review of Law and Social Science* 8, 85–104.
- Bénabou, R., A. Falk, and J. Tirole (2018). Narratives, Imperatives, and Moral Reasoning.
- Bicchieri, C. and E. Xiao (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making* 22(2), 191–208.
- Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature* 516(7529), 86–89.
- Cohn, A., M. A. Maréchal, and T. Noll (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies* 82(4), 1289–1308.
- Dana, J., G. Loewenstein, and R. Weber (2011). Ethical Immunity: How people violate their own moral standards without feeling they are doing so. In D. De Cremer and A. E. Tenbrunsel (Eds.), *Behavioral business ethics: Shaping an emerging field*, pp. 197–216. New York: Psychology Press.
- Decety, J. and D. C. Batson (2009, 08). *Empathy and Morality: Integrating Social and Neuroscience Approaches*, pp. 109–127.
- Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory* 175(175), 248–264.

- Feltovich, N. (2018). The interaction between competition and unethical behaviour : experimental evidence. *Experimental Economics*. Forthcoming.
- Gino, F., S. Ayal, and D. Ariely (2009). Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological Science* 20(3), 393–8.
- Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review* 108(2).
- Goette, L., D. Huffman, S. Meier, and M. Sutter (2012). Competition between organizational groups : Its impact on altruistic and anti-social motivations. *Management Science* 58(5), 948–960.
- Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association* 11(3), 495–524.
- Loewenstein, G. (2000). Emotions in economic theory and economic behavior.
- Maggian, V. and M. C. Villeval (2016). Social preferences and lying aversion in children. *Experimental Economics* 19(3), 663–685.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3), 531–542.
- Moll, J., R. De Oliveira-Souza, and R. Zahn (2008). The neural basis of moral cognition: sentiments, concepts, and values. *Annals of the New York Academy of Sciences* 1124, 161–80.
- Niederle, M. and L. Vesterlund (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *Quarterly Journal of Economics* 122(3), 1067–1101.
- Rick, S. and G. Loewenstein (2008). The role of emotion in economic behavior. In *Handbook of emotions.*, 3rd ed., pp. 138–156.
- Schwieren, C. and D. Weichselbaumer (2010). Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology* 31(3), 241–253.
- Smith, V. L. and B. J. Wilson (2019). *Humanomics*, pp. i–i. Cambridge Studies in Economics, Choice, and Society. Cambridge University Press.
- Tangney, J. P., J. Stuewig, and D. J. Mashek (2007). Moral emotions and moral behavior. *Annual Review of Psychology* 58(1), 345–372. PMID: 16953797.

