Universitat Autònoma de Barcelona

# Essays on Social Pressure: Choice and (Dis)honesty

**Dilan Okcuoglu**

Universitat Autònoma de Barcelona

Doctoral Thesis

# Essays on Social Pressure: Choice and (Dis)honesty

*Author:*
Dilan Okcuoglu

*Advisor:*
Prof. Jordi Brandts

*Tutor:*
Prof. Caterina Calsamiglia

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in the*

International Doctorate in Economic Analysis
Departament d'Econòmia i Historia Econòmica

2017

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to Jordi Brandts for accepting to be my advisor at a critical stage of my PhD and for his continuous support and guidance from ever since.

I thank Miguel A. Ballester for the time and effort he spent on this dissertation. I have always admired at his academic brilliance and enthusiasm.

I thank each and every faculty member of IDEA for, sometimes, just sharing a moment of delightful conversations in the corridors of the department, but mostly for the inspiring research environment they create together. Especially, I am grateful to Pedro Rey Biel for his continuous encouragement and contribution to the present work. I thank Antonio Miralles for having his door open all the time and for his sincere interest in every work of mine. I am thankful to Jordi Massó, Tomás Rodríguez Barraquer, Xavier Martínez-Giralt and Johannes Gierlinger for their valuable comments.

For the experiment paper, I am indebted to Salvador Barberà and to my advisor Jordi Brandts for helping me to find the necessary financial resources. I thank José María Ortiz Gómez who helped me to organize and run the sessions. I would like to thank him, Francesco Gallio and Javier A. Rodríguez Camacho for the translations of the instructions. I am also indebted to Pedro Rey Biel who provided helpful feedback from the very beginning of the project as well as financial support to run pilot sessions. I thank my colleagues Jesús, Bartosz, Shahir, Pablo, Tingting and Aldo for helping us in the lab.

I would like to take this opportunity to express my deep gratitude to the past and present directors of the IDEA program, for their assuring help and guidance, but especially to Susanna Esteban and Luca Gambetti for their caring and emotional support in the most stressful times of this process. I am sincerely thankful to Àngels and Mercè for bearing my countless interruptions and making everything easier not only for me but for everyone in the department.

I feel very lucky for having the chance to become not only colleagues but also friends with Alberto, Alex, Benjamin, Edgardo, Francesco C, Francesco G, Guillem, Isabel, Javi, John, Yuliya and Tugce from the very beginning of this adventure. My friends Bartosz, Danisz, Jesus have made the last year of the PhD delightful. Javi deserves a special *thank you* for being a wonderful flatmate, friend, and the incredible editor

for this dissertation.

Here comes the most difficult part as I am not capable of finding the right words to express my gratitude for my family. Several years of being apart in different countries and yet I have always felt their warmth. I am indebted to my parents for supporting me in all of my pursuits and for their unconditional love. My sister and my best friend Gulten and my fiancee Adnan, each time I have stumbled your love and encouragement made it possible to continue, I couldn't survive without you.

I gratefully acknowledge the financial support from the grant for trainee research staff (PIF-UAB) and MOVE which made the PhD work possible.

# Contents

# Chapter 1

# Introduction

Human behavior is strongly affected by social factors such as norms or social pressure, and economic decision making is not an exception. The impact of self and social image considerations is well documented in decisions ranging from contract enforcement to labour productivity. Understanding the social and moral dimensions of decisions remains an important topic in economic research. Particularly for its potential as an alternative or complement to the material incentives or sanctions which are often used to foster desired behavior in agents. This dissertation includes different approaches, theoretical models and experimental designs, to provide insights regarding the impact of such concerns on both individual and collective choice.

In chapter 2, we are interested in the micro-foundations of such concerns, and we explore it by extending the standard choice model to incorporate the social environment in decision making. We define different social environments, each with different social pressure levels a decision maker may feel while deciding. As proposed by experimental findings from different fields of research, the model assumes that individuals are more likely to comply with the norms when the social pressure is higher.

This assumption will lead us to answer two important questions: (i) whether we can distinguish choices that are driven by intrinsic preferences from those that are guided by image concerns and (ii) whether we can collect any information regarding the norm that applies to the choice problem.

The underlying mechanism of choice under social pressure is very intuitive. The decision maker has image concerns and is thus averse to deviating from the norms, since such a deviation results in a negative social image. Social pressure determines the strength of those concerns and, therefore, the individual's degree of compliance with the norms. Hence, choices at low and high pressure levels are more likely to be driven by: intrinsic preferences and by image motivations, respectively. By tracing each refinement in the decisions we can recover the hidden preferences of the decision maker and the norm.

We provide an axiomatization of the model and suggest a general form of utility

representation that coincides with the intuition behind a range of models with norm compliance. A decision maker who faces a choice problem first forms the best option by taking into account her intrinsic tastes and what the norm dictates. Then she chooses the closest alternative to that option. The degree of compromise between individual preferences and the norm depends on the pressure level at the time of decision-making. The key point of our representation is characterized by a monotone function which captures the intuition of this compromise in the following way: as social pressure increases the decision maker, if not more, can not exhibit less concern for social image. Thus, the best option should move towards the norm.

Our results are novel for three reasons: First of all, to the best of our knowledge, our model is the first attempt to develop a choice theoretical approach that allows norm compliance in a general sense such that, we are not considering any particular norm that imposes a certain behavior unlike other papers on choice theory regarding norms. Moreover, the model doesn't require any compliant behavior a priori. In fact, if such compliance exists, it is detected by the choice data. Second, it provides a very general form of utility function that allows us to distinguish the intrinsic valuations from the socially motivated ones. Finally, it suggests another methodology to elicit, although partially, the norm that applies to a particular choice problem.

In chapter 2, the choice problem is defined on a particular choice set, which is the closed interval $[0, 1]$. Although many choice problems can be defined on such sets, one can naturally be interested in choice problems that may fall outside of this domain. Chapter 3 represents the choice problem under different social pressure levels where there is no such natural distance between different alternatives. Moreover, as opposed to the partial revelation of the norm in the previous chapter, in chapter 3 we present a multi-agent setup where the full revelation is possible.

We first take a step backward in the analysis and propose a simple measure of compliance based on the preferences of the decision maker and the norm, which is basically a relation that orders each alternative from the most socially desirable to the least. We define compliance rate as the level of agreement between the individual's preference and the norm. We then study the relation between this agreement and the existence of a family of preferences that comply with the norm as the social pressure increases.

As the preferences in our set-up are not observable, we revisit the choice outcomes and study the properties of the observed behavior that would imply that the individual is taking decisions under social pressure, which refers to that the choice is more compliant as the social pressure increases. By analyzing the choices of many individuals and suggesting a straightforward property on the collective data,

which basically postulates that there is only one norm that guides the behavior, we completely unveil the corresponding norm.

Although the standard economic analysis neglects normative advice, as an immediate implication of being part of a society, individuals ask each others' opinions when they face a decision that can be morally and/or socially inappropriate. The impact of such advises remains unexplored although it is a simple and an easy way of fostering good (or bad) behavior, which may have important implications.

In the fourth chapter we investigate and document such an effect of advice using experimental data in situations where people have to lie to earn higher payoffs, comparing decisions with and without advice. In line with past studies we find that a significant number of people avoid lying in both cases which suggests that either self or social image considerations are at work. Strikingly, we show that a great number of people ask for non-informative recommendation and it affects dishonest deeds. While advice influence lying to the full extent possible upward, partial lying is affected in a downward direction.

To distinguish whether advice helps to cope with the psychological (self-image considerations) or reputational costs (social-image), we compare behavior in one treatment where advice asking is confidential with behavior where advice solicitation is public. While we don't find any evidence for social-image considerations, we argue that the fact that lying fully is more prevalent in the presence of advice whereas, in contrast, partial lying decreases suggest that advice ease the psychological cost of lying. The patterns in following a given advice confirm those findings even further. Moreover, we provide evidence that non-incentivised participants (advisors) can make decisions similar to those who are incentivised (decision makers), even when the decisions posses socially inappropriate attributes.

# Chapter 2

# Revealed Image: Choice under Social Pressure

## 2.1 Introduction

A couple of years ago a best-selling Turkish author published a book called *Love*. The cover was pink. After a short while, another version of it was released, this time with a grey cover. The reason was that the male readers were complaining about the pink edition embarrassing them whenever they read the book in public. Reading a pink-covered book, called *Love* became a problem once it was observable.[1] This anecdote highlights two important features of the decision process: choice items, be it a tangible product or an action, may inhere social meanings, which turns them into means of reflecting some aspects of our personal or social traits. This raises social image concerns associated to the consumption of such items. Second, exactly for that reason, the decision problem can be tackled differently depending on the social context.

In the economic literature, the notion that a concern for social image influences decisions is documented and analysed in different contexts (Bernheim 1994; Bagwell and Bernheim 1996; Tadelis 2011), as well as the effect of social setups on our choices (Dana et al. 2007; Soetevent 2005). Yet, in the specific sphere of the decision theory, social concerns have not received enough attention.[2] The present paper proposes a decision procedure where the decision maker cares about not only material consumption but also how her choice is perceived and interpreted by others. We study this behaviour by developing an extended choice model incorporating different social environments as a part of the choice data.

Decisions where such concerns play an important role are ubiquitous, embedded in choices as volunteering, effort in the workplace or product selection. Therefore, the evaluation of those goods or actions are sensitive to social factors such as the mere

[1] Similar marketing strategies can be observed in different countries, too, for different reasons. For example in United Kingdom, Harry Potter was released in two versions: adult vs. junior

[2] See for example Baigent and Gaertner 1996; Gaertner and Xu 1999; Tutić 2015 for a few exceptions

observation by others. For instance, consuming a “green" product can signal one's ecological awareness. For this reason, a decision maker may prefer an eco-friendly brand for the coffee standing on her office desk, whereas the same awareness may not be shown for the frying oil she has in her kitchen cupboard. Or an employee may surf less on the internet in an open-concept office than in one with cubicles. Indeed, not only casual observation but also sizeable experimental data documents evidence for “audience effect", thus image concerns, especially when behaving in a certain way is clearly interpreted as one's generous or altruistic predispositions.[3]

Motivated by those findings and observations, the main objective of this paper is to propose a rational choice model that allows for the presence of social pressure. To our knowledge, this effect has not been studied in a decision theoretical set up before. We provide a revealed-preference foundation for a general form of utility representation of the choice behaviours that are guided not only by monetary motives but also non-pecuniary motivations.

More formally, a choice problem in our setting consists of a pair $(A, p)$ where $A$ is the feasible set of alternatives and $p \in [0, 1]$ represents the different levels of social pressure. As usual, choice correspondence assigns a subset of feasible alternatives to each choice problem. At each pressure level the decision maker forms a set of choosable alternatives.

In particular, we consider a decision maker who lives in a society which adheres to some norm. In our context a norm is a prescription of behaviors which indicates what should or shouldn't be done (Akerlof and Kranton 2010). Therefore, a norm can be considered a reference point to evaluate the appropriateness of each option in the grand set (Krupka and Weber 2013). The deviation from prescribed behavior brings about a negative image.

We present choice under social pressure as an intuitive and simple decision mechanism where pressure influences the degree to which the decision maker cares about her image, thus, how consistent is her choice with the prescribed behaviour. The only assumption that is embedded in our model is that with the increasing level of social pressure; choices, if not converging, can not diverge from the prescribed behaviour. That leads us to distinguish those choices that are driven by intrinsic valuations from the ones that are taken by social influence. Moreover, with our data we can partially deduce the norm that applies to the choice problem at hand. In our setup, if the decision maker doesn't modify her choices with increasing levels of social pressure, this implies one of the following: either the choice doesn't contain

[3]For instance, the effort one exerts in a public goods context may increase with audience (Filiz-Ozbay and Ozbay 2013) or the fraction of the participants in a dictator game who shares half of the prize is increasing once the actions are scrutinized (Andreoni and Bernheim 2009)

any social meaning i.e., there is no norm that applies to the choice problem or she doesn't feel enough pressure to modify her decisions.

The approach we follow is axiomatic. We introduce to the model a consistency property across different pressure levels by requiring that if an alternative is revealed preferred to another at low and high levels of social pressure then it should be revealed preferred to that alternative at an intermediate level. This captures the gist of the monotonicity of the compliance with the norms. Another axiom is a reflection of a rationality requirement in the classical model, i.e., *WARP*. One of the other two properties links the choice behavior across different social pressure levels while the other simply states that the preferences satisfy the usual convexity condition. We also impose a standard continuity assumption on choice correspondences.

Those properties together characterize the choice behavior under social pressure. Our utility representation coincides with the intuition behind a range of models with norm compliance. A decision maker who faces a choice problem first forms the best option by taking into account her intrinsic tastes and what the norm dictates. Then she chooses the closest alternative to that option. The degree of the compromise between individual preferences and the norm depends on the pressure level at the time of decision making. The key point of our representation is characterized by a monotone function which captures the intuition of this compromise in the following way: as social pressure increases the decision maker, if not more, can not exhibit less concern for social image, thus, the best option should move towards the norm.

The rest of the paper is organized as follows. First in a subsection we present a review of the literature. Then the following section introduces basic set up: The five axioms and the characterization of choice behavior under different social pressure levels. In the section 2.3., we introduce two applications of our model. In the first application we consider a dictator game where the dictator has to make the sharing decisions at different pressure levels. In this setup the divisions vary as documented in many experimental studies. In a second application, we consider pressure generated by the choices of the other individuals in the following way: each decision maker observes the choices of other individuals and pays attention to those who are conforming with the norm to a greater extent than she does. This creates social pressure on her and leads a modification towards a behavior more concurrent with the norm . We show that there are two steady state equilibria where everyone conforms more. Section 2.4., concludes and proofs are left to an appendix.

### 2.1.1 Literature Review

The importance of non-pecuniary motivations in relation to economic behaviour has been drawn attention since the seminal work of Becker (1957) which introduces a

taste for discrimination and altruism, among other motivations. Since then, several works have been discussing the effect of social motivations, in particular norms, on economic decisions. However, to the best of our knowledge, there is no attempt to include norms in choice based analysis of economic behaviour.

Part of this literature analyses the effect of norms that are generated by the threat of social sanctions which are called ''social norms". Akerlof (1980) is an early attempt to incorporate the norms in economic analysis. He provides an explanation of unemployment when the fair wage is a social custom which may generate a loss of reputation in case it is violated. In the line with experimental evidence, Fehr and Gächter (2000) also analyze how norm violators are punished in public good games. In a natural field experiment conducted by Balafoutas and Nikiforakis (2012), we can see that even in one shot anonymous games people try to enforce behavior compatible with the existing norms.[4]

Another strand of papers focuses on the behavioral implications of the concern for a social image under a game theoretical approach. In our paper, on the contrary, we assume that the individuals do not interact strategically, but they still care about their social image. Bernheim (1994) shows that when prestige is sufficiently important, many individuals conform to a behavior that is prescribed by the social norm. This effect is studied to explain different economic phenomena like conspicuous consumption (Bagwell and Bernheim 1996), donations and altruistic behaviors (Glazer and Konrad 1996; Andreoni 1990). A few examples that show how changes in the decision making environment affect the individuals are by Bohnet and Frey (1999) and Rege and Telle (2004). They establish that the fraction of individuals who equally share their prize even when having all bargaining power, increases whenever their behaviors become more observable. In line with these findings, Andreoni and Bernheim (2009) present the same effect and argue that the hypothesis stating that people care to be perceived as fair explains better the 50-50 division norm than the assumption that people care about fairness itself.

There is also a growing literature where identity is incorporated to the economic analysis starting with Akerlof and Kranton (2000). The individual "internalizes" the norms of a group and she identifies herself with the group. The more a person adheres to the group, the more likely is that she behaves accordingly.[5] In this case, internal sanctions such as shame, guilt, etc., play an important role for the enforcement of norms. In our model we don't impose that the norms are necessarily

[4]See Elster 1989 for a general discussion on the contrast of rational and norm guided behaviors. Posner and Rasmusen (1999) provides an elaborated analysis of how norms operate through a variety of sanctions.

[5]See Benjamin et al. 2010 to see how social identities affect risk and investment choices or Chen and Li (2009) for its effect on social preferences.

internalized.

## 2.2 Model

### 2.2.1 Set up

We consider $X$ to be the grand set of alternatives. We posit that $X$ is the closed interval $[0, 1]$. Some interpretations of $X$ can be in line with those in Bernheim (1994) e.g., it may represent the fraction of time or money one spends on a specific activity. We believe that many decision making problems can be defined in this domain. For instance, $x$ can be interpreted as the proportion of the time a worker shirks during the working hours, the share of income one donates or even it may represent the color spectrum of a book where $x = 0$ stands for pink and $x = 1$ for black.

We consider a decision maker who takes decisions in different environments which are identified by different social pressure levels.[6]. We model pressure with the set $P = [0, 1]$. One can view $p = 0$ as the standard individual choice problem, on the other hand $p = 1$ can be the extreme interpretation such as the act is televised.

Let $\mathcal{X}$ be the set of all nonempty closed subsets of $X$. Choice problems in our setting consist of a pair $(A, p)$ where $A \in \mathcal{X}$ and $p \in P$. The set of all choice problems is denoted by $\mathcal{C}(X)$. By a choice correspondence we mean a map $c : \mathcal{C}(X) \to \mathcal{X}$ such that $c(A, p) \subseteq A$ for all $(A, p) \in \mathcal{C}(X)$. [7]

A choice correspondence is a non-empty valued mapping by definition. We interpret a choice problem as follows: The decision maker faces a menu $A$. In addition, she becomes aware of the extent of the pressure she is exposed to. From each choice problem, the decision maker identifies a nonempty subset of available alternatives as the set of choosable elements. Obviously these subsets can be different for the same feasible set if different pressure levels are considered.

To illustrate, consider the “dilemma" of a Turkish man. He may personally be

[6]Although one can easily relate social pressure with the observability of the choice behavior, we abstain from making any assumption on the source of this pressure. It may be derived from the number, as well as identity, of the observers. In the literature, some authors identified different methods to create different degrees of “audience" effect e.g., the contributions are made observable (Andreoni and Petrie 2004) or individuals are put in different rooms with different visibility (Ariely et al. 2009)

[7]We follow the approach that views $c(A, p)$ as a reflection of the choice objects among which one alternative will do fine for the decision maker (Kreps 1988). Hence, our decision maker is indifferent among the any alternatives that she chooses. However, the choice correspondence is also interpreted as the set of elements that are choosable under any of some possible circumstances that is not included in the description of choice behavior (Rubinstein 2012).

indifferent between the black and the pink cover whenever he is at home but he suddenly switches to $x = 1$, which, again, can be interpreted as the black cover, once he is in public transportation. Or think of the behavioral changes observed in the dictator game experiments that we mentioned before. The dictators on average shares less in a double blind treatment, for instance, than in the treatment with some observers.

### 2.2.2 Main Axioms

We identify the minimal set of axioms that characterize the class of choices where the decision maker is driven by concerns for social image, therefore, for norm compliance. Our first axiom is an immediate reflection of the weak axiom of revealed preference i.e., WARP. In the context of our study, we can write the weak axiom as follows:

**Axiom 2.1.** *(WARP given pressure level) Let $p \in P$ and $A, B \in \mathcal{X}$. Consider any pair $x, y \in X$ such that $x, y \in A \cap B$ and $x \in c(A, p)$. If $y \in c(B, p)$ then $x \in c(B, p)$.*

Axiom 2.1. guarantees that some sort of maximization happens at each pressure level $p$. That is, for a given pressure level, the choices of the decision maker can be viewed as the outcome of the maximization of a complete preorder.[8]

We now introduce an axiom which posits that the decision maker has single-peaked preferences at each pressure level. As in the usual interpretation of single-peakedness, we assume that the decision maker has an ideal point, a peak, at a given $p$ and the alternatives that are further from that peak are less preferred.

**Axiom 2.2.** *(Single-Peakedness) Take any $p \in P$ and any different alternatives $x, y$ and $z \in X$. If $y \in c(xy, p)$ and $z \in c(xz, p)$ then both $y$ and $z$ belong to either $[0, x)$ or $(x, 1]$*

Given the structure of the grand choice set we have, this property becomes a simple and intuitive requirement. Recall that in our setup, the norm points out an option in $X$ as the most appropriate choice and at each pressure level the decision maker chooses the best option as a result of a compromise between her intrinsic tastes and the norm. This property simply posits that all other alternatives are evaluated based on their distance to that best option. However, it is useful to make the following remark: the distance here is not necessarily the standard Euclidean distance on $[0, 1]$. To illustrate it better, consider again a dictator's allocation decision. Suppose she is a completely selfish individual in the standard sense, thus, her intrinsic utility increases with her own consumption. At the same time she cares about how she is perceived by other people. In a social environment, where her behavior is not

[8] A binary relation is a preorder if it is reflexive and transitive

completely anonymous and the norm is a $50-50$ division, the best option for her is to give away 30% of her endowment. Single-peakedness ensures that 40% is considered better than 50%. However, she may either be indifferent between 20% and 40% or prefer one to the other.[9]

The above two axioms don't have any bite when the pressure varies. The following property, however, builds a connection between choices under different pressure levels.

**Axiom 2.3.** *(Consistency in Compliance) Let $A \in \mathcal{X}$ and $p, p'$ and $p''$ with $p \leq p' \leq p''$. If $x \in c(A,p)$ and $x \in c(A,p'')$ then $x \in c(A,p')$. Moreover if $\{x\} = c(A,p)$ and $\{x\} = c(A,p'')$ then $x = c(A,p')$*

We can view this property, in contrast to Axiom 2.1., as a consistency requirement across different pressure levels. Our decision maker is consistent in terms of compliance with the norm. That is, if she chooses one alternative over another both at low and a high level of social pressure, then we can postulate that if she hasn't already changed her choice at the low level, then either her personal preferences and the norms coincide, so that she doesn't feel the need to review her choice, or the pressure level is not high enough to make her to change her choice in compliance with the norm. Notice that in both scenarios there is no reason for her to modify her choice in an intermediate level.

The next property, as Axiom 2.3., also imposes a restriction on choice behavior across different pressure levels. Specifically, it requires that, if the decision maker moves a way from her most preferred alternative by some amount in result of a change in social pressure, then all the other alternatives that were indifferent will move in the same amount that of best alternative and will remain indifferent.

**Axiom 2.4.** *(Preference Equivalence) For any $p, p' \in P$, let $y = c(X,p)$ and $\{x,z\} = c(xz,p)$ for some $x, z \in X$. If $y + k = c(X,p')$ then $\{x+k, z+k\} = c((x+k)(z+k), p')$*

In other words, this postulate implies that what matters, in terms of the desirability of an alternative, is the relative distance of this alternative to the most preferred one. Since we work on an uncountable alternative space, we will introduce a continuity requirement in order to get a complete characterization of the choice behavior under different pressure levels. Of course, we need to introduce a metric on $\mathcal{X}$ and we mean Hausdorff metric whenever we refer a metric on $\mathcal{X}$.

[9] We can interpret this as follows as well: the decision maker has a different type at each pressure level e.g., at low levels she is selfish, that is she only cares about what she likes. However, at higher levels she turns to a prosocial person and, now, she is also concerned about norms. She attains the maximum utility whenever she acts as her type and each deviation from it decreases her well-being

**Axiom 2.5.** *(Continuity) Let* $(A_m, p_m) \to (A, p)$ *with* $A_m \xrightarrow{H} A$ *and* $p_m \xrightarrow{E} p$. *If* $x_m \in A_m$ *such that* $x_m \to x$ *and* $x_m \in c(A_m, p_m), \forall m = 1, 2, 3...$ *then* $x \in c(A, p)$

The situation is analogous to that in the standard choice. We require our decision maker to choose the similar alternatives in similar choice problems.

### 2.2.3 Result

Now we are ready to derive the main result of this paper. The following theorem characterizes those choice correspondences that satisfy the five axioms above.

**Theorem 2.1.** *A choice correspondence* $c$ *on* $\mathcal{X}$ *satisfies Axiom2.1-2.5 if and only if there exists a strictly quasi-concave utility function* $u : \mathbb{R} \to \mathbb{R}$ *with a maximum at* $0$ *and a monotone mapping* $b : P \to X$ *such that*

$$c(A, p) = \underset{x}{\operatorname{argmax}}\, u(x - b(p))$$

Axiom 2.1, with continuity, requires the existence of a continuous function that represents such choice behaviors. The strict quasi-concavity of the utility function is due to the single-peakedness of the choice correspondence.

We interpret the choice behavior that is characterized in Theorem 2.1. as follows: At each pressure level $p$, the decision maker first forms her peak point which is the alternative that is assigned by the bliss function $b$. The underlying mechanism behind $b$ can be considered as the interaction between social norms and her intrinsic valuation of the alternatives. Absent any social pressure, for example, we can expect that the decision maker will choose her most preferred alternative. However, increasing levels of social pressure may lead a compromise between social norms and individual tastes if any conflict exists between them. Hence, we anticipate that they compromise to a greater extent as the level of pressure increases, i.e., the peak converges to an alternative that is socially more appropriate. This is captured by the monotonicity of the bliss function. Once this peak is determined than our decision maker with such a choice correspondence solves her choice problems by maximizing a utility function $u$ which is maximized at $b(p)$.

## 2.3 Applications

### 2.3.1 Dictator game

In the classical dictator game, one player, the dictator, decides an allocation of a fixed amount of money i.e., $x \in [0, 1]$ between herself and another player, the recip-

ient, who is merely passive. The standard game theoretical analysis requires that the dictator shares nothing with the recipient. However, a mounting evidence from experimental studies documents that a significant number of subjects, indeed, pass a positive amount to the other player. The leading theories put forward altruism, warm glow giving or fairness concerns to explain such behaviors (Fehr and Schmidt 1999; Andreoni 1989). But also, as mentioned before, another group of studies established that this sharing decision varies with the anonymity of the decision making process and invoked the image concerns as the driven motivation behind sharing a positive amount.

Now suppose that the dictator is concerned about, both, how fair she is perceived and her monetary consumption. As suggested in our model, this concern depends on the social pressure. Her choice behavior is described by the utility function as follows:

$$u_1(x) = 1 - |x - b(p)| \text{ such that } b(p) \leq 1/2 \text{ for all pressure levels}$$

Player 2, the receiver, is the standard utility maximizer with $u_2(x) = x$. Implicit in this utility representation, we assume that the norm for a fair image is a $50 - 50$ division which is in line with the empirical data observed both in the field and in the lab.[10] Player 1 places value on monetary payoff $(1 - x)$, but also on the degree to which she adheres to the norm. It obviously depends on the pressure level. For example, $b(p)$ can have the following form:

$$b(p) = p \text{ for } p \leq 0,5 \text{ and } b(p) = 1/2 \text{ for } \geq 1/2$$

The greater is the pressure the more she complies with the norms. For instance, if the pressure is relatively low i.e., $p = 0,1$ the dictator proposes only 10% of her endowment, on the other hand whenever the pressure is sufficiently high she complies with the norm and offers the half of the prize.

Our model is an attempt to account for the behavioral pattern that are observed in the dictator games that are already replicated several times. Moreover, with our model we can distinguish, to a degree, the intrinsic valuations from "other-dependent" utilities. That is, if the dictator would choose $0,5$ at each treatment, e.g., different pressure levels, then one could claim that she genuinely cares about fairness. On the other hand, if her behaviour is as described above then we can infer that she is more concerned about her image.

[10] See Andreoni and Bernheim 2009 for a wide range of contexts where such $50 - 50$ division is common.

### 2.3.2 Endogenous pressure in small groups

In this section we study pressure as a consequence of the actions of a group of people who may share the same environment e.g., classmates or colleagues in the same office. Throughout the paper, we took pressure as an exogenous variable but in some situations what actually socially intimidating is how deviant one's choice from the norm compared to the other people's deviance. In this way the choice of this individual can be salient within the group and she may update and conform to the norm more. We consider a setup where everyone can observe each other's behavior which in turn may influence each other's actions through image considerations (Bernheim 1994). We first present the basic setup and, then, show that there is a unique equilibrium in one of which, everyone completely conforms with the norm and in the other equilibrium everyone but the most conformist decision maker is compromising more with the norm.

Formally, $X = [0, 1]$ is the choice set and there are $N$ individuals. For the ease of the demonstration suppose the norm is $x = 1$.[11] Our analysis is based on two assumptions:

- Choosers care about their social image or in other words, they behave as our model suggests, that is

$$c_i(A, p) = \underset{x}{\operatorname{argmax}}\, u(x - b_i(p))$$

  where $u$ is strictly quasi-concave.

- Pressure is generated by the choices of those who are closer to the norm, i.e.,

$$b_i(p) = x_i^0 + \alpha(\sum_{N_i} x_j - x_i)$$

  where $x_i^0$ is the choice that the decision maker $i$ would choose without any pressure; $x_j, x_i$ are the choices of individual $j$ and $i$, respectively and $N_i = \{j \in N \backslash \{i\} \text{ such that } x_j \geq x_i\}$. The propensity to pressure i.e., $\alpha$, is assumed to be same across individuals and is defined on the unit interval.

An individual $i$ observes the choices of others and pays attention to those who are conforming with the norm to a greater extent than she does. Then she modifies her choice. If there is a $j \in N$ who is affected by this modification, also changes

[11] It is a mirror solution for $x = 0$ and for any intermediate value such that $x \in (0, 1)$, the solution is very similar. It only requires a change for the set of individuals whose choices may put pressure on the others

her choice and so on and so forth. An equilibrium is an allocation where no one is willing to make a further change given the choices of other decision makers.

We should make two remarks here. First, notice that algebraically $x_i$ can result in a value greater than 1 but we assume that this value corresponds to 1 since the choice set is bounded above. Second, if any two decision makers make a modification such that they end up at the same choice, we assume that they continue pressuring each other so that they don't move back to their original choices.

**Definition** An *equilibrium* is an allocation $\{x_1^*, x_2^*, ....x_N^*\}$ such that

$$x_i^* = \arg\max_x u(x - b_i(p))$$

where $b_i(p) = x_i^* = x_i^0 + \alpha(\sum_{N_i} x_j^* - x_i^*)$

**Proposition 2.1.** *Given $\alpha$ and the intrinsic valuations $\{x_1^0, x_2^0..., x_N^0\}$ equilibrium exists and it is unique. Moreover the equilibrium outcome $\{x_1^*, x_2^*, ....x_N^*\}$ holds weakly the same order of the intrinsic valuations, that is, $x_1^0 \leq x_2^0 \leq ... \leq x_N^0$ implies $x_1^* \leq x_2^* \leq ... \leq x_N^*$*

The choices of decision makers who are closer to the norm create a social pressure on those whose choices are deviating from the norm more. This pressure leads modifications in decisions such that each agent, perhaps except the one whose choice is already closest to the norm, comply with the norm to a greater extent. Depending on the parameter $\alpha$ and initial or intrinsic choices, this mechanism may lead even to a full conformity where everyone chooses the alternative that norm dictates which is 1 in this setup.

## 2.4 Concluding remarks

This paper analyses the effect of social motivations, concerns for image in particular, on the decision making process. We argue that many choices require an interaction of the two motivations: to satisfy the intrinsic tastes and as well as to comply with the norms if any applies to the choice problem at hand. Therefore, the extent of this interaction is crucial to have a better understanding of economic behavior whose social component is mostly ignored in such analysis. We want to highlight two important contributions of our paper.

First of all, we axiomatically characterize a model in which decision maker 's behavior is affected by social pressure. To our knowledge, ours is the first attempt in choice theory. Our model leads us an intuitive representation of choice behavior which captures both the effect of this pressure and the degree to which the decision

maker complies with norms. Our characterization is general enough to be applied to a wide range of a specific phenomena.

Second, our model let us to identify norm, if any exists and up to some extent, that applies to the choice set at hand. Particularly, we argue that one can infer the norm, up to a degree, by looking at choice behavior that varies across different social pressure levels. We stipulate that the decision making is guided more by individual tastes at low levels of pressure and directed more towards a prescribed behavior for the higher values. Therefore, any modification in choices from low to high pressure levels corresponds to the influence of norms and, hence, is traced back partially to most socially appropriate behavior, hence to the norm itself. We also think that our model is suitable for several application.

As a future work, we consider several other possible applications. Consider the following example. Assume that $n \geq 2$ package stores compete for a location on a street as in the classical Hotelling (1929) problem.[12] Suppose there is a norm against alcohol consumption and these stores face customers who are concerned about this norm. Therefore, the potential customers don't want to be seen by other people when they enter the shop. The stores face two conflicting decisions: on one hand, they want to maximize their market share by reaching as many consumers as they can, but on the other hand, high accessibility may lead the customer to prefer the store less since it is more probable to be seen by somebody else. An interesting question could be the optimal location of a store in this set up.

We believe that there are also several directions that need to be explored. For one thing, we consider a single agent setup in the main characterization of the problem. However, given the nature of the influence, which is social, the relevant question could be what are the necessary and sufficient properties that the choice data should satisfy to infer the norm not partially but completely from decisions of several individuals.

## 2.5 Appendix

***Proof of Theorem 2.1.*** Before we start to prove the theorem, let us first introduce more notation on preference orders. We denote the complete preorder at any social pressure level $p$ with $\succeq_p$. The asymmetric (strict) and the symmetric (indifference) parts of this relation are denoted by $\succ_p$ and $\sim_p$, respectively. We omit the necessary part of the theorem. To prove the only if part of the theorem, suppose that the choice correspondence $c$ satisfies the five axioms.

[12]These are the stores that mainly sells alcoholic beverages

We will first show that we can elicit a monotone bliss function that is generated by choosing the most preferred alternative in $c$.

**Step 1**: Define $b : T \to X$ such that $b(p) = c(X, p)$

**Claim 1**: $b$ is a single-valued correspondence, i.e., b is a function.

*Proof of Claim 1:* We work on the full information set up. Hence for each $p \in P$, $c(X, p)$ is well defined. Moreover suppose for a contradiction that there are two distinct alternatives $x$ and $y$ such that $b(p) = x$ and $b(p) = y$. Hence $x, y = c(X, p)$. Without loss of generality, suppose $x < y$. By the completeness of $\mathbb{R}$, there exists a $z \in X$ such that $x < z < y$. By Axiom 2.1., we have $x \in c(xz, p)$ and $y \in c(yz, p)$. By Axiom 2.3., we should have either $x < y < z$ or $z < x < y$ but $z \in (x, y)$ which is the desired contradiction.

**Claim 2** $b$ is monotone.

*Proof of Claim 2:* Suppose for a contradiction, it is not. Therefore, for some $p, p'$ and $p''$ such that $p < p' < p''$, we have $b(p) < b(p')$ but $b(p'') < b(p')$.[13] Take any $x, y \in (max\{b(p), b(p'')\}, b(p'))$. Wlg, suppose that $x < y$. By Axiom 2, we have $x = c(xy, p)$ and $x = c(xy, p'')$ but $y = c(xy, p')$ which is obviously contradicting with Axiom 2.3.

**Claim 3**: $b$ is continuous.

*Proof of Claim 3:* Let $A_m = X$ for all $m = 1, 2, 3...$ and $p_m \in P$ with $p_m \to p$. Assume that $b(p_m) \to b^*$. By Axiom 2.5., $b^* = c(X, p)$. Therefore $\lim_{p_m \to p} b(p_m) = b^* = b(p)$ from which we can conclude that $b$ is continuos.

**Step 2:** Axiom 2.1. implies that $c$ satisfies WARP at each $p \in P$. By the classical argument based on WARP, we know that there exists a preference relation that can be constructed from binary choices. Therefore, at each pressure level, $\succeq_p$ defined by

$$x \succeq_p y \text{ if and only if } x \in c(xy, p).$$

The symmetric and indifference relations are defined in the standard way from the weak part.

**Step 3:** We will prove that Axiom 2.5. implies that $\succeq_p$ is upper semicontinous (USC) and lower semicontinuos (LSC) . Take any $x \in X$ and any sequence $(y_m) \in X$ such that $y_m \succeq_p x$ and $y_m \to y$. We will show that $y \succeq_p x$ as well. Since $y_m \succeq_p x$ for each $m = 1, 2, 3..$, we have $y \in c(xy_m, p)$ for all $m$. By Axiom 2.5., we get

[13]The case $b(p) > b(p')$ and $b(p'') > b(p')$ is the symmetric case which we omit here.

$y \in c(xy, p)$ as well. Hence $y \succeq_p x$ which proves that $\succeq_p$ is upper semicontinuous. Lower semicontinuity can be verified similarly so we omit the proof here. By Debrue Theorem (Mas-Colell et al., 1995) we know that there exists a continuos utility function $u$ on $X$ such that

$$c(A, p) = \underset{x \in A}{\operatorname{argmax}}\, u(x)$$

**Step 4:** In this step, we will show that the function $u$ is strictly quasi-concave. We will first demonstrate that Axiom 2.2. implies that the preorder $\succeq_p$ is strictly convex.[14] Suppose $c$ satisfies Axiom 2.2.. For a contradiction suppose $\succeq_p$ is not strictly convex. Hence for some $x, y, z \in X$ such that $y \succeq_p x$ and $z \succeq_p x$, there exists $\alpha \in (0, 1)$ such that $x \succeq_p (1-\alpha)y + \alpha z$. But notice that $y \succeq_p x$ implies that $y \in c(xy, p)$ and $z \succeq_p x$ implies that $y \in c(xy, p)$. Hence by Axiom 2.2., we have either $y, z \in [0, x)$ or they are both in $(x, 1]$. Wlg, assume that the former case is true. By transitivity of $\succeq_p$, we have $y \succeq_p (1-\alpha)y + \alpha z$ and $z \succeq_p (1-\alpha)y + \alpha z$. Obviously $(1-\alpha)y + \alpha z \in (y, z)$ since $\alpha \in (0, 1)$ which is the desired contradiction with Axiom 2.2. The strict convexity of $\succeq_p$ implies that the utility function $u(x)$ is strictly quasi-concave (see Mas-Colell et al. 1995).

Now we will define the set of indifferences at each pressure level $p$ as follows:

$$I(p) = \{(x, y)\} \in X \times X \text{ such that } x \sim_p y\}$$

We will show that there exists a pressure level $\tilde{p}$ such that $I(\tilde{p})$ is the superset of the each indifference set of any other pressure level

**Claim 4:** $\exists \tilde{p}$ such that $I(\tilde{p}) \supseteq I(p)$ for all $p \in P$.

*Proof of Claim 4:* We need to study two cases:

• Suppose the decision maker chooses 0 over 1 at some pressure levels and for some other levels we observe that she chooses 1 over 0. Then by Axiom 2.5., it is easy to verify that she is indifferent between 0 and 1 at some pressure level. Let $p^* \in P$ be the pressure level such that $\{0, 1\} = C(01, p^*)$. Let $u^*$ be the utility function that represents the choice behavior at $p^*$. By the previous steps, we know that we can write the maximization problem as follows:

$$c(A, p^*) = \underset{x}{\operatorname{argmax}}\, u^*(x - (b(p^*))$$

[14] In fact the two properties are equivalent in the sense that Axiom 2.2. implies that the preference relation $\succeq_p$ is single-peaked and on a closed subset of $\mathbb{R}$ with the continuity of $\succeq_p$, single peakedness is equivalent to strict convexity. For the sake of the brevity, we leave the proof of sufficiency of the strict convexity to the reader.

By the continuity and strict quasi concavity of $u^*$, it can be easily verified that $u^*$ is strictly increasing to the left and strictly decreasing to the right of $b(p^*)$. Notice also that $u^*(0) = u^*(1)$. Moreover by continuity, we know that the image of the intervals $[0, b(p^*)]$ and $[b(p^*), 1]$ under $u^*$ are connected and also they are equivalent i.e., $[u^*(0), u^*(b(t^*)] = [u^*(1), u^*(b(p^*)]$. By Intermediate Value Theorem, we know that for all $\alpha \in [u^*(0), u^*(b(p^*))], \exists y \in [0, b(p^*)]$ and $z \in [b(p^*), 1]$ such that $u^*(y) = u^*(z)$, hence, $y \sim z$. Therefore, for each $y \in [0, b(p^*)]$, we know the indifferent counterpart of that alternative in $[b(p^*), 1]$ and vice versa. Hence we will show that by Axiom 2.4., we can represent any choice behavior at any pressure level with $u^*$.

Take any $p \in P$. By Axiom 2.4., we know $y + (b(p) - b(p^*)) \sim_p z + (b(t) - b(t^*))$. Hence it is just a horizontal shift of $u^*$ by $(b(p) - b(p^*))$ The new representation will be

$$c(A, p) = \underset{x}{\operatorname{argmax}}\, u^*(x - (b(p) - b(p^*) - b(p^*)), \text{which is } u^*(x - (b(p))$$

Now, instead, suppose the decision maker always prefers either 0 over 1 or 1 over 0.

For simplicity, the rest of the proof is organized assuming that she chooses 0 over 1. The other case is just analogous.

We denote the utility function that represents the choice behavior at any pressure level $p$ by $u_p$. Since 0 is always preferred to 1 then $u_p(0) > u_p(1)\ \forall p \in P$. Hence at all pressure levels, we have $u_p(b(p)) > u_p(0) > u_p(1)$. By Intermediate Value Theorem, at each $p$, there exists an alternative in $[b(p), 1]$, i.e., $y_p$, such that $u_p(0) = u_p(y_p)$. Following similar arguments above, it is immediate to see that the greater is $y_p$, the bigger is the set of indifferences $I(p)$. Hence, $\tilde{p}$ is the pressure level such that $y_{\tilde{p}} = max\{y_p : u_p(y_p) = u_p(0)\}$. Therefore we can write the maximization problem as follows:

$$c(A, \tilde{p}) = \underset{x}{\operatorname{argmax}}\, u_{\tilde{p}}(x - (b(\tilde{p}))$$

By Axiom 2.4., for those alternatives $x, y$ such that $x, y \in [0, y_{\tilde{p}}]$, we can write the representation as

$$c(A, p) = \underset{x}{\operatorname{argmax}}\, u_{\tilde{p}}(x - (b(p) - b(\tilde{p}) - b(\tilde{p})), \text{which is } u_{\tilde{p}}(x - (b(p))$$

And any strictly decreasing function, let's say $f$, would represent the utility representation of the alternatives that are in $[y_{\tilde{p}}, 1]$. Therefore the utility representation for this case is as follows:

$$c(A, p) = \begin{cases} \operatorname{argmax} u_{\tilde{p}}(x - (b(p)) & \text{if } x \in [0, y_{\tilde{p}}] \\ \operatorname{argmax} f(x) & \text{otherwise} \end{cases}$$

and we are done. ■

***Proof of Proposition 2.1***

We can order the decision makers according to the distance of their choices to the norm. Suppose agent N is the decision maker whose choice is closest to 1, agent N-1 is the decision maker whose choice is closest to agent N and so on and so forth. Without any modification in decisions, there is no pressure on agent N and she is the only decision maker who puts pressure on agent N-1. This pressure, depending on the initial choices of the two agents and $\alpha$, will lead one of the following three allocations: i) agent N-1 moves towards agent N but chooses a smaller alternative ii) two agents choose the same alternative or iii) agent N-1 chooses a bigger alternative. If the two first cases apply to each decision maker than it is easy to see that an equilibrium $x_1^* \leq x_2^* \leq .....x_N^*$ will be attained. It will be enough to show that for any $i, j \in N$ if $x_i^0 \leq x_j^0$ then $x_i^* \leq x_j^*$. Suppose not, that is, $x_i^* > x_j^*$. This implies $\sum_{N_j} x_i^*$ is bigger than $\sum_{N_i} x_j^*$. By $b_i(p) = x_i^* = x_i^0 + \alpha(\sum_{N_i} x_j^* - x_i^*)$, we can rewrite the equilibrium allocations as $x_i^* = \frac{x_i^0 + \sum_{N_i} x_j^*}{1+\alpha}$ and $x_j^* = \frac{x_j^0 + \sum_{N_j} x_i^*}{1+\alpha}$. Since $x_i^0 \leq x_j^0$ and $\sum_{N_i} x_j^* \leq \sum_{N_j} x_i^*$, the contradiction is immediate. Moreover, given that $b_i(p)$ is the unique peak alternative by strictly quasi-concavity of $u$, this equilibrium is unique.

On the other hand, if the third case applies, that is agent N, now, is pressured by agent N-1 then she will also modify her choices. This may lead one of the cases above and by the argument that follows, we know that $x_N^* < x_{N-1}^*$ can not happen. It is true for any pair of decision makers, thus, either they reach a point where they choose the same alternative or each modification makes the decision makers move towards 1. Depending on $\alpha$ and initial allocations, they all may comply with the norm completely that is $x_i^* = 1$ for all $i \in \{1, 2, 3..N\}$ or again attain an allocation such that $\{x_1^*, x_2^*, ....x_N^*\}$ with $x_k^* \neq x_l^*$ for some $k, l \in \{1, 2, 3..N\}$ ■

# Chapter 3

# Norm-revelation from individual decisions: Choice and Compliance

## 3.1 Introduction

Choices, even those that at first sight might appear to be driven only by individual tastes, can be cast under the spell of social norms. From charitable donations to fashion choices, a variety of decisions convey information about the status, generosity, and cultural background of the decision maker. This information is relevant for two of its main characteristics: it is often in plain sight for others to observe, and cannot be unbundled from its underlying decision process, even when it is apparently not connected to it. What do we choose to reveal when deciding to wear a particular shirt, and how much of those decisions are implicit, when not determined by the social spaces we inhabit? Hence, the decision maker's concerns for the social interpretation of her choices can be expected to play a role in a wide range of choices we make on daily basis.

However, even if some cultural and social patterns might be thought to shape said choices, there is ample evidence that norm compliance shows inconstancy (Andreoni and Bernheim 2009; Bertrand and Morse 2016; Bursztyn and Jensen 2015; Filiz-Ozbay and Ozbay 2014; Funk 2010).[1] Even if a norm prescribes the same type of action, the same person may be observed being very generous in one context while in some other she may act selfish. Indeed, in chapter 1 we provide a general utility representation for choice behaviors of this sort, where the decision maker (DM, henceforth) updates her decisions with changing degrees of social pressure. If the DM strives for maintaining a positive (social) image, she may feel more compelled to comply with the norms when her choices are more exposed to social scrutiny than when they are kept anonymous; that is, when the DM feels more social pressure. For instance, generosity during a church service might be observed more often in a DM's behavior than when she finds an online petition or crosses a homeless person down the street.

[1] See Bursztyn and Jensen 2016 for a review of the literature on the influence of social pressure in the field studies

At the individual level, if the data is taken at face value, these situations would be considered in the literature within the preference reversal phenomenon (Grether and Plott 1979). However, the additional information embedded in the choice problem – in this case social pressure levels – let us glace at some sort of optimization principle behind the choices. In this paper we aim to illustrate such a situation. Namely, when and how we can deduce both DM's preferences and the underlying norm from observed choices and pressure levels.

Formally, we consider a DM who lives in a society which adheres to some norm. We endow the norm with a strict linear relation that orders the alternatives from the most accepted to the least, and we call it the *social code*. That is, we consider a norm of injunctive nature which tells individuals what "one ought to do" in situations where it applies (Elster 2000). In our setup the DM is characterized by a family of preferences that comply with the social code at greater degrees as the social pressure increases. We define the degree of compliance for each social pressure level as the set of the pairs of alternatives that are both included or excluded in the social code and in the corresponding preference. We denote this as the compliance set. We define the extent of compliance as the weighted sum of the pairs in the compliance set. That is, we count each pair in this set with a positive weight associated to each pair. We call this sum the *weighted compliance rate*. We interpret this rate as the level of the agreement between the DM's preference and the social code at a given pressure level. In a first result we show that we can model our DM with preferences that comply with the code more if the weighted compliance rate is increasing for every collection of positive weights.

Our main analysis is based on the choice behavior of a group of individuals who form a society with no strategic interaction among them. For this purpose we first study the choice patterns of a unique DM and suggest two intuitive properties on the choice outcomes. One is a reflection of the weak axiom of revealed preference in our framework, which simply states the consistency of the choice behavior at each level of social pressure. The second property requires that if the individual chooses one alternative from a set of available alternatives both at a low and at a high pressure level, then she sticks at this choice also for an intermediate level. Our second result states that the DM's choice behavior satisfies these properties if we can find a social code and a set of preferences that, on top of describing her choice, is more compliant with the social code as the social pressure increases. We show that it is enough to have an individual's choices in order to partially elicit the social code that guides the decisions.

We assume that the choice behavior of each DM in this society satisfies these properties. Moreover, we suggest a property for the aggregated choice data of the indi-

viduals, simply stating that the observed changes in the choices of the DM's across different pressure levels do not conflict. That is, if one individual modifies her choice in favor of another alternative, then we can't see any other DM who does the contrary. To put it in other words, all DM's in this society try to comply with the same set of norms.

Building on the assumption that our data is heterogeneous enough, our main result states that the each DM's choice behavior satisfies the two properties postulates, and that the changes in the choice behaviors do not conflict if we can find a social code and an a profile of preferences that, on top of describing the choice of each DM, more comply with the social code as the pressure increases. With choice outcomes from multiple DMs, we show that the social code is completely uncovered.

The rest of the paper is organized as follows: The following section introduces the basic set up for individual decision making under social pressure where the primitives are preferences. In section 3.3., we switch to choice behavior and characterize it under social pressure with two intuitive properties. Section 3.4. presents the multiple-agent model. Section 3.5. concludes and all proofs are relegated to an appendix.

## 3.2 Model

Let $X$ be a finite set of $k$ objects which we view as the universal grand set of alternatives. The alternatives in $X$ are denoted by $x_1, x_2...x_k$. We analyze the behavior of a DM under different pressure levels. We model the social pressure level as a continuous variable $T = [0, 1]$ where a value close to zero corresponds to low and a number close to one to high levels of pressure. Therefore, we will often refer to $t = 0$ as no pressure and $t = 1$ as full pressure levels.

The DM's preferences over the alternatives in $X$ depend on the pressure level under which she evaluates them. Formally, we denote by $R^t$ the preference that describes the views over the alternatives of the DM at pressure level $t$. As usual, we assume preferences are complete and transitive binary relations over $X$.

Following the standard interpretation, $x_i R^t x_j$ is interpreted as $x_i$ is at least as good as $x_j$ for our DM when the level of pressure is $t$.[2] As usual, the behavior of our individual can be described by the maximization of her preference relation. When facing a subset of alternatives $A \subseteq X$ under pressure level $t$, the DM picks up the

[2]The strict and indifference relations associated with $R^t$ are denoted by $P^t$ and $I^t$, respectively. Hence, $x_i P^t x_j$ if and only if $x_i R^t x_j$ and $\neg x_j R^t x_i$ and $x_i I^t x_j$ if and only if $x_i R^t x_j$ and $x_j R^t x_i$

$R^t - maximal$ alternatives in $X$

$$\max(A, R^t) = \{x_i \in X : \nexists x_j \in X \text{ for which } x_j P^t x_i\}$$

We now explain how the preferences of our DM $\{R^t\}^{t \in T}$ are related to each other. A norm prescribes certain behaviors, which indicate what is "approved" or "disapproved" within a social context. We assume it is possible to compare the extent to which every alternative in $X$ is approved or disapproved by the society. Let $\succ$ be a complete binary relation on the alternative set $X$. We denote $x_i \succ x_j$ whenever $x_i$ is more socially accepted than $x_j$. We also assume the norm indicates exactly which alternative is more accepted compared to any other alternative. In other words, $\succ$ is also asymmetric, thus taking the form of a strict linear order.[3] We will refer to $\succ$ as a social code.

Now we are ready to describe how the preferences of our DM depend on social pressure. The more the DM is exposed to social pressure, the more she is willing to comply with the social code. It would be natural to define the compliance of our DM at a given pressure level $t$ as the set of pairs whose order in $R^t$ coincides with the order in $\succ$. But notice that, by doing so, we wouldn't capture the difference in the level of compliance between behaviors that comply with the social code partially, and those that do to a full extent.

To illustrate this point, let $X$ consist of two alternatives: $x_i$ and $x_j$ and suppose that $x_i \succ x_j$. Consider the following two scenarios: In scenario 1, the DM is indifferent between $x_i$ and $x_j$ and in scenario 2, she strictly prefers $x_i$ to $x_j$. A proper definition of compliance requires that the preference in scenario 2 complies more with the code than the preference in scenario 1. Thus, we define compliance at a given pressure level as follows:

**Definition 3.1.** *Let $t$ be a pressure level and $\succ$ a social code. The Compliance Set at pressure level $t$, $\sigma^t$, is the set of all pair of alternatives $(x_i, x_j) \in X \times X$ such that*

$$x_i \succ x_j \iff x_i R^t x_j$$

Note that these are the pairs that are both included or excluded in the preference relation of the DM and the social code $\succ$. In our former example, the DM in scenario 1 only complies with $(x_i, x_j)$, while the DM in scenario 2 complies both with$(x_i, x_j)$ and $(x_j, x_i)$. We can now formalize the idea of increasing compliance. When the level of pressure increases, the compliance set of DM's preferences expands. This is captured by our next definition.

[3]We say that a binary relation is a strict linear order whenever it is complete, asymmetric and transitive.

**Definition 3.2.** *The collection of* $\{R^t\}^{t\in T}$ *is said to be* $\succ$ *−monotone if for all* $t, t' \in T$ *such that* $t \leq t'$*, we have* $\sigma^t \subseteq \sigma^{t'}$*.*

The notion of code-monotone preferences captures two important features: First, it describes how individual behavior converges to the prescribed behavior as the pressure level increases. In other words, the presence of social pressure triggers the enforcement of the norms. Second, this convergence is in a very strong sense. Code-monotonicity implies that if our DM feels the need to modify her preferences because the social code says so, then for an increasing level of pressure it is impossible to observe her to switch back to the own preference. For instance, a long-haired male who gets a haircut because his family disapproves it, would certainly not grow it back if his high school is similarly opposed to such a hairstyle.

In many applications we may be interested in using a cardinal notion of compliance. Notice that $\succ$ *−monotone* preferences imply an increasing compliance rate with respect to pressure, where the compliance rate is defined as the cardinality of the compliance set. However, the other implication is not satisfied.[4]

Proposition 3.1. shows that $\succ$ *−monotone* preferences can be reformulated in terms of compliance rates if the compliance rates are defined as a weighted sum of the alternatives in the compliance set. This weighted sum, which we denote as the weighted compliance rate, must be increasing in $t$ for all possible non-negative weight values. Before proceeding with Proposition 3.1., we first present the formal definition of a weighted compliance rate.

**Definition 3.3.** *Let* $t$ *be a pressure level and* $\succ$ *a social code. Given a collection of positive weights* $\{w_{ij}\}$*, the Weighted Compliance Rate at pressure level* $t, \phi_w(\sigma^t)$ *is as follows:*

$$\phi_w(\sigma^t) = \sum_{(x_i,x_j)} w_{ij} I(ij)$$

*with* $I(ij) = 1$ *for* $(x_i, x_j) \in \sigma^t$ *and* $I(ij) = 0$ *for* $(x_i, x_j) \notin \sigma^t$*.*

We count the number of alternatives in the compliance set, but now we attach a different value to each of the pairs in it. We are now ready to present Proposition 3.1.

**Proposition 3.1.** *A collection of preferences* $\{R^t\}^{t\in T}$ *is* $\succ$ *−monotone if and only*

[4]For an example illustrating this point, suppose that for pressure levels $t \in [0, 0.2]$, let the preference of the DM is $cP^t bP^t a$. For levels $t' \in (0.2, 0.5]$, her preferences are $bP^{t'} cP^{t'} a$. Finally, for the pressure levels $t'' \in (0.5, 1]$, the preferences are $aP^{t''} cP^{t''} b$. Let the code be $a \succ b \succ c$. Thus, the compliance rates of $P^t, P^{t'}$ and $P^{t''}$ with respect to the code $\succ$ are increasing with the values 0, 2 and 4, respectively. However, it is obvious that these preferences are not $\succ$ *−monotone*. Notice that the pair $(b, c)$ is in the compliance set of $P^t$ but not in that of $P^{t''}$

*if for every collection of positive real weights* $\{w_{ij}\}$*, the Weighted Compliance Rate* $\phi_w(\sigma^t)$ *is increasing with respect to* $t$*.*

It is natural for the weighted compliance rate of $\succ$ $-monotone$ set of preferences to be increasing in $t$, since the compliance set expands with the increasing levels of pressure. Aside from this, Proposition 1 states that, if the weighted compliance rate of a set of preferences is increasing for every collection of weights, then these preferences should be $\succ$ $-monotone$.

## 3.3 Individual Choice

Until now we have described the economic agent only through preferences. Preferences are very useful concepts to define the DM's attitude toward alternatives; however, they are never expressed directly. Thus, in this section we switch our analysis from preferences to the choice behavior, and show which regularities over choices guarantee the existence of a social code $\succ$ and a set of $\succ$ $-monotone$ preferences that explain every single choice of our DM.

Let $\mathcal{X}$ be the set of all non-empty subsets of $X$. We define a choice problem with pressure as $(A, t)$ with a menu $A \in X$, paired with a pressure level $t \in T$. We interpret a choice problem as follows: The DM faces a menu $A$. In addition, she becomes aware of the extent of the social pressure, $t$. From each choice problem $(A, t)$ a subset of $A$ in $\mathcal{X}$. That is $c : \mathcal{X} \times T \to \mathcal{X}$ with $c(A, t) \subseteq A$ for all $A \in \mathcal{X}$ and $t \in T$.

Our first axiom is an immediate reflection of the weak axiom of revealed preference *i.e.*, the WARP. In the context of our study, we can intuitively reformulate the weak axiom as follows:

**Axiom 3.1.** *(WARP given pressure level) Let* $t \in T$ *and* $A, B \in \mathcal{X}$*. Consider any pair* $x, y \in X$ *such that* $x, y \in A \cap B$ *and* $x \in c(A, t)$*. If* $y \in c(B, t)$ *then* $x \in c(B, t)$

This property ensures that for a given pressure level the choice correspondence $c$ satisfies the *Weak Axiom of Revealed Preferences*. That is, if the DM reveals that she prefers alternative $x$ to alternative $y$, then it is impossible to observe that she reveals the contrary if the pressure level has not changed. Thus, at each pressure level $t$ the choice behavior of the DM is rationalizable. In particular, for every $t \in T$ and $A \in X$, there exists a preference relation $R^t$ such that, $c(A, t) = \max(A, R^t)$

*Axiom 3.1* guarantees that some sort of utility maximization happens at each pressure level $t$. The following axiom, however, builds a connection between choices under different pressure levels.

**Axiom 3.2.** *(Consistency in Compliance) Let* $A \in \mathcal{X}$ *and* $t, t'$ *and* $t''$ *with* $t \leq t' \leq$

$t''$. *If* $x \in c(A,t)$ *and* $x \in c(A,t'')$ *then* $x \in c(A,t')$. *Moreover if* $\{x\} = c(A,t)$ *and* $\{x\} = c(A,t'')$ *then* $x = c(A,t')$

We can view this property, in contrast to Axiom 3.1., as a consistency requirement across different pressure levels. Our DM is consistent in terms of her compliance with the code. That is, if she chooses one alternative over another both at low and a high level of pressure, then we can postulate that if she hasn't already changed her choice at the low level, then either her personal preferences and the social code coincide – so that she doesn't feel the need to review her choice – or the pressure level is not high enough to make her to change her choice in compliance with the code. Notice that in both scenarios there is no reason for her to modify her choice in an intermediate level.

The following theorem states that the choice behavior of a DM satisfies these two basic properties if we can find a social code $\succ$ and a set of a rational, $\succ$ $-monotone$ preferences that explains her choices. Formally:

**Theorem 3.1.** *A choice behavior c satisfies Axiom 3.1. and Axiom 3.2. if and only if there exists a code* $\succ$ *and a family of* $\succ$ $-monotone$ *preferences* $\{R^t\}^{t \in T}$ *such that* $c(A,t) = \max(A, R^t)$.

This result identifies a pattern of choice behavior for different pressure levels. Its intuitive appeal is quite straightforward: The more pressure you feel, the more you converge to a socially accepted behavior.

The theorem's sufficiency proof is established through a series of simple steps. First, we show that Axiom 3.1. guarantees that we can construct a set of rational preferences describing the choice behavior of the DM at each pressure level. We then start to elicit a code from the observed behavior. We know that the only motivation for our DM to modify her choices is to comply with the social code. Thus, her changes across different pressure levels reveal information about this code. We then introduce a binary relation that links these pairs of alternatives, such that one of the alternatives in this pair improves upon the other. An improvement is defined based on the relative positions of the alternatives in the revealed preferences when pressure increases. Therefore, an alternative $x$ is said to improve upon another alternative whenever its relative position with respect to $y$ increases from $R^0$ to $R^1$. We show that this improvement relation is transitive and antisymmetric. Hence, it is a partial order. Later, by invoking Szpjlran's Theorem we build a code by extending this partial order. Notice that the elicitation is not complete. Last step shows that Axiom 3.2. implies that the compliance sets expand when the level of pressure increases. Therefore, we can conclude that these preferences are code-monotone.

## 3.4 Multi agent Setup

In the previous section we showed that the choice behavior of a DM who satisfies the two axioms can be described by a family of rational, $\succ$ $-monotone$ preferences. Moreover, we claim that we can reveal some information about the norms that are accepted by this society, just by looking at individual behavior across different pressure levels.

In this section we take the analysis one step further. We intend to explore the choice behaviors of a group of individuals under social pressure, among which there is no strategic interaction although they are all affected by the same social code. Therefore, in addition to the assumptions that each DM's choice behavior satisfies the two basic axioms, we also assume that the aggregated data satisfies a consistency property, which we refer as *Social Coherence* (SC).

This property postulates that there are no two individuals in this society who try to comply with different social codes when facing the same choice. We show that the if the choice behavior of each DM satisfies Axiom 3.1. and Axiom 3.2., and the collection of these choices satisfies SC, we can find the social code $\succ$ that applies to the choice problem in hand and a profile of $\succ$ $-monotone$ set of preferences that explains the choices of each individual. It is worth to note that the full revelation of the social code is possible in this setup.

Before proceeding with the analysis, let us first introduce the additional necessary notation for this section. We denote the society by $S$, and assume that there are $N$ individuals in $S$. As in the previous section, we can observe the choice behavior of each $i \in S$ across different pressure levels. Let us denote the choice correspondence of individual $i$ by $c_i$. We can consider our DM in the previous sections as a member of the society $S$.

With the following axiom, we basically assume that, as previously discussed, all individuals in society $S$ comply with the same social code. This axiom links the choice behavior of different DM's through the changes across different pressure levels.

**Axiom 3.3.** *(Social Coherence) Let $x, y \in X$ and $i, j \in S$ such that $y \in c_i(xy,t) \cap c_j(xy,t')$ and $x \in c_i(xy,t') \cap c_j(xy,t)$. If $c_i(xy,t) \neq c_i(xy,t')$ then $c_j(xy,t) = c_j(xy,t')$*

*Social coherence* can be interpreted as follows: For any pair $x, y$ in $X$; if $x$ improves upon $y$ for some individual in this society, then the opposite can not happen for someone else.

In order to prove the main result of this section we consider a relatively heteroge-

neous society, in the sense that it comprises agents with different tastes. This will ensure that we have sufficient amount of data in order to compare the order of each pair at low and high pressure levels. Moreover, some individuals are effected by social pressure so that they modify their behavior at higher social pressure levels. Formally, we assume that the following condition holds:

**Richness Assumption.** *For every triple of alternatives $x, y$ and $z$ there exists an individual $i$ and two pressure levels $t$ and $t'$ such that $c_i(S,t) \neq c_i(S,t')$ for all $S \subseteq \{x,y,z\}$*

Through this condition we assume that, at least for one individual in society $S$, the social code and her personal desires differ. Moreover, such DM is affected by the pressure she feels upon. Hence, she changes his choice behavior with increasing level of social pressure.

The following theorem states that a profile of choice behaviors that satisfy Axiom 3.1., Axiom 3.2. and SC can be characterized by a profile of rational, $\succ$ *−monotone* set of preferences. Formally we represent it as follows:

**Theorem 3.2.** *Given the richness assumption, $\{c_i\}_{i\in S}$ satisfies Axiom 3.1., Axiom 3.2. and Social Coherence if and only if there exists a code $\succ$ and a profile of $\succ$ −monotone preferences $\{R^t\}_{i\in S}^{t\in T}$ such that $c_i(A,t) = \max(A, R_i^t)$*

The interpretation is similar to that of Theorem 3.1., except that in Theorem 3.2. we can elicit the code completely.

The proof of this theorem follows a very similar logic to that of Theorem 3.1.. The main difference involves the construction of the social code $\succ$. To do so, we first define the binary relation that links the pairs in which one of the alternatives shows an improvement upon the other for each individual $i$ in $S$. We then set the code as the union of these individual improvements. The Richness assumption help us to show that $\succ$ is complete and SC implies that it is antisymmetric. We finally prove by using both Richness assumption and SC that $\succ$ is transitive. This concludes that $\succ$ is a code.

## 3.5 Concluding Remarks

In this paper we study the decision making process of an agent who is exposed to different degrees of social pressure. Our analysis is based on the observation that, with increasing levels of pressure, norm compliance is higher. The primitives of our first analysis are preferences. We formulate norm as a strict relation that order the alternatives depending on the level of their approval by the norm. The

weighted compliance rate is then introduced in order to capture the cardinality of the agreement between the DM's preference and the social code. We showed that the DM is described by a family of preferences that comply with the code more as the social pressure increases if, the weighted compliance for any collection of positive weights increases too.

We apply our analysis to the choices of the DM as well. We show that the choice behavior of the DM satisfies two consistency properties (one is defined at each pressure level and the other is defined across different levels) if we can find a social code and a set of preferences that describe the choices and comply with the code more as the pressure increases. We note that the individual choice behavior reveals partial information about the code. We then introduce our DM into a multiagent set up. We finally showed that given sufficiently heterogeneous choice data, the individuals comply with the same norm and satisfies the two basic properties if, again, each individual is described by increasingly compliant preferences. With multiple agents, however, we showed that the full revelation of the code is possible.

The present paper suggests several avenues for future research. First, it would be interesting to further elaborate the multiagent set up. In section 4, we treat society as a cohesive social unit. However several decision making set-ups indicate that there are different social categories and, hence, different set of norms associated to these categories. If we allow each category to correspond to a different type, then the analysis of the choice behavior of several agents under different pressure levels may reveal us the social types in such society. Obviously, SC property and Richness assumption should be revisited in this setup and the boundaries of identification strategies should be explored more.

Related to multiple social categories, another future research path would be the study of the interaction of different social types under different social pressure level. Adding a strategical interaction obviously brings in a new complexity to our model, but at the same time it allows us to better understand the micro-foundations of in-group vs out-group behaviors.

## 3.6 Appendix

***Proof of Proposition 3.1.*** We first start to prove the "only if" part of the proposition. Assume that $\{R^t\}^{t\in T}$ is a set of $\succ$ $-monotone$ preferences that is $\sigma^t \subseteq \sigma^{t'}$ whenever $t \leq t'$. Since the weighted compliance rate puts a positive weights over all the elements in the compliance set, we can immediately conclude that $\phi_w(\sigma^t) \leq \phi_w(\sigma^{t'})$ for $t \leq t'$.

To show the other implication suppose that $\phi_w(\sigma^t)$ is increasing with respect to

$t$, that is $\phi_w(\sigma^t) \leq \phi_w(\sigma^{t'})$ whenever $t \leq t'$. We will demonstrate that $\sigma^t \subseteq \sigma^{t'}$. Suppose for a contradiction that it is not. Hence there exists a pair $(x_l, x_m) \in \sigma^t$ but $(x_l, x_m) \notin \sigma^{t'}$. Now consider a weighted compliance rate with the following weight function $w_{ij} = 1$ where for all pairs of $i, j$ either $i \neq l$ or $j \neq m$ and $w_{lm} = n^2$. Hence, it is straightforward to see that $\phi_w(\sigma^t) \geq \phi_w(\sigma^{t'})$ ■

***Proof of Theorem 3.1.*** We begin by proving the "only if" part of the theorem. Take a choice behavior $c$ that satisfies Axiom 3.1. and Axiom 3.2. We will first show that we can elicit a collection of rational and $\succ -$*monotone* preference relations that generate the choice behavior $c$ through maximization. We abuse notation by often suppressing set delimiters *e.g.* writing $c(xy, t)$ instead of $c(\{x, y\}, t)$

- Axiom 3.1. implies that $c$ satisfies WARP at each $t \in T$. By the classical argument based on WARP, we know that there exists a preference relation that can be constructed from binary choices. Therefore, at each pressure level, $R^t$ is defined by

  $$xR^t y \text{ if and only if } x \in c(xy, t)$$

  The strict relations associated with $R^t$ will be identified as

  $$xP^t y \text{ if and only if } x = c(xy, t)$$

  and the indifference will be as

  $$xI^t y \text{ if and only if } \{x, y\} = c(xy, t)$$

- We now start to define a code. In the model described, a change in choice behavior is originated from increasing pressure. Whenever the pressure level increases, the DM complies with the norms more. By Axiom 3.1., we know that each revealed preference $R^t$ is well-behaved, in particular $R^0$ and $R^1$, the revealed preferences at no pressure and full pressure levels, respectively. Hence we can elicit a part of the code from these extreme levels. To do so, we construct a binary relation, say $\theta$, where $x\theta y$ will be described as "$x$ represents an improvement over $y$". An alternative $x$ improves upon $y$ whenever $y$ is revealed at least as good as $x$ in $R^0$ and $x$ is revealed at least as good as $y$ in $R^1$, given that one of the weak relations is strict. Formally $\theta$ is defined as follows:

  $$\theta = \{(x, y) \in X \times X \mid (yP^0 x \text{ and } xR^1 y) \text{ or } (yI^0 x \text{ and } xP^1 y)$$

- Now we will show that $\theta$ is transitive. Take $x, y, z \in X$ such that $x\theta y$ and $y\theta z$. We will show that $x\theta z$ as well. Notice that an improvement of $x$ upon $y$ implies that $yR^0x$ and $xR^1y$ with at least one of the weak relations is strict i.e., either $yP^0x$ or $xP^1y$. Same applies for $y\theta z$ that is either $zP^0y$ or $yP^1z$. Hence, by transitivity of $R^0, R^1, P^0$ and $P^1$, together with the conditions above, we have one of the following three cases: (i) $zP^0x$ and $xR^1z$, (ii) $zR^0x$ and $xP^1z$ and finally (iii) $zP^0x$and $xP^1z$. These all three cases define the improvement relation $\theta$ and we have $x\theta$. This proves that $\theta$ is transitive.

- It is obvious that $\theta$ is not necessarily a complete relation defined on $X$ since it doesn't include the pairs that do not represent an improvement upon each other. However, we have proved that it is transitive. Moreover, notice that it is antisymmetric by definition. Thus $\theta$ is a partial order. By Szpilrajn's theorem, we know that every partial order on a nonempty set can be extended to a linear order on the same set. Thus we denote $\succ$ as one of the extensions of $\theta$ and we call it the revealed code of the choice behavior $c$.

- Now we will show that the set of the revealed preferences $\{R^t\}^{t\in T}$ are $\succ -monotone$. Take some pressure levels $t, t'$ such that $t \leq t'$ and suppose that the pair $(x, y)$ is in the compliance set at level $t$ i.e., $(x, y) \in \sigma^t$. We will show that $(x, y) \in \sigma^{t'}$ as well. Notice that $(x, y) \in \sigma^t$ by definition implies either (i) ($x \succ y$ and $xR^ty$) or (ii) ($y \succ x$ and $yP^tx$). Assume the first case applies. $x \succ y$ means that $x$ represents an improvement upon $y$ thus, we have $xR^1y$. Hence we have both $x \in c(xy, t)$ and $x \in c(xy, 1)$, and Axiom 3.2. implies $x \in c(xy, t')$, therefore, $(xR^{t'}y)$. This concludes that $(x, y) \in \sigma^{t'}$. Now consider the other case i.e., $y \succ x$ and $yP^tx$. By following the similar argument above, an improvement of $y$ over $x$ implies $xR^0y$. Note that $xR^0y$ and $yP^tx$ is equivalent to $x \in c(xy, 0)$ and $y = c(xy, t)$, respectively. Axiom 3.2. implies that $y = c(xy, t')$, which is equivalent to $yP^{t'}x$. Hence we get $(x, y) \in \sigma^{t'}$.

To prove the other implication assume that there exists a code $\succ$ and $\succ -monotone$ family of preferences $\{R^t\}^{t\in T}$ such that $c(A, t) = \max(A, R^t)$. this means that $c$ is rationalizable at each $t \in T$ i.e., $c(\cdot, t)$ satisfies WARP, hence $c$ satisfies Axiom 3.1.

To show that c satisfies Axiom 3.2., take pressure levels $t, t'$ and $t''$ with $t \leq t' \leq t''$. Suppose that for some $A \in \mathcal{X}$, we observe that the choice behavior is as follows: $x \in c(A, t)$ and $x \in c(A, t'')$. Suppose for a contradiction that $x$ is not chosen at the intermediate level i.e., $x \notin c(A, t')$. This implies that there exists another element $z$ which is strictly preferred to $x$ i.e., $zP^{t'}x$. Moreover, $x \in c(A, t)$ and $x \in c(A, t'')$ imply that $x$ is at least as good as all other options in $A$. Therefore we have $xR^tz$

and $xR^{t''}z$. If the code orders $x$ and $z$ as $x \succ z$, then with $xR^{t'}z$, we know that $(x, z)$ is in the compliance set of pressure level $t$; i.e., $(x, z) \in \sigma^{t'}$ But, since $zP^{t'}x$, we get $(x, z) \notin \sigma^{t'}$. This contradicts with the assumption of $(xR^{t'}y)$ being $\succ -monotone$. On the other hand, if the code orders $x$ and $z$ as $z \succ x$, then by $zP^{t'}x$, we have $(z, x) \in \sigma^{t'}$. This generates a contradiction to $\succ -monotone$ preference relations since $xR^{t''}z$ implies $(x, z) \notin \sigma^{t''}$.

For the case in which $x$ is the unique choice of the DM at pressure levels $t$ and $t''$, we will show that it has to be the unique choice at $t'$ as well. Let $\{x\} = c(A, t) = c(A, t'')$. Therefore $x$ is strictly preferred to all $y$ in $A$ i.e., $xP^t y$ and $xP^{t''}y$. Suppose for a contradiction that $\{x\} \neq c(A, t')$. Therefore there exists another alternative $z$ in $A$ such $zR^{t'}x$. If the code orders $x$ and $z$ as $x \succ z$, then, with $xP^t y$ for all $y \in A$, we have $(x, z) \in \sigma^t$. But notice that since $zR^{t'}x$, we get $(x, z) \notin \sigma^{t'}$. This contradicts to the assumption $\succ -monotone$ preferences. Lastly, if $z \succ x$, then by $zR^{t'}x$, we have $(x, z) \in \sigma^{t'}$. However, since $x$ is the unique choice at $t''$, $z \succ x$ implies $(x, z) \notin \sigma^{t''}$. This is again the desired contradiction.■

***Proof of Theorem 3.2.*** We first show the "only if" part. Take a collection of choice behaviors $\{c_i\}_{i \in S}$ such that each $c_i$ satisfies Axiom 3.1., Axiom 3.2. and SC. We will show that we can elicit a profile of rational and $\succ -monotone$ preference relations $R_i^t$ for each $i \in S$ and at each pressure level $t$, that generates the choice behavior $c_i(\cdot, t)$.

- We follow the steps in the proof of Theorem 3.1. and construct $R_i^t$ in the same way, that is

$$xR_i^t y \text{ if and only if } x \in c_i(xy, t)$$

- Now we define a binary relation describing changes by using the choice data. We start, as in the proof of Theorem 3.1., by defining the set of changes for DM i as follows:

$$\theta_i = \{(x, y) \in X \times X \mid (yP_i^0 x \text{ and } xR_i^1 y) \text{ or } (yI_i^0 x \text{ and } xP_i^1 y)$$

  Now consider the union of all these changes. Formally: $\succ = \bigcup_{i \in N} \theta_i$.

- We now show that $\succ$ is a code. Namely, that it is complete, asymmetric and transitive.

  By the Richness assumption we know that for each triple of alternatives $x, y$ and $z$, there exists one individual who has modified her choice behavior for some pressure levels $t, t'$. Hence for any $x, y \in X$, there is one individual such that $c_i(xy, t) \neq c_i(xy, t')$. Without loss of generality, assume that

$t \leq t'$. Therefore for these pressure levels, we have four cases, without loss of generality, in which the choice at $t$ involves $x$:

1. $x = c_i(xy, t)$ and $\{x, y\} = c_i(xy, t')$
2. $x = c_i(xy, t)$ and $y = c_i(xy, t')$
3. $\{x, y\} = c_i(xy, t)$ and $y = c_i(xy, t')$
4. $\{x, y\} = c_i(xy, t)$ and $x = c_i(xy, t')$

Given the revealed preferences that are defined in the proof of Theorem 3.1., it is straightforward to see that the first three cases together with Axiom 3.2. imply an improvement of $y$ upon $x$ while the the fourth one implies the contrary.

Therefore for each $x, y \in X$, there exists an individual $i \in S$ such that either $x\theta_i y$ or $y\theta_i x$. Hence, we can conclude that either $x \succ y$ or $y \succ x$. It is straightforward to see that by SC $\succ$ is asymmetric. Now we will show that it is transitive. Take three alternatives $x, y, z \in X$ such that such that $x \succ y$ and $y \succ z$. Notice that these comparisons imply that there exists $i, j \in S$ such that $x\theta_i y$ and $y\theta_j z$. If $i = j$, then it is immediate from the proof of Theorem 3.1. that $x\theta_i z$. Now suppose that $i \neq j$. By the Richness Assumption we know that there exists $l \in S$ such that $c_l(S, t) \neq= c(S, t'), S \subseteq \{x, y, z\}$ for some $t \leq t''$. Axiom 3.2. implies that $c_l(S, 0) = c_(S, t) \neq c_l(S, t') = c_l(S, 1)$. Let $S = \{x, y\}$ and $\{y, z\}$. Since $x\theta_i y$ and $y\theta_j z$, respectively, SC guarantees both that $x\theta_l y$ and $y\theta_l z$. By the transitivity of $\theta_l$ we get $x\theta_l z$ This proves that $x \succ z$. Notice that the proof of Theorem 3.1. applies to any possible extension of the partial order $\theta_i$. Since the social code $\succ$ that we have construct in step 2 is also one of these extensions, the rest of the proof applies exactly as in Theorem 3.1.. Hence we omit it here. ■

# Chapter 4

# Dishonesty through others: Advice and Deception

## 4.1 Introduction

Over the last decade dishonesty and its consequences, ranging from devastating global financial crises to profit loss due to petty theft in the workplace, have drawn increasing attention from researchers in economics and other fields. While misconducts like deception or corruption prevail in "real life", data and everyday interactions seem to show that a considerable amount of people abstain from engaging in unethical behaviour even if it is inconsequential. Therefore, an important body of research has been trying to understand what honesty entails and how it can be captured in market settings. Or, in a broader sense, how a (socially) appropriate behaviour would be explained in cases where the economic agent paradigm may seem to predict its non-existence.

One strand of the literature has focused on the underlying reasons behind this phenomenon, especially in the context of lying decision. The intrinsic cost of lying (Gneezy 2005; Erat and Gneezy 2012; López-Pérez and Spiegelman 2013), guilt aversion (Charness and Dufwenberg 2006), self and social image concerns (Mazar et al. 2008; Bryan et al. 2013) are some of the explanations that have been put forth and studied so far. Another line of research has explored the resistance of (dis)honest deeds to the changes in social or institutional circumstances which in return, allows us to identify and understand the relevance of the aforementioned underlying reasons better. For instance, how making decision for the group (Conrads et al. 2013), the existence of intermediaries (Erat 2013) or avoiding information about the outcomes of one's own actions (Dana et al. 2007) would affect the prevalence of (socially) appropriate behaviour.

However, an important aspect of social life has mostly been overlooked in the investigation of (dis)honesty. Many "real life" choices where individuals face the dilemma of engaging in dishonest acts and earning higher payoffs, are taken after consulting with others. In the standard economic analysis this consultation process is analysed

in settings where there is some form of information flow from a better informed agent (advisor) to a less informed one (advisee). On the other hand, the psychology and organisational decision making literature has considered non-informational dimensions of advice. Namely, by studying how emotions (Gino et al. 2012), responsibility attributions (Harvey and Fischer 1997)) or perceptions (Brooks et al. 2015) influence advice seeking and advice following.[1]

Motivated by those observations and findings, we are interested in advice in what could be called a "moral context", where there is no expertise or informational value embodied in advice itself. Hence, the main question of the present paper is whether such normative advice has any effect on the execution of a dishonest act. Particularly, in situations where one has to lie in order to obtain material gains.

In line with the existing literature on (dis)honesty, we conjecture that in such context advice may have a two-fold effect: First, it may decrease the discomfort one suffers in the act of lying. In this sense, advice could be considered as an emotional support of sorts; a mechanism that helps individuals to maintain a positive self-image while acting unethically. Second, as noted in political and management literature, it may help the decision maker to manage their social image by signalling the others, either those at the other end of the interaction or independent third parties, that the decision is taken under someone else's influence, although this person has no agency on the decision making process (Harvey and Fischer 1997; Vreeland 1999).

To address those questions, we adapt the deception game in Gneezy et al. (2013). In the baseline treatment, a sender is assigned a number between one and six and has to report it to a receiver through a message. The sender's payoff increases with the number reported, while the receiver's earning depends on whether the message contains the true assigned number. In other words, the sender is incentivised to misreport, i.e., to lie.

In contrast to Gneezy et al. (2013)'s game, to be able to amplify the ethical responsibility underlying the decision, we mute the role of the receiver, whose payoff in the present work is only determined by the sender's actions. A more important modification is included in the form of two different treatments, in which we introduce an advisor whose payoff doesn't align with none of the players, and who is consulted by the sender before sending the message. The advice is offered upon request and senders can take any action after receiving the advice, effectively overlooking it. Advice comes in the form of "send the message that 'the assigned number is...' "

In one variant of the treatments, which we call private-advice treatment (PRT), ad-

[1]See Schotter 2003 for a review of naive advice in economic experiments and Bonaccio and Dalal 2006 for a review of how advice is provisioned in organisational decision making

vice is asked confidentially. Only the sender and the advisor who is consulted know whether a recommendation is asked. In public-advice treatment (PUT), however, the other participants can observe when a sender asks for advice. That is, in PUT advice-asking is scrutinised. Hence, our design allows us to test and measure the effect of advice on agents' lying decisions. In addition, by comparing the two treatments, we are able to distinguish the different motivations behind advice-asking. As mentioned before, whether advice is used to cope with the psychological cost of lying (advice asked in PRT and PUT) or as a tool to deal with reputational concerns (advice in PUT).

There are a few key points we should highlight before we proceed. Both senders and advisers make their decisions following a strategy elicitation method. This implies that they have to send a message for each number between 1 and 6 that could be potentially assigned to them. As in Fischbacher and Föllmi-Heusi (2013) and papers following their design, we also can distinguish among different lying patterns. In the present work, senders who choose to report six for each state are said to be lying "fully" and those who do not tell the truth but don't tell a full lie either, lie partially.

We find that 39.1% of the senders lie fully, while 48.3% tell a partial lie. Our results are in line with the existing experimental deception games, where it is well documented that people tend to disguise their lies by not lying fully. That is, choosing to tell a "lie" which does not maximise their pay-offs despite being untruthful. However, the literature's findings are somewhat conflicting. On one hand, partial lying has been related to reputational concerns, and it is shown to occur less when decisions are more visible to others Gneezy et al. (2016). On the other hand, Fischbacher and Föllmi-Heusi (2013) and Mazar et al. (2008) find no effect of anonymity on partial lying, and Mazar et al. (2008) in their self-concept maintenance theory argues that people cheat not to the maximum extent to maintain an honest self image while they can still lie for material gain. Our results are in line with self-concept maintenance. While we find no difference across PRT and PUT, we show that lying fully is more prevalent in the presence of advice whereas, in contrast, partial lying decreases as advice become available.

Despite the fact that there is no informational motivation behind advice, we see a high rate of advice-asking (74.2%). On the other hand, the fraction of senders who follow, sending a message exactly recommended by the advisor, is surprisingly low (28.3%). Our results, however, suggest that rate of advice-following is sensitive to the content of the advice. A larger fraction choose to follow in case of an advice of lying fully than an advice of reporting truthfully. We observe an opposite, but not a significant, relation in rate of advice-following when we restrict our attention to the advice of truth-telling vs. partial lying. However, the majority of the senders

who don't follow an advice of partial lying, tell a different type of partial lie and a considerable amount lie fully.

Our design also allows us to analyse how a non-incentivised advisor makes decision. Traditionally much focus has been put on professional advice, where an advisor's payoff is either negatively or positively aligned with the advisee's actions. However in many situations people rely on the opinions of colleagues, friends or parents, who may have important impact on the final decision but with no money at stake. Keeping in mind the effect of advice on sender's behaviour's, we compare advisors to the senders in our baseline, where the decision is not altered by the advice. Advisors' decisions are classified exactly as the senders'. Therefore, an advisor can lie fully, partially, or be honest with the advice she gives. Our results suggest that there is no difference in the decisions of lying fully or partially, across advisors and senders.

The contributions of this paper are the following: Normative advice, which occurs naturally in many real life situations, has not merited much attention in economics except Schram and Charness (2015). Our paper contributes to this line of research. It provides stronger evidence for the impact of normative advice since, unlike Schram and Charness (2015), there is no uncertainty on the nature of the dishonest behaviour in our context. We show that a considerable amount of people appeal to advice even there is no accountability for the decisions taken by the players, which, suggests even further that advice is used to mitigate the psychological cost of dishonest act. In addition, we show that non-incentivised participants (advisors) can make decisions similar to those who are incentivised (senders), even when the decisions posses socially inappropriate attributes.

The rest of the paper is organised as follows. In section 4.1.1, we review the literature on lying and advice briefly. Section 4.2 describes the design and the logistics of our experiment in detail. Section 4.3 discusses the predictions of lying, advice-asking/following and advising lying. Section 4.4 presents our main results while a discussion of post-experiment data and conclusion is given in Section 4.5.

### 4.1.1 Literature Review

The literature on lying is extensive both inside and outside the domain of traditional economic science. The phenomenon causing people to lie less than the predictions of canonical models of economic decision making has drawn the attention of empirical and theoretical researchers, who have suggested that people exhibit an aversion to lying (Gneezy 2005; Gibson et al. 2013; López-Pérez and Spiegelman 2013; Kartik et al. 2007; Dufwenberg 2016). While an important body of research tries to understand the nature of this aversion by changing, for instance, the incentives (Erat and Gneezy 2012; Kajackaite and Gneezy 2017) or distorting the strategic character of

the decision making (Mazar et al. 2008; Fischbacher and Föllmi-Heusi 2013), another branch of studies focus on different dimensions of (dis)honest acts, such as: whether and how communication means (Lundquist et al. 2009), framing (Cohn et al. 2014) or the content of the lie (Cappelen et al. 2013) affect one's misbehaviour.

We can divide the deception literature in two main groups. One group uses strategic information games where the decision maker may deceive another participant who, in most of the cases, affect the final outcome. In this setting, motivations such as guilt aversion or strategic honesty are also suggested as an explanation to the observed abstention from lying (Charness and Dufwenberg 2006; Charness and Dufwenberg 2010; Sutter 2009). The second group focuses on how participants report private information such as the outcome of a die roll or coin flip. Hence, the decision maker may only deceive the experimenter (see Abeler et al. 2016 for a meta study of 72 studies which use the die-roll experiment introduced by Fischbacher and Föllmi-Heusi (2013). In this second group, mainly image considerations are focal issues. We use a variant of the games from the first group instead of the second one because, as stated in Gneezy et al. (2013), we also believe that the identity of the affected party might influence the decision making. We want to highlight the morality of the situation and not to give a natural justification for the sender's lying ("nobody is hurt by my decision").

This paper provides experimental evidence on the effect of an important aspect of social interactions: the role of advice on lying behaviour. Our design and analysis is based on a common observation that has emerged in the existing literature: There is a positive fraction of participants who tell the truth (ethical type) and a positive fraction who lies maximally (economical type). But there is also a sizeable proportion of individuals who "misbehave" but not to the full extent possible (Mazar et al. 2008; Fischbacher and Föllmi-Heusi 2013; Gneezy et al. 2013; Abeler et al. 2016) and, moreover, a significant group whose dishonesty is sensitive to the aforementioned manipulations.

Our postulate that advice may influence lying by lowering the psychological cost associated to it is connected to the studies of self-justification, where individuals appeal to others' opinions or advice in order to maintain a honest self-view while engaging in dishonest deeds. (Shalvi et al. 2011a; Shalvi et al. 2015; Shalvi et al. 2011b). Experimental economics, mostly, accounts for partial lying by referring to such explanations, particularly to the notion of self-concept maintenance which was coined by Mazar et al. (2008). In this vein, partial lying can be considered as another internal justification that may ease the threat to one's self-image as a honest person (Fischbacher and Föllmi-Heusi 2013; Khalmetski and Sliwka 2017). This threat itself may create an emotional distress and trigger to seek advice despite

it is useless. Indeed, some research in advice literature show that people appeal to advice more when they feel more anxious (Gino et al. 2012).

However, the findings on lying partially are mixed. For instance, Gneezy et al. (2016) explain partial lying with reputational considerations since partial lying is less prevalent as the choices can be verified by the experimenter than when they are unobservable. It is well documented in experimental and field studies, that social-image (reputation) may have important effects on economical decisions (Andreoni and Bernheim 2009; Samek and Sheremeta 2014; Bursztyn and Jensen 2017). For deception games, though, the results are not conclusive. For instance, Ven and Villeval (2015) and Fischbacher and Föllmi-Heusi (2013) finds no significant effect of scrutiny on dishonest act but, in contrast, studies like Gneezy et al. (2016) and Hao and Houser (2017) suggest the opposite. Although, from advice literature, we know that the advice seeker may interpret advice as a way to share the responsibility (Harvey and Fischer 1997; De Wit et al. 2017), we do not provide any evidence that social-image concerns are in action in advice seeking in our setting.

To our knowledge, the only paper that studies the effect of normative advice on decision is Schram and Charness (2015). Advice in their context comes from a group of participants with the intention of creating a shared understanding of what is an appropriate distribution of some amount of money where this appropriateness is not clear a priori. It is difficult for agents to decide which distribution is fair. Accordingly, advice still has some informational value. In contrast, in our setting there is no ambiguity about the immoral nature of the behaviour. Moreover, unlike the advisor group in Schram and Charness (2015), in our experiment advice is formulated by a single participant, eliminating the potential of creating a perception that lying or not lying is a norm.

With respect to actions that the decision maker may take, the relevant studies are scarce. To the best of our knowledge, only two papers (Erat 2013; Sánchez-Pagés and Vorsatz 2009) expand the binary, lie-not lie choice space. Sánchez-Pagés and Vorsatz (2009) adds the option of remaining silent into the message set by which they aim to distinguish between preferences for truth telling and a dislike for lying. Erat (2013) employs a delegation decision, where the principal (sender) can hire an agent to lie for herself.

## 4.2 Experimental Design

This paper focuses on advice, its effect on lying decision and the underlying reasons for searching for advice in such settings. We adopt as a stepping-stone the one-shot version of the sender-receiver game developed in Gneezy et al. (2013), with some

modifications explained below.

There are two types of players in the baseline (a sender and a receiver), and three in the treatments (a sender, receiver, and adviser).

The baseline sender-receiver game is a close variant of Gneezy et al. (2013) which allows us to compare the lying behaviour of our subject pool in the absence of advice. Each sender (Participant A) is matched with a receiver (Participant B), who receives a message from the sender about a state $s$, that is assigned to the sender. The assigned state or the number is an outcome of a virtual die roll i.e., $s \in \{1, 2, 3, 4, 5, 6\}$. The message should be in the form of: "The assigned number is $r$" where $r \in \{1, 2, 3, 4, 5, 6\}$. The sender's payoff only depends on the number she sends in the message according to the following function: $\pi_A = 8 + 2r$.

On the other hand, unlike Gneezy et al. (2013), in our design the receiver is a completely passive player and doesn't make any decision. By doing so we intend to capture a setting where the sender can't legitimise lying by the possible actions the receiver could take. In their paper, the receiver's payoff is determined as a result of whether she follows the message and whether the message contains the true assigned number. In such context, given that the payoff structure is common knowledge, a sender can lie expecting that a receiver who understands the incentives wouldn't follow the message. This may shrink the responsibility a sender feels about her own actions. Nevertheless, we too, employ another participant as a recipient instead of the experimenter or a hypothetical player which, otherwise, could also provide some ground to self-justifying lying. Moreover, by making receiver inert, we simplify the decision even further and decrease the chances of asking advice for informational motivations. All in all, in our setup the recipient payoff is 8 if the message contains the assigned state i.e., $s = r$ and 0 otherwise.

In the treatments, we consider the lying/truth telling behaviour of the sender when she can ask the opinion of the advisor (Participant C), to learn what the advisor would do when confronted with the same decision problem. Advice is introduced in the instructions as follows.

*"We imagine that it may be difficult for you, as a Participant A, to decide which message to send in this situation. Therefore we give you the opportunity to ask advice from another participant, Participant C, to learn what s/he would do being in your position."*[2].

A sender is matched with a receiver and also with an advisor. The sender plays

[2]Although the contexts are different enough, we try to keep the phrasing in the instructions as close as possible to the one in Schram and Charness (2015), since it is the only paper that studies the effect of moral advice in subsequent behaviour.

the sender-receiver game as described above, with the same payoff function but before delivering any decision, she is given the option of consulting to the advisor she is matched with. The advice comes in the form of: " Send the number that the assigned number is...". In our context advice is optional and non-binding. After receiving the advice, the sender makes his decision and sends any number she wants to send. Given that in our design advice is moral and there is no "good" or "better' decision, we don't incentivise advice: the advisor's payment is not aligned neither with the sender's nor the receiver's payoff. Therefore, advisor gets a fixed amount of 8.

The difference between the two treatments hinges on the prior that advice is considered either as another source of "moral wiggle room", which may help to avoid the disutility inflicted upon a honest self-image from lying, or as a channel to signal to a third person that the decision was taken under the influence of advisor. Therefore, in PRT the advice-asking process is kept as confidential as possible. The senders are asked to mark on a tick-box (☑) if they want to receive advice. Hence, no one except the advisor could know whether the sender has demanded any advice. On the other hand, in PUT we alter the visibility of advice-asking in the way that the sender has to raise hand if she is appealing any recommendation. In the experiment room, in PUT sessions, the subjects were arranged in a way that the recipients, the experimenters, and the other senders could see those who wanted advice.

In a one-shot, strategy-elicitation method in which both the senders and the advisors have to make contingent decisions, the senders send a message and the advisor gives an advice for each integer between 1 and 6. [3]

### 4.2.1 Experimental Procedure

The experiment was conducted using pen and paper at the Autonomous University of Barcelona with a total of 236 subjects who were recruited through ORSEE (Greiner 2004). The main and first part of the experiment took place in October 2016, in eight sessions with 174 subjects. The second part was conducted in five sessions between February and March 2017, applying minor changes as detailed in the following paragraphs. The whole experiment was implemented with a between-subject design, in which a pre-determined number of subjects were allocated to each treatment and subjects were randomly assigned roles. Table 4.1 summarises the number of participants, their roles and the treatments.

[3]Considering the nature of the interaction, which doesn't involve any strategic thinking, we believe there is no particular reason for senders to send different messages in a repeated-version of our design. More importantly, lying in a repeated sender-receiver game may confound with other reasons such as sender's depleting self-control due to truth-telling in the previous periods. See Gneezy et al. 2013 for the repeated version of the sender-receiver game without the advisor)

Table 4.1: *Treatments and roles*

| *Treatments* | *#of Senders* | *#of Recievers* | *#of Advisors* | *Type of Advice* |
|---|---|---|---|---|
| Baseline | 25 | 25 | — | No Advice |
| PRT | 32 | 32 | 32 | Private Advice |
| PUT | 30 | 30 | 30 | Public Advice |

All sessions were run in the same classroom. Upon entering the classroom, the participants chose a piece of paper which determined their experimental id number as well as their role. They were informed that they were matched with one participant in the baseline sessions and with two other participants in the treatments, who were present in the same session.

The instructions were read out in Spanish and the participants were told to read them once again on their own, and several examples were worked out before the actual experiment started. The design and instructions mostly followed Gneezy et al. (2013), since we essentially used a modified version of their sender-receiver game. However, our game was one-shot and an advisor was included in the decision making process in the two treatment sessions. Subjects were told that all decisions would be kept anonymous and the payments would be done in private by someone who didn't know the content of the experiment (see the English translation of the Instructions in the Appendix).

In the baseline treatment senders were provided with a "message sheet" on which they wrote down their messages for each possible assigned number. As stated in the instructions, a number between one and six had been assigned by the computer to each sender before the session, according to which the payments were carried out. The senders were informed about this number after they delivered their choices. In the treatments, before taking any decisions, senders had to decide whether they wanted advice and they asked for it in the way instructed by each treatment, either putting a tick on "advice sheet" (PRT), which was distributed to senders along with the decision sheet or raising hand (PUT). Advisors stated their advice in the same advice sheet, which was later returned to the corresponding sender. The same procedure in the baseline treatment was applied afterwards. Each sheet was passed from participant to participant in envelope. At the end of the sessions, subjects were asked to fill out a short questionnaire.

The payments of senders and receivers were calculated based on the message sent for the number assigned by the computers. They were paid in private and in cash right after we finished the sessions. All subjects in the first part of the experiment received a 5 euro participation fee.

After a preliminary analysis of the data, we found interestingly high rate of responses

which we interpreted as anomalies since they were not consistent neither with any social-preference model nor self-interested behaviour. In addition, except one sender and one advisor who sent one for each state, those choices were not even consistent in the "less is better" mindset. One could argue that the advisors, whose payments are constant, just didn't put enough effort to comprehend the decision structure and responded randomly. However, a surprising amount of senders whose decisions affected both their own and others' payoffs, seemed to play indiscriminately too. We presumed that this result reflected a lack of understanding of the decisions and the payoff structure associated to them.

Therefore, we conducted extra sessions with 62 students to collect more data. The second part of the experiment was conducted in February and March 2017, whose sessions were run as described above apart from applying some modifications to the instructions and performing an understanding test on senders and advisors, which was solved after reading the instructions. The modifications in the instructions involved changes like the order of sentences or paragraphs, extra examples explaining the strategy-method and associated incentives and the repetition of the last point throughout the text, since, we experienced in the earlier sessions that the participants had difficulties to understand contingent decision making.[4] Moreover, the participation fee in the second part was 3 euros due to budget limitations. The details of the changes made in the instructions and the understanding test can be found in the Appendix.

The subjects were told that the sessions wouldn't start unless every participant correctly finished the test. Subjects' responses were checked by the experimenters and any question was answered carefully. Although every participant who had to make a decision in the experiment passed the understanding test, we continue to observe anomalies in choices similar to those taken in the first part of the experiment. Nevertheless, the test and the changes led to a significant improvement on mistake rate overall.

Table 4.2 shows the rate of the anomalies in decisions in the first and the second wave of the sessions.

[4]See Brandts and Charness 2011 for a discussion of the comparison of strategy elicitation method versus direct response method.

Table 4.2: *Frequency of anomalies*

| *Treatments* | **With Test** | | **Without Test** | |
|---|---|---|---|---|
| | Senders | Advisors | Senders | Advisors |
| *Baseline* | 44,4 % (8/18) | — | 42,9 % (3/7) | — |
| *PRT* | 25 % (6/24) | 37,5 % (9/24) | 12,50 % (1/8) | 0 % (0/8) |
| *PUT* | 31,8 % (7/22) | 36,4 % (8/22) | 12,50 % (1/8) | 12,50 % (1/8) |
| *Total* | 32,8 % (21/64) | 37 % (17/46) | 21,7 % (5/23) | 6,3 % (1/16) |

Without the test we find that 38 out of 110 (34,5 %) and with test, 6 out of 39 (15,4 %) of the senders and advisors made decisions that cannot be explained by the predictions of any behavioural models that we know of. A chi-square test shows that the difference is statistically different ($\chi^2 = 5.08$, $p = 0.024$). Strikingly, this effect doesn't persist for senders.[5]. The rate of the unexplained choices decreases significantly for advisors. ($p = 0.025$, Fisher's exact test). This pattern partially confirms our initial presumption. Advisors, with fixed payoffs and no direct effect on the others' payments, may have slacked off trying to understand the instructions in the sessions without the understanding control. But, it seems the test forced advisors to think and grasp the decision structure. This easily explains the decrease in the advisors' haphazard behaviours.

On the other hand, the persistence in sender's failure suggests two possible explanations which are not necessarily mutually exclusive. Either the test was not able to measure what it was aiming to i.e., subjects' understanding of the decision form or, as discussed in Chou et al. (2009), what might seem as subject irrationality is a consequence of mis-implementation of the experiment which resulted in lack of engagement with it.

## 4.3 Predictions

Before we proceed with our hypothesis regarding the lying decisions of senders and advisors, it is worth to note and discuss a few key features of our design and experimental procedures. First, there is no uncertainty on the nature of the moral behaviour. Unlike Schram and Charness (2015), the senders know which decision is morally accepted and which is not. It is clear that a player, except when assigned to 6, should lie to gain more material payoff. Secondly, advice is given only by another participant and not by a group, which let us avoid confounding effects from induced

[5]It decreases from $32,81\%$ to $21,74\%$ but the two mistake rates are not significantly different. ($p = 0,426$, Fisher's exact test.)

norms in the lab. Third, there is no punishment; therefore, advice can not be used to shift the blame of a selfish or immoral act on a third party. Fourth, advice is given only when it is demanded by the sender.

We distributed two sheets along with two envelopes in the treatments: one for the senders' decision regarding which message to send and the other for advice-asking. Although we intended to obscure the identity and the decision of the senders as much as we could (the payment was carried forward in a separate room by someone who didn't know the content of the experiment, sheets were delivered in closed envelopes etc.), notice that we used individual level data without applying a double-blind procedure. The senders know that the payments were going to be calculated by the experimenter based on the their choices on the message sheet. Therefore in PRT, at the moment of the calculation, the experimenter couldn't know which senders had asked for advice since it was in the other envelope which was irrelevant for the final payoffs. In PUT, on the other hand, the experimenter could observe the senders who asked for advice and the senders were aware of this mere but salient difference before the experiment started.

One set of inferences in the present work are based on the premise that the senders may want to signal a third party, – to the experimenter, in this case – to show that the decision was not taken alone and in order to lessen the social stigma associated with deceptive behaviour. In the literature it is well established that striving for a positive self and social image may cause important alterations in behaviour that serve one's self-interest. (Ariely et al. 2009; Andreoni and Bernheim 2009; Grossman 2015). People search, consciously or unconsciously, for justifications or rationales that may help them maintain a self or social view which could be deemed as honest, generous etc., even when they are acting against to such image (Ayal and Gino 2011; Barkan et al. 2015; Drugov et al. 2014).

To sum up, we suggest two explanations for advice seeking in the set-up: First, advice is either a mechanism that helps the participants to cope with their "ethical discomfort" and/or maintain a positive self-image after engaging in dishonest act. Or, second, it is used for signalling a positive social image to a third party. Given the features of our experiment and the findings of the existing literature on the relation of partial lying and self-concept maintenance (Fischbacher and Föllmi-Heusi 2013; Mazar et al. 2008), we propose the following hypotheses.

If advice serves as a justification to lie without feeling too bad about oneself, then partial lying looses its functionality, to some extent, in the existence of advice. Therefore, the first hypothesis, we test is that advice affects partial and full lying in the opposite directions. We argue that under self-concept maintenance assumption

(SCMA) and given that advice are mainly in the direction of lying, lying fully is more prevalent in PRT and PUT compare to baseline and lying partially is *less* prevalent in PUT and PRT compare to the baseline.

Remember that in PRT senders cannot signal to a third-party whether the advice is taken or not, unlike in PUT. Therefore, advice in PRT cannot be used with reputational concerns. Hence, our second hypothesis compares the lying decisions across treatments to see whether advice serves as a tool to manage one's social image. This hypothesis implies that lying rate, both when full and partial cases, should be higher in PUT. We argue that under social-image maintenance assumption (SIMA), regardless of the advice content, lying partially and fully are more prevalent in PUT compare to baseline and PRT. Moreover, it should be same across PRT and the baseline.

In our design advice is voluntary and doesn't provide information for better decisions. A self-interested sender without any image concern, self or social, maximises the material gain, hence, doesn't need any advice. Similarly, pure moral subjects wouldn't ask for advice either, they always tell the truth. Therefore, implicit in our design (because advice is on request) we suggest that those who ask advice indeed want to lie fully but need some sort of confirmation for the reasons we discussed so far. Therefore, both under SCMA and SIMA, the rate of advice-following should be highest upon an advice of lying fully. Under SIMA, the content of the advice is irrelevant since neither the experimenter nor any other participant can see the advice type. Hence, we argue that following rate for advice of partial lying and truth telling should be lower than for advice of lying fully.

On the other hand, not following an advice of truth-telling can be a stronger threat to one's self image. However, as we discussed earlier, if individuals that are willing to lie ask for advice, then there is another force in play. In this case, even if they don't follow an advice of truth-telling, they may lie at least lie partially after receiving such advice. Therefore, we can't tell whether the following rate for truthful advice are higher than for the advice of partial or full lying. With regards to advisors' behaviour, we don't have an a priori hypothesis. According to research on advice giving, individuals make decisions differently when they give recommendations to others than when they make the decisions for themselves. When they choose for themselves they weight the attributes more uniformly than when they give advice (Kray 2000; Kray and Gonzalez 1999). Therefore, advisors may weight the moral dimension of the choice more than the monetary one. Moreover, with the lack of any monetary benefit of lying, advisors who are even slightly averse to lying would avoid suggesting lying. Consequently, fewer advisors may choose to advice lying than the senders choosing to send a deceitful message. On the other hand, since advisors

don't hold any agency in decision making, they may feel more detached and they may weight the pecuniary dimension of the decision making, hence, more advisors may lie than the senders.

## 4.4 Results

We believe that there is no way to know whether the participants didn't understand the rules of the game and, consequently, ended up with responses that are inconsistent with the theory, or whether they deliberately decided to behave as they did, even at the cost of their own earnings. Given that there is an improvement after the test, we can argue that the lack of understanding has a role in unexplained decisions. Nonetheless, the existence of such decisions even after completing the test correctly leads us to the conclusion that such role is at most partial.[6] Therefore, besides the analysis of the whole data in the following section, we decided to provide another result section in the appendix where we exclude the participants whose choices exhibit anomalies.

The main interest of the present paper is centered on the senders' decisions: lying, advice-asking, and if it is asked, the decision of whether to follow the given advice. At the same time, advisors' lying behaviour provides us a useful source to discuss the relevance of the potential motivations behind truth telling or lying in this context. Thus, we subdivide each section of the results into a section where we analyse the decisions of the senders and another section where we look at the data from the advisors.

### 4.4.1 Lying patterns

In total there are 25 senders in our baseline treatment, 32 in PRT and 30 senders in PUT. A self-interested subject should report 6 for each possible assigned number, a lying behaviour that we classify as *lying fully*. A participant who chooses to tell the truth should send a message with a number matching the one assigned to him i.e., 1 for 1, 2 for 2 etc. Though truthful, this behaviour is detrimental to his utility, for he would obtain a higher payoff when claiming 6 even when the number assigned to him is another one.

We observe these two types of behaviour, but also subjects that lie partially. In our setting a sender who doesn't lie fully nor tells the truth is considered as a partial liar. So a response such as $6, 6, 3, 4, 5, 6$ and $4, 4, 4, 4, 5, 6$ for the states $1, 2, 3, 4, 5$ and 6, respectively are both considered as partial lying. Notice that our definition of partial

[6] Our experiment is only one-shot and the subjects had enough time to think about their decisions. Moreover, given that the decision is not cognitively loaded, that, it doesn't require high level of attention or intelligently challenging, we can rule out the possibility of erroneous responses.

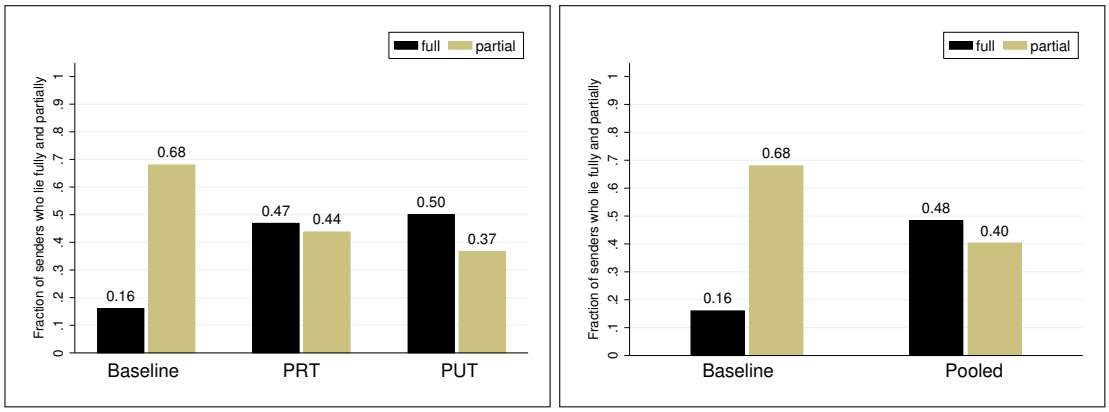
lying is broader than in Gneezy et al. (2013) or in Fischbacher and Föllmi-Heusi (2013). For example, in our setting a sender who sends a message of $6, 6, 3, 4, 5, 6$ is considered as lying partially, although she lies fully for the low states and tells the truth for the higher ones.

It should be noted that all the unexplained responses are classified as partial lying in this section. Therefore a decision profile such as $4, 5, 2, 6, 1, 3$ is also considered partial lying. Since the advisors also have to make six conditional decisions, one for each state, we classify advice in a manner entirely analogous to the messages. Therefore, an advisor who advice to send a message in the form of $6, 6, 3, 4, 5, 6$ advice to lie partially while an advice of sending a message as $6, 6, 6, 6, 6, 6$ is advising lying fully.

The first hypothesis we test is that the lying decision is affected by the existence of advice. Fig. 4.1a displays a graph showing the fraction of the liars for each treatment. The graph suggests that lying fully is lower for the baseline, compared to PRT and PUT. On its side, partial lying demonstrates a reverse relation i.e., higher in PRT and PUT compare to the baseline.

In pairwise comparisons, there are significant differences in lying fully across the baseline and the two treatments. In the baseline 16% (4 of 25) of the senders lie fully. It increases to 47% (15 of 32) in PRT ($p = 0,023$, Fisher's exact test) and to 50% (15 of 30) in PUT ($p = 0,011$, Fisher's exact test). The difference across the baseline and treatments shows a downward trend for partial lying. In the baseline, 68% (17 of 25) of the senders chooses to lie partially. It decreases to 43.8% (14 of 32) in PRT and to $36,7\%$ (11 of 30) in PUT. While a chi-square test shows that the difference across the baseline and PRT is statistically insignificant ($\chi^2 = 3.327$, $p = 0.068$), it is significant when we compare the baseline and PUT ($\chi^2 = 5.35$, $p = 0.021$). We find no difference in neither lying partially nor lying fully across PRT and PUT. This suggests that the way how advise is asked, private vs. public, has no effect on lying decision and that advice indeed is not employed to manage social image.

Given that we didn't find any statistical difference in lying decision across PRT and PUT, we pool these two treatments. The differences in lying fully and partial lying are even more significant where the test statistics for partial lying is $p = 0.019$ ($\chi^2 = 5.465$) and for lying fully we get $p = 0,007$ (Fisher's exact test). Fig 4.1b shows the difference in lying rates between the baseline and the pooled treatments.




(a) Fraction of senders who lie in the baseline and treatments

(b) Fraction of senders who lie in the baseline and pooled treatments

Figure 4.1

Table 4.3 shows the results of logistic regressions where the dependent variables are lying fully, models (1)-(3), and partial lying, models (4)-(6). In model (1) and (4), regressions for treatment effects are presented. The baseline treatment is the reference group. The results echo our findings from cross-tabulation for the treatment effects. The coefficients of PRT and PUT are positive and significant at the 5% level in model (1) which confirms that lying fully is more prevalent in the treatments compare to baseline. According to marginal effects, the probability of lying fully increases by 31% and 34% being in PRT and PUT treatments, respectively. The treatments effect partial lying in the opposite direction. It is more likely that a sender tells a partial lie in the baseline compared to the two treatments. However, we get a significant estimate for such an effect for PUT but not for PRT. The marginal effects show that being in PUT decreases the probability of lying partially by 34%.

In models (2) and (5), we add three control dummies in the logit analysis: female, work and study where female takes 1 for female participants and work is 1 if the participant works in any job besides studying. The study field is divided into two subgroups: 1 stands for any major in social sciences or humanities and 0 stands for the rest of the fields such as engineering, natural sciences etc. Although women and students who also work are more likely to tell a full lie, the effects are not significant. Additionally, it seems that students from a natural sciences field are more prone to lie fully but again the test statistic is not significant. The control dummies do not have any significance either when the dependent variable is lying partially. However, we see that working and studying together and the study field have an opposite effect on partial lying than lying fully.

In models (3) and (6), we test whether the advice type has any effect on the propen-

sity to lie. Advice full is 1 whenever sender receives an advice of lying fully and it takes zero otherwise. Similarly. Advice partial takes 1 if the advice is lying partially and 0 otherwise. An advice of full lying has a significant and positive effect on the likelihood of lying fully, while an advice of partial lying has an opposite but insignificant effect. According to marginal effects, receiving an advice of lying fully increases the probability of lying fully by 49%. We find no evidence for the effect of an advice of partial lying. However, the test statistics suggest that, although not significant, it has adverse influence on partial lying decisions. Those regression results, with the advice types, are consistent with the findings we will explain more in detail in section 4.4.2.

Table 4.3: **Logistic regression analysis: lying fully and lying partially**

| | Lying fully | | | Lying partially | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| PRT | 1.533** | 1.411** | | -1.005* | -1.209** | |
| | (0.650) | (0.675) | | (0.558) | (0.605) | |
| PUT | 1.658** | 1.542** | | -1.300** | -1.315** | |
| | (0.656) | (0.694) | | (0.572) | (0.616) | |
| Female | | 0.595 | | | 0.213 | |
| | | (0.519) | | | (0.491) | |
| Work | | 0.0600 | | | -0.815* | |
| | | (0.494) | | | (0.488) | |
| Study | | -0.892* | | | 0.859* | |
| | | (0.491) | | | (0.485) | |
| Advice full | | | 2.303** | | | -1.435 |
| | | | (0.987) | | | (0.887) |
| Advice partial | | | 1.050 | | | -0.336 |
| | | | (0.892) | | | (0.724) |
| Constant | -1.658*** | -1.562** | -1.609** | 0.754* | 0.761 | 0.336 |
| | (0.546) | (0.719) | (0.775) | (0.429) | (0.638) | (0.586) |
| $N$ | 87 | 83 | 46 | 87 | 83 | 46 |
| pseudo $R^2$ | 0.074 | 0.104 | 0.108 | 0.049 | 0.094 | 0.049 |

Dependent variable of models (1)-(3) are lying fully and of (4)-(6) is lying partially which take 1 in case of lying of that type and 0 otherwise. In models (1) and (4), explanatory variable is treatment where the baseline is the reference group. In models (2) and (5), we add control groups: female (1 for women 0 for men), work (1 if student also works 0 otherwise) and study (1 for social sciences & humanities 0 other fields). In models (3) and (6) we regress lying on advice type where advice full is 1 if the sender receives an advice of lying fully and 0 otherwise. Similarly, advice partial is 1 if the sender receives an advice of partial lying and 0 otherwise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ and numbers in parentheses are standard errors.

### 4.4.2 Asking for advice

Interestingly there is a high rate of advice-asking among the senders. In PRT 78, 12% (25 out of 32) and in PUT 70% (21 out of 30) choose to ask someone else's opinion. Although there is a slightly decrease when the advice-asking is made more public, this difference is not statistically significant ($\chi^2 = 0.534, p = 0.465$). See the table 4.4 for advice-asking rates in treatments and in total.

Table 4.4: The fraction of senders who asks advice

| | |
|---|---|
| PRT | 25/32 (78.1%) |
| PUT | 21/30 (70%) |
| Total | 46/62 (74.2%) |

To reiterate, we classify the advice exactly in the same way as the senders' lying behaviour: advice that tell the advisee to lie fully i.e. advising to send six for each possible assigned number, advice of telling the truth i.e., advising to send the true assigned number for each state and those that advice to lie partially. Again, we should note that advice that seem unexplained are considered as advising to lie partially. We consider a sender follows an advice if she takes a decision which is stated as in the advice. Thus, a sender who receives an advice that recommends her to lie fully may not follow it but still lies, if she chooses to lie partially.

Table 4.5 summarises the following and lying decision of the senders based on the advice type they receive. We see that only 13 out of 46 (28.3%) of the senders follow the advice given to them. We do not expect any treatment effect on rate of advice-following and data confirms this intuition, the rates are almost same between PRT and PUT, where 28% (7 of 25) of the senders in PRT and 28, 6% (6 of 21) in PUT follow the advice given to them ($p = 0.966$, $\chi^2 = 0.0018$). However, it is important to see whether there are any differences in lying choices across the advice type i.e., "tell the truth" vs. "lie". When we control for the advice type, whether it is an advice of lying or truth telling, the fraction of the senders that choose to follow the given advice depicts differences. The rate of advice-following across the advice of don't lie and lie fully is statistically different. It increases from 16.7% (2 of 12) to 66.7% (8 of 12) when advice changes from don't lie to lie fully ($p = 0.036$ Fisher's exact test).

We see an opposite pattern when we look at advice that suggest to lie partially. This time, the rate of advice-following decreases as an advice of lying is received. It falls from 16.7% (2 of 12) to 13.6% (3 of 22), however, this difference is not statistically different ($p = 1$ Fisher's exact test). Yet, notice that all the advice that

seem delivered haphazardly are considered in this category. This opposite pattern and the difference in rate of advice-following across partial lying and lying fully may be due to the fact that senders didn't follow an advice that seem nonsense. However, although the results are also inconclusive, in the appendix we see this opposite direction in following rate for partial lying vs. truth-telling comparison.

The difference is highly significant when we compare the following rates for the advice of partial lying and lying fully. The following rate for lying fully is 66.7% (8 of 12) and for partial lying is 13.6% (3 of 22) ($p = 0.005$, Fisher's exact test).

Table 4.5: *Rate of advice-followings and Lying decision conditional on advice type*

| | Following Decision | Lying Decision | | |
|---|---|---|---|---|
| advice Type | Follow Rate | Full | Partial | Truth-telling |
| Advice Full | 66.7% (8/12) | 66.7% (8/12) | 25% (3/12) | 8.3% (1/12) |
| Advice Partial | 13.6% (3/22) | 36.4% (8/22) | 50% (11/22) | 13.6% (3/22) |
| Advice Truth | 16.7% (2/12) | 16.7% (2/12) | 58.3 % (7/12) | 25% (3/12) |

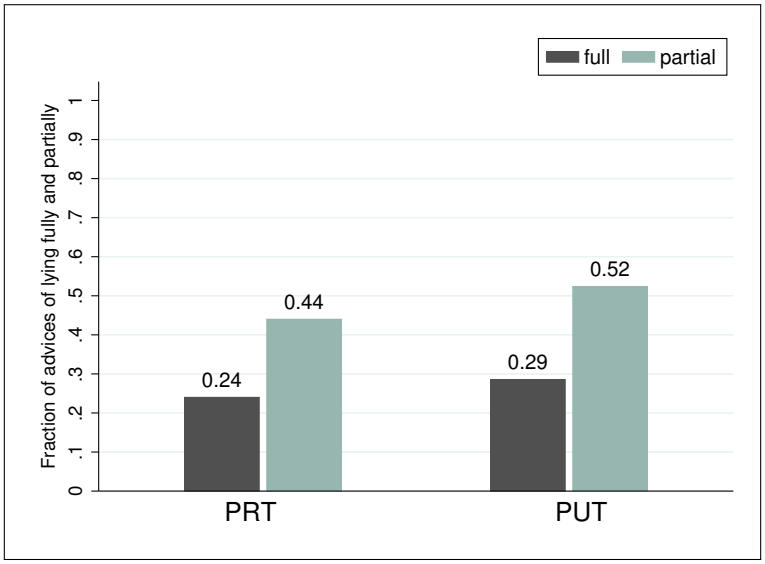
The table shows the rate of advice-following and lying decision of the senders conditional on the advice type they receive. For instance, 8 of 12 who receive an advice of lying fully follow this advice, hence, lie fully. 3 of 12 lie partially and 1 of 12 tells the truth. The difference in the last row between rate of advice-following and truth-telling when an advice of truth telling is received comes from one participant's advice type. It was stated only for one state and reporting this state truthfully. The advisee (sender) didn't lie but didn't follow this advice either.

As seen in table 4.5, we see differences in lying behaviour too. Those who were told to lie fully follow that advice and lies more. This is in line with the regression results we found in 4.4.1. Only 1 player, out of 12, chooses to send a truthful message despite being advised to lie fully. Three of those lie partially. We see that the low following rate for partial lying doesn't indicate being more truthful. Most of those who doesn't follow an advice of partial lying tell a different kind of partial lie (50%, 11/22) and a considerable amount lies fully (36.4%, 8/22). 58.3% of the participants who are advised to tell the truth tells a partial lie while only 16.7% chooses to lie fully. The differences we observe in lying behaviour conditional on the advice type are not supported statistically though except for lying fully. We find statistical evidence that those who receive an advice of truth-telling and those who receive an advice of lying fully behave differently when it comes to decide whether to tell a full lie ($p = 0.005$, Fisher's exact test).

### 4.4.3 Advice Patterns

Fig. 4.2 summarises fractions of the advisors, for each treatment, who advised to either lie fully or to lie partially. As the graph suggests, although there is an increasing trend for advice of each lie type, however, the difference is not significant. This suggests that the way how the advice is asked, whether it is more observable or not, doesn't affect advisors' decisions.

Figure 4.2: Fraction of advisors who advice to lie fully and partially in treatments




We found that advice has a significant effect on senders' lying decision and also that, there is no significant difference in lying behaviour across PRT and PUT, neither for senders nor for advisors. Therefore, we find it reasonable to pool the data from both treatments for both roles. We compare the lying decisions of advisors, pooled for PRT and PUT, with those of the senders from baseline and also those from the pooled PRT and PUT. [7] We find evidence that advisors and the senders in the pooled treatment behave differently. While only 26.09% of advisors lie fully, 48.39% of senders choose to tell a full lie ($\chi^2 = 5.5258$, $p = 0.019$). But this significance doesn't hold when we compare senders in baseline to the advisors where 16% of the senders lie fully ($p = 0.39$, Fisher's exact test).

Difference in the fraction of partial liars across advisors and senders, neither in the pooled treatments nor in the baseline, is not significant. The percentage of advice of partial lying is 47.83% and the percentage of the senders who lie partially in baseline is 68%, while it is 40.32% in the pooled data.

[7] We say that an advisor lies fully if she gives an advice of lying fully, lies partially if she gives an advice of lying partially etc.

## 4.5 Discussion and Conclusion

After each session was completed, we asked the participants to fill a short questionnaire. Apart from demographic-related questions such as age, gender, field of study and whether they work too, we asked each participant what they think about the other participants' decisions.[8]

The elicitation is not incentivised, therefore, we should note that the discussion related to the post-questionnaire in this section is inconclusive. Particularly, we cannot distinguish whether the association between the elicited beliefs and the decisions is subject to false consensus effect (Ross et al. 1977), which states that people tend to overestimate the degree to which their own actions or beliefs agree with the actions or beliefs of the others. Hence, we describe the data and register any association between the beliefs and the actions without any causal inference.

In the baseline and in two treatments we ask every participant whether they think that the majority of the participants A (senders) have sent a number greater than the assigned number i.e., lied. $87,4\%$ of senders, $75,6\%$ of recipients and $79,03\%$ of advisors believe that the the majority of the senders have lied. Among the senders who lie, either partially or fully, $90,8\%$ believe that the majority of senders did the same, while among those who sent a truthful message only $63,6,\%$ believe that the majority have lied. We find a weak positive correlation between lying and believing that the majority of the senders also sent that a false message (Spearman non-parametric Corr. Coeff=0,27, p-value=0,011). This correlation disappears when we look at the choices and the beliefs of the advisors. The proportion of advisors who think the majority of the senders lied is $85,3\%$ among those who advise to lie, partial or fully, and it is $83,3\%$ among those who doesn't advise lying (Spearman non-parametric Corr. Coeff=0.02, p-value=0.87).

In the two treatments we add another question related to lying behaviour, but this time focused on the advisors. We ask whether the participants think that the majority of the participants C (advisors) have advised to send a number greater than the assigned number i.e., lied. 51.6% of senders, 48.4% of recipients and 40.3% of advisors stated that they believe that the majority of the advisors have advised a lie, either a partial or a full one. Among the advisors who have advise to lie, fully or partially, $55,9\%$ believe that the majority of the advisors advised to lie too and only 8.3% of the the advisors who have advised not to lie believe the same. We reveal a stronger correlation between believing that the majority of the advisors advised lying and advising lying (Spearman non-parametric Corr. Coeff=0.421, p=0.004).

[8] 59% of the participants are women, 51% work beside studying, the average age is 23.7 and the 51.3% has a social science or humanities background

We also ask participants their opinions regarding advice-asking behaviour in treatments. Our analysis reveals that there is a moderate correlation between advice-asking and believing that the majority of the senders asked for advice. Spearman correlation is $0.5161 (p = 0.000)$. However, there is no evidence for any correlation between advice-asking and believing that the majority of the advisors advised lying (Spearman non-parametric Corr. Coeff=0.093, p-value=0,473).

Overall, our results show that advice indeed influence the lying decision of the individuals. Although the dishonest nature of the decision setup presented here is unambiguous and the choice problem doesn't require any help or expertise, we have observed that a substantial fraction of the participants ask for advice ( 74.2%). We proposed two motivations for advice seeking in such context and design two treatments to distinguish among them: Either advice is regarded as a psychological support that helps the individual to maintain a positive self-view or it is used to manage how one's lying behaviour is perceived by others. We have only found evidence for the first motivation. Moreover, we argue that the increasing lying fully and decreasing partial lying in the presence of advice suggest that advice replace the "function" of partial lying which is considered, in the existing lab research, a lying pattern that help individual to appreciate the pecuniary reward without giving up holding a favourable self-concept.

We have seen that the influence of advice doesn't arise from the mere existence of the advisor rather the content of the advice. We found that an advice of lying fully is followed at the highest rate and there are significant differences in lying behaviour across those who receive a recommendation of lying fully and telling the truth. Participants are more inclined to lie fully upon receiving an advice of lying fully. However, we don't observe such difference for partial lying. One possible explanation is that the advice that seem to been given randomly, which are considered as advice of partial lying, could be disregarded. Nevertheless, in the appendix where we provide a robustness check of our all results with the data in which we eliminate all such decisions, the absence of any difference persists. We argue that lying partially even after receiving an advice of truth-telling leaves some room to uphold a positive self-image. Indeed, when we look at closer to the choices of senders who receive a recommendation of reporting truthfully, we see that only a small fraction chooses to tell a full lie (16.7%) while more than the majority (58.3%) lies partially.

With respect to the advisors behaviour, although we see that more advisors (26.1%) than senders (16%) lie fully and less advisors (47.8%) than senders (68%) lie partially, the differences are not significant. For the moment we can only speculate on why uncompensated participants lie at a similar rate as the compensated ones in a moral context. Two conflicting patterns may have canceled each other here.

On one hand, some advisors, for not being the final decision maker, may feel less responsible for the decision and, hence, are concerned less about lying. This may increase the lying rate. On the other hand, in line with the advice literature (Kray 2000; Kray and Gonzalez 1999), some of the advisors may weight the social aspect of the decision and lie less, which pulls the lying rate downward.

## 4.6 Appendix

# Results with Restricted Data

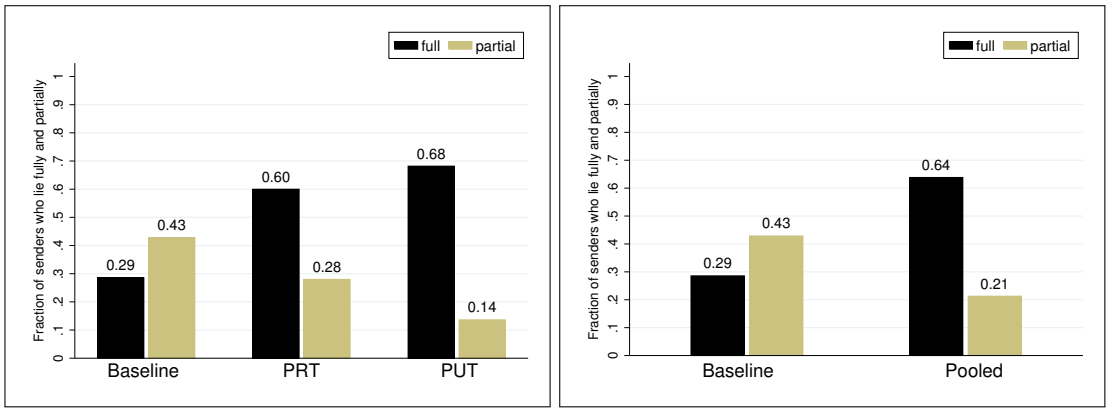
Once we exclude the senders and advisors whose responses seem unexplained – *i.e.* might be attributed to cognitive or motivational factors outside our control –, we have 14, 25 and 22 senders in the baseline, PRT and PUT, respectively. 23 advisors participated in PRT and 21 in PUT.

Fig. 4.3a shows the fractions of senders who choose to lie both fully and partially. Compared to the results we found before, the trend in lying decisions is similar. Particularly, lying fully increases from baseline to treatments and partial lying shows a reverse relation.

Comparing the two cases pairwise, we found no significant difference in lying partially between baseline and the two treatments. Partial lying decreases from 42.86%(6/14) to 28%(7/25) from baseline to PRT ($p = 0.482$, Fisher's exact test) and to 13.64%(3/22) from baseline to PUT ($p = 0.111$, Fisher's exact test). Lying fully is 28.57%(4/14) in baseline and increases to 60%(15/25) in PRT. The difference is not statistically significant ($p = 0.096$, Fisher's exact test). It increases significantly to 68.18%(15/22) in PUT ($p = 0.039$, Fisher's exact test). The difference in any types of lying across PRT and PUT continues to be insignificantly different. Once we pool PRT and PUT, the difference between the baseline and the pooled treatments becomes slightly more significant for lying fully ($p = 0.031$, Fisher's exact test), whereas the results are inconclusive for partial lying.

The advice-asking rate across PRT and PUT is not different in the restricted data either. Table 4.6 shows the fraction of the senders, for each treatment and the total, who ask for advice.

The rate of advice-following, overall, increases to 43.48% (10 0f 23) which suggests that the majority of the advice that appeared to be nonsensical, was not followed by the senders (83.3 %, 15 of 18). Yet, we don't observe any difference across PRT and PUT. The significance in the difference in rate of advice-following across an advice of lying and truth telling disappears.




(a) Fraction of senders who lie in the baseline and treatments

(b) Fraction of senders who lie in the baseline and pooled treatments

Figure 4.3

Table 4.6: The fraction of senders who asks advice

| | Fraction of senders asking advice |
|---|---|
| PRT | 21/25 (84%) |
| PUT | 14/22 (63, 64%) |
| Total | 35/47 (74.47%) |

Although the decrease is notable, from 72.73% (8 of 11) in case of receiving an advice of lying fully to 25% (2 of 8), it is not significant ($p = 0.07$, Fisher's exact test). Most remarkably, the reverse relation in the rate of advice-following when considering advice of partial lying holds in the restricted data too. Comparing an advice of lying partially to an advice of truth telling, the rate of advice-following increases from 0% (0 of 4) when the sender gets an advice of partial lying to 25% (2 of 8). The difference is not statistically significant though. When we compare following rate for advice of partial lying and for lying fully (from 0% (0 of 4) to 72.73% (8 of 11)), we see that the difference is significant ($p = 0.026$, Fisher's exact test).

Lying behaviour is affected by the advice type in a similar way as we found in the whole data set. Those who receive an advice of lying fully tell a full lie more than the senders who receive a different type of advice. Again, we observe that although the rate of advice-following is 0 for partial lying, the majority of the senders still choose to lie but in a different form than they are advised to. None of the differences in lying behaviour we observe is statistically significant. See Table Table 4.7 for the summary of these results.

Table 4.7: Rate of advice-following and Lying decision conditional on advice type-Restricted sample

| | Following Decision | Lying Decision | | |
|---|---|---|---|---|
| advice Type | Follow Rate | Full | Partial | Truth-telling |
| Advice Full | 72.7% (8/11) | 72.7% (8/11) | 18.2% (2/11) | 9.09% (1/11) |
| Advice Partial | 0% (0/4) | 25% (1/4) | 50% (2/4) | 25% (1/4) |
| Advice Truth | 25% (2/8) | 25% (2/8) | 37.5% (3/8) | 37.5% (3/8) |

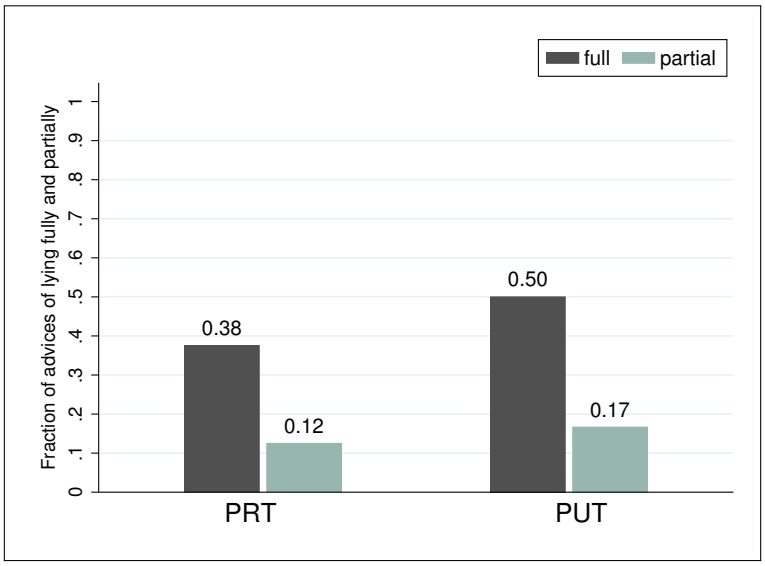
The table shows the rate of advice-following and lying decision of the senders conditional on the advice type they receive. For instance, 8 of 11 who receive an advice of lying fully follow this advice, hence, lie fully. 2 of 11 lie partially and 1 of 11 tells the truth. The difference in the last row between rate of advice-following and truth-telling when an advice of truth telling is received comes from one participant's advice form. It was stated only for one possible assigned number and that single advice was reported thruthfully. The advisee (sender) didn't lie but didn't follow this advice either.

As Fig. 4.4 depicts the fraction of advisors who advice to lie, fully and partially, increases in PUT compare to PRT. The difference is higher when we look at the advice that suggest to lie fully and lower for partial lying. Yet none of the differences, in the restricted data either, are statistically different. We confirm here, again, that there is no affect of scrutinising the advice process on advice pattern. Therefore, to replicate the analysis in the results section, we pool the senders and advisors for PRT and PUT and compare advisors to senders in the baseline and in the pooled treatments. A general observation is that, considering lying fully, senders lie more frequently than advisors when we look at the pooled data.The difference in the fraction of advisors and senders who choose to lie fully, however, is not significant, where, 63.83% of the senders and 42.86% of the advisors lie fully. For partial lying, we see an opposite relation: 21.28% of the advisors and 14.29% of the senders lie partially. This difference is not significant either.

On the other hand when we compare the senders in the baseline group to the pooled advisors, we see that advisors advise lying fully more than the senders choosing to lie fully. Compare to the advisors' lying rates stated above; the rates of lying fully and partial lying for senders in baseline are 28.6% and 42.9% respectively.

Table 4.8, presents the logit regressions of the restricted data. We have similar results to those reported in section 4.4. There is no change on the sign of the coefficients, however, we obtain fewer significant coefficients. The treatment effect for PRT for lying fully doesn't exists in none of (1)-(3) models. Similarly, we lose the significance for PUT for partial lying in (1) but we believe that those results are due to smaller sample size. We could be worried about our inferences if we saw

Figure 4.4: Fraction of advisors who advice to lie fully and partially in treatments




any change in the sign of the coefficient after excluding some of the observations. Although the coefficients are larger for PUT and smaller for PRT, we don't observe such change not even for partial lying which is expected to be more influenced by the exclusion of the data.

Table 4.8: **Logistic regression analysis: lying fully and lying partially**

| | Lying fully | | | Lying partially | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| PRT | 1.322* | 1.148 | | -0.657 | -0.880 | |
| | (0.719) | (0.748) | | (0.700) | (0.728) | |
| PUT | 1.678** | 1.655** | | -1.558* | -1.555* | |
| | (0.748) | (0.815) | | (0.823) | (0.847) | |
| Female | | 0.876 | | | 0.292 | |
| | | (0.603) | | | (0.662) | |
| Work | | -0.450 | | | -0.578 | |
| | | (0.599) | | | (0.631) | |
| Study | | -0.286 | | | 0.104 | |
| | | (0.575) | | | (0.634) | |
| Advice full | | | 2.079** | | | -0.993 |
| | | | (1.061) | | | (1.070) |
| Advice partial | | | 4.14e-17 | | | 0.511 |
| | | | (1.414) | | | (1.238) |
| Constant | -0.916 | -1.012 | -1.099 | -0.288 | -0.0291 | -0.511 |
| | (0.592) | (0.838) | (0.816) | (0.837) | (0.816) | (0.730) |
| $N$ | 61 | 58 | 23 | 61 | 58 | 23 |
| pseudo $R^2$ | 0.070 | 0.092 | 0.171 | 0.056 | 0.071 | 0.060 |

Dependent variable of models (1)-(3) are lying fully and of (4)-(6) is lying partially which take 1 in case of lying of that type and 0 otherwise. In models (1) and (4), explanatory variable is treatment where the baseline is the reference group. In models (2) and (5), we add control groups: female (1 for women 0 for men), work (1 if student also works 0 otherwise) and study (1 for social sciences & humanities 0 other fields). In models (3) and (6) we regress lying on advice type where advice full is 1 if the sender receives an advice of lying fully and 0 otherwise. Similarly, advice partial is 1 if the sender receives an advice of partial lying and 0 otherwise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ and numbers in parentheses are standard errors.

## Instructions for Participant A (Control)

### Introduction

Thank you for participating in our experiment. Please read the instructions carefully. If you follow them, you may earn extra cash beside the 5 euros you will be paid for having arrived at the experiment on time.[1] From now on any kind of communication between participants is strictly forbidden. If you have any questions or need help, please raise your hand and wait until one of the researchers assists you. NOT FOLLOWING THE INDICATED RULES WILL RESULT IN EXCLUSION FROM THE EXPERIMENT WITHOUT ANY PAYMENT.

Each participant has an experiment identity number (ID). Please note that you have to keep it until the end of the experiment otherwise you can't be paid. You will be randomly matched with two other participants. None of you will be informed of the identity of the other participants at any point in time.

You will be paid in cash and in private at the end of today's session by someone who doesn't know the content of the game. The ID numbers are distributed randomly so researchers neither track your identity nor your choices from the ID numbers. Your decisions will be anonymous.

### Roles

There will be two types of participants with two different roles in this experiment: Participant A and Participant B. The roles are randomly assigned to each participant. You are a Participant A. You are matched with one Participant B who also present in this session.

The experiment is about individual decision making and only Participants A is going to make decisions. Participant B's payoff will be determined by your choices. This is not an exam, thus, there are no correct or false answers. Don't think, therefore, that we expect a particular conduct on your side. Keep in mind that your decisions will affect the amount of the money you will earn.

The computer has randomly assigned a number between 1 and 6 to each Participant A i.e., to you before this session started.[2] You will send a message about that number to the Participant

[1] The participation fee was 3 euros in the second group of experiment.
[2] We add the following sentence here: Participant A will know this number at the end of the session

B you are matched with.[3] Participant B won't know this number. Your and Participant B's payment depend on which number you will send in this message. The details of the payment structure are explained below.

**Decisions and Payments in detail**[4]

We will ask you to send a message for each number between 1 and 6 that could have been assigned to you. In other words, you will take six conditional decisions before knowing which number has been assigned to you.[5] Below you can see the decision table on **Message sheet** where you will later write down your decisions.[6]

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — |

You will learn the number that was assigned to you by the computer at the end of the session. Your and participant B's earnings will be calculated based on the number you send in your message for that specific assigned number.[7]

You will receive 8 Euros plus the doubled amount of the number you will send in your message. For example, if you will send a message saying that the assigned number is 1 then you will receive 8 plus 2 that is 10 Euros, if you will send a message saying that the assigned number is

[3] From this point, we erase the whole paragraph and we add the following new version: "But we won't tell this number to participant B. In the following paragraphs, we will explain in detail how participant A sends a message about this number without knowing it and how the numbers in this message will affect the payments of both participants A and B."

[4] Decisions and payments were divided in two sections.

[5] A new paragraph starts at this point and we add the following explanation after the first sentence: "To put it more simply, instead of knowing the number assigned by the computer, we ask participant A to send a message for each number between 1 and 6 as if those numbers were assigned by the computer

[6] We changed the look of the message table. (see ...) and we add the following sentence here: "Participant A should consider each box on the following table as a separate message s/he is sending to participant B"

[7] This part is moved under the following paragraph, starting with "you will earn 8 euros..." Moreover, we explain how the payments are realized in more detail with the following sentences: "For example, if computer has assigned 1 to participant A, then we will check the message for 1 in the message sheet of this participant and calculate her/his earning as well as the participant B's earnings s/he is matched with based on the number sent for 1. If computer has assigned 2 then we will check the message for 2 on the message sheet and so on.

2, you will receive 8 plus 4 which is 12 Euros and so on.[8] On the other hand, participant B's payment depends on whether your message contains the true assigned number. That is, if the number you will send in your message will be the true assigned number then participant B will earn 8 Euros, if not, s/he will receive 0 Euros.

### Game Begins[9]

1. Now, you are ready to take your decisions. As we mentioned earlier, you will have to send a message for every possible assigned number.
2. Please write down your ID on "identification of participant ..." on **Message sheet.** You should decide which number you want to send in your message for every possible assigned number. Put the message sheet into the envelope. We will recollect those envelopes one by one.
3. We will inform you about the number that was assigned to you by the computer. Your and Participant B's earnings will be calculated based on the number you have sent in your message for that specific assigned number.
4. While we are determining your and Participant B's earnings, please write down your ID and fill out the questionnaire. And leave it on the desk with the written part facing down.
5. Once the experimenter tells you that you can leave, you should direct yourself to the person who will pay your earnings.

## Instructions for Participant B (Control)

### Introduction

Thank you for participating in our experiment. Please read the instructions carefully. If you follow them, you may earn extra cash beside the 5 euros you will be paid for having arrived at the experiment on time. From now on any kind of communication between participants is

[8] The paragraph is divided here and we add the following sentence: "We want to remind you that participant A must send a message for each number between 1 and 6"

[9] Before this section, we add a paragraph that explains the understanding test that we have done only in the second group: "Before we start the experiment, we would like to test your understanding of the decisions you will take and how they affect your and the others' payments. Given that participant B is passive in this experiment, only participants A will solve the test. Please read the instructions on yourself once again. Participants A can start to answer the understanding test as soon as they are ready. Remember that we will not proceed until every participant A has done the understanding test correctly.
After completing the test, please raise your hand. One of the assistants will check your answers."

strictly forbidden. If you have any questions or need help, please raise your hand and wait until one of the researchers assists you. NOT FOLLOWING THE INDICATED RULES WILL RESULT IN EXCLUSION FROM THE EXPERIMENT WITHOUT ANY PAYMENT.

Each participant has an experiment identity number (ID). Please note that you have to keep it until the end of the experiment otherwise you can't be paid. You will be randomly matched with two other participants. None of you will be informed of the identity of the other participants at any point in time.

You will be paid in cash and in private at the end of today's session by someone who doesn't know the content of the game. The ID numbers are distributed randomly so researchers neither track your identity nor your choices from the ID numbers. Your decisions will be anonymous.

**Roles**

There will be two types of participants with two different roles in this experiment: Participant A and Participant B. The roles are randomly assigned to each participant. You are a Participant B. You are matched with one Participant A who also present in this session.

The experiment is about individual decision making and only Participants A is going to make decisions. Your payoff will be determined by participant A's choices. This is not an exam, thus, there are no correct or false answers. Don't think, therefore, that we expect a particular conduct on your side. Keep in mind that your decisions will affect the amount of the money you will earn.

The computer has randomly assigned a number between 1 and 6 to each Participant A before this session started. Participant A, who you are matched with, will send a message about that number to you. You won't know this number. Your and Participant A's payment depend on which number s/he will send in this message. The details of the payment structure are explained below.

**Decisions and Payments in detail**

We will ask participant A to send a message for each number between 1 and 6 that could have been assigned to her/him. In other words, participant A will take six conditional decisions before knowing which number has been assigned to her/him. Below you can see the decision table on **Message sheet** where participant A will later write down your decisions.

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — |

Participant A will learn the number that was assigned to her/him by the computer at the end of the session. Your and participant A's earnings will be calculated based on the number s/he sends in the message for that specific assigned number.

Participant A will receive 8 Euros plus the doubled amount of the number s/he will send in the message. For example, if s/he will send a message saying that the assigned number is 1 then participant A will receive 8 plus 2 that is 10 Euros, if s/he will send a message saying that the assigned number is 2, s/he will receive 8 plus 4 which is 12 Euros and so on. On the other hand, your payment depends on whether the participant A's message contains the true assigned number. That is, if the number participant A will send in the message will be the true assigned number then you will earn 8 Euros, if not, you will receive 0 Euros

### Game Begins

1. Please wait while participants A are takin their decisions.
2. While we are determining your and Participant A's earnings, please write down your ID and fill out the questionnaire. And leave it on the desk with the written part facing down.
3. Once the experimenter tells you that you can leave, you should direct yourself to the person who will pay your earnings.

## Instructions for Participant A (PRT and PUT)

### Introduction

Thank you for participating in our experiment. Please read the instructions carefully. If you follow them, you may earn extra cash beside the 5 euros you will be paid for having arrived at the experiment on time. From now on any kind of communication between participants is

strictly forbidden. If you have any questions or need help, please raise your hand and wait until one of the researchers assists you. NOT FOLLOWING THE INDICATED RULES WILL RESULT IN EXCLUSION FROM THE EXPERIMENT WITHOUT ANY PAYMENT.

Each participant has an experiment identity number (ID). Please note that you have to keep it until the end of the experiment otherwise you can't be paid. You will be randomly matched with two other participants. None of you will be informed of the identity of the other participants at any point in time.

You will be paid in cash and in private at the end of today's session by someone who doesn't know the content of the game. The ID numbers are distributed randomly so researchers neither track your identity nor your choices from the ID numbers. Your decisions will be anonymous.

Roles

There will be three types of participants with three different roles in this experiment: Participant A, Participant B and Participant C. The roles are randomly assigned to each participant. You are a Participant A. You are matched with one Participant B and one Participant C who also present in this session.

The experiment is about individual decision making and only Participants A and C are going to make decisions. Participant B's payoff will be determined by the choices of Participants A and C. This is not an exam, thus, there are no correct or false answers. Don't think, therefore, that we expect a particular conduct on your side. Keep in mind that your decisions will affect the amount the of money you will earn.

The computer has randomly assigned a number between 1 and 6 to each Participant A i.e., to you before this session started. You will send a message about that number to the Participant B you are matched with. Participant B won't know this number. Your and Participant B's payment depend on which number you will send in this message. Participant C's earning is constant. The details of the payment structure are explained below.

We imagine that it may be difficult for you, as a Participant A, to decide which message to send in this situation. Therefore, we give you the opportunity to ask advice from the Participant C you are matched with, to learn which number s/he would send being in your position.

**Decisions and Payments in detail**

We will ask you to send a message for each number between 1 and 6 that could have been assigned to you. In other words, you will take six conditional decisions before knowing which number has been assigned to you. Similarly, Participant C, if s/he is asked, will form an advice for each possible assigned number too.[10] Below you can see the decision table on **Message sheet** where you will later write down your decisions.

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — |

In the same way, participant C will take six conditional decisions. Participant C will write her/his advice, if asked for, on the table on **Advice sheet** as appears below. [11]

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — |

You will learn the number that was assigned to you by the computer at the end of the session. Your and participant B's earnings will be calculated based on the number you send in your message for that specific assigned number.

You will receive 8 Euros plus the doubled amount of the number you will send in your message. For example, if you will send a message saying that the assigned number is 1 then you will receive 8 plus 2 that is 10 Euros, if you will send a message saying that the assigned number is

[10] We add the following message: "Without knowing the number assigned by the computer, participant C will write an advice for each number between 1 and 6.

[11] We changed the look of the message table. (see ...) and we add the following sentence here: "Participant C should consider each box on the following table as a separate message s/he is sending to participant B"

2, you will receive 8 plus 4 which is 12 Euros and so on. On the other hand, participant B's payment depends on whether your message contains the true assigned number. That is, if the number you will send in your message will be the true assigned number then participant B will earn 8 Euros, if not, s/he will receive 0 Euros. Participant C's payment is constant and doesn't depend on the decisions taken here, s/he will earn 8 Euros.

**Game Begins**

1. Now you are ready for making your decision. As we mentioned before, you will send a message for every possible assigned number.
2. But first you should decide whether you want to ask advice from the Participant C you are matched with. If you will ask for advice, participant C will tell you which number s/he would send for each possible assigned number as you will do for the message you send.
3. Now please write down your ID number on the **Advice sheet** where you see "participant A's ID ..." You will see the box next to the sentence "I want advice". Please tick the box (☑) if you want advice, otherwise leave it empty. Put this sheet in the envelope and we will collect them one by one and pass them to the participants C.
4. If you have asked for advice, participant C will write his/her advice on "Send the message that the assigned number is...." for every possible assigned number. Otherwise s/he will leave it blank. We will give this paper back to you in envelope.
5. Now open the envelope even if you haven´t asked advice and leave it on your desk with the written part facing down.
6. It is time to decide which message to send. Please write down your ID on the **Message sheet**. You should write down the number you want to send in your message for every possible assigned number on this sheet. Put it in the envelope. We will collect them one by one.

*(In PUT treatment items numbers 3, 4,5 and 6 we have the following paragraph divided into three item lists): (3)Now please write down your ID number on the* ***Advice sheet*** *where you see "participant A's ID...". If you want advice, please put the advice sheet into the envelope and raise your hand. We will pass it to the participant C you are matched with. Participant C will write his/her advice on "Send the message that the*

*assigned number is..." for every possible assigned number. We will give this paper back to you in the envelope. You can open and see the advice you are given. (4) If you don't want advice, you can decide which number you want to send in your message for every possible assigned number and write them down on message sheet. (5) Participant A who has asked for advice, after receiving it, will do the same as those participants A who haven't asked for advice. That's, you should write down the number you want to send for every possible assigned number.)*

7. We will inform you about the number that was assigned to you by the computer. Your and Participant B's earnings will be calculated based on the number you have sent in your message for that specific assigned number.
8. While we are determining your and Participant B's earnings, please write down your ID and fill out the questionnaire. And leave it on the desk with the written part facing down.
9. Once the experimenter tells you that you can leave, you should direct yourself to the person who will pay your earnings.

## Instructions for Participant B (PRT and PUT)

### Introduction

Thank you for participating in our experiment. Please read the instructions carefully. If you follow them, you may earn extra cash beside the 5 euros you will be paid for having arrived at the experiment on time. From now on any kind of communication between participants is strictly forbidden. If you have any questions or need help, please raise your hand and wait until one of the researchers assists you. NOT FOLLOWING THE INDICATED RULES WILL RESULT IN EXCLUSION FROM THE EXPERIMENT WITHOUT ANY PAYMENT.

Each participant has an experiment identity number (ID). Please note that you have to keep it until the end of the experiment otherwise you can't be paid. You will be randomly matched with two other participants. None of you will be informed of the identity of the other participants at any point in time.

You will be paid in cash and in private at the end of today's session by someone who doesn't know the content of the game. The ID numbers are distributed randomly so researchers neither track your identity nor your choices from the ID numbers. Your decisions will be anonymous.

**Roles**

There will be three types of participants with three different roles in this experiment: Participant A, participant B and participant C. The roles are randomly assigned to each participant. You are a participant B. You are matched with one participant A and one participant C who also present in this session.

The experiment is about individual decision making and only Participants A and C are going to make decisions. Your payoff will be determined by the choices of Participants A and C. This is not an exam, thus, there are no correct or false answers. Don't think, therefore, that we expect a particular conduct on your side.

The computer has randomly assigned a number between 1 and 6 to each participant A before this session started. Participant A will send a message about that number to the participant B s/he is matched with i.e. to you. You won't know this number. Your and participant A's payments depend on which number participant A will send in this message. Participant C's earning is constant. The details of the payment structure are explained below.

We imagine that it may be difficult for a participant A to decide which message to send in this situation. Therefore, we give the participant A the opportunity to ask advice from the participant C s/he is matched with to learn which number the participant C would send being in participant A's position.

**Decisions and Payments in detail**

We will ask participant A to send a message for each number between 1 and 6 that could have been assigned to her/him. In other words, participant A will take six conditional decisions before knowing which number has been assigned to her/him. Similarly, participant C, if s/he is asked for, will form an advice for each possible assigned number too. Below you can see the decision table on **Message sheet** where participant A will later write down her/his decisions.

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| The message I send to participant B is: "The assigned number is" ___ | The message I send to participant B is: "The assigned number is" ___ | The message I send to participant B is: "The assigned number is" ___ | The message I send to participant B is: "The assigned number is" ___ | The message I send to participant B is: "The assigned number is" ___ | The message I send to participant B is: "The assigned number is" ___ |

In the same way, participant C will take six conditional decisions. Participant C will write her/his advice, if asked for, on the table on **Advice sheet** as appears below.

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Send the message that the assigned number was ___ | Send the message that the assigned number was ___ | Send the message that the assigned number was ___ | Send the message that the assigned number was ___ | Send the message that the assigned number was ___ | Send the message that the assigned number was ___ |

Participant A will learn the number that was assigned to her/him by the computer at the end of the session. Your and participant A's earnings will be calculated based on the number participant A sends to you in the message for that specific assigned number.

Participant A will receive 8 Euros plus the doubled amount of the number s/he will send in the message. For example, if s/he will send a message saying that the assigned number is 1 then participant A will receive 8 plus 2 that is 10 Euros, if s/he will send a message saying that the assigned number is 2, s/he will receive 8 plus 4 which is 12 Euros and so on. On the other hand, your payment depends on whether the participant A's message contains the true assigned number. That is, if the number participant A will send in the message will be the true assigned number then you will earn 8 Euros, if not, you will receive 0 Euros. Participant C's payment is constant and doesn't depend on the decisions taken here, s/he will earn 8 Euros.

**Game Begins**

1. The participant A that is matched with you will send you a message for every possible assigned number. But first s/he should decide whether s/he wants to ask advice from participant C. If participant A will ask for advice, participant C will tell her/him which

number s/he would send for each possible assigned number as participant A will do for the message s/he sends.

2. There is a box next to the sentence "I want advice" on **advice sheet**. If participant A asks for advice he will tick the box (☑), otherwise s/he will leave it empty. Participant A will put the advice sheet into the envelope and we will pass those envelopes to participants C one by one.
3. If participant A has asked for advice, participant C will write his/her advice on "Send the message that the assigned number is...." for every possible assigned number. Otherwise participant C will leave it blank. We will give this paper back to participant A in envelope.
4. Participant A will open the envelope even if s/he hasn't asked advice and leaves it on the desk with the written part facing down.
5. Participant A will open the envelope and read its content, even if she has not asked for advice. S/he will leave the advice sheet on the desk with the written part facing down. Participant A, then, will decide which number s/he wants to send for each possible assigned number and will write them down on **Message sheet.**

   *((In PUT treatment, instead of item numbers 2 and 3 we have the following paragraph divided into three item lists): (2) If participant A wants advice, s/he will put the advice sheet into the envelope and raise her/his hand. We will pass it to participant C. Participant C will write her/his advice on "Send the message that the assigned number is..." for every possible assigned number. We will give this paper back to participant A in the envelope. (3) If participant A doesn't want advice, s/he can decide which number s/he wants to send in the message for every possible assigned number and write them down on message sheet. (5) Participant A who has asked for advice, after receiving it, will do the same as those participants A who haven't asked for advice. That's, participant A should write down the number s/he wants to send for every possible assigned number.)*
6. We will inform participant A about the number that was assigned to her/him by the computer. Your and Participant A's earnings will be calculated based on the number participant A has sent in the message for that specific assigned number.

7. While we are determining your and Participant A's earnings, please write down your ID and fill out the questionnaire. And leave it on the desk with the written part facing down.
8. Once the experimenter tells you that you can leave, you should direct yourself to the person who will pay your earnings.

## Instructions for Participant C (PRT and PUT)

### Introduction

Thank you for participating in our experiment. Please read the instructions carefully. If you follow them, you may earn extra cash beside the 5 euros you will be paid for having arrived at the experiment on time. From now on any kind of communication between participants is strictly forbidden. If you have any questions or need help, please raise your hand and wait until one of the researchers assists you. NOT FOLLOWING THE INDICATED RULES WILL RESULT IN EXCLUSION FROM THE EXPERIMENT WITHOUT ANY PAYMENT.

Each participant has an experiment identity number (ID). Please note that you have to keep it until the end of the experiment otherwise you can't be paid. You will be randomly matched with two other participants. None of you will be informed of the identity of the other participants at any point in time.

You will be paid in cash and in private at the end of today's session by someone who doesn't know the content of the game. The ID numbers are distributed randomly so researchers neither track your identity nor your choices from the ID numbers. Your decisions will be anonymous.

### Roles

There will be three types of participants with three different roles in this experiment: Participant A, participant B and participant C. The roles are randomly assigned to each participant. You are a participant C. You are matched with one participant A and one participant B who also present in this session.

The experiment is about individual decision making and only Participants A and C are going to make decisions. Participant B's payoff will be determined by the choices of Participants A and C. This is not an exam, thus, there are no correct or false answers. Don't think, therefore, that we expect a particular conduct on your side.

The computer has randomly assigned a number between 1 and 6 to each participant A before this session started. Participant A will send a message about that number to the participant B s/he is matched with. Participant B won't know this number. Participant A and participant B's payment depend on which number participant A will send in this message. Your earning is constant. The details of the payment structure are explained below.

We imagine that it may be difficult for a Participant A, to decide which message to send in this situation. Therefore, we give the participant A the opportunity to ask advice from the participant C s/he is matched with, i.e. from you to learn which number you would send being in participant A's position.

**Decisions and Payments in detail**

We will ask participant A to send a message for each number between 1 and 6 that could have been assigned to her/him. In other words, participant A will take six conditional decisions before knowing which number has been assigned to you. Similarly, you, if you are asked for, will form an advice for each possible assigned number too. Below you can see the decision table on **Message sheet** where participant A will later write down her/his decisions.

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — | The message I send to participant B is: "The assigned number is" — |

In the same way, you will take six conditional decisions. You will write your advice, if asked for, on the table on **Advice sheet** as appears below.

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — | Send the message that the assigned number was — |

Participant A will learn the number that was assigned to her/him by the computer at the end of the session. Participant A's and participant B's earnings will be calculated based on the number participant A sends in the message for that specific assigned number.

Participant A will receive 8 Euros plus the doubled amount of the number s/he will send in the message. For example, if s/he will send a message saying that the assigned number is 1 then participant A will receive 8 plus 2 that is 10 Euros, if s/he will send a message saying that the assigned number is 2, s/he will receive 8 plus 4 which is 12 Euros and so on. On the other hand, participant B's payment depends on whether the participant A's message contains the true assigned number. That is, if the number s/he will send in the message will be the true assigned number then participant B will earn 8 Euros, if not, participant B will receive 0 Euros. Your payment is constant and doesn't depend on the decisions taken here, you will earn 8 Euros.

**Game Begins**

1. The participant A that is matched with you will send a message for every possible assigned number. But first s/he should decide whether s/he wants to ask advice from you. If you participant a will ask for advice, you will tell her/him which number you would send for each possible assigned number as participant A will do for the message s/he sends.
2. Now please write down your ID number on the **Advice sheet** where you see "participant C's ID ..." If the box next to the sentence "I want advice" is ticked (☑), please write down for every possible assigned number, which number you would send in the message if you were in participant A's position. Otherwise, don't write anything on that paper, leave it empty. Participant A will put the advice sheet into the envelope and we will pass those envelopes to participants C i.e., to you, one by one.
3. Participant A will open the envelope and read its content, even if she has not asked for advice. S/he will leave the advice sheet on the desk with the written part facing down. Participant A, then, will decide which number s/he wants to send for each possible assigned number and will write them down on **Message sheet.**

*(In PUT treatment, instead of item numbers 2 and 3 we have the following paragraph divided into three item lists): (2) Now please write down your ID number on the* ***Advice***

## Message Sheet

**Participant A's identification: .....**

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| The message I send to participant B is: "The assigned number is" __ | The message I send to participant B is: "The assigned number is" __ | The message I send to participant B is: "The assigned number is" __ | The message I send to participant B is: "The assigned number is" __ | The message I send to participant B is: "The assigned number is" __ | The message I send to participant B is: "The assigned number is" __ |

---

## Advice Sheet (PRT)

**Participant A's identification: .....**

□

**Participant C's identification: .....**

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ |

# Advice Sheet (PUT)

**Participant A's identification: .....**

##########

**Participant C's identification: .....**

| If the assigned number is | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ | Send the message that the assigned number was __ |

---

**Questionnaire:**

**Identification of the participant: .....**

1. Age: .....
2. Sex: .....
3. What do you study? .....
4. Do you also work? .....
5. Do you think that the majority of the participants A asked advice?

   ☐ YES ☐ NO

6. How do you think the majority of the participants A made their decisions? Did the majority send a message with a number greater than the assigned number?

   ☐ YES ☐ NO

7. How do you think the majority of the participants C made their decisions? Did the majority send an advice with a number greater than the assigned number? [1]

   ☐ YES ☐ NO

[1] For the baseline, we had the same questionnaire except questions 5and 7, which are eliciting beliefs regarding advice asking behavior

# Understanding Test

Please notice that the numbers we use both in the instructions and in the understanding test are for illustrative purposes. Remember that a virtual die has been used before the session has started and it is equally likely to roll 1,2,3,4,5 and 6. Now please fill in the blank in the following question.

***Question:***

If the computer assigns 3, then the message sent for _____ in Message Sheet will be checked. If you send

- 1 in that message, you will receive _____euros and Participant B will receive _____euros.
- 2 in that message, you will receive _____euros and Participant B will receive _____euros.
- 3 in that message, you will receive _____euros and Participant B will receive _____euros.
- 4 in that message, you will receive _____euros and Participant B will receive _____euros.
- 5 in that message, you will receive _____euros and Participant B will receive _____euros.
- 6 in that message, you will receive _____euros and Participant B will receive _____euros.

# Bibliography


Abeler, Johannes, Daniele Nosenzo, and Collin Raymond (2016). "Preferences for truth-telling". In:

Akerlof, George A (1980). "A theory of social custom, of which unemployment may be one consequence". In: *The quarterly journal of economics* 94.4, pp. 749–775.

Akerlof, George A and Rachel E Kranton (2000). "Economics and identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753.

– (2010). *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press.

Andreoni, James (1989). "Giving with impure altruism: applications to charity and Ricardian equivalence". In: *The Journal of Political Economy*, pp. 1447–1458.

– (1990). "Impure altruism and donations to public goods: a theory of warm-glow giving". In: *The economic journal*, pp. 464–477.

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". In: *Econometrica* 77.5, pp. 1607–1636.

Andreoni, James and Ragan Petrie (2004). "Public goods experiments without confidentiality: a glimpse into fund-raising". In: *Journal of public Economics* 88.7, pp. 1605–1623.

Ariely, Dan, Anat Bracha, and Stephan Meier (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially". In: *The American Economic Review*, pp. 544–555.

Ayal, Shahar and Francesca Gino (2011). "Honest rationales for dishonest behavior". In: *The social psychology of morality: Exploring the causes of good and evil. Washington, DC: American Psychological Association*, pp. 149–66.

Bagwell, Laurie Simon and B Douglas Bernheim (1996). "Veblen effects in a theory of conspicuous consumption". In: *The American Economic Review*, pp. 349–373.

Baigent, Nick and Wulf Gaertner (1996). "Never choose the uniquely largest A Characterization". In: *Economic Theory* 8.2, pp. 239–249.

Balafoutas, Loukas and Nikos Nikiforakis (2012). "Norm enforcement in the city: a natural field experiment". In: *European Economic Review* 56.8, pp. 1773–1785.

Barkan, Rachel, Shahar Ayal, and Dan Ariely (2015). "Ethical dissonance, justifications, and moral behavior". In: *Current Opinion in Psychology* 6, pp. 157–161.

Becker, Gary (1957). *The Economics of Discrimination*. University of Chicago Press.

Benjamin, Daniel J, James J Choi, and A Joshua Strickland (2010). "Social Identity and Preferences". In: *American Economic Review* 100.4, pp. 1913–1928.

Bernheim, B Douglas (1994). "A theory of conformity". In: *Journal of political Economy*, pp. 841–877.

Bertrand, Marianne and Adair Morse (2016). "Trickle-down consumption". In: *Review of Economics and Statistics* 98.5, pp. 863–879.

Bohnet, Iris and Bruno S Frey (1999). "Social distance and other-regarding behavior in dictator games: Comment". In: *American Economic Review*, pp. 335–339.

Bonaccio, Silvia and Reeshad S Dalal (2006). "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences". In: *Organizational Behavior and Human Decision Processes* 101.2, pp. 127–151.

Brandts, Jordi and Gary Charness (2011). "The strategy versus the direct-response method: a first survey of experimental comparisons". In: *Experimental Economics* 14.3, pp. 375–398.

Brooks, Alison Wood, Francesca Gino, and Maurice E Schweitzer (2015). "Smart people ask for (my) advice: Seeking advice boosts perceptions of competence". In: *Management Science* 61.6, pp. 1421–1435.

Bryan, Christopher J, Gabrielle S Adams, and Benoît Monin (2013). "When cheating would make you a cheater: Implicating the self prevents unethical behavior." In: *Journal of Experimental Psychology: General* 142.4, p. 1001.

Bursztyn, Leonardo and Robert Jensen (2015). "How Does Peer Pressure Affect Educational Investments?" In: *The quarterly journal of economics* 130.3, pp. 1329–1367.

– (2016). *Social Image and Economic Behavior in the Field: Identifying, Understanding and Shaping Social Pressure*. Tech. rep. National Bureau of Economic Research.

– (2017). "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure". In: *Annual Review of Economics* 9.1.

Cappelen, Alexander W, Erik Ø Sørensen, and Bertil Tungodden (2013). "When do we lie?" In: *Journal of Economic Behavior & Organization* 93, pp. 258–265.

Charness, Gary and Martin Dufwenberg (2006). "Promises and partnership". In: *Econometrica* 74.6, pp. 1579–1601.

– (2010). "Bare promises: An experiment". In: *Economics Letters* 107.2, pp. 281–283.

Chen, Yan and Sherry Xin Li (2009). "Group identity and social preferences". In: *The American Economic Review*, pp. 431–457.

Chou, Eileen et al. (2009). "The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person ?guessing? game". In: *Experimental Economics* 12.2, pp. 159–179.

Cohn, Alain, Ernst Fehr, and Michel André Maréchal (2014). "Business culture and dishonesty in the banking industry". In: *Nature* 516.7529, pp. 86–89.

Conrads, Julian et al. (2013). "Lying and team incentives". In: *Journal of Economic Psychology* 34, pp. 1–7.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness". In: *Economic Theory* 33.1, pp. 67–80.

De Wit, Frank RC et al. (2017). "Whether power holders construe their power as responsibility or opportunity influences their tendency to take advice from others". In: *Journal of Organizational Behavior*.

Drugov, Mikhail, John Hamman, and Danila Serra (2014). "Intermediaries in corruption: an experiment". In: *Experimental Economics* 17.1, pp. 78–99.

Dufwenberg, Martin (2016). "Lies in Disguise-A Theoretical Analysis of Cheating". In:

Elster, Jon (1989). "Social norms and economic theory". In: *The Journal of Economic Perspectives*, pp. 99–117.

– (2000). "Social norms and economic theory". In: *Culture and Politics*. Springer, pp. 363–380.

Erat, Sanjiv (2013). "Avoiding lying: the case of delegated deception". In: *Journal of Economic Behavior & Organization* 93, pp. 273–278.

Erat, Sanjiv and Uri Gneezy (2012). "White lies". In: *Management Science* 58.4, pp. 723–733.

Fehr, E. and Simonachter Gächter (2000). "Cooperation and punishment in public goods experiments". In: *The American Economic Review* 90.4, pp. 980–994.

Fehr, Ernst and Klaus M Schmidt (1999). "A theory of fairness, competition, and cooperation". In: *Quarterly journal of Economics*, pp. 817–868.

Filiz-Ozbay, Emel and Erkut Y Ozbay (2013). "Effect of an audience in public goods provision". In: *Experimental Economics*, pp. 1–15.

– (2014). "Effect of an audience in public goods provision". In: *Experimental Economics* 17.2, pp. 200–214.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013). "Lies in disguise?an experimental study on cheating". In: *Journal of the European Economic Association* 11.3, pp. 525–547.

Funk, Patricia (2010). "Social incentives and voter turnout: evidence from the Swiss mail ballot system". In: *Journal of the European Economic Association* 8.5, pp. 1077–1103.

Gaertner, Wulf and Yongsheng Xu (1999). "On the structure of choice under different external references". In: *Economic Theory* 14.3, pp. 609–620.

Gibson, Rajna, Carmen Tanner, and Alexander F Wagner (2013). "Preferences for truthfulness: Heterogeneity among and within individuals". In: *The American Economic Review* 103.1, pp. 532–548.

Gino, Francesca, Alison Wood Brooks, and Maurice E Schweitzer (2012). "Anxiety, advice, and the ability to discern: feeling anxious motivates individuals to seek and use advice." In: *Journal of personality and social psychology* 102.3, p. 497.

Glazer, Amihai and Kai A Konrad (1996). "A signaling explanation for charity". In: *The American Economic Review*, pp. 1019–1028.

Gneezy, Uri (2005). "Deception: The role of consequences". In: *The American Economic Review* 95.1, pp. 384–394.

Gneezy, Uri, Bettina Rockenbach, and Marta Serra-Garcia (2013). "Measuring lying aversion". In: *Journal of Economic Behavior & Organization* 93, pp. 293–300.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel (2016). "Lying Aversion and the Size of the Lie". In:

Greiner, Ben (2004). "An online recruitment system for economic experiments". In:

Grether, David M and Charles R Plott (1979). "Economic theory of choice and the preference reversal phenomenon". In: *The American Economic Review* 69.4, pp. 623–638.

Grossman, Zachary (2015). "Self-signaling and social-signaling in giving". In: *Journal of Economic Behavior & Organization* 117, pp. 26–39.

Hao, Li and Daniel Houser (2017). "Perceptions, intentions, and cheating". In: *Journal of Economic Behavior & Organization* 133, pp. 52–73.

Harvey, Nigel and Ilan Fischer (1997). "Taking advice: Accepting help, improving judgment, and sharing responsibility". In: *Organizational Behavior and Human Decision Processes* 70.2, pp. 117–133.

Hotelling, Harold (1929). "Stability in Competition". In: *The Economic Journal* 39, pp. 41–57.

Kajackaite, Agne and Uri Gneezy (2017). "Incentives and cheating". In: *Games and Economic Behavior* 102, pp. 433–444.

Kartik, Navin, Marco Ottaviani, and Francesco Squintani (2007). "Credulity, lies, and costly talk". In: *Journal of Economic theory* 134.1, pp. 93–116.

Khalmetski, Kiryl and Dirk Sliwka (2017). "Disguising Lies-Image Concerns and Partial Lying in Cheating Games". In:

Kray, Laura and Richard Gonzalez (1999). "Differential weighting in choice versus advice: I'll do this, you do that". In: *Journal of Behavioral Decision Making* 12.3, p. 207.

Kray, Laura J (2000). "Contingent weighting in self-other decision making". In: *Organizational behavior and human decision processes* 83.1, pp. 82–106.

Kreps, David M (1988). *Notes on the Theory of Choice*. Vol. 2. Westview press Boulder.

Krupka, Erin L and Roberto A Weber (2013). "Identifying social norms using coordination games: Why does dictator game sharing vary?" In: *Journal of the European Economic Association* 11.3, pp. 495–524.

López-Pérez, Raúl and Eli Spiegelman (2013). "Why do people tell the truth? Experimental evidence for pure lie aversion". In: *Experimental Economics* 16.3, pp. 233–247.

Lundquist, Tobias et al. (2009). "The aversion to lying". In: *Journal of Economic Behavior & Organization* 70.1, pp. 81–92.

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green, et al. (1995). *Microeconomic theory*. Vol. 1. Oxford university press New York.

Mazar, Nina, On Amir, and Dan Ariely (2008). "The dishonesty of honest people: A theory of self-concept maintenance". In: *Journal of marketing research* 45.6, pp. 633–644.

Posner, Richard A and Eric B Rasmusen (1999). "Creating and enforcing norms, with special reference to sanctions". In: *International Review of Law and Economics* 19.3, pp. 369–382.

Rege, Mari and Kjetil Telle (2004). "The impact of social approval and framing on cooperation in public good situations". In: *Journal of public Economics* 88.7, pp. 1625–1644.

Ross, Lee, David Greene, and Pamela House (1977). "The "false consensus effect": An egocentric bias in social perception and attribution processes". In: *Journal of experimental social psychology* 13.3, pp. 279–301.

Rubinstein, Ariel (2012). *Lecture notes in microeconomic theory: the economic agent*. Princeton University Press.

Samek, Anya Savikhin and Roman M Sheremeta (2014). "Recognizing contributors: an experiment on public goods". In: *Experimental Economics* 17.4, pp. 673–690.

Sánchez-Pagés, Santiago and Marc Vorsatz (2009). "Enjoy the silence: an experiment on truth-telling". In: *Experimental Economics* 12.2, pp. 220–241.

Schotter, Andrew (2003). "Decision making with naive advice". In: *The American Economic Review* 93.2, pp. 196–201.

Schram, Arthur and Gary Charness (2015). "Inducing social norms in laboratory allocation choices". In: *Management Science* 61.7, pp. 1531–1546.

Shalvi, Shaul, Michel JJ Handgraaf, and Carsten KW De Dreu (2011a). "Ethical manoeuvring: Why people avoid both major and minor lies". In: *British Journal of Management* 22.s1.

Shalvi, Shaul et al. (2011b). "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior". In: *Organizational Behavior and Human Decision Processes* 115.2, pp. 181–190.

Shalvi, Shaul et al. (2015). "Self-Serving Justifications: Doing Wrong and Feeling Moral". In: *Science* 24.2, pp. 125–130.

Soetevent, Adriaan R (2005). "Anonymity in giving in a natural context?a field experiment in 30 churches". In: *Journal of public Economics* 89.11, pp. 2301–2323.

Sutter, Matthias (2009). "Deception through telling the truth?! Experimental evidence from individuals and teams". In: *The Economic Journal* 119.534, pp. 47–60.

Tadelis, Steven (2011). "The power of shame and the rationality of trust". In:

Tutić, Andreas (2015). "Revealed norm obedience". In: *Social Choice and Welfare* 44.2, pp. 301–318.

Ven, Jeroen Van de and Marie Claire Villeval (2015). "Dishonesty under scrutiny". In: *Journal of the Economic Science Association* 1.1, pp. 86–99.

Vreeland, James Raymond (1999). "The IMF: lender of last resort or scapegoat". In: *Midwest Political Science Association Meeting, Chicago, IL*, pp. 15–17.