Thesis submitted to the University of East Anglia
for the Degree of Doctor of Philosophy

Experimental Essays on Singling Out, Emotions, and Electoral Decisions

Fabio Galeotti

27th of March 2014

Word count: 68,039

Acknowledgements

I would first like to thank my first supervisor Prof. Daniel Zizzo, for his precious advice, and support during my PhD. I am also thankful to my second supervisor Dr. Abhijit Ramalingman for his valuable comments, suggestions, and encouragement. I am also grateful to Prof. Elke Renner, and Prof. Enrique Fatas, who have kindly accepted to be discussants of this thesis. A special thank goes to the other PhD students (Abdul, Alex, Fred, Gianni, Jack, Jiwei, Saeed, and Zoe), who shared the office with me during all these years, and who help to make this journey a more enjoyable experience. The financial support from the School of Economics and the CBESS is gratefully acknowledged. Last, but not least, I thank Marie-Cécile, and my family in Italy for unconditional support and encouragement during all these years. This thesis is dedicated to them and to my little baby Francesca.

Fabio Galeotti

University of East Anglia

March 2014

Abstract

This PhD thesis is a collection of three independent essays in the area of experimental economics. The first investigates the effects of singling out an individual on trust and trustworthiness. The second essay studies the role played by the punishment technology, and the experience and cultural background of the subjects in driving emotions and behavior in power-to-take game experiments. The third essay investigates the preferences of voters over the trustworthiness and competence of candidates in public elections.

Contents

Introduction

This doctoral thesis consists of three independent essays which utilize experimental methods to investigate different topics relevant to economics. There is a general link between the first and the third chapters as they both use the trust game as a vehicle of research to examine the preferences of people for the individual characteristics of economic agents. However, besides this connection, each chapter focuses on a different theme and deals with different research questions. Hence, each chapter can be read as a standalone piece of work.

The first essay investigates what are the effects of singling out an individual on trust and trust fulfilling. Singling out occurs whenever a subject, who has specific attributes that make him or her potentially different from the others, ceases to be an ordinary and usual person, and becomes a distinct one in the eyes of other subjects. More precisely, singling out can be defined as an inter-group situation in which one group is a singleton group made up of a single individual - the singled out individual -, whose social identity is perceived as different by a second group that is larger in size. The starting point of our investigation is the fact that singling someone out is a pervasive phenomenon of social and economic life, both at a micro and macro level. For example, people are singled out in organizational and workplace settings, as well as countries and firms are selected and treated differently in political and economic alliances. Despite the fact that singling a member out is a common phenomenon of social and economic groups, it has received no attention so far in economics. In the first essay, we try to fill this gap focusing on the implications of singling out an individual, at the micro level, in the context of trust games. This investigation is particularly important for organizations. For example, an employee may be object of social recognition (or reproof) for his or her desirable (or undesirable) socio-economic characteristics or because he or she belongs to a privileged (or disadvantaged) group. As result, he or she may lose the status of ordinary member, and acquires a positive (or negative) aura of uniqueness in the eyes of his or her colleagues or superiors. Similarly, a worker who is promoted or sanctioned may start to be perceived and treated differently by his or her colleagues. All these situations may entail positive or negative consequences in terms of group cohesion and trust, which, in turn, may affect the productivity of an organization.

To conduct this investigation, we run a lab experiment. We artificially induce a status of being singled out in the lab by asking the subjects to express their preferences for the other participants based on the individual characteristics of the subjects. We assign either a positive or a negative frame to the condition of being singled out. Under a positive frame, subjects singled out the *most preferred match* in the experiment, whereas under a negative frame, they singled out the *least preferred match* in the experiment. We also manipulate the subjects' responsibility in the selection of the singled out individual. Identification effects are controlled by varying the extent to which the status of being singled out is identified by non-singled out subjects. Furthermore, we test the implications of having a random assignation of such status under a neutral frame. Finally, we conducted some ex-post analyses where we ruled out the possibility that singled out subjects behaved differently because of the individual characteristics that made them singled out. As such, this study provides evidence of the 'pure' effect of singling out an individual from a group, both on the singled out side and on the non-singled out side. The main findings are that the effect of singling out on trust is negligible but its effect on trustworthiness is negative when it is significant. In particular, singled out subjects, under a negative frame, return less than non-singled out subjects, while those under a positive frame behave bimodally either selecting very low or very high return rates. In addition, we find that non-singled out subjects negatively discriminate the singled out subject but only when they are not responsible for his distinct status. Finally, trustworthiness decreases if there is a singled out subject.

In the second essay of this thesis, we investigate an important methodological issue that characterizes the study of emotions and behavior in power-to-take game (PTTG) experiments, and whose implications have not been so far consider in the literature. More specifically, we study the extent to which the punishing behavior observed in previous PTTG experiments is explained by a non-constant "fine-to-fee ratio"[1] instead of negative emotions, and, in particular, anger, irritation, and contempt. In the PTTG, this parameter is in fact increasing with the offence. This means that subjects may punish simply because it is cheaper to punish and not because they experience negative feelings. Hence, previous PTTG experiments might have overstated the role played by negative emotions on the punishment decision (Bosman and van Winden, 2002; Bosman *et al*., 2005; Ben Shakhar *et al*., 2007). It is thus important to investigate whether the findings of previous studies on the PTTG are driven by this potential confound and to what extent the punishment behavior observed can be truly attributed to emotions.

In addition to this, we also investigate the impact of the cultural background and experience of the subjects on emotions and behavior in the PTTG. The psychological and anthropological literature has in fact shown that there are cultural differences in the elicitation and manifestation of emotions (see, e.g., Mesquita and Frijda, 1992). In addition, it is reasonable to expect that subjects with more experience of the environment and the dynamics of laboratory experiments are more aware of what they should expect in an economic experiment and, therefore, they might experience less strong emotions and/or be better able to cope with their emotional urges than inexperienced subjects. The previous literature on the PTTG has not examined these important issues.

To carry out this investigation, we design an experiment, building on the 2002 EJ seminal paper by Bosman and van Winden and the following literature on the PTTG, where we vary the extent to which the punishment technology is characterized by a variable or constant "fine-to-fee" ratio. In addition, we run separate sessions for UK students and non-UK students, and control, in a systematic way, for the experience of the subjects in our statistical analysis. The main findings of the second chapter indicate that a large part of the punishment behavior observed in previous PTTG studies is explained by the technology of punishment adopted, and that the role played by emotions is overstated. When the potential confound is removed from the

[1] The 'fine-to-fee' ratio is defined as "the income reduction for the targeted subject relative to the cost for the subject who requested the punishment" (Casari, 2005:107).

punishment technology, negative emotions do still play an important role, but much smaller. With respect to the experience and the cultural background of the subjects, we find that previous experience mediates how contempt impacts on the decision to punish, and that non-UK students experience similar emotions to UK students, but generally take more resources from the counterpart than UK students.

In the last chapter of the thesis, we present an experiment aimed at measuring the extent to which voters care about the competence and trustworthiness of candidates in public elections,[2] and at establishing whether one of these characteristics matters more than the other. Despite the topic is very important for both economists and political scientists, whose aim is to understand the voters' decision making in public elections, only little attention has been so far paid to how people weigh the two characteristics – trustworthiness and competence – that define the quality of a public official. In particular, one of the most significant current discussions in political economy is whether voters care only about the state of the economy when they make their voting decisions. In our experiment, we are able to shed light on this by looking at whether there exist any biases in the voting behavior of people towards a specific characteristic of the candidates or whether people simply favor the contender who provides the highest expected payoff, irrespectively of her or his competence or trustworthiness. In addition, the results of this study may provide an explanation of why democracies may at times suffer from dishonesty and corruption at the public level. If people in fact display a rational and profit-maximizing voting behavior or a preference for competence over trustworthiness, the existence of corruption and dishonesty in modern democracies may be explained by people's voting preferences.

To conduct this investigation, we run a lab experiment where we ask voters to select a public official, on the competence and trustworthiness of which their final payoffs depend. We measure the competence of the candidates in a real effort task and their trustworthiness in a trust game, and

[2] We define trustworthiness as the attitude of the potential public official to fulfil the trust that her electors have placed on her. Competence refers instead to the ability of the potential public official to get the job done.

provide this information to voters when they make their voting decision. By looking at cases where there is a competence-trustworthiness trade-off, we can then measure the extent to which competence and trustworthiness matter in electoral decisions. We find that, in general, most voters tend to select the candidate rationally, based on who provides the highest expected payoff irrespectively of trustworthiness and competence, but there is a bias towards caring about trustworthiness when the difference in expected payoffs between the two candidates is small enough. The findings of this chapter provide evidence of the fact that voters mostly behave rationally and care only about their final monetary payoffs quite independently of which trade-off between trustworthiness and competence they face. In other words, it is the final state of the economy that matters the most for voters. This may explain why people may be willing to support untrustworthy candidates and why democracies may at times suffer from dishonesty and corruption at the public level. When however the candidates are not that dissimilar in terms of their contribution to the welfare of the voters, the information about the trustworthiness of the candidates becomes crucial to determine which candidate will be elected. In these occasions, trustworthiness seems to be the aspect that matters more.

Chapter 1: What Happens If You Single Out? An Experiment[3]

1. Introduction

We present the results of an experiment to test the effect of *singling out* an individual on trust and trustworthiness. The act of singling out an individual from a group based on his or her socio-economic categories (e.g. gender, race, age, income, political view) is a pervasive phenomenon of economic and social life. It occurs whenever a subject, who has specific attributes that make him or her potentially different from the others, ceases to be an ordinary and usual person, and becomes a distinct one in the eyes of the other people. More precisely, singling out can be defined as an inter-group situation in which one group is a singleton group made up of a single individual - the singled out individual -, whose social identity is perceived as different by a second group that is larger in size.

Someone can be singled out because he or she possesses some desirable qualities; therefore, the status of being singled out can be associated with a positive social standing. For instance, in organizational and workplace settings, an employee may be object of social recognition or appraisal for his or her desirable socio-economic characteristics (e.g. age or experience) or because he or she belongs to a privileged group, such as in the case, reported by Heikes (1991), of white-male nurses working in all-female environment. As result, he or she loses the status of ordinary member, and acquires a positive aura of uniqueness in the eyes of his or her colleagues or superiors. A subject can also be singled out by others because he or she possesses undesirable qualities; therefore, the status of being singled out can coincide with a negative social standing. In line with the previous example, an employee may be singled out by his or her colleagues or a superior for reproach or because he or she is disliked within the team or belongs to socially disadvantaged minorities, such as women or racial/ethnic minorities. The literature in social psychology offers several examples of *solo* or *token* individuals who have been singled out in the workplace by

[3] This chapter is based on a paper co-authored with Prof. Daniel Zizzo.

colleagues or superiors because of their undesirable qualities (e.g. see Kanter, 1977; Yoder and Aniakudo, 1997; Niemann and Dovidio, 1998; DePaulo and Morris, 2006). When the status of being singled out is attributable to something undesirable about the socio-economic characteristics of the subject, the latter may also be object of social exclusion, marginalization, stigmatization, negative stereotypes, bullying, or, more generally, negative discrimination (e.g. Heatherton *et al.*, 2000; Abrams *et al.*, 2005). In other words, the status of being singled out can be a precondition for these social mechanisms.

While in some cases there may be a consensus among the members of the group on who is the singled out individual, in others the latter is selected by a specific individual, such a manager in an organizational setting:[4] an agent who has the right and power to enforce the status of being singled out.

This chapter presents an experiment designed to test specifically the behavioral implications of singling out. We do so in the context of trust games. A trust game is a standard stylized setup used in the economic literature to study trusting behavior and trustworthiness. Economists are aware of the importance that both trust and trustworthiness play in economic interactions, especially with respect to the formation of social capital (e.g., Putman, 2000). In particular, they reduce the costs of transacting (Frank, 1988), promote efficiency in markets (Arrow, 1974), improve cooperation (Smith *et al.*, 1995) and increase firms' ability to adapt to complexity and change (Korsgaard *et al.*, 1995). Trust and trustworthiness are also considered to be "at the core of group life" (Hogg *et al.*, 2005, p. 193). In particular, they play a fundamental role in ensuring the stability of a group or a team, and, therefore, are important in organizational and workplace settings. Anything that perturbs the stability of the group or team may affect the way in which the members trust and fulfill each other's trust. As singling out may have important consequences in terms of group and team cohesion in organizations, it may then also affect trust and trustworthiness.

[4] Other examples are a teacher in a classroom or a superior in an army force (for instance, think of the overweight, and bumbling marine soldier who was named 'gomer pyle' by the drill instructor, in the Kubrick's movie *Full Metal Jacket*).

For this reason, the trust game appears to be a natural environment where to test, as a starting point, the economic implications of singling a member out in a group.

To test these implications, we artificially induced a status of being singled out in the lab. Under a positive frame, the *most preferred match* in the experiment is singled out, whereas, under a negative frame, the *least preferred match* in the experiment is singled out. We controlled for identification effects by varying the extent to which the status of being singled out could be identified by non-singled out subjects. That is, in certain sessions, the singled out individual was identified with a mark (i.e. an asterisk), and, therefore, recognizable by the other subjects, whereas in other sessions the singled out participant was not marked with an asterisk, and, therefore, could not be identified by the other participants. Furthermore, we tested the implications of having a random assignation of such status under a neutral frame. In particular, rather than having the status of being singled out assigned by the participants, it is randomly assigned by the computer and, therefore, is not associated to being the least or most preferred match. We also investigated the effects of singling a member out under a negative frame when one specific individual rather than the whole group is responsible of such decision. To check the robustness of our results against what we refer as the *individual characteristics hypothesis*, we tested and rejected the possibility that singled out subjects behaved differently because of the individual characteristics that made them singled out. As such, this study provides evidence of the 'pure' effect of singling out an individual from a group, controlling for the individual characteristics of the singled out individual.

Our key finding is that singling out individuals has a negligible effect on trust and is potentially negative in terms of trustworthiness. More specifically, we find that singled out subjects in the negative framework return considerably less than non-singled out subjects, probably because they do not feel any bond with the other members but anger and resentment for the attribution of a lower status. In contrast, the singled out subjects in the positive framework display a bimodal behavior, returning either more or less, probably depending on whether they perceives themselves as insiders or outsiders. We also find that non-singled out subjects return substantially less to the singled out subject but only when they do not feel responsible of the distinct status of this person. Finally, we find that trustworthiness generally decreases if there is a singled out subject. Section 2 briefly reviews some of the related literature. Section 3 describes the experimental design. Section 4 presents the alternative behavioral conjectures about the implications of singling an individual out. Section 5 reports the main results. Section 6 provides a discussion of the findings and concludes.

2. Related Literature

In the economic literature, we did not locate any papers that specifically analyze the economic implications of singling an individual out based on socio-economic categories.[5] However, an area of economic research somewhat related to our study is the one that examines experimentally the impact of group identity. In our experiment, we manipulate the social identity and status of one member, the singled out subject (and, therefore, indirectly that of the other members, the non-singled out subjects), within the reference group, thus creating *de facto* two potential distinct units: a majority group and a singleton group, i.e. a group represented by only one individual. Hence, our manipulation may have implications in line with the main findings of the economic literature on group identity. Although there are several experimental works that looked at group identity in economics (e.g. Hargreaves-Heap and Varoufakis, 2002; Tan and Bolle, 2007; Chen and Li, 2009; Hargreaves-Heap and Zizzo, 2009; Chen and Chen, 2011), none of them has considered a case of social fragmentation like the one implemented in our experiment, in the context of

[5] There is a recent and interesting literature, in particular on public good games, that looked at how "conferring status" to one or few members of the group affects plays of the game. However, none of these studies considered the implications of a subject being disliked or liked, hence singled out by the others, within a group. In particular, in that literature, status is usually conferred to subjects who obtained the highest/lowest score in unrelated quizzes, and implies additional changes on how, for example, information is transmitted to the players (e.g. Eckel *et al*., 2010); or it is conferred at the end of a play to top contributors as a form of incentive to stimulate competition and promote cooperation (e.g. Pan and Houser, 2011). Other early studies explored more broadly the implications of social status, for instance, in markets (e.g. Ball *et al*., 2001), and bargaining games (e.g. Ball and Eckel, 1998). However, differently to our study and the literature described above, status was not conferred to only one individual but to groups of subjects.

a trust game. Two recent papers of this literature are however particularly relevant for our study. The first is Tsutsui and Zizzo (forthcoming). In this study, the authors investigated the role played by majority versus minority groups, and high status versus low status groups in the context of trust games.[6] They observed that minority and low status subjects dislike being in such condition, and discriminate generally less. The second study is Chakravarty and Fonseca (2012), who studied the effects of social fragmentation and group identity on public good contributions. While their vehicle of research (i.e. a six-player public good game) differs from the one used in our experiment (i.e. a two-player trust game), in one treatment they induce a social fragmentation resembling that of our experiment (i.e. one subject experiencing solo status). They found that minority group subjects contribute more to the public good than majority group and middle-sized group subjects.[7]

Another stream of research that is to some extent related to our study is the one that looks at status, social recognition and ranking as a form of incentive in organizations (e.g. Eriksson *et al.*, 2012; Neckermann *et al*, 2012; Charness *et al*., forthcoming). This type of mechanisms might evolve in or conceal a 'singling out' phenomenon insofar as the allocation of the immaterial award or the implementation of the ranking is not based solely on performance but it is, for example, based on subjective evaluation (see, for instance, Neckermann *et al.*, 2012) or the awarded individual starts to be perceived differently by the others. All these studies have only considered the impact of such managerial tools on performance, without looking at their potential side effects (in terms of inducing 'singling out') on trust and trustworthiness in the workplace. On similar grounds, singling out may be also related to leadership, insofar as the leader acquires uniqueness in the eyes of the others. Most of the economic literature on leadership focused on the implications of having a leader making a public decision (e.g.

[6] The group size varied from 4 subjects (minority) to 8 subjects (majority). Group status was manipulated by labeling the high status group in terms of Blue group, whereas the low status group in terms of subjects who do not belong to any group (Tsutsui and Zizzo, forthcoming).

[7] These results might be affected by reputational effects. In particular, largely in treatment 5-1, and weakly in treatment 4-2, minority group individuals are easily detectable even if the software randomized the display order of the individual contributions.

contribution to a public good) or sending a public message before the decisions of the other players (e.g. Güth *et al*., 2007; Potters *et al*., 2007; Gäcther *et al*., 2012). Some of these studies also compared different mechanisms of appointing the leader (e.g. Rivas and Sutter, 2011; Brandts *et al*., 2011; Arbak and Villeval, 2013). None of these studies have however considered the potential implications in terms of group cohesion from having a leader whose social identity is perceived as different by his or her followers.

Finally, some of the behavioral implications which might result from singling an individual out might also be linked to psychological phenomena which has been studied in the psychological research with respect to social exclusion/inclusion, marginalization, stigmatization, and stereotyping (for an overview of this literature, see Heatherton *et al*., 2000; Abrams *et al*., 2005). This literature usually focuses on attitudes rather than behavior (e.g., Sekaquaptewa and Thompson, 2002; Thau *et al*., 2007), and extensively uses deception as a way to manipulate the behavior when the latter is the object of interest (e.g., Twenge *et al*., 2001; Twenge *et al*. , 2007; Derfler-Rozin *et al*., 2010).

3. **Experimental Design**

A. Outline

The experiment was conducted between March and July 2011 at the University of East Anglia with a total of 324 subjects divided into 54 sessions; there were 6 subjects per session.[8] The participants were mostly students with a variety of different backgrounds.[9] The experiment was fully computerized with the z-Tree software (Fischbacher, 2007). Subjects received both computerized and printed instructions at the beginning of each

[8] We ran sessions with only 6 players in order to have enough variation in characteristics between sessions and minimize the possibility that specific characteristics were systematically associated to the singled out subject. Further advantages of running the sessions with only 6 players were that of reducing the chances of subjects knowing each other; minimizing the likelihood of subjects seeing and interacting with each other as they arrived, as they were immediately seated, avoiding people queuing at the entrance of the lab; and minimizing the likelihood of the subjects seeing each other as they left; partitions ensured that subjects did not see each other during the experiment.

[9] Details of the socioeconomic background of the experimental participants, and experimental instructions can be found in the appendix.

experimental task. The presentation of the experimental instructions was as neutral as possible avoiding terms such "trust", "truster", or "trustee". The experiment employed a fictional currency, the *experimental credit*, which was converted to pounds at the end of the experiment at the rate of 20 UK pence per experimental credit. Subjects earned on average £11.78 (around 18-19 US dollars), including a show-up fee of £1.50. Earnings were paid privately and anonymously at subjects' stations at the end of the experiment. Each session lasted around 35 minutes. Subjects were allowed to participate in no more than one session.

The experiment consisted of seven treatments, described below: the baseline (B), the black sheep treatment (BS), the golden sheep treatment (GS), the random sheep treatment (RS), the privately informed black sheep treatment (PIBS), the privately informed golden sheep treatment (PIGS), and the authority and black sheep treatment (ABS). We ran eight sessions per treatment (seven in PIBS and PIGS).

B. Beginning of the Experiment and Ranking Phase

In each session, subjects were randomly assigned to computer terminals, which were separated by partitions in order to avoid facial or verbal communication between subjects. After being assigned to computer stations, subjects were asked to fill in a questionnaire with their personal information. In particular, we asked subjects to indicate their gender, age, current university status, country of origin, whether their main field of studies was related to Economics or not, their religion, whether they used Facebook or not, their current relationship status, and whether they smoked or not. After completing the questionnaire, subjects received the instructions for the first experimental task (i.e. trust game).

Once everyone had finished reading the instructions, each subject was informed about these characteristics for the other participants, and asked to rank them according to how much she or he would like to be matched with them in the experiment (from the most preferred match to the least preferred match).[10] Without informing the participants *ex ante*, the

[10] Ties in the ranking were not allowed.

computer allocated to each subject a certain number of points corresponding to the rank assigned by each other participant to that specific individual (i.e. five points for being ranked first, four for being ranked second, and so on), and ordered the subjects from the participant with the most points (*the most preferred match*) to the one with the least points (*the least preferred match*).[11] In other words, the computer applied a Borda count to the individual rankings in order to determine a consensus-based preference ordering of the participants.

A disadvantage of the Borda count is that it may induce strategic behaviors or a false revelation of own preferences. This is not a problem in our experiment because subjects were not informed *ex ante* about the aggregation procedure and why they had to rank the other participants (i.e. selection of the singled out subject). Furthermore, we are not interested in the results of the Borda count *per se*, but only as a framing tool to induce singling out.

C. The Baseline (B) Treatment

The experimental treatments differed in what followed the ranking phase.[12] We first describe the B treatment. After all the subjects submitted their rankings, they were not told how the computer processed these data. The participants simply proceeded to the next phase. In particular, they filled in a control questionnaire designed to check their understanding of the instructions. Clarifications were individually given to subjects with incorrect answers. The experimental task was a standard Berg *et al*. (1995)'s trust game. In this set-up, a truster (the *first mover*) must decide how much to invest/keep of an endowment X (48 experimental credits in our case). Calling the amount invested T, the investment gives $3 \times T$. This investment's return is sent to the trustee (the *second mover*) who must decide how to share the received amount with the truster. If she keeps Y, the total payoffs will be $(X - T) + 3 \times T - Y$, for the truster, and Y for the trustee, respectively. T measures the amount of trust, $3 \times T - Y$ measures

[11] Ties in the ranking were dealt with by the computer with a random draw.

[12] A critical reader might argue that the ranking phase might change later trust game play. However, all we are interested in this chapter is across-treatment differences, and these cannot be explained by the ranking phase, which equally preceded all treatments.

trust fulfilling and therefore, trustworthiness. There were 4 rounds of the trust game. In each round, each subject was matched with a different co-participant (absolute stranger matching). In this way, we avoided reputation building, which could stem from re-matching the same subjects more than once. Each subject was also randomly assigned the role of trustee for half the time and truster for the other half. The randomness of the order in which roles were assigned to subjects enabled us to rule out any effect due to playing first as truster or trustee. In addition, any across-treatment differences cannot be explained by the fact that individuals played both roles since this equally occurred in all treatments. No information about a co-participant (e.g. participant's ID, gender, nationality, and so on) was revealed to the subjects. At the end of each round, each subject was informed about the decision of the counterpart and the experimental credits that he or she could earn if the round were to be selected for the payment.

Once subjects completed this experimental task, a new set of instructions for an incentivized individual task was given. This new task was a standard Holt and Laury (2002) questionnaire in the domain of gains.[13] The aim of the task was to measure risk attitude by counting the number of times subjects chose the safer option. The task details are in the experimental instructions. A further questionnaire was then given, in two parts. The first part was the 17-item Social Desirability Scale (Stöber, 2001). This scale measures the desire to present oneself in a positive light. For each item of the scale, a subject has to decide if the statement describes himself/herself or not (true-or-false type of scale). The second part of the questionnaire was the Rosemberg's (1965) self-esteem scale, widely used in psychology to measure state self-esteem. For each of the ten items of the scale, a subjects has to indicate how much he or she agree with the statement on a four-point scale (from strongly agrees to strongly disagree). Final payments were based on the earnings of one randomly chosen trust game round, plus the earnings from the Holt and Laury (2002) task.

[13] Houser *et al*. (2010) found no evidence for order effects from having the Holt and Laury task played after the trust game and vice versa.

D. Other Treatments

Now we turn to the description of the other treatments. These were identical to the B treatment except in what follows. In the black sheep (BS treatment), after all the subjects submitted their rankings, the computer explained how it processed these data to determine which one was considered the *least preferred match* in the experiment, i.e. the participant that everyone else least wanted to interact with in the experiment. Subjects were also told that whenever a participant was matched with the least preferred match during the experiment, the least preferred match would be identified with a mark (i.e. an asterisk). The real identity of the least preferred match was not revealed at any point of the experiment.[14] Through this procedure, we artificially induced an identifiable status of being singled out which was based on consensually undesirable attributes.

The golden sheep (GS) treatment was similar to the BS treatment except that the computer explained how it processed the data on the individual rankings to determine which participant was considered the *most preferred match* in the experiment, i.e. the participant that everyone else most wanted to interact with in the experiment. Each subject was then informed whether he or she was the most preferred match. The most preferred match was identified with an asterisk during the trust game. In contrast to the BS treatment, the aim of the GS treatment was to induce an artificial identifiable status of being singled out which was based on consensually desirable attributes.

In the privately informed black sheep, PIBS (privately informed golden sheep, PIGS), treatment, after the ranking phase, the computer informed the subjects about how the least (most) preferred match was selected from the individual rankings, and told the least (most) preferred match about his or her status. However, the singled out participant was not

[14] Our subjects were recruited using the ORSEE software (Greiner, 2004) from the UEA subject pool of over 1,000 potential participants, thus ensuring a systematic randomization of the participants while at the same time minimizing the probability that among the 6 subjects there were acquaintances that could pick up on specific combinations of characteristics to identify co-participants. Importantly, while information was provided on subjects during the ranking phase, once a singled out subject was picked up, we did not reveal his or her individual characteristics on the screen. Overall, there is therefore good reason to believe that subjects did not know who the singled out subject was.

marked with an asterisk during the experiment. In other words, the singled out subject could not be identified by the other participants during the trust game. This treatment was designed to disentangle the "pure" effect of being singled out, which comes from the personal recognition of the singled out subject to be consensually disliked (liked), from the effect of being identifiable as the singled out subject by the others.

In the authority and black sheep (ABS) treatment, after the ranking phase, the computer informed the subjects that the individual ranking of one randomly selected participant from the experiment (i.e. the *authority*) determined who was considered the least preferred match in the experiment. In particular, the least preferred match was the participant that the authority least wanted to interact with in the experiment according to his or her individual ranking. As in the BS treatment, the least preferred match was identified with an asterisk. Subjects were also told whether they were the authority or not, and that the authority could not be matched with the least preferred match during the experiment. This is because we wanted to isolate the behavioral reaction of the singled out subject towards those who were not responsible of his or her status. Note that in the ABS treatment we were not interested in identifying and inducing a real status of being authority. Indeed, what we refer as the authority is simply an individual randomly selected by the computer. Here, we simply wanted to investigate the effects of singling a member out, under a negative frame, when a specific individual rather than the whole group is responsible of such decision. This treatment, together with the RS treatment, enabled us to test the implications on trust and trustworthiness of shifting the responsibility for the condition of the singled out individual to someone else.

Finally, in the random sheep (RS) treatment, subjects were simply told that one of them was going to be randomly singled out by the computer and identified with an asterisk for the rest of the experiment. Hence, here the status of being singled out was not associated to a negative or a positive social standing. This treatment was designed to pick up the effect of a random attribution of distinctiveness under a neutral frame, and when no participant was responsible for such attribution.

As noted earlier, these treatments were identical to the B treatment except for the points noted above, e.g. subjects filled in an initial control questionnaire designed to check their understanding of the instructions, and had a Holt and Laury (2002) and a psychological questionnaire at the end. Table 1 summarizes the main features of our treatments. In BS, we artificially induced a negative identifiable status of being singled out in order to study its implications on trust and trust fulfilling within a group of individuals. GS was identical to BS except that we artificially induced a positive identifiable status of being singled out. In PIBS and PIGS, we controlled for the possible effect that being identifiable as singled out subject has on the behavior of this latter individual. In other words, we removed the effect of identification. RS tested whether the identification mark *per se* affects the behavior of the singled out subject, when no other subject is responsible for his or her status. Finally, in ABS we investigated the implications of having a singled out member within a group of subjects when a specific subject is the only one to blame for the status of the singled out individual.

Table 1: Features of the Experimental Treatments

Treatment	Sessions	Asterisk	Social standing	Being responsible
B	8	NO	NO	-
BS	8	YES	Negative	YES
GS	8	YES	Positive	YES
RS	8	YES	Neutral	NO
PIBS	7	NO	Negative	YES
PIGS	7	NO	Positive	YES
ABS	8	YES	Negative	NO

4. **Behavioral conjectures**

To understand the consequences of singling an agent out, we consider the following individual's utility function that linearly depends on the own (π_O) and counterpart's (π_C) monetary payoffs:[15]

$$U_O(\pi_O, \pi_C) = w_O\pi_O + w_C\pi_C$$

[15] This utility function was proposed by Charness and Rabin (2002) to capture social preferences and extended by Chen and Li (2009) to incorporate group identity.

w_O and w_C are the weights that the individual puts on the own and counterpart's payoffs respectively, with $w_C = 1 - w_O$. We assume that the weight that the individual places on the own payoffs depends on a series of elements (e.g. fairness, reciprocity) including the social distance of the two individuals, by which we mean the degree of demographic similarity between the two agents (Buchan *et al*., 2006).[16]

To formalize this, let I_O be the social distance of the individual with respect to the counterpart and $\theta_O$all other elements that affects the weight. We can rewrite the utility function as

$$U_O(\pi_O, \pi_C) = w_O(I_O, \theta_O)\pi_O + [1 - w_O(I_O, \theta_O)]\pi_C$$

where $\partial w_O / \partial I_O \leq 0$ and $\partial w_O / \partial \theta_O \geq 0$ for any I_O and θ_O. θ_O is a parameter that captures the distribution and other-regarding preferences of the individual, other than those related to social distance. θ_O can, for example, identify the charity concern of the individual when his or her payoff is higher than his or her counterpart's payoff or the envy when his or her payoff is lower than his or her counterpart's payoff (see Charness and Rabin, 2002; Chen and Li, 2009); it can also capture reciprocity concerns. If $\partial w_O / \partial \theta_O = 0$, the individual does not care about other-regarding preferences not captured by I_O.

If $\partial w_O / \partial I_O = 0$, the individual does not care about social distance; the more negative $\partial w_O / \partial I_O$ is, the more the individual cares about social distance and the more he or she will weigh the own payoff compared to the counterpart's payoff. In terms of our manipulation, this means that non-singled out subjects will weigh the payoff of the singled out subject less compared to other non-singled out individuals. This is because they will perceive the singled out subject as socially distant from them. As a result, non-singled out subjects will treat the singled out subject worse than other

[16] Two individuals may differ in many demographic characteristics, such as age, gender, ethnicity, social status, class, etc. Social distance captures the extent to which they are overall different. It can be also measured in terms of group identity (Charness *et al*., 2007; Buchan *et al*., 2006). In particular, two individuals are socially closer if they belong to the same social group and more distant if they do not belong to the same group (Buchan *et al*., 2006).

individuals (see, for example, Hargreaves Heap and Zizzo, 2009), especially if the subject is singled out because of his or her undesirable qualities. At the same time, the presence of a singled out subject will strengthen the feelings of in-group inclusion of the other members (Pickett and Brewer, 2005), thus inducing reciprocal favoritism among those individuals. Similarly, the member who has been singled out will give and return less compared to other individuals because he or she no longer feels any bond with the other members. Hence, the first two conjectures (*social distance conjectures*) are as follows.

Conjecture 1. When the condition of singled out is made salient, non-singled out members will give and return less (more) to the singled out individual (to other non-singled out individuals).

Conjecture 2. The singled out individual will be less trusting and trustworthy compared to other individuals.

Conjecture 1 may be sensitive to whether the responsibility of singling a member out is attributable or not to the non-singled out individual interacting with the singled out member. In particular, in the former case, non-singled out subjects might experience guilt and distress for the singled out subject's condition. This is what Charness (2000, p. 375), in a different context, called the *responsibility-alleviation effect*, i.e. a mitigation of "internal impulses toward honesty, loyalty, or generosity" because of "shifting the responsibility for an outcome to an external authority". In the context of singling out, this means that non-singled out individuals, who can be held responsible for the condition of the singled out individual, might feel guilty, and, therefore, less inclined to treat him or her worse than the others. In other words, this psychological effect counterbalances the effect of perceiving the singled out individual as an outsider (Conjecture 1). The result is that, when the non-singled out individuals can be held responsible for the status of the singled out individual, the two effects may cancel out and therefore we should observe no discrimination. On the other hand, when they cannot be held responsible, there is no guilt involved, and, therefore, we should observe negative discrimination that stems from perceiving the singled out individual as an outsider. Hence, in complement to Conjecture 1, we also consider the following conjecture (*responsibility conjecture*).

Conjecture 3. Non-singled out members interacting with the singled out individual will give and return less to him/her than non-singled out individuals, only when they are *not* responsible for his or her condition.

In opposition to Conjecture 2, it is also possible that the member, who has been singled out because of his or her undesirable qualities, will trust and fulfill trust more in order to demonstrate his or her social value to the others. Such behavior would be consistent with some studies in social psychology showing that individuals who are at the risk of exclusion engage in pro-social behavior in order to reconnect with the others (e.g. Derfler-Rozin *et al*., 2010). In terms of our utility function, this means that the individual who is consensually disliked in the group will weigh the payoff of the counterpart more than other individuals will do. In a similar fashion but for a different reason, the weight that the consensually liked member put on the payoff of the counterpart may be larger than the weight put by other members. This is because he or she may perceive his or her relationship with the others as an intra-group relationship where he or she fulfill a special role, with a greater responsibility for the wealth of the group due to the higher status' attribution. Hence, consistently with some of the findings of the psychological research on leadership theory (e.g. Hogg, 2001), the positively singled out individual may display in-group favoritism towards the other individuals and, therefore, be less selfish or adopt pro-social behaviors. Taking these considerations into account, we can devise the following conjecture, opposed to Conjecture 2, regarding the behavior of the singled out individual (*social standing conjecture*):

Conjecture 4. The negatively (positively) singled out individual will be more trusting and trustworthy compared to other individuals.

Note that all the aforementioned conjectures are related to the pure effect of singling out and abstract from the individual characteristics of the singled out individuals. As our experiment induces the status of being singled out artificially, we are able to provide a powerful test of these conjectures ruling out any explanations that stem from the particular individual characteristics of the singled out individual. In particular, we can test for Conjecture 1, by comparing the behavior of non-singled out subjects towards the singled out individual and other non-singled subjects with the behavior of baseline subjects. If Conjecture 1 is sensitive to the responsibility-alleviation effect (Conjecture 3), we should observe a drop in trust and/or trustworthiness of non-singled out subjects towards the singled out individual only in ABS and RS. Finally, Conjectures 2 and 4 can be tested by looking at the behavior of the singled out subject compared to baseline subjects. On this respect, our control treatments PIBS and PIGS allow us to test the "pure" effect of being singled out, removing any effects that stem from being identifiable as the singled out subject by the others.

5. **Experimental Results**

A. Bivariate Tests

Our focus in this chapter is on the results of the trust games.[17] The *giving rate* identifies the proportion of endowment that the truster transfers to the trustee, while the *return rate* measures the amount returned by the trustee to the truster as a fraction of the total amount received from the truster (i.e. three times the amount given by the truster).

Cooperation towards singled out subjects (Conjectures 1 and 3). Tables 2 and 3 show average giving and return rates for each experimental treatment, while Figure 1 displays giving and return rates for each experimental treatment. Note that, while Table 3 provides information on most and least preferred subjects in the baseline for comparison with the other treatments, most and least preferred subjects in this treatment were not singled out.

[17] We have also analyzed matching preferences as revealed in our unincentivized initial ranking phase. These are presented in the appendix. Throughout the chapter, except where otherwise specified, session averages are used as the unit of observation for bivariate statistical tests (the reported *p* values are two tailed), and individual averages as the unit of observation for the regression analysis.

Table 2: Giving and return rates to singled out and non-singled out

	B	BS	GS	RS	PIBS	PIGS	ABS	Tot.
Giving rate	0.400	0.387	0.398	0.402	0.367	0.405	0.400	0.394
To Singled Out	--	0.354	0.318	0.284	--	--	0.387	0.336
To non-Singled Out*	--	0.422	0.421	0.456	--	--	0.432	0.414
Return Rate	0.244	0.138	0.210	0.204	0.162	0.195	0.184	0.191
To Singled Out	--	0.216	0.198	0.142	--	--	0.100	0.165
To non-Singled Out*	--	0.147	0.212	0.201	--	--	0.246	0.206

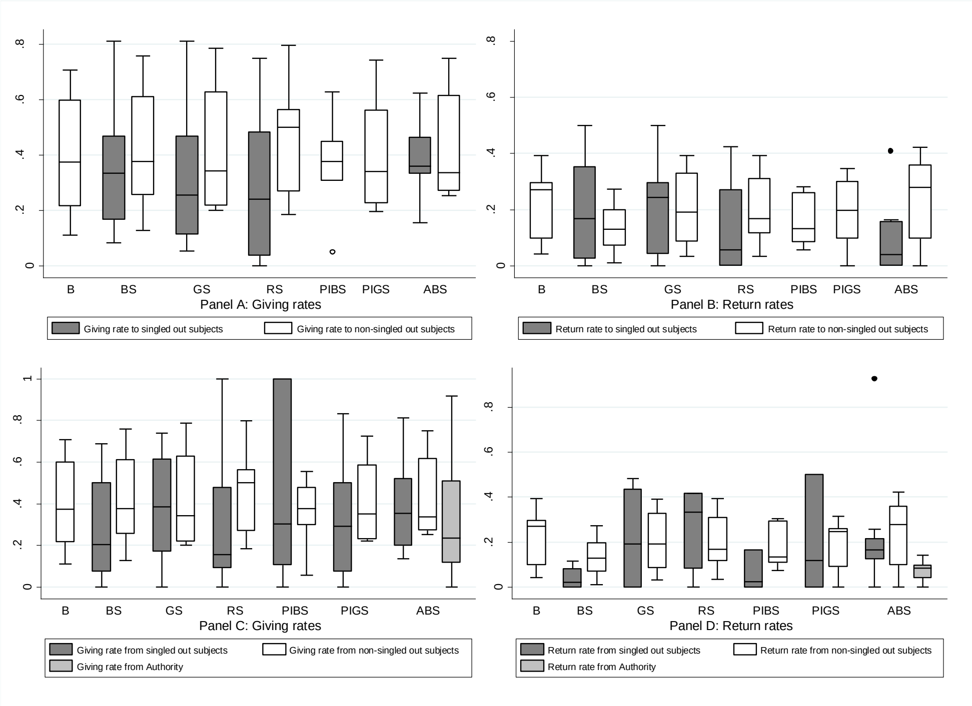
Notes: *Giving/return rate of non-singled out subjects to non-singled out subjects. Tot.: Total.

Table 3: Giving and return rates from singled out, non-singled out, and authority

	B	BS	GS	RS	PIBS	PIGS	ABS	Tot.
Giving rate	0.400	0.387	0.398	0.402	0.367	0.405	0.400	0.394
From 'Singled Out'*	0.500[a] 0.413[b]	0.280	0.385	0.307	0.412	0.329	0.387	0.349
From non-Singled Out**	--	0.422	0.421	0.456	0.358	0.420	0.432	0.414
From Authority	--	--	--	--	--	--	0.331	0.331
Return rate	0.244	0.138	0.210	0.204	0.162	0.195	0.184	0.191
From 'Singled Out'	0.320[a] 0.283[b]	0.037	0.211	0.260	0.073	0.204	0.211	0.164
From non-Singled Out**	--	0.147	0.212	0.201	0.185	0.193	0.246	0.206
From Authority	--	--	--	--	--	--	0.078	0.078

Notes: *Giving rate from most preferred and least preferred subjects to non-singled out subjects; most preferred or least preferred subjects were singled out in all treatments *except* the B treatment, for which values are provided as controls. **Giving/return rate of non-singled out subjects to non-singled out subjects. [a] Giving/return rate from the least preferred baseline subject. [b] Giving/return rate from the most preferred baseline subject. Tot.: Total. B = baseline.

Figure 1: Giving and return rates by session per each treatment



Notes: the middle bar refers to the median value; the edges of the box correspond to the 25th and 75th percentile; whiskers extend to 1.5 times the inter-quartile range; circles identify any other observation.

Result 1. Subjects who have not been singled out gave less on average to singled out than non-singled out individuals, in all the treatments where the status of being "singled out" was revealed to all the participants (BS, GS, RS and ABS).

Result 1 achieves statistical significance (Wilcoxon p = 0.011) in aggregate, and in the RS treatment where the difference is largest (Wilcoxon p = 0.036). This preliminary evidence for Conjecture 1 will be verified in the regression analysis presented later.

Turning to the return rate, there is no evidence of discrimination in aggregate (Wilcoxon p = 0.399). A closer examination of the RS and ABS treatments suggests that return rates in the two treatments exhibit a similar pattern. We find no statistically significant evidence of a different mean return rate to singled out subjects between the two treatments (Mann-Whitney p = 0.676). The singled out subject is selected by the computer (i.e. through a random draw) in RS, and by the authority (i.e. the participant whose individual ranking determined who was considered the least preferred match) in ABS: in both treatments subjects matched with the singled out individual were not responsible for her/his status.

To test the responsibility conjecture (Conjecture 3), we analyzed the mean return rate of RS and ABS sessions pooled together. In 10 out of 15 sessions of RS and ABS,[18] the mean return rate was lower when interacting with a singled out than with a not singled out subject (Wilcoxon $p = 0.038$). In contrast, the mean return rate to singled out subjects did not differ from the mean return rate to non-singled subjects in both BS and GS treatment (Wilcoxon $p = 0.176$ and 0.866 respectively). This analysis leads to the following result that supports the responsibility conjecture with respect to trustworthiness.

Result 2. There is preliminary evidence that, when subjects were not responsible of the distinct status of the singled out individual, they were less trustworthy towards him or her. When they were responsible, trustworthiness was the same as towards non-singled out subjects.

Of course, the return rate may depend on the amount sent by the truster, i.e. the giving rate of the trustee's counterpart. This is because of several psychological reasons such as inequality aversion (Fehr and Schmidt, 1999), reciprocity (Falk and Fischbacher, 2006), and trust responsiveness (Guerra and Zizzo, 2004). This problem will be controlled for in the regression analysis presented later in the chapter.

We can compare giving and return rates towards singled out and non-singled out subjects[19] against giving and return rates in the B treatment in order to determine whether the discrimination is positive or negative, i.e. whether they are treated better or worse than baseline subjects. The aggregate mean giving rate to singled out subjects did not differ from that in the B treatment (Mann-Whitney $p = 0.417$). This result is also robust across treatments. The aggregate mean return rate to singled out subjects did not

[18] One session displayed identical mean return rates, whereas four sessions displayed higher mean return rates when interacting with a singled out individual. Note also that one observation in the RS treatment is missing because in one session the singled out individual did not trust at all the other subjects. This explains why we have 15 observations instead of 16.

[19] Giving and return rates towards non-singled out individuals do not include those from singled out individuals.

differ in aggregate from that in the B treatment (Mann-Whitney p = 0.214). However, if we compare the mean return rates of RS and ABS with the return rates of the B treatment, we mildly reject the hypothesis that return rates to singled out subjects are the same as in the baseline (Mann-Whitney p = 0.059). This will be investigated further in the regression analysis since, as noted earlier, the return rate may depend on the giving rate received by the trustee.

Behavior of singled out subjects (Conjectures 2 and 4). We now consider the behavior of singled out subjects. A first result concerns the trusting behavior of singled out subjects. In 30 out of 46 sessions mean giving rates from singled out subjects were lower than non-singled out subjects (Wilcoxon p = 0.040).

Result 3. In aggregate, there is evidence that singled out subjects trust other people less than non-singled out subjects.

Relatively to trust, this preliminary result appears to support Conjecture 2 against Conjecture 4. Also note that, as shown from Table 3, least and most preferred subjects in the B treatment did not exhibit this pattern,[20] which is consistent with Result 3 as in this treatment most and least preferred subjects were not singled out.

We do not detect any statistically significant difference in mean return rates between singled out and non-singled out subjects (Wilcoxon p = 0.214). As shown by Figure 1 (Panel D), two different behavioral patterns seem at work respectively in BS and PIBS treatments, and GS and PIGS treatments relative to the return rate of singled out subjects.[21] There is preliminary evidence that singled out least preferred matches returned significantly less than subjects who had not been singled out (Wilcoxon p = 0.003)[22] and baseline subjects (Mann-Whitney p = 0.004), including

[20] If we focus on the baseline subjects, and compare the behavior of both the least and most preferred baseline matches with the behavior of the other baseline participants, we do not detect any statistically significant difference in mean giving rates (Wilcoxon p = 0.161 and 0.263 respectively).

[21] Return rates in BS and PIBS display a similar pattern (Mann-Whitney p = 0.464), as do the return rates in GS and PIGS (Mann-Whitney p = 0.952).

[22] This result is robust if we conduct a test on each treatment separately. The Wilcoxon signed-ranks p-values are respectively 0.017 and 0.063 for BS and PIBS.

specifically baseline least preferred matches (Mann-Whitney $p = 0.002$); they also returned less than subjects in RS (Mann-Whitney $p = 0.009$) and ABS sessions (Mann-Whitney $p = 0.007$).

Result 4. Least preferred singled out matches, i.e. singled out subjects under a negative frame, return significantly less than non-singled out subjects.

Result 4 is in contrast to Conjecture 4 but in line with Conjecture 2. In contrast, differences in central tendency between the return rates of the most preferred matches and those of non-singled out subjects were not significant (Wilcoxon $p = 0.798$). Equally, there is no evidence that mean return rates from the most preferred matches in GS and PIGS treatments were different from the mean return rate of the B treatment (Mann-Whitney $p = 0.696$), nor from other treatments (Mann-Whitney $p > 0.100$). Hence, we would be tempted to reject both Conjectures 2 and 4. However, a closer look at the distribution of the mean return rates (Figure 2) reveals why the statistical tests for GS and PIGS provide null results. In BS and PIBS singled out subjects display only low return rates (between 0 and 0.117 in BS, and 0 and 0.167 in PIBS), while Figure 2 shows a different, bimodal pattern for GS and PIGS, with either high or low return rates.[23] We can therefore derive the following result, which, relatively to trustworthiness, appears to partially support both Conjectures 2 and 4.

Result 5. Most preferred singled out matches, i.e. singled out subjects under a positive frame, behave bimodally, with either high or low return rates.

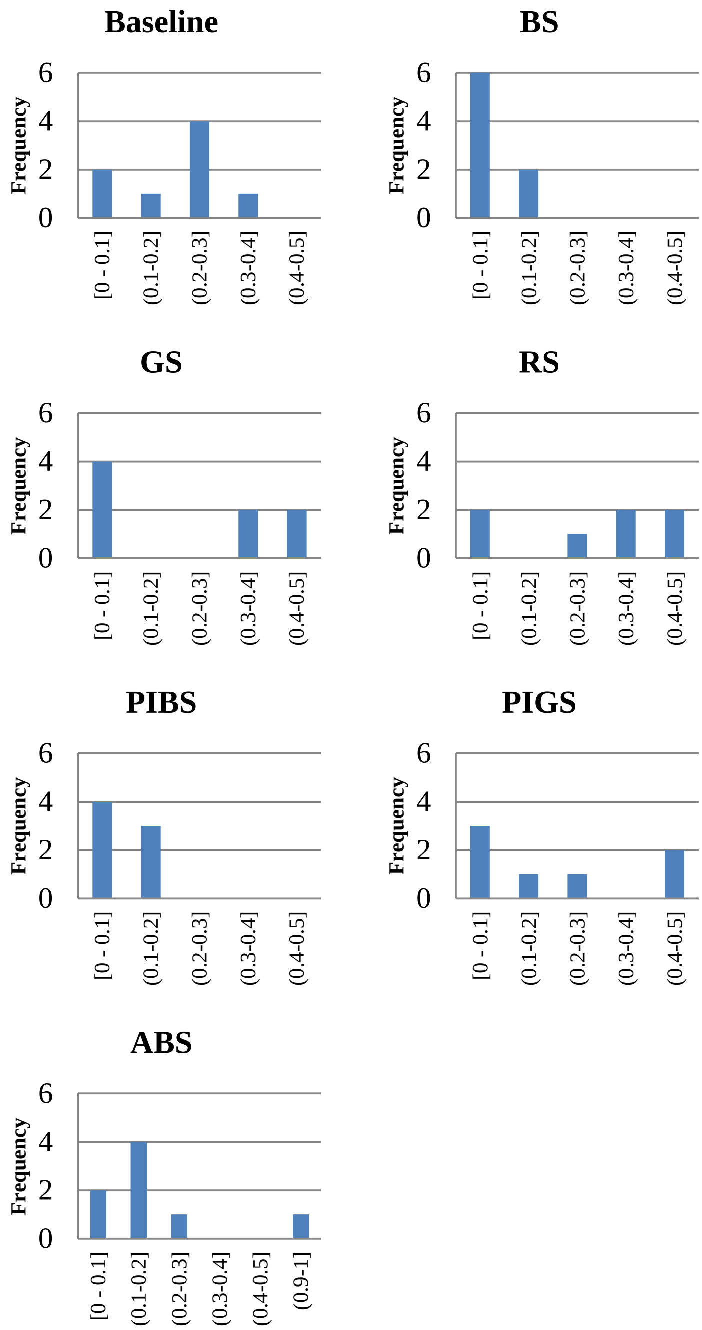
Result 5 implies that the variance of return rates of most preferred matches should be higher than that of return rates of baseline subjects; Siegel-Tukey tests support this ($p = 0.007$).[24] Conversely, return rates variance by least preferred matches (BS and PIBS) is not statistically

[23] BS and PIBS return rates have a standard deviation respectively of 0.049 and 0.082, while in B return rates are smoothly distributed between 0.042 and 0.392 with a standard deviation of 0.125. In contrast, the return rates lie between 0 and 0.483 in GS and between 0 and 0.5 in PIGS, with a standard deviation respectively of 0.224 and 0.226.

[24] A treatment-by-treatment comparison gives similar results (the *p*-values are 0.001 and 0.076 for respectively B versus GS and B versus PIGS).

different from that in the B treatment (p = 0.271),[25] and equally the variance of return rates by most preferred subjects in the B treatment does not differ from that of other baseline subjects (p = 0.164).

Figure 2: Frequencies of Return Rates From Singled Out Subjects



[25] A treatment-by-treatment comparison again gives similar results (the p-values are 0.561 and 0.611 for respectively B versus BS and B versus PIBS).

Behavior of authorities. In the ABS treatment, the mean giving rate of authorities did not significantly differ from the mean giving rate of singled out subjects (Wilcoxon p = 0.575) or of non-singled subjects (Wilcoxon p = 0.327). In contrast, we do have evidence that authorities returned significantly less than singled out subjects (Wilcoxon p = 0.068) and non-singled out subjects (Wilcoxon p = 0.030). Furthermore, while the mean giving rate of authorities did not differ from the baseline (Mann-Whitney p = 0.401), we do find strongly significant evidence of lower return rates from authorities relative to the baseline (Mann-Whitney p = 0.045).

Result 6. There is preliminary evidence that the subject randomly assigned the role of authority returned significantly less because of being assigned to the role of authority.

B. Regression Analysis

In the regression analysis we treat each individual as the unit of observation.[26] We employ Tobit regressions[27] with clustered robust standard errors in order to control for the possible non-independence of the observations within a same session.[28] Regressions 1-2 in Table 4 and regressions 3-4 in Table 5 use the mean giving and return rate respectively to non-singled out subjects as dependent variable, while regressions 5-6 in Table 6 and regressions 7-8 in Table 7 use respectively the giving and return

[26] The *i* observation on the giving/return rate corresponds to the average giving/return rate of the *i*-th subject over the two rounds of the trust game where the subject played as truster/trustee (note from the design section that each subject played two times in the role of truster and two times in the role of trustee in a random order). We use individual averages in order to control for the non-independence of the observations at individual level. Note in fact that we have two potential levels of non-independence: at individual and session level. By taking the individual mean giving and return rates, we control for the first. This simplifies our estimation, reducing potential endogeneity issues and providing more comparable results with previous studies (e.g. Hargreaves-Heap and Zizzo, 2009; Tsustui and Zizzo, in press). In the appendix, we also report the results of Tobit regressions with clustered robust standard errors at subject level or session level, and multilevel mixed-effects linear regressions, where the unit of observation is the round. The results of these regressions are qualitatively similar to those reported in the chapter. However, they do not fully control for the non-independence of the observations or the censored nature of the data, leaving the regressions reported in the chapter as a better estimation option.

[27] Giving and return rates lie between 0 and 1.

[28] We have also conducted random effects regressions, which however, generally failed to pass the Hausman diagnostic test. The only regressions that did pass the Hausman diagnostic test are on the giving rate to non-singled out and singled out subjects. Their results broadly replicate those in the chapter.

rate to singled out subjects.[29] This distinction between giving/return rate to non-singled out subjects and giving/return rate to singled out subjects allows us to test whether singled out subjects discriminate against non-singled out subjects and/or the reverse relative to the Baseline treatment. In particular, they allow us to separate individual decisions made when a subject was matched with a singled out individual and with a non-singled out individual.

The regressions employ dummy variables for the experimental treatments, either individually (BS = 1 for BS treatment observations, and similarly for GS, RS, PIBS, PIGS and ABS) or in combination (RS+ABS = 1 for RS or ABS treatments observations,[30] and similarly for BS + PIBS and for GS + PIGS). In regressions 2 and 4 we use a single dummy variable ("All Treatments") for all treatments with a singled out subject; in regressions 6 and 8 we employ a single dummy variable ("Asterisk") for all treatments where the singled out was identified by an asterisk. In regressions 1-4 we also employ dummy variables, one for each treatment, which take value 1 if the subject was a singled out subjects in BS and PIBS or alternatively in GS and PIGS;[31] we also use a further dummy variable for the authority (Authority = 1 for authorities). In the return rate regressions (Table 5 and 7), an extra explanatory variable is the giving rate received by the trustee.[32] All the regressions include demographic variables, such as age, gender (=1 for men), economics background (=1 if applicable), nationality (UK=1 for UK subjects, and India =1 for Indian subjects), religion (Christian = 1 for Christian subjects, Muslim = 1 for Muslim subjects),

[29] Note that each subject could not interact with a singled out individual (or any other subject) more than once. Hence, each observation of regressions 7-8, in a treatment where the singled out was identified by an asterisk, is not an average across two rounds but corresponds to the giving/return rate of a non-singled out subject interacting with a singled out subject.

[30] The singled out subject was selected by the computer (i.e. through a random draw) in RS, and by the authority (i.e. the participant whose individual ranking determined who was considered the least preferred match) in ABS: in both treatments subjects matched with the singled out individual were not responsible for her/his status. In a bivariate test, we find no statistically significant evidence of a different mean giving/return rate to singled out subjects between the two treatments (Mann-Whitney $p > 0.1$). Hence, we employ a unique dummy to identify both RS and ABS treatment in regressions 5 and 7 (i.e. giving and return rate to singled out subjects).

[31] Giving/return rates of singled out subjects in BS and PIBS display a similar pattern (Mann-Whitney $p > 0.1$), as do the giving/return rates in GS and PIGS (Mann-Whitney $p > 0.1$).

[32] As we already pointed out, the return rate may depend on the amount sent by the truster, i.e. the giving rate of the trustee's counterpart.

whether the subject smokes or not (Smoker = 1 for smoker subjects), whether the subject is a MPhil/PhD student or not (PhD = 1 for MPhil/PhD students) and relationship status (Single = 1 for subjects who were not in a relationship or were unmarried), and psychological measures (i.e. social desirability, self-esteem, and risk attitude).[33] In the regressions of Table 5, we added interaction terms between Authority and the psychological measures.[34]

Table 4. If we first consider the findings regarding the *giving rate to non-singled out subjects*, none of the treatment dummies is statistically significant. Among the dummy variables identifying the giving rates from singled out subjects, none of them is statistically significant. This brings us to the following result that does not seem to support Conjecture 1 with respect to trusting behavior.

Result 7. The presence of singled out subjects did not affect mean giving rates to non-singled out subjects.

Table 5. Turning to the *return rate to non-singled out subjects*, and in line with previous findings on trust games, the giving rate from the truster are found to positively affect the return rate of the trustee (p = 0.000). We also replicate our previous findings that the least preferred match, in line with Conjecture 2, and the authority returned significantly less than other subjects ($p < 0.01$).

[33] We do not introduce a dummy variable for Facebook use since only 7 subjects out of 324 in our sample stated that they do not use Facebook. Also, since the risk elicitation task was administered at the end of the experiment, it is possible that subjects made more or less risky choices depending on their expected earnings obtained in the trust game. If such a bias exists, we should observe a correlation between our measure of risk aversion and the expected payoffs from the trust game. However, this correlation is low and not significant (Spearman $\rho = 0.013$, $p = 0.818$). Hence, we conclude that there is no evidence of systematic bias in measuring risk aversion. Finally, we have also tried other specifications where we have included interactions of the dummies for most and least preferred subjects with dummies related to those attributes (i.e. UK, single, PhD) which were more likely to characterize least and most preferred subjects (see section E of the appendix, and section 4 of the chapter). None of these interaction terms resulted significant.

[34] In all the regressions, all the psychological variables as well as the 'Trust Rate as Trustee' variable are centered in order to control for the high correlation between the independent variables. In other words, we subtract the mean from every observation. For a discussion, see Marquardt (1980).

Table 4: Regressions on Giving Rate to non-singled out subjects

	Regression 1			Regression 2		
	b	se	p	b	se	p
All Treatments				-0.01	0.1	0.919
BS+PIBS	-0.059	0.11	0.594			
GS+PIGS	0.016	0.11	0.887			
RS	0	0.12	0.997			
ABS	0.022	0.11	0.847			
Singled out in RS	-0.081	0.08	0.301	-0.066	0.12	0.585
Singled out in ABS	-0.001	0.07	0.984	0.031	0.08	0.684
Singled out in BS and PIBS	-0.02	0.12	0.868	-0.066	0.12	0.576
Singled out in GS and PIGS	-0.098	0.06	0.128	-0.071	0.08	0.36
Selector	-0.103	0.15	0.494	-0.073	0.13	0.579
Risk Aversion	-0.005	0.01	0.596	-0.007	0.01	0.496
SDS17 Score	-0.015	0.01	0.117	-0.015	0.01	0.12
RSE Score	-0.004	0	0.37	-0.004	0	0.408
Age	0.009	0.01	0.175	0.009	0.01	0.16
Gender	0.037	0.04	0.321	0.033	0.04	0.385
Economics	-0.113**	0.05	0.022	-0.111**	0.05	0.027
UK	-0.051	0.05	0.268	-0.053	0.05	0.251
India	-0.206***	0.07	0.002	-0.199***	0.07	0.003
Christian	-0.07*	0.04	0.087	-0.076*	0.04	0.056
Muslim	-0.334****	0.09	0	-0.333****	0.09	0
Single	0.084*	0.04	0.051	0.089**	0.04	0.041
Smoker	0.019	0.08	0.804	0.029	0.08	0.705
PhD	0.035	0.08	0.651	0.037	0.08	0.64
Constant	0.221	0.19	0.253	0.228	0.18	0.212
Obs	324			324		
Pseudo R-sqr	0.091			0.086		
Df	302			305		
Prob > F	0			0		

Notes: Tobit regression with clustered robust standard errors. * p < 0.1, ** p<0.05,* ** p<0.01, **** p<0.001

Table 5: Regressions on Return Rate to non-singled out subjects

	Regression 3			Regression 4		
	b	se	p	b	se	p
Trust Rate as Trustee	0.406****	0.06	0.000	0.409****	0.06	0.000
All Treatments				-0.076*	0.04	0.052
BS+PIBS	-0.107**	0.05	0.026			
GS+PIGS	-0.078*	0.05	0.091			
RS	-0.041	0.06	0.487			
ABS	-0.047	0.07	0.496			
Singled out in RS	0.105	0.11	0.359	0.143	0.10	0.141
Singled out in ABS	0.109	0.13	0.398	0.138	0.10	0.185
Singled out in BS and PIBS	-0.129***	0.05	0.007	-0.160***	0.05	0.001
Singled out in GS and PIGS	0.064	0.08	0.420	0.060	0.07	0.418
Authority	-0.170****	0.05	0.000	-0.143****	0.04	0.001
Risk Aversion	0.004	0.01	0.574	0.002	0.01	0.742
SDS17 Score	-0.005	0.01	0.379	-0.006	0.01	0.332
RSE Score	-0.007*	0.00	0.094	-0.006	0.00	0.127
Authority ×SDS17	0.048***	0.02	0.003	0.048***	0.02	0.002
Authority ×RSE	0.004	0.01	0.471	0.003	0.01	0.514
Authority ×Risk Aversion	-0.008	0.02	0.653	-0.006	0.02	0.711
Age	0.003	0.01	0.630	0.003	0.00	0.594
Gender	0.011	0.03	0.718	0.009	0.03	0.761
Economics	-0.083**	0.04	0.049	-0.081**	0.04	0.047
UK	-0.060	0.05	0.185	-0.060	0.05	0.193
India	-0.058	0.06	0.323	-0.056	0.06	0.350
Christian	-0.002	0.04	0.958	-0.007	0.04	0.845
Muslim	-0.118	0.08	0.161	-0.122	0.08	0.146
Single	0.067**	0.03	0.024	0.067**	0.03	0.021
Smoker	-0.012	0.04	0.758	-0.009	0.04	0.817
PhD	-0.013	0.06	0.821	-0.010	0.06	0.856
Constant	0.151	0.14	0.273	0.152	0.12	0.217
Obs	307			307		
Pseudo R-sqr	0.393			0.382		
Df	281			284		
Prob > F	0.000			0.000		

Notes: Tobit regression with clustered robust standard errors. * p < 0.1, ** p<0.05,* ** p<0.01, **** p<0.001

Table 6: Regressions on Giving Rate to singled out subjects

	Regression 5			Regression 6		
	b	se	p	b	se	p
Asterisk				-0.127	0.11	0.27
BS	-0.067	0.14	0.638			
GS	-0.127	0.15	0.404			
PIBS	-0.07	0.12	0.57	-0.071	0.12	0.566
PIGS	-0.013	0.12	0.917	-0.012	0.12	0.921
RS+ABS	-0.157	0.13	0.222			
Risk Aversion	-0.008	0.01	0.568	-0.006	0.01	0.63
SDS17 Score	-0.007	0.01	0.512	-0.007	0.01	0.527
RSE Score	-0.006	0.01	0.367	-0.006	0.01	0.376
Age	0.013	0.01	0.149	0.013	0.01	0.156
Gender	-0.022	0.06	0.7	-0.023	0.06	0.677
Economics	-0.15*	0.08	0.064	-0.156*	0.08	0.056
UK	-0.036	0.07	0.607	-0.035	0.07	0.611
India	-0.220**	0.09	0.014	-0.215*	0.09	0.013
Christian	-0.049	0.06	0.421	-0.045	0.06	0.459
Muslim	-0.434***	0.14	0.002	-0.431***	0.14	0.002
Single	0.061	0.06	0.298	0.059	0.06	0.323
Smoker	0.058	0.11	0.593	0.053	0.11	0.628
PhD	-0.119	0.09	0.208	-0.121	0.09	0.203
Constant	0.18	0.23	0.437	0.189	0.23	0.408
Obs	196			196		
R-sqr	0.086			0.084		
Df	178			180		
Prob > F	0.015			0.009		

Notes: Tobit regression with clustered robust standard errors. * p < 0.1, ** p<0.05,* ** p<0.01, **** p<0.001

Table 7: Regressions on Return Rate to singled out subjects

	Regression 7			Regression 8		
	b	se	p	b	se	p
Trust Rate as Trustee	0.437****	0.08	0.000	0.416****	0.07	0.000
Asterisk				-0.130***	0.05	0.005
BS	-0.016	0.09	0.858			
GS	-0.063	0.06	0.297			
PIBS	-0.070	0.04	0.109	-0.076*	0.04	0.087
PIGS	-0.082	0.06	0.157	-0.091	0.06	0.119
RS+ABS	-0.201****	0.06	0.000			
Risk Aversion	-0.004	0.01	0.606	-0.002	0.01	0.766
SDS17 Score	0.005	0.01	0.433	0.004	0.01	0.496
RSE Score	-0.005	0.00	0.330	-0.005	0.00	0.231
Age	0.007	0.01	0.245	0.007	0.01	0.239
Gender	-0.027	0.04	0.448	-0.020	0.04	0.575
Economics	-0.048	0.05	0.323	-0.063	0.05	0.195
UK	0.031	0.05	0.532	0.024	0.05	0.623
India	-0.077	0.05	0.140	-0.098*	0.05	0.054
Christian	-0.048	0.04	0.199	-0.049	0.04	0.195
Muslim	-0.138*	0.08	0.098	-0.094	0.09	0.312
Single	0.032	0.04	0.371	0.036	0.04	0.326
Smoker	0.018	0.05	0.717	0.013	0.05	0.799
PhD	0.037	0.07	0.586	0.031	0.07	0.657
Constant	0.056	0.14	0.695	0.047	0.15	0.754
Obs	179			179		
Pseudo R-sqr	0.491			0.438		
Df	160			162		
Prob > F	0.000			0.000		

Notes: Tobit regression with clustered robust standard errors. * p < 0.1, ** p<0.05,* ** p<0.01, **** p<0.001

Result 8. Least preferred singled out subjects, and authorities, were less trustworthy than baseline subjects.

Authority × SDS17 is statistically significant ($p < 0.05$). This implies that authorities with a high score in the SDS17 questionnaire return proportionally more; this suggests that they perceive that social pressure is put on them (Zizzo and Fleming, 2011).

The aggregate treatments dummy of regression 4 is mildly significantly negative ($p = 0.052$). In regression 3, all the coefficients of treatment variables are negative. However, only the coefficients of BS+PIBS and PG+PIGS are strongly and mildly statistically significant respectively. This evidence brings us to the next result that, relatively to trustworthiness, is strikingly in opposition to Conjecture 1.

Result 9. There is some evidence that the presence of a singled out subject reduced return rates to non-singled out subjects. This is particularly prominent in the treatments where the singled out subject was the least preferred match.

Table 6. In the regressions on the giving rate to singled out subjects, all the coefficients of the treatments dummies are negative, but statistically not significant.

Result 10. Once covariates are controlled for, giving rates towards singled out subjects were not different from those towards baseline subjects.

Hence, although the sign of the coefficients in the regressions points to a negative effect, there is no statistical significant evidence that the singled out individual was trusted less than other individuals (i.e. Conjecture 1 does not hold with respect to trust).

Table 7. In the regressions on the return rate to singled out subjects, all the treatment dummies in regression 7 are negative, though only the coefficient on RS+ABS is statistically significant ($p < 0.001$). An F test restricting all the dummies corresponding to treatments where the singled out can be identified (BS, GS, RS, ABS) to 0 is rejected ($p < 0.01$). In addition, the Asterisk dummy of regression 7 is statistically significant ($p < 0.01$).

Result 11. Subjects who were not responsible of the distinct status of the singled out subject returned about 20% less to this person.

This result provides strong evidence in favor of Conjecture 3 with respect to trustworthiness.

In all the regressions, we control for demographic and psychological variables, and some of them turned out to be statistically significant.[35]

[35] Most notably, participants with a background in economics gave less to both non-singled and singled out subjects; a similar behavioral pattern is observed for Muslim subjects and Indian subjects; while single participants gave and returned more to non-singled out subjects. Our psychological measures had limited power to explain trust game behavior. First, the risk attitude of the subjects, as measured in the Holt and Laury (2002) task, did not relate to trusting or trust fulfilling behavior, as already found in Tsutsui and Zizzo (forthcoming), Lönnqvist *et al*. (2010), and Houser *et al*. (2010). Second, although self-esteem is a key concept of social identity theory, our measure of self-esteem does not

6. Discussion and Conclusions

Our experiment was the first to look at singling out in an economic setting and was run under a minimal and artificial manipulation. While obviously research with natural group has merits, the artificial set-up is a good one to tackle a new area, reduces potential confounding such as natural group stereotyping and minimizes any potential experimenter demand effects. However, it is possible that stronger or less artificial manipulations (e.g., associating a status of being singled out with a specific socio-economic characteristic) may provide additional insights in the understanding of how this social phenomenon works in the real world. Furthermore, we studied the implications of singling someone out in the specific setting of trust games. It is likely that singling someone out also affects the standard results of other experimental environments of interest to within-firm cooperation, such as public goods games, weakest link games and so on; it might also affect individual preferences for time or work productivity. This is why additional research should take place to confirm the robustness of our results and further our understanding of the implications of singling out.

One potential explanation of our findings might be that there is not a behavior change because of singling out. Rather, the singled out subjects may behave differently because individual characteristics information made them singled out (i.e., least preferred or most preferred) *and* implies that they behaved differently. The focus here is on individual characteristics information which was transmitted (e.g., age or PhD), and which may make the sample of singled out subjects different from the sample of non-singled out subjects. This *individual characteristics hypothesis* is within the realm of possibility in the least preferred and most preferred singled out subjects treatments precisely because the choice of such least preferred and most preferred singled out subjects may be non-random, as information is

contribute to explain the behavior in our trust game. Third, other than in relation to authorities as just remarked on, the social desirability scale was unrelated to both trusting and trust fulfilling behavior. Since this measure is a proxy for experimenter demand effects (see Zizzo and Fleming, 2011), the fact that it does not correlate with trust or trustworthiness in our experiment, and that our key results above are robust to controlling for it, inspires confidence for the robustness of our results to potential experimenter demand effects. In the appendix, we also analyze the time trend of the key variables.

provided to subjects to choose their ranking of subjects they would wish to be matched with. We are able to control for this hypothesis in three complementary ways. *First*, in section 3, we provided evidence on how the behavior of the least or most preferred matches in the B subjects, who were *not* singled out, was different from that of the least and most preferred matches in BS/PIBS and GS/PIGS treatments respectively, who *were* singled out. Specifically, the behavioral patterns identified by Result 4 and 8 – such as the lower trustworthiness of singled out subjects – were not replicated by looking at least/most preferred matches in the B subjects. This implies that the act of singling out as opposed to the individual characteristics making a subject least or most preferred determined those results. *Second*, we control directly for various individual characteristics in the regression analysis presented in Tables 4 and 5, thus enabling us to identify the effects of singling out as separate to that of being a singled out subject. *Third*, and more fundamentally, the appendix (section E) shows that the sample of singled out subjects does not differ from the sample of corresponding non-singled out subjects in almost all of the individual characteristics. As shown in the appendix, PhD students were more likely to be least preferred, but the PhD dummy is statistically insignificant in all regressions in Tables 4-7, implying that it made no difference.[36] Equally, UK subjects were more likely to be most preferred matches (see appendix), but the UK dummy is again statistically insignificant in Tables 4-7, implying that they did not behave any differently from everyone else. Finally, single subjects were *less* likely to be most preferred matches, and what we find is that they were more trusting and trustworthy towards non-singled out subjects according to Tables 4 and 5. However, we have not found any evidence that high status subjects (i.e., the most preferred matches) are less trusting and trustworthy towards non-singled out subjects, and so this potential individual characteristic effect does not turn out to be relevant. It is, of course, anyway controlled for in the regression analysis.

[36] Further evidence of the irrelevance of this variable for our findings on low status subjects (i.e. least preferred matches) is that the results of the regressions do not change if we drop the observations corresponding to PhD students selected as the least preferred match (3 observations).

As we can rule out the individual characteristics hypothesis as an explanation of our results, we conclude that singling out appears to matter as such. Insofar as we could glean from our experiment, we found no evidence suggesting that singling out is beneficial, at least with respect to trust or trustworthiness. We found it is irrelevant for trust and potentially disruptive for trustworthiness.

Trustworthiness of non-singled out subjects towards singled out subjects. In line with Conjecture 1, we found an overall reduction of trustworthiness of non-singled out subjects towards singled out subjects, and one largely focused on the RS and the ABS treatments. In these treatments, subjects returned around 20% less to the singled out subject compared to the baseline (Result 6), and over twice as large an effect relative to the other treatments. This effect holds even while controlling for covariates such as behavioral reciprocity (based on how much trustees were given by trusters) and any potential experimenter demand (as proxied by our social desirability scale measure). Our interpretation is that, in RS and ABS, the responsibility of choosing the singled out subject shifted to someone else, and, therefore, any concern for the singled out subject's condition was mitigated if not removed. Conversely, in the case of the other treatments, such concern could be present, as subjects may have felt responsible for the singled out subject, thus reducing the negative effect of perceiving the singled out individual as an outsider. This interpretation is in line with a *responsibility-alleviation effect* (Charness, 2000) identified in Conjecture 3. A second explanation, which is still linked with a responsibility-alleviation argument, is that, in treatments such as BS, subjects might have thought to have made mistakes in the selection process of the singled out subject, and, therefore, did not want to take actions which could have harmed a "blameless" person. In other words, here the responsibility is not to have consciously singled someone out, but rather to have made mistakes in singling someone out.[37]

[37] Note that people might be in general more careful in evaluating top ranked choices than lower ranked choices (Hausman and Ruud, 1987), and, therefore, commit increasing mistakes with the latter.

Trustworthiness of non-singled out subjects towards non-singled out subjects.[38] In contrast to Conjecture 1, we found no evidence that singling out works as a bonding tool for other group members leading to greater in-group cooperation: relative to the baseline, there is no significant increase in trustworthiness from non-singled out subjects towards other non-singled out subjects. If anything, there is evidence the presence of a singled out subject reduced return rates to non-singled out subjects, notably by around 10-11% in the treatments where the singled out subject was the least preferred match (Result 9 and Table 5). There are different possible reasons for this; we mention two. First, the presence of an a-prototypical member in the group may jeopardize the distinctiveness of the in-group as far as the singled out subject is not excluded from the group (Hogg *et al*., 2005). Second, non-singled out subjects may blame other non-singled out subjects for some responsibility for the singling out of a specific subject.

Trustworthiness of singled out subjects towards non-singled out subjects. It mattered for singled out subjects whether they were singled out for being the least preferred match or otherwise. In contrast to Conjecture 4 and in agreement with Conjecture 2, return rates by least preferred matches strikingly decreased from the 24% of the baseline to single digits (4-7%) as per Table 3. According to the regressions in Table 5, once covariates are taken into account, subjects who were seen as least preferred matches were less trustworthy by as much as 16%. It was not the act of being marked as low status that caused this reaction, since it occurred even in PIBS, when only least preferred matches knew they were the least preferred, and they knew that this was the case. We can also exclude the fact of being singled out as having such an effect *per se*, since we do not find the same large effect outside BS and PIBS. Rather, it was the fact of being considered by other subjects as undesirable that appears to have elicited the negative reaction. Many psychological studies show that people who have been excluded appear to engage in anti-social behaviors (e.g. Twenge *et al.*, 2001). Our manipulation does not imply exclusion. Nevertheless, some of

[38] We obviously do not have a discussion of the trustworthiness of singled out subjects towards other singled out subjects since there was a single singled out subject in each session, and so no other singled out subject that each singled out subject could interact with.

the underlying psychological forces motivating excluded people to engage in self-defeating social behaviors might also be the same that trigger the anti-social behavior of the singled out subject. Anger, resentment, and reciprocity might be the driving forces of such retaliatory behavior.

Most preferred singled out matches, i.e. singled out subjects under a positive frame, behaved bimodally, with either high or low return rates (Result 5 and Figure 2). This result supports at the same time both Conjecture 2 and 4. One interpretation of this result is that it reflects some of the mixed results of the psychological research on leadership theory. Highly prototypical subjects should display more distinct group behaviors, and, therefore, more in-group favoritism (Hogg, 2001). However, the status-based gap between the highly prototypical subject and the rest of the group may transform an intra-group relationship between the consensually liked subject and the others into an inter-group relationship (Hogg, 2001). Due to this, the singled out subject may behave in a more anti-social way toward the lower status subjects. Our results suggest that both behavioral patterns may describe the singled out subject's decision whether to fulfill trust. Which behavior turns out to happen probably depends on whether the most preferred singled out individual perceives himself or herself as an insider or outsider.

Another interpretation of our bimodality finding is that also it may be explained by the particular beliefs of most preferred matches, quite independently of group identity concerns. The most preferred singled out subject might have believed to be selected by others because of strategic reasons rather than niceness (i.e. he or she was considered the most exploitable subject in the groups), and, therefore, he or she might have behaved antisocially in response to such attribution (in the spirit of McCabe *et al*., 2003). Alternatively, the fact that he or she is considered most preferred might support a belief that trust is being placed on him or her, and that he or she should feel let down if trust is not fulfilled; this would lead to more pro-social behavior (e.g., Dufwenberg and Gneezy, 2000; Battigalli and Dufwennberg, 2007). Our design does not allow us to disentangle these alternative explanations, and it would therefore be worthwhile to investigate this in further research.

Trustworthiness of *authorities* *towards* *non-singled out* *subjects*.[39] ABS treatment authorities had a return rate of just around 7-8%. The reduction in trustworthiness is moderated by experimenter and social demand: subjects more sensitive to social pressure such as experimenter demand return comparatively more, as shown by the significance of SDS17 × Authority in the Table 5 regressions. Nevertheless, Table 5 also shows that the effect persists when controlling for our social desirability scale measure that we employ to control for experimenter demand.

We should point out that the aim of our experiment was not to study the behavior of the authority. We simply wanted to investigate the effects of singling out when a specific individual rather than the entire group is responsible for the lower-status attribution. Different conjectures might explain why the authority did not fulfill trust. We mention two. First, authorities might have felt that, since their co-participants had been assessed as comparatively worthy matches, they should be more generous in their giving. Second, they may have felt entitled to keep more money because he or she had already a service to everyone else by helping identify the least preferred match.

Our starting point was the fact that singling out is a pervasive phenomenon of economic and social life. We found that singling out individuals does not carry any benefit in terms of trust and has a negative effect for trustworthiness. Obviously, further research is needed and singling out may yet have benefits for organizations – if, for example, it is connected to social rewards and therefore can be used to elicit greater work productivity –. However, if you are a manager and you are considering singling out someone for blame and praise, you may wish to bear in mind that this, and especially the former, may disrupt the social glue holding the team together.

[39] We do not discuss the trustworthiness of authorities towards singled out subjects since, as noted earlier, authorities were not matched with singled out subjects. Also, non-singled out subjects did not know they were matched with an authority, and so could not condition their behavior on a subject being marked as an authority.

References

Abrams, D., M. A. Hogg and J. M. Marques eds. 2005. *The Social Psychology of Inclusion and Exclusion*. New York, NY: Psychology Press.

Arbak, E. and M. Villeval (2013), Voluntary leadership: motivation and influence. *Social Choice and Welfare*, 40(3), 635-662.

Arrow, K. 1974. *The Limits of Organizations*. New York: Norton.

Ball, S. B., and C. Eckel. 1998. The economic value of status. *Journal of Socio-Economics*, 27, 495-514.

Ball, S. B., C. Eckel, P. J. Grossman, and W. Zame. 2001. Status in Markets. *The Quarterly Journal of Economics*, 116 (1), 161-188.

Battigalli, P. and M. Dufwenberg. 2007. Guilt in Games. *American Economic Review Papers & Proceedings*, 97(2), 170-76.

Berg, J., J. Dickhaut and K. McCabe. 1995. Trust, Reciprocity, and Social-History. *Games and Economic Behavior*, 10(1), 122-42.

Brandts, J., D. Cooper, and R. Weber. 2011. Legitimacy, social distance, and leadership in the turnaround game. Florida State University Working Paper.

Buchan, N. R., E. J. Johnson, and R. T. A. Croson. 2006. Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences. *Journal of Economic Behavior & Organization*, 60(3), 373-398.

Chakravarty, S. and M. A. Fonseca. 2012. The Effect of Social Fragmentation on Public Good Provision: an Experimental Study. *Discussion Papers 1207, Exeter University*.

Charness, G. 2000. Responsibility and Effort in an Experimental Labor Market. *Journal of Economic Behavior & Organization*, 42(3), 375-84.

Charness, G., E. Haruvy, and D. Sonsino. 2007. Social distance and reciprocity: An Internet experiment. *Journal of Economic Behavior & Organization*, 63(1), 88-103.

Charness, G., D. Masclet and M. Villeval. 2013. The dark side of competition for status. Forthcoming in *Management Science*.

Charness, G., and M. Rabin. 2002. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*. 117(3), 817-869.

Chen, R. and Y. Chen. 2011. The Potential of Social Identity for Equilibrium Selection. *American Economic Review*, 101(6), 2562-89.

Chen, Y. and S. X. Li. 2009. Group Identity and Social Preferences. *American Economic Review*, 99(1), 431-57.

Croson, R. 2000. Thinking like a Game Theorist: Factors Affecting the Frequency of Equilibrium Play. *Journal of Economic Behavior & Organization*, 41(3), 299-314.

DePaulo, B. M. and W. L. Morris. 2006. The Unrecognized Stereotyping and Discrimination against Singles. *Current Directions in Psychological Science*, 15(5), 251-54.

Derfler-Rozin, R., M. Pillutla and S. Thau. 2010. Social Reconnection Revisited: The Effects of Social Exclusion Risk on Reciprocity, Trust, and General Risk-Taking. *Organizational Behavior and Human Decision Processes*, 112(2), 140-50.

Dufwenberg, M. and U. Gneezy. 2000. Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior,* 30(2), 163–182.

Eriksson, T., L. Mao, and M. Villeval. 2012. Losing Face. Mimeo.

Falk, A. and U. Fischbacher. 2006. A Theory of Reciprocity. *Games and Economic Behavior*, 54(2), 293-315.

Eckel C. C., E. Fatas, and R. Wilson. 2010. Status and Cooperation in Organizations. *Journal of Public Economic Theory*, 12 (4), 737-762.

Fehr, E. and K. M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3), 817-68.

Fischbacher, U. 2007. Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, 10(2), 171-78.

Frank, R.H. 1988. *Passions within Reason: The Strategic Role of Emotions*. New York: Norton.

Guerra, G. and D. J. Zizzo. 2004. Trust Responsiveness and Beliefs. *Journal of Economic Behavior & Organization*, 55(1), 25-30.

Gächter, S., D. Nosenzo, E. Renner, M. Sefton. 2012. Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-by-example. *Economic Inquiry*, 50(4), 1465-7295.

Greiner, B. 2004. The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. University of Cologne, Working Paper Series in Economics 10.23, 63-104.

Güth, W., M. V. Levati, M. Sutter, and E. van der Heijden. 2007. Leading by Example with and without Exclusion Power in Voluntary Contribution Experiments. *Journal of Public Economics*, 91(5–6), 1023-1042.

Hargreaves-Heap, S. P. H. and Y. Varoufakis. 2002. Some Experimental Evidence on the Evolution of Discrimination, Co-Operation and Perceptions of Fairness. *Economic Journal*, 112(481), 679-703.

Hargreaves-Heap, S. P. H. and D. J. Zizzo. 2009. The Value of Groups. *American Economic Review*, 99(1), 295-323.

Hausman, J. A. and P. A. Ruud. 1987. Specifying and Testing Econometric-Models for Rank-Ordered Data. *Journal of Econometrics*, 34(1-2), 83-104.

Heatherton, T. F., M. R. Hebl, J.G. Hull and R. E. Kleck eds. 2000. *The Social Psychology of Stigma*. New York, NY: Guilford Press.

Heikes, E. J. 1991. When Men Are the Minority - the Case of Men in Nursing. *Sociological Quarterly*, 32(3), 389-401.

Hogg, M. A. 2001. A Social Identity Theory of Leadership. *Personality and Social Psychology Review*, 5(3), 184-200.

Hogg, M. A., Fielding, K. S., and Darley, J. 2005. Fringe Dwellers: Processes of Deviance and Marginalization. D. Abrams, Hogg, M. A., and Marques, J. M., *The Social Psychology of Inclusion and Exclusion.* New York: Psychology Press, 191-210.

Holt, C. A. and S. K. Laury. 2002. Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644-55.

Houser, D., D. Schunk, and J. Winter. 2010. Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior and Organization*, 74(1–2), 72-81.

Kanter, R. M. 1977. Some Effects of Proportions on Group Life - Skewed Sex-Ratios and Responses to Token Women. *American Journal of Sociology*, 82(5), 965-90.

Korsgaard, M. A., D. M. Schweiger and H. J. Sapienza. 1995. Building Commitment, Attachment, and Trust in Strategic Decision-Making Teams -

the Role of Procedural Justice. *Academy of Management Journal*, 38(1), 60-84.

Lönnqvist, J., M. Verkasalo, G. Walkowitz and P. C. Wichardt. 2010. Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison. Working Paper SSRN eLibrary.

Marquardt, D. W. 1980. A Critique of Some Ridge-Regression Methods - Comment. *Journal of the American Statistical Association*, 75(369), 87-91.

McCabe, K. A., M. L. Rigdon, and V. L. Smith. 2003. Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52(2), 267-275.

Neckermann, S., R. Cueni and B. S. Frey. 2012. Awards at Work. Working Paper SSRN eLibrary.

Niemann, Y. F. and J. F. Dovidio. 1998. Relationship of Solo Status, Academic Rank and Perceived Distinctiveness to Job Satisfaction of Racial/Ethnic Minorities. *Journal of Applied Psychology*, 83(1), 55-71.

Pan X.S. and D. Houser. 2011. Competition for Trophies Triggers Male Generosity. *PLoS ONE*, 6(4), e18050.

Pickett, C. L. and M. B. Brewer. 2005. The Role of Exclusion in Maintaining Ingroup Inclusion. D. Abrams, M. A. Hogg and J. M. Marques, *The Social Psychology of Inclusion and Exclusion.* New York: Psychology Press, 89-111.

Potters, J., M. Sefton, and L. Vesterlund. 2007. Leading-by-example and signaling in voluntary contribution games: an experimental study. *Economic Theory*, 33(1), 169-182.

Putnam, R. D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon & Schuster.

Rivas, M. F., and M. Sutter. 2011. The benefits of voluntary leadership in experimental public goods games. *Economics Letters*, 112(2), 176-178.

Rosenberg, M. 1965. *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Sekaquaptewa, D. and M. Thompson. 2002. The Differential Effects of Solo Status on Members of High- and Low-Status Groups. *Personality and Social Psychology Bulletin*, 28(5), 694-707.

Smith, K. G., S. J. Carroll and S. J. Ashford. 1995. Intraorganizational and Interorganizational Cooperation - toward a Research Agenda. *Academy of Management Journal*, 38(1), 7-23.

Stöber, J. 2001. The Social Desirability Scale-17 (Sds-17) - Convergent Validity, Discriminant Validity, and Relationship with Age. *European Journal of Psychological Assessment*, 17(3), 222-32.

Tajfel, H. and J. Turner. 1979. An Integrative Theory of Intergroup Conflict, W. Austin and S. Worchel, *The Social Psychology of Intergroup Relations.* Monterey, CA: Brooks Cole, 33-47.

Tan, J. H. W. and F. Bolle. 2007. Team Competition and the Public Goods Game. *Economics Letters*, 96(1), 133-39.

Thau, S., K. Aquino and P. M. Poortvliet. 2007. Self-Defeating Behaviors in Organizations: The Relationship between Thwarted Belonging and Interpersonal Work Behaviors. *Journal of Applied Psychology*, 92(3), 840-47.

Tsutsui, K. and D. J. Zizzo. Group Status, Minorities and Trust. Forthcoming in *Experimental economics*.

Twenge, J. M., R. F. Baumeister, C. N. DeWall, N. J. Ciarocco and J. M. Bartels. 2007. Social Exclusion Decreases Prosocial Behavior. *Journal of Personality and Social Psychology*, 92(1), 56-66.

Twenge, J. M., R. F. Baumeister, D. M. Tice and T. S. Stucke. 2001. If You Can't Join Them, Beat Them: Effects of Social Exclusion on Aggressive Behavior. *Journal of Personality and Social Psychology*, 81(6), 1058-69.

Yoder, J. and P. Aniakudo. 1997. "Outsider within" the Firehouse - Subordination and Difference in the Social Interactions of African American Women Firefighters. *Gender & Society*, 11(3), 324-41.

Zizzo, D. J. and P. Fleming. 2011. Can Experimental Measures of Sensitivity to Social Pressure Predict Public Good Contribution? *Economics Letters*, 111(3), 239-42.

Appendix to Chapter 1: What Happens If You Single Out? An Experiment

A. Experimental instructions
B. Background information on participants
C. Background information on singled out subjects and authorities
D. Ranking phase results
E. Test for random sampling of the least and most preferred subjects
F. Time trend of key variables
G. Additional Regression Analysis

A. Experimental Instructions

ALL TREATMENTS (beginning of the experiment)

Introduction

This is an experiment on decision making. During the experiment, you are not allowed to communicate with other participants. Please raise your hand if you have any questions at any point in the experiment.

The experiment consists of **four rounds**. In addition to these four rounds, there is an individual task at the end of the experiment.

There are six participants in the experiment, all of which have received the same set of instructions as you have.

Round Decisions

In each round, you will be matched at random with a **different** participant (the **coparticipant**). Therefore, you will never be matched with the same coparticipant twice.

Give/Return Decisions: In each round, you or your coparticipant will be designated to move first. The **First Mover** will begin by receiving 48 credits. He or she will decide how many credits (if any) to give to the other person and how many (if any) to keep.

All the credits given get multiplied by **3** before they are received by the **Second Mover**. The Second Mover then decides how much (if any) to keep and how much (if any) to return to the First Mover.

Role: you will have the role of First Mover for two random rounds, and that of Second Mover for the other two random rounds.

Round earnings: The decisions that you and your coparticipants make in each round will determine the amounts you gain as round earnings.

Information: In each round you will learn about your coparticipant's decision and about your round earnings.

Payments

At the end of the experiment a **winning round** is chosen at random from the four rounds, and you will be paid according to your earnings of this round. Your earnings in the winning round will be converted into pounds at the rate of 20 pence per experimental credit. Your final earnings will also include additional earnings that you can gain from the individual task at the end of the experiment. Please remain seated until we come to your desk to give you the money.

Please raise your hand if you have any questions.
BS TREATMENT (ranking phase)

Before proceeding with the experiment, please rank the coparticipants according to how much you would like to be matched with them in the experiment (from the most preferred match to the least preferred match).
For each member, indicate his/her rank from 1 (the most preferred match) to 5 (the least preferred match). In other words, please assign a rank of 1 to the coparticipant that you would most like to be matched with in the experiment, assign a rank of 2 to the second most preferred match, 3 to the third most preferred match, and so on. Ties in the ranking are not allowed.

Explanation of the least preferred match

The computer has received the individual ranking of each of you. It will now aggregate all the rankings and will determine which participant is considered the *least preferred match* in the experiment, i.e. the participant that everyone else least wants to interact with in the experiment according to the following rule.

The computer will give each participant a certain number of points corresponding to the position in which he or she has been ranked by the coparticipants. Specifically, each participant will receive five points each time he or she has been ranked first by another participant, four for being ranked second, three for being ranked third, two for being ranked fourth and one for being ranked last.

The participant with the least points will be selected by the computer as the least preferred match in the experiment.

In each round, if a participant is matched with the least preferred match, the least preferred match will be marked with an asterisk (*).

Please raise your hand if you have any questions.

GS TREATMENT (ranking phase)

Before proceeding with the experiment, please rank the coparticipants according to how much you would like to be matched with them in the experiment (from the most preferred match to the least preferred match).
For each member, indicate his/her rank from 1 (the most preferred match) to 5 (the least preferred match). In other words, please assign a rank of 1 to the coparticipant that you would most like to be matched with in the experiment, assign a rank of 2 to the second most preferred match, 3 to the third most preferred match, and so on. Ties in the ranking are not allowed.

Explanation of the most preferred match

The computer has received the individual ranking of each of you. It will now aggregate all the rankings and will determine which participant is considered the *most preferred match* in the experiment, i.e. the participant that everyone else most wants to interact with in the experiment according to the following rule.

The computer will give each participant a certain number of points corresponding to the position in which he or she has been ranked by the coparticipants. Specifically, each participant will receive five points each time he or she has been ranked first by another participant, four for being ranked second, three for being ranked third, two for being ranked fourth and one for being ranked last.

The participant with the most points will be selected by the computer as the most preferred match in the experiment.

In each round, if a participant is matched with the most preferred match, the most preferred match will be marked with an asterisk (*).

Please raise your hand if you have any questions.

PIBS TREATMENT (ranking phase)

Before proceeding with the experiment, please rank the coparticipants according to how much you would like to be matched with them in the experiment (from the most preferred match to the least preferred match).
For each member, indicate his/her rank from 1 (the most preferred match) to 5 (the least preferred match). In other words, please assign a rank of 1 to the coparticipant that you would most like to be matched with in the experiment, assign a rank of 2 to the second most preferred match, 3 to the third most preferred match, and so on. Ties in the ranking are not allowed.

Explanation of the least preferred match

The computer has received the individual ranking of each of you. It will now aggregate all the rankings and will determine which participant is considered the *least preferred match* in the experiment, i.e. the participant that everyone else least wants to interact with in the experiment according to the following rule.

The computer will give each participant a certain number of points corresponding to the position in which he or she has been ranked by the coparticipants. Specifically, each participant will receive five points each time he or she has been ranked first by another participant, four for being ranked second, three for being ranked third, two for being ranked fourth and one for being ranked last.

The participant with the least points will be selected by the computer as the least preferred match in the experiment.

A participant does **not** know whether and when he or she is matched with the least preferred match during the experiment.

Please raise your hand if you have any questions.

PIGS TREATMENT (ranking phase)

Before proceeding with the experiment, please rank the coparticipants according to how much you would like to be matched with them in the experiment (from the most preferred match to the least preferred match).
For each member, indicate his/her rank from 1 (the most preferred match) to 5 (the least preferred match). In other words, please assign a rank of 1 to the coparticipant that you would most like to be matched with in the experiment, assign a rank of 2 to the second most preferred match, 3 to the third most preferred match, and so on. Ties in the ranking are not allowed.

Explanation of the most preferred match

The computer has received the individual ranking of each of you. It will now aggregate all the rankings and will determine which participant is considered the *most preferred match* in the experiment, i.e. the participant that everyone else most wants to interact with in the experiment according to the following rule.

The computer will give each participant a certain number of points corresponding to the position in which he or she has been ranked by the coparticipants. Specifically, each participant will receive five points each time he or she has been ranked first by another participant, four for being ranked second, three for being ranked third, two for being ranked fourth and one for being ranked last.

The participant with the most points will be selected by the computer as the most preferred match in the experiment.

A participant does **not** know whether and when he or she is matched with the most preferred match during the experiment.

Please raise your hand if you have any questions.

RS TREATMENT (ranking phase)

Ranking

Before proceeding with the experiment, please rank the coparticipants according to how much you would like to be matched with them in the experiment (from the most preferred match to the least preferred match).
For each member, indicate his/her rank from 1 (the most preferred match) to 5 (the least preferred match). In other words, please assign a rank of 1 to the coparticipant that you would most like to be matched with in the experiment, assign a rank of 2 to the second most preferred match, 3 to the third most preferred match, and so on. Ties in the ranking are not allowed.

Asterisk

The computer will now **randomly** select a participant from the experiment. This participant will be referred to as the *denoted match*.

In each round, if a participant is matched with the denoted match, the denoted match will be marked with an asterisk (*).

Please raise your hand if you have any questions.

ABS TREATMENT (ranking phase)

Explanation of the least preferred match

The computer has received the individual ranking of each of you. It will now randomly select a participant from the experiment. This participant will be referred to as the *authority*.

The individual ranking of the authority will determine who is considered the *least preferred match* in the experiment, i.e. the participant that the authority least wants to interact with in the experiment according to his/her individual ranking.

In each round, if a participant is matched with the least preferred match, the least preferred match will be marked with an asterisk (*****).

The authority will not be matched with the least preferred match for the rest of the experiment.

Please raise your hand if you have any questions.

Screens:

You have not been randomly selected by the computer and, therefore, you are not the authority.

Please wait to be informed whether you are the least preferred match or not.

You have been randomly selected by the computer as the authority.

The participant that you have indicate as the least preferred match is going to be the least preferred match in the experiment and will be marked with an asterisk (*****).

You will not be matched with the least preferred match for the rest of the experiment.

ALL TREATMENTS (beginning of individual task)

Instructions for the individual task

You now need to make 10 decisions. Each decision is a paired choice between two options ("Option A" and "Option B").
Only one of these 10 decisions will be used in the end to determine your earnings. You will only know which one at the end of the experiment.

How will these decisions affect your earnings for this part of the experiment? After you have made all of your decisions, the computer will randomly select which of the 10 decisions will be used to determine your earnings. In relation to this decision, the computer will then randomly select the outcome based on the probabilities assigned to the option you chose.

As an example, assume that, for the randomly selected decision, the option to the left pays 10 credits with a 10% chance and 5 credits with a 90% chance, while the option to the right pays 8 credits with a 20% chance and 4 credits with an 80% chance. Assume that you chose the option to the left for this decision; then there is a 10% chance that you will earn 10 credits and a 90% chance that you will earn 5 credits.

Please raise your hand if you have any questions.

B. Background of Experimental Participants

Gender	Frequency	Percent
Female	167	51.54
Male	157	48.46
Total	324	100.00

Degree	Frequency	Percent
INTO	3	0.93
Bachelor	166	51.23
PG diploma	1	0.31
Master	101	31.17
MPhil/PhD	40	12.35
Staff	5	1.54
Other	8	2.47
Total	324	100.00

	Obs.	Mean	St. Dev.	Min	Max
Age	324	23.52	4.82	18	59

Economics	Frequency	Percent
No	264	81.48
Yes	60	18.52
Total	324	100.00

Religion	Frequency	Percent
No religion	163	50.31
Buddhist	14	4.32
Christian	106	32.72
Hindu	18	5.56
Jewish	3	0.93
Muslim	14	4.32
Other	6	1.85
Total	324	100.00

Facebook	Frequency	Percent
No	7	2.16
Yes	317	97.84
Total	324	100.00

Relationship Status	Frequency	Percent
Single	168	51.85
Engaged	6	1.85
In a relationship	128	39.51
Married	22	6.79
Total	324	100.00

Smoker	Frequency	Percent
No	291	89.81
Yes	33	10.19
Total	324	100.00

Country	Frequency	Percent
Azerbaijan	1	0.31
Belgium	1	0.31
Brazil	1	0.31
Brunei Darussalam	1	0.31
Bulgaria	2	0.62
China	22	6.79
Egypt	1	0.31
Estonia	1	0.31
Finland	1	0.31
France	5	1.54
Germany	3	0.93
Greece	2	0.62
Guyana	1	0.31
Hong Kong	7	2.16
India	24	7.41
Indonesia	2	0.62
Iraq	1	0.31
Italy	3	0.93
Japan	2	0.62
Jersey	1	0.31
Jordan	1	0.31
Kenya	2	0.62
Laos	1	0.31
Latvia	2	0.62
Lithuania	3	0.93
Malaysia	3	0.93
Mexico	5	1.54
Nepal	1	0.31
Netherlands	2	0.62
Nigeria	8	2.47
Pakistan	2	0.62
Palestine	1	0.31
Philippines	1	0.31
Poland	3	0.93
Romania	2	0.62
Russian Federation	1	0.31
Slovenia	1	0.31
Spain	2	0.62
Sri Lanka	2	0.62
Switzerland	1	0.31
Taiwan	1	0.31
Thailand	4	1.23
Turkey	2	0.62
Ukraine	1	0.31
United Kingdom	157	48.46
United States	9	2.78
Vietnam	22	6.79
Zambia	1	0.31
Zimbabwe	1	0.31
Total	324	100.00

C. Background of Singled Out Subjects and Authority Subjects

Type	Treatment	Gender	Age	Degree	Country	Economics	Religion	Facebook	Relationship Status	Smoker
Least Preferred	BS	Male	20	Bachelor	Thailand	YES	No Religion	YES	Single	NO
Least Preferred	BS	Male	18	Bachelor	UK	NO	Christian	YES	Single	NO
Least Preferred	BS	Male	33	MPhil/PhD	Mexico	NO	Other	YES	Married	NO
Least Preferred	BS	Female	28	MPhil/PhD	France	NO	Christian	YES	Relationship	NO
Least Preferred	BS	Male	20	Bachelor	Laos	YES	Buddhist	YES	Relationship	NO
Least Preferred	BS	Male	25	MPhil/PhD	Palestine	YES	Muslim	YES	Single	NO
Least Preferred	BS	Female	23	Master	Vietnam	YES	No Religion	YES	Single	NO
Least Preferred	BS	Male	21	Bachelor	UK	NO	Christian	YES	Single	YES
Least Preferred	PIBS	Female	19	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Least Preferred	PIBS	Female	26	Master	Hong Kong	NO	No Religion	YES	Relationship	NO
Least Preferred	PIBS	Female	21	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Least Preferred	PIBS	Male	18	Bachelor	UK	NO	Christian	YES	Relationship	YES
Least Preferred	PIBS	Male	18	Bachelor	Greece	NO	No Religion	YES	Single	NO
Least Preferred	PIBS	Male	22	Bachelor	UK	NO	No Religion	YES	Single	YES
Least Preferred	PIBS	Female	21	Bachelor	UK	NO	Christian	YES	Relationship	NO
Most Preferred	GS	Male	20	Bachelor	UK	YES	No Religion	YES	Relationship	NO
Most Preferred	GS	Female	20	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Most Preferred	GS	Male	18	Bachelor	UK	NO	Christian	YES	Single	YES
Most Preferred	GS	Female	32	MPhil/PhD	UK	NO	No Religion	YES	Married	NO
Most Preferred	GS	Female	22	Bachelor	UK	NO	No Religion	YES	Relationship	YES
Most Preferred	GS	Female	27	Master	Sri Lanka	NO	Buddhist	YES	Relationship	NO
Most Preferred	GS	Female	21	Bachelor	UK	YES	No Religion	YES	Relationship	NO
Most Preferred	GS	Female	18	Bachelor	UK	NO	Christian	YES	Relationship	NO
Most Preferred	PIGS	Male	30	MPhil/PhD	Pakistan	NO	Muslim	YES	Married	NO
Most Preferred	PIGS	Male	23	Master	UK	NO	No Religion	YES	Relationship	NO
Most Preferred	PIGS	Female	19	Bachelor	UK	NO	No Religion	YES	Relationship	YES
Most Preferred	PIGS	Female	21	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Most Preferred	PIGS	Female	23	Master	Vietnam	NO	No Religion	YES	Single	NO
Most Preferred	PIGS	Male	25	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Most Preferred	PIGS	Female	33	Master	UK	NO	Christian	NO	Single	NO

Type	Treatment	Gender	Age	Degree	Country	Economics	Religion	Facebook	Relationship Status	Smoker
Random	RS	Male	37	MPhil/PhD	Mexico	NO	Christian	YES	Married	NO
Random	RS	Female	20	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Random	RS	Female	19	Bachelor	UK	NO	Christian	YES	Relationship	NO
Random	RS	Female	19	Bachelor	Brazil	YES	Christian	YES	Single	NO
Random	RS	Male	22	Master	UK	NO	No Religion	YES	Single	NO
Random	RS	Male	20	Bachelor	Netherlands	NO	Muslim	YES	Single	NO
Random	RS	Male	21	Bachelor	UK	NO	No Religion	YES	Single	NO
Random	RS	Female	21	Bachelor	UK	NO	Christian	YES	Relationship	NO
Authority	ABS	Female	22	Master	India	NO	Hindu	YES	Relationship	NO
Authority	ABS	Male	22	Bachelor	UK	NO	No Religion	YES	Relationship	NO
Authority	ABS	Male	29	Master	Thailand	YES	Buddhist	YES	Single	NO
Authority	ABS	Female	26	MPhil/PhD	France	NO	No Religion	YES	Relationship	YES
Authority	ABS	Female	28	Master	India	NO	Hindu	YES	Married	NO
Authority	ABS	Female	24	Master	Nigeria	NO	Christian	YES	Relationship	NO
Authority	ABS	Female	24	Master	China	NO	No Religion	YES	Single	NO
Authority	ABS	Female	21	Bachelor	UK	NO	No Religion	YES	Single	NO
Least Preferred	ABS	Male	28	MPhil/PhD	Italy	NO	No Religion	YES	Relationship	NO
Least Preferred	ABS	Female	24	Master	Greece	YES	No Religion	YES	Single	NO
Least Preferred	ABS	Female	22	Master	India	NO	Hindu	YES	Engaged	NO
Least Preferred	ABS	Male	25	Master	Japan	YES	No Religion	YES	Married	NO
Least Preferred	ABS	Female	23	Master	China	YES	No Religion	NO	Relationship	NO
Least Preferred	ABS	Male	19	Bachelor	Estonia	NO	Other	YES	Single	NO
Least Preferred	ABS	Female	43	Master	Kenya	YES	Christian	YES	Married	NO
Least Preferred	ABS	Female	22	Master	UK	NO	Christian	YES	Single	NO

D. Data from the Ranking Phase

We first consider the data on the individual rankings. The econometric model that we use is the rank-ordered logit model (see Beggs *et al*., 1981; Hausman and Ruud, 1987), which is a generalization of the conditional logit regression model introduced by McFadden (1974). This model allows us to analyze the data that we obtained by asking subjects to rank the other people in the lab given some personal details of these people.[40]

For each subject we collected five observations. Each observation is a rank (from 1 to 5) on one of the other five participants of an experimental session.[41] Therefore, we have overall 1620 observations. We ran two regressions, using as dependent variable the ranks given by each subject.[42] Table 1D shows the results of the estimation. The first regression (Regression A) uses as explanatory variables only alternative-specific variables, i.e. variables that describe the individuals who had been ranked by the subjects. These independent variables are age, gender (=1 for men), economics background (=1 if applicable), nationality (Uk=1 for UK subjects), religion (Christian = 1 for Christian subjects, and Muslim = 1 for Muslim subjects), university status (PhD = 1, for PhD students), whether the individual smokes or not (Smoker = 1 for smokers), and relationship status (Single = 1 for subjects who were not in a relationship or were unmarried).

The second regression (Regression B) also includes interactions of observation-specific variables with case-specific variables. In particular, we

---

[40] One limitation of this approach is that the rank-ordered logit model assumes independency between the relative preference for a subject j over a subject k and the current choice set (Allison and Christakis, 1994). Although this assumption might be somewhat implausible, we cannot relax it without incurring difficult problems of either computation or identification; given this limitation, our results should be more correctly interpreted as an approximation of a more complex phenomenon (Allison and Christakis, 1994). Note also that this problem of interpretation arises generally for all multinomial logistic models.

[41] Alternatives are ranked from "most preferred" to "least preferred". Therefore, $rank_i = 1$ corresponds to the most preferred match for subject *i*, $rank_i = 2$ to the second most preferred match for subject *i*, and so on. In other words, the variable *rank* used here corresponds to the actual rank by "attractiveness", and not the points assigned by the computer (5 points for most preferred, 4 for second most preferred, etc.). Note that the rank-ordered logit model may give different estimated coefficients between a specification by attractiveness and one by unattractiveness. Therefore, we also tried a rank-ordered logit model by reversing the rank order. However, the results did not differ.

[42] The model takes into account the fact that all observations related to one individual, i.e. each of the five ranked alternatives, are linked together.

create the variable Single×SINGLE, which takes value 1 when both the individual who ranked and the individual who was ranked were single, 0 otherwise; the variable Gender×GENDER which takes value 1 when both individuals were male, 0 otherwise; the variable Uk×UK which takes value 1 when both individuals were from UK; the variable Economics×ECONOMICS which takes value 1 when both individuals were students in Economics; PhD×PHD which takes value 1 when both individuals were PhD students; Smoker×SMOKER which takes value 1 when both individuals were smokers; Christian×CHRISTIAN which takes value 1 when both individuals were Christian; and the variable Age×AGE which is an interaction of two continuous variables, the age of the subjects who ranked and the age of the ranked subject[43]. The interaction variables allow us to control whether different socioeconomic characteristics of those who ranked, relatively to the socioeconomic characteristics of those who had been ranked, affected the rank decision. A comparison of the log likelihoods of the two models suggests that including interaction variables among the explanatory variables improves significantly the fit of the model (likelihood ratio test, p = 0.001). Hence, different subjects appear to have different preferences regarding the rank to allocate to other subjects.

Result A1. Focusing on variables which have significant coefficients (at the 5% level) in regression A, subjects preferred to be matched with English female participants, whose educational background is not in Economics, and who do not smoke.

This result provides a basis of comparison for the next results. In particular, if we move to regression B, we obtain the following results.

Result A2. Female subjects were more likely to assign lower ranks to male subjects than female (the β of Gender is significantly negative), while male subjects did not seem to discriminate between male and female subjects (the β of Gender×GENDER is not significant). This explains why female subjects were generally more preferred than male.

[43] In the regression, we did not include interaction variables for which we had few observations or we did not have any *a priori* belief about their importance in explaining the dependent variable.

Table 1D: Rank-ordered logit model on individual rankings

	Regression A			Regression B		
	b	se	p	b	se	p
Gender	-0.168**	0.07	0.014	-0.201**	0.09	0.029
Age	-0.005	0.01	0.511	-0.110***	0.05	0.014
Economics	-0.289***	0.09	0.002	-0.300***	0.10	0.003
Uk	0.156**	0.08	0.047	0.234**	0.11	0.035
PhD	0.100	0.12	0.397	-0.005	0.13	0.970
Christian	0.069	0.07	0.352	-0.046	0.09	0.605
Muslim	-0.298	0.18	0.098	-0.289	0.18	0.112
Single	-0.101	0.07	0.166	-0.274***	0.11	0.010
Smoker	-0.286**	0.12	0.013	-0.381***	0.12	0.002
Single×SINGLE				0.302**	0.14	0.030
Gender×GENDER				0.094	0.13	0.482
Age×AGE				0.004**	0.00	0.017
Uk×UK				-0.137	0.15	0.346
Economics×ECONOMICS				-0.057	0.25	0.819
PhD×PHD				0.922**	0.36	0.011
Smoker×SMOKER				0.794**	0.38	0.036
Christian×CHRISTIAN				0.364**	0.16	0.026
Obs	1620			1620		
Log Likelihood	-1530.359			-1517.503		
Prob > F	0.000			0.000		

* p<0.1, **p<0.05, *** p<0.01, **** p<0.001. *Notes*: case-specific variables are in small caps; alternative-specific variables are written in lowercase letters.

Result A3. The older a subject was the less she or he preferred young people compared to old.

This effect is captured by the predictor variables Age (−) and Age×AGE (+). In particular, the marginal effect of a subject k 's Age on the utility function of a subject j is conditional to the value of AGE_J. The utility function can be written as $U_j = \left(-0.110 + 0.004 \times \text{AGE}_\text{J}\right) \times Age_k$. For older subjects (AGE_J > 27), the marginal utility becomes positive, whereas for more young subjects (AGE_J < 27), it is negative. This implies that older subjects were more likely to assign higher ranks to older subjects, while more young subjects were more likely to assign higher ranks to more young subjects.

Result A4. Subjects who did not study Economics preferred to be matched with subjects with no background in Economics ($\beta_{\text{Economics}}$ is significantly negative). Instead, subjects who studied Economics did not seem to take into account in their rank decision whether a co-participant had an Economic background or not.

Result A5. Subjects who were not from UK were more likely to assign higher ranks to UK subjects than non UK (the β of UK is significantly positive).

Result A6. There is evidence that PhD students preferred to be matched with other PhD students (the β of PhD×PHD is significant and positive).

Result A7. With respect to religious affiliation, Christians showed a significant preference for subjects of their same religious affiliation (the β of Christian×CHRISTIAN is significant and positive).

Result A8. Subjects who were in a relationship or were married assigned low ranks to single subjects (β_{Single} is significantly negative), whereas single subjects more likely preferred to be matched with other single individuals (the β of Single×SINGLE is significant and positive).

Result A9. Subjects who did not smoke did not prefer to be matched with smokers ($\beta_{Smoker} < 0$), while the latter were more likely to assign high ranks to smokers ($\beta_{Smoker \times SMOKER} > 0$) than non smokers.

Most of the results are not surprising since they seem to reflect phenomena of the real world or behavioral attitudes of experimental subjects which have been widely documented.[44] For instance, most of our results seem to be consistent with the socio-psychological literature on group identity. In particular, people generally prefer to interact with individuals who share similar socioeconomic characteristics, whereas they might dislike people who are different. This might explain why Christians, PhD students, female subjects, and singles gave high ranks to fellow members of the same social group or why old subjects tended to prefer old individuals over young and *vice versa*. Some of the evidence might also reflect sociological patterns which are explained in the literature on social

[44] For example, Holm and Engseld (2005) showed that female subjects are more popular than male.

exclusion and stigmatization. For example, evidence in social psychology showed that singles (e.g. DePaulo and Morris, 2006) and smokers (e.g. Goldstein, 1991) are usually stereotyped and discriminated in societies. Furthermore, we cannot exclude the possibility that some subjects ranked their co-participants according to some strategic criteria, in particular, if they had some expectations that the ranking would have an effect on the experimental tasks. For instance, women are usually considered more cooperative and socially sensitive (e.g. Eagly, 1995). This belief might have motivated some female participants to allocate higher ranks to female than male subjects. Similarly, some participants might have believed that students in Economics are greedier and less cooperative, and, therefore, had been reluctant to assign high ranks to them. Finally, we cannot completely exclude the possibility that some subjects who shared similar characteristics knew each other from outside the lab and, therefore, assigned higher rank to themselves. However, we believe that this possibility is very unlikely. First, we ran sessions with only 6 players in order to minimize the probability that among the 6 subjects there were acquaintances. Second and most importantly, we recruited subjects using the software ORSEE (Greiner, 2004), which ensured a systematic randomization of the participants**.** In particular, we randomly recruit students from the whole university, and, therefore, it was very unlikely that, for instance, two PhD students from the same department participated in the same session. Third, interactions between participants were minimized before, during and after the experiment. Hence, if two subjects knew each other, they could identify themselves only from the characteristics displayed in the computer screen during the ranking phase, and, unless for certain outstanding cases, it may be not have been an easy task to do if there were other participants with similar characteristics.

References Not in Chapter

Allison, P. D. and N. A. Christakis. 1994. "Logit-Models for Sets of Ranked Items." *Sociological Methodology*, 24(1), 199-228.

Beggs, S.; S. Cardell and J. Hausman. 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics*, 17(1), 1-19.

Eagly, A. H. 1995. "The Science and Politics of Comparing Women and Men." *American Psychologist*, 50(3), 145-58.

Goldstein, J. 1991. "The Stigmatization of Smokers - an Empirical-Investigation." *Journal of Drug Education*, 21(2), 167-82.Holm, H. and P. Engseld. 2005. "Choosing Bargaining Partners - an Experimental Study on the Impact of Information About Income, Status and Gender." *Experimental Economics*, 8(3), 183-216.

McFadden, D. L. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior," P. Zarembka, *Frontiers in Econometrics.* New York: Academic Press, 105–42.

E. Test for random sampling of the least and most preferred subjects

The act of choosing singled out subjects is in itself non-random. Therefore, the least preferred subjects might have behaved differently because of the specific characteristics that made them the least preferred subjects. One way of controlling for this problem is to conduct a logistic regression analysis in which the dichotomous dependent variable – whether or not a subject was considered by the whole group the least preferred match – is modelled as a function of the socio-economic characteristics of the subjects.[45] The model also accounts for the session-level clustered nature of the sample. This analysis allows us to test whether the characteristics of the consensually least preferred subjects significantly differ from those of the other subjects who were not consensually least preferred. Table 1E presents the results of the regression.[46]

Table 1E: Logistic regression analysis (least preferred subjects)

	Regression 1		
	b	se	p
UK	-0.198	0.59	0.737
PhD	2.425**	0.94	0.010
Christian	-0.499	0.66	0.450
Muslim	-0.215	1.30	0.869
Single	-0.126	0.54	0.814
Age	-0.270	0.18	0.128
Economics	0.964	0.78	0.215
Gender	-0.173	0.61	0.777
Smoker	1.665	1.14	0.144
Constant	4.196	3.90	0.282
Obs	90		
R-sqr	0.135		
Prob > F	0.177		

Logistic regression with clustered robust standard errors. * p < 0.1, ** p<0.05,* ** p<0.01, **** p<0.001

[45] We were unable to obtain logistic model estimates of the impact of 'India', and other marginal variables (e.g. other types of religious affiliation) because there were not enough occurrences of these attributes among the participants and they characterized either only singled out or non-singled out subjects (e.g. there were not Indian subjects among the least preferred subjects). Therefore, we did not include these variables in the regression.

[46] Data correspond to the socio-economic characteristics of the participants of the BS and PIBS treatments (i.e. treatments where participants selected the consensually least preferred subject). In the regression, we treat each individual as the unit of observation.

In the regression, the coefficient of PhD is statistically significantly positive. None of the other variables appear to be significant. Therefore, we can conclude that PhD students were more likely to be consensually selected as the least preferred match. This evidence indicates that the least preferred subjects might have behaved differently because they were more likely PhD students. In the regression analysis of the chapter, we control for that by including a dummy for PhD students, which, however, results statistically insignificant. We also tried regressions where we included an interaction term between the dummy for PhD students and the dummy for least preferred subjects. The coefficient of this interaction variable was not statistically significant. Finally, the results of the regressions in the chapter do not change if we drop the observations corresponding to PhD students selected as the least preferred match (3 observations).

A similar analysis can be conducted for the consensually most preferred subjects. In this case, the dichotomous dependent variable is whether or not a subject was considered the most preferred match by the whole group. Table 2E presents the results of the logistic regression.[47]

Table 2E: Logistic regression analysis (most preferred subjects)

	Regression 2		
	b	se	p
UK	2.264**	0.88	0.010
PhD	0.448	1.01	0.658
Christian	-0.319	1.05	0.761
Muslim	0.656	1.54	0.671
Single	-1.547**	0.71	0.029
Age	0.073	0.09	0.406
Economics	0.159	1.05	0.880
Gender	-1.066	0.75	0.158
Smoker	0.069	0.93	0.940
Constant	-3.745	2.03	0.065
Obs	90		
R-sqr	0.231		
Prob > F	0.052		

Logistic regression with clustered robust standard errors. * p < 0.1, ** p<0.05,* ** p<0.01, **** p<0.001

[47] Data correspond to the socio-economic characteristics of the participants of the GS and PIGS treatments (i.e. treatments where participants selected the consensually most preferred subject). In the regression, we treat each individual as the unit of observation.

The regression analysis suggests that UK subjects were more likely to be consensually selected as the most preferred match, whereas single subjects were less likely to be selected as the most preferred match, other things being equal. This evidence indicates that the most preferred subjects might have behaved differently because they were more likely UK and less likely single. In the regression analysis of the chapter, we control for that by including a dummy for UK subjects, and one for single subjects. The UK dummy is statistically insignificant in all regressions, implying that it made no difference. On the other hand, single subjects were more trusting and trustworthy towards non-singled out subjects according to Tables 3 and 4 of the chapter. However, we have not found any evidence, and made no claim, that high status subjects are less trusting and trustworthy towards non-singled out subjects, and so this potential individual characteristic effect does not turn out to be relevant. It is, of course, anyway controlled for in the regression analysis. We also tried regressions where we included an interaction term between the dummy for UK subjects and the dummy for most preferred subjects, and an interaction variable between the dummy for single subjects and the dummy for most preferred subjects. The coefficients of these interaction variables were not statistically significant. Hence, we can conclude that the behavior of the least preferred subjects is not explained by their specific individual characteristics.

Overall, while the sampling of the individual rankings is not random (as per section D of this appendix), given the aggregation procedure and the heterogeneity of the rankings across individuals, we found that the socio-economic characteristics of both the least and most preferred subjects were mostly random or did not explain the results on the effects of singling out in the trust game.

F. Time trend of key variables

We also looked at how giving and return rates evolved over time. Tables 1F and 2F show average giving and return rates per round. Giving rates only slightly decreased over time, but the trend is not statistically significant (Spearman ρ = -0.044, p = 0.658); return rates statistically significantly decreased over time (ρ = -0.167, p < 0.001). Specifically, giving and return rates from singled out subjects did not show any significant time trend (ρ = -0.084 and ρ = -0.097 respectively, p > 0.1). The same applies to giving rates (ρ = -0.067, p = 0.392) and return rates (ρ = -0.194, p < 0.001) specifically of non singled out subjects when interacting with other non-singled out subjects. Finally, giving rates from non singled out subjects to singled out subjects significantly increased over time (ρ = 0.150, p = 0.008), whereas return rates from non singled out subjects to singled out subjects did not exhibit any time trend (ρ = 0.076, p = 0.701).[48]

Table 1F: Giving and return rates to singled out and non-singled out over time

	1	**2**	**3**	**4**	**Tot.**
Giving rate	0.393	0.406	0.395	0.385	0.394
To Singled Out	0.240	0.325	0.454	0.328	0.336
To non-Singled Out*	0.421	0.437	0.405	0.394	0.414
Return Rate	0.231	0.213	0.150	0.167	0.191
To Singled Out	0.240	0.131	0.077	0.216	0.165
To non-Singled Out*	0.253	0.214	0.176	0.171	0.206

Notes: *Giving/return rate of non-singled out subjects to non-singled out subjects.

Table 2F: Giving and return rates from singled out, non-singled out, and authority over time

	1	**2**	**3**	**4**	**Tot.**
Giving rate	0.393	0.406	0.395	0.385	0.394
From Singled Out	0.428	0.310	0.326	0.340	0.349
From non-Singled Out*	0.421	0.437	0.405	0.395	0.414
From Authority	0.422	0.358	0.389	0.097	0.331
Return Rate	0.231	0.213	0.150	0.167	0.191
From Singled Out	0.154	0.281	0.104	0.141	0.164
From non-Singled Out*	0.253	0.214	0.176	0.171	0.206
From Authority	0.083	0.157	0.049	0.067	0.078

Notes: *Giving/return rate of non-singled out subjects to non-singled out subjects.

[48] In all these tests, we controlled for the non-independence (at session and individual level) of the observations.

G. Additional regression analysis

In this section, we report some of the regressions of the chapter employing giving and return rate per period as the unit of observation. In these regressions we employ different estimation techniques. While these regressions are provided for the convenience of the reader and the results broadly replicate those in the chapter, they do not fully control for the non-independence of the observations or the censoring nature of the data, leaving the regressions reported in the chapter as a better estimation option. In tables G1-G12 we report the results of the regressions on return rates to non-singled out and singled out subjects.[49] In table G13, we report the results of the regressions on giving rates to non-singled out subjects, controlling for the trustworthiness of the co-participant in the previous round and its interaction with a singled out co-participant.[50] Note that in the regressions of table G13 we lose half of the observations since we include a lag. In addition, we cannot avoid an endogenity problem.

[49] We also conducted similar regressions for the giving rate. We do not report them here as we did not find any treatment effects for trust and they do not add anything to the regressions of the chapter.

[50] Note in fact that the trustworthiness of the co-participant in the previous round (and its interaction with a singled out co-participant) may influence the trusting decision in the current round.

Table G1: Tobit Random Effects at subject level on Return Rates to non-singled out subjects

	Regression G1		Regression G2	
	b	p	b	p
Trust Rate as Trustee	0.395****	0	0.394****	0
All Treatments			-0.08*	0.063
BS+PIBS	-0.118**	0.019		
GS+PIGS	-0.06	0.225		
RS	-0.075	0.204		
ABS	-0.048	0.456		
Singled out in RS	0.107	0.334	0.114	0.27
Singled out in ABS	0.066	0.522	0.098	0.29
Singled out in BS and PIBS	-0.167**	0.039	-0.204**	0.008
Singled out in GS and PIGS	0.027	0.721	0.047	0.514
Authority	-0.194*	0.071	-0.164*	0.092
Risk Aversion	0	0.964	-0.002	0.794
SDS17 Score	-0.004	0.458	-0.004	0.456
RSE Score	-0.007*	0.067	-0.007*	0.075
Authority×SDS17	0.038	0.333	0.037	0.355
Authority×RSE	0.007	0.645	0.007	0.641
Authority×Risk Aversion	-0.006	0.887	-0.004	0.924
Age	0.002	0.69	0.002	0.65
Gender	0.016	0.622	0.012	0.7
UK	-0.068*	0.068	-0.069*	0.064
India	-0.04	0.498	-0.035	0.55
Christian	-0.029	0.399	-0.035	0.302
Muslim	-0.109	0.225	-0.101	0.258
Single	0.063*	0.056	0.066**	0.042
Economics	-0.101**	0.019	-0.098**	0.023
Smoker	-0.007	0.884	-0.001	0.985
PhD	-0.004	0.942	-0.006	0.909
Constant	0.141	0.193	0.136	0.198
Obs	484		484	
Prob > F	0		0	
ll	-161.86		-162.95	

Table G2: Tobit Random Effects at session level on Return Rates to non-singled out subjects

	Regression G3		Regression G4	
	b	p	b	p
Trust Rate as Trustee	0.391****	0	0.389***	0
All Treatments			-0.072	0.137
BS+PIBS	-0.116**	0.036		
GS+PIGS	-0.049	0.365		
RS	-0.075	0.248		
ABS	-0.028	0.684		
Singled out in RS	0.109	0.269	0.11	0.244
Singled out in ABS	0.021	0.817	0.052	0.53
Singled out in BS and PIBS	-0.171**	0.018	-0.198***	0.005
Singled out in GS and PIGS	0.023	0.721	0.038	0.551
Authority	-0.218**	0.021	-0.188**	0.032
Risk Aversion	0	0.941	-0.002	0.786
SDS17 Score	-0.002	0.709	-0.002	0.712
RSE Score	-0.006*	0.068	-0.006*	0.074
Authority×SDS17	0.031	0.379	0.029	0.405
Authority×RSE	0.009	0.493	0.009	0.483
Authority×Risk Aversion	0.003	0.935	0.006	0.876
Age	0.001	0.848	0.001	0.769
Gender	0.017	0.535	0.015	0.601
UK	-0.06*	0.076	-0.062*	0.068
India	-0.037	0.473	-0.035	0.508
Christian	-0.031	0.321	-0.035	0.252
Muslim	-0.115	0.157	-0.109	0.176
Single	0.048	0.107	0.05*	0.09
Economics	-0.102***	0.009	-0.100**	0.011
Smoker	-0.027	0.558	-0.022	0.624
PhD	-0.01	0.833	-0.014	0.768
Constant	0.163	0.107	0.155	0.121
Obs	484		484	
Prob > F	0		0	
ll	-170.63		-171.95	

Table G3: Tobit Random Effects at subject level on Return Rates to non-singled out subjects controlling for decision period

	Regression G5		Regression G6	
	b	p	b	p
Trust Rate as Trustee	0.411****	0	0.411****	0
All Treatments			-0.085**	0.045
BS+PIBS	-0.122**	0.015		
GS+PIGS	-0.066	0.18		
RS	-0.082	0.16		
ABS	-0.055	0.388		
Singled out in RS	0.126	0.252	0.131	0.201
Singled out in ABS	0.072	0.484	0.102	0.267
Singled out in BS and PIBS	-0.159**	0.047	-0.194**	0.011
Singled out in GS and PIGS	0.026	0.728	0.046	0.522
Authority	-0.201*	0.062	-0.172*	0.077
Risk Aversion	0	0.981	-0.001	0.858
SDS17 Score	-0.004	0.473	-0.004	0.473
RSE Score	-0.007*	0.069	-0.006*	0.077
Authority×SDS17	0.04	0.309	0.039	0.329
Authority×RSE	0.007	0.622	0.007	0.618
Authority×Risk Aversion	-0.01	0.811	-0.008	0.845
Age	0.002	0.637	0.002	0.596
Gender	0.013	0.667	0.01	0.744
UK	-0.069*	0.061	-0.071*	0.058
India	-0.047	0.425	-0.042	0.47
Christian	-0.029	0.382	-0.035	0.29
Muslim	-0.114	0.201	-0.106	0.232
Single	0.058*	0.075	0.062*	0.057
Economics	-0.104**	0.014	-0.102**	0.018
Smoker	-0.003	0.95	0.003	0.951
PhD	-0.006	0.904	-0.009	0.866
Period	-0.083****	0	-0.084****	0
Constant	0.275**	0.015	0.271**	0.015
Obs	484		484	
Prob > F	0		0	
ll	-155.01		-156	

Table G4: Tobit Random Effects at session level on Return Rates to non-singled out subjects controlling for decision period

	Regression G7		Regression G8	
	b	p	b	p
Trust Rate as Trustee	0.406****	0	0.404****	0
All Treatments			-0.078*	0.099
BS+PIBS	-0.119**	0.026		
GS+PIGS	-0.055	0.299		
RS	-0.082	0.193		
ABS	-0.036	0.599		
Singled out in RS	0.127	0.196	0.127	0.176
Singled out in ABS	0.026	0.774	0.056	0.49
Singled out in BS and PIBS	-0.163**	0.021	-0.190***	0.006
Singled out in GS and PIGS	0.022	0.735	0.037	0.555
Authority	-0.224**	0.018	-0.194**	0.026
Risk Aversion	0	0.987	-0.001	0.834
SDS17 Score	-0.002	0.7	-0.002	0.704
RSE Score	-0.006*	0.07	-0.006*	0.075
Authority×SDS17	0.032	0.351	0.03	0.378
Authority×RSE	0.01	0.463	0.01	0.451
Authority×Risk Aversion	-0.002	0.953	0	0.99
Age	0.001	0.77	0.001	0.69
Gender	0.015	0.603	0.012	0.675
UK	-0.061*	0.07	-0.062*	0.063
India	-0.043	0.406	-0.04	0.438
Christian	-0.033	0.286	-0.037	0.222
Muslim	-0.123	0.127	-0.117	0.145
Single	0.046	0.116	0.049*	0.096
Economics	-0.105***	0.007	-0.102***	0.008
Smoker	-0.021	0.644	-0.017	0.715
PhD	-0.011	0.822	-0.015	0.754
Period	-0.084***	0.001	-0.085***	0.001
Constant	0.295***	0.006	0.288***	0.007
Obs	484		484	
Prob > F	0		0	
ll	-165.52		-166.82	

Table G5: Multilevel Random Effects on Return Rates to non-singled out subjects

	Regression G9		Regression G10	
	b	p	b	p
Trust Rate as Trustee	0.243****	0	0.242****	0
All Treatments			-0.057**	0.045
BS+PIBS	-0.086**	0.01		
GS+PIGS	-0.049	0.137		
RS	-0.058	0.144		
ABS	-0.013	0.765		
Singled out in RS	0.071	0.346	0.074	0.293
Singled out in ABS	0.026	0.71	0.069	0.266
Singled out in BS and PIBS	-0.094*	0.058	-0.121***	0.009
Singled out in GS and PIGS	0.041	0.404	0.05	0.289
Authority	-0.163**	0.019	-0.119*	0.057
Risk Aversion	-0.002	0.727	-0.003	0.5
SDS17 Score	-0.003	0.438	-0.003	0.416
RSE Score	-0.005*	0.065	-0.004*	0.074
Authority×SDS17	0.029	0.282	0.028	0.307
Authority×RSE	0.003	0.761	0.003	0.758
Authority×Risk Aversion	-0.003	0.907	-0.001	0.96
Age	0.001	0.616	0.002	0.463
Gender	0.021	0.322	0.017	0.408
UK	-0.031	0.211	-0.032	0.199
India	-0.043	0.281	-0.039	0.325
Christian	-0.014	0.53	-0.018	0.414
Muslim	-0.067	0.259	-0.067	0.254
Single	0.038*	0.08	0.04*	0.062
Economics	-0.054**	0.048	-0.051*	0.065
Smoker	0.001	0.967	0.008	0.816
PhD	-0.006	0.859	-0.009	0.802
Constant	0.197***	0.006	0.184***	0.009
Obs	484		484	
Prob > F	0		0	
ll	56.349		61.945	

Table G6: Multilevel Random Effects on Return Rates to non-singled out subjects controlling for decision period

	Regression G11		Regression G12	
	b	p	b	p
Trust Rate as Trustee	0.252****	0	0.251****	0
All Treatments			-0.060**	0.036
BS+PIBS	-0.088***	0.009		
GS+PIGS	-0.052	0.115		
RS	-0.061	0.121		
ABS	-0.015	0.717		
Singled out in RS	0.081	0.282	0.083	0.238
Singled out in ABS	0.025	0.711	0.068	0.267
Singled out in BS and PIBS	-0.093*	0.059	-0.119**	0.01
Singled out in GS and PIGS	0.042	0.4	0.049	0.287
Authority	-0.166**	0.017	-0.123**	0.049
Risk Aversion	-0.001	0.752	-0.003	0.526
SDS17 Score	-0.003	0.462	-0.003	0.441
RSE Score	-0.004*	0.068	-0.004*	0.077
Authority×SDS17	0.029	0.279	0.028	0.302
Authority×RSE	0.003	0.75	0.003	0.746
Authority×Risk Aversion	-0.004	0.882	-0.002	0.934
Age	0.002	0.561	0.002	0.412
Gender	0.019	0.361	0.016	0.453
UK	-0.031	0.208	-0.032	0.195
India	-0.045	0.248	-0.042	0.287
Christian	-0.015	0.499	-0.019	0.39
Muslim	-0.069	0.235	-0.07	0.229
Single	0.035*	0.097	0.038*	0.077
Economics	-0.055**	0.043	-0.052*	0.058
Smoker	0.003	0.932	0.009	0.784
PhD	-0.008	0.815	-0.011	0.758
Period	-0.049****	0.001	-0.049****	0.001
Constant	0.274****	0	0.262****	0
Obs	484		484	
Prob > F	0		0	
ll	58.413		64.083	

Table G7: Tobit Random Effects at subject level on Return Rates to singled out subjects

	Regression G13		Regression G14	
	b	p	b	p
Trust Rate as Trustee	0.416****	0	0.416****	0
Asterisk			-0.087*	0.095
BS	-0.017	0.85		
GS	-0.001	0.995		
PIBS	-0.111**	0.026	-0.115**	0.022
PIGS	-0.069	0.158	-0.075	0.129
RS+ABS	-0.153**	0.017		
Risk Aversion	-0.01	0.202	-0.008	0.281
SDS17 Score	0.005	0.401	0.005	0.433
RSE Score	-0.006	0.168	-0.006	0.155
Age	0.009	0.115	0.009	0.109
Gender	0.001	0.97	0.002	0.955
Economics	-0.055	0.256	-0.062	0.202
UK	0.052	0.233	0.046	0.291
India	-0.05	0.469	-0.062	0.372
Christian	-0.07*	0.07	-0.072*	0.065
Muslim	-0.190**	0.033	-0.163*	0.064
Single	0.044	0.263	0.043	0.271
Smoker	0.015	0.772	0.014	0.794
PhD	0.023	0.724	0.02	0.762
Constant	-0.064	0.654	-0.061	0.671
Obs	266		266	
Prob > F	0		0	
ll	-64.168		-81.725	

Table G8: Tobit Random Effects at session level on Return Rates to singled out subjects

	Regression G15		Regression G16	
	b	p	b	p
Trust Rate as Trustee	0.416****	0	0.392****	0
Asterisk			-0.079	0.186
BS	-0.017	0.85		
GS	-0.001	0.995		
PIBS	-0.111**	0.026	-0.109*	0.085
PIGS	-0.069	0.158	-0.074	0.241
RS+ABS	-0.153**	0.017		
Risk Aversion	-0.01	0.202	-0.006	0.495
SDS17 Score	0.005	0.401	0.006	0.338
RSE Score	-0.006	0.168	-0.005	0.203
Age	0.009	0.115	0.008	0.165
Gender	0.001	0.97	-0.006	0.877
Economics	-0.055	0.256	-0.064	0.193
UK	0.052	0.233	0.043	0.328
India	-0.05	0.469	-0.066	0.349
Christian	-0.07*	0.07	-0.061	0.128
Muslim	-0.190**	0.033	-0.147*	0.094
Single	0.044	0.263	0.032	0.42
Smoker	0.015	0.772	-0.005	0.92
PhD	0.023	0.724	0.015	0.819
Constant	-0.064	0.654	-0.031	0.831
Obs	266		266	
Prob > F	0		0	
ll	-74.659		-86.47	

Table G9: Tobit Random Effects at subject level on Return Rates to singled out subjects controlling for decision period

	Regression G15		Regression G16	
	b	p	b	p
Trust Rate as Trustee	0.420****	0	0.419****	0
Asterisk			-0.089*	0.089
BS	-0.017	0.854		
GS	-0.01	0.911		
PIBS	-0.109**	0.028	-0.112**	0.024
PIGS	-0.071	0.148	-0.076	0.12
RS+ABS	-0.151**	0.018		
Risk Aversion	-0.01	0.201	-0.009	0.27
SDS17 Score	0.006	0.377	0.005	0.406
RSE Score	-0.006	0.175	-0.006	0.158
Age	0.009	0.122	0.009	0.114
Gender	-0.001	0.97	0	0.992
Economics	-0.054	0.264	-0.061	0.208
UK	0.052	0.233	0.047	0.285
India	-0.049	0.473	-0.061	0.379
Christian	-0.069*	0.071	-0.071*	0.068
Muslim	-0.188**	0.034	-0.163*	0.064
Single	0.043	0.272	0.043	0.276
Smoker	0.013	0.811	0.011	0.841
PhD	0.023	0.731	0.019	0.77
Period	-0.05	0.145	-0.054	0.121
Constant	0.022	0.886	0.03	0.847
Obs	266		266	
Prob > F	0		0	
ll	290.834		-79.288	

Table G10: Tobit Random Effects at session level on Return Rates to singled out subjects controlling for decision period

	Regression G17		Regression G18	
	b	p	b	p
Trust Rate as Trustee	0.420****	0	0.419****	0
Asterisk			-0.089*	0.089
BS	-0.017	0.854		
GS	-0.01	0.911		
PIBS	-0.109**	0.028	-0.112**	0.024
PIGS	-0.071	0.148	-0.076	0.12
RS+ABS	-0.151**	0.018		
Risk Aversion	-0.01	0.201	-0.009	0.27
SDS17 Score	0.006	0.377	0.005	0.406
RSE Score	-0.006	0.175	-0.006	0.158
Age	0.009	0.122	0.009	0.114
Gender	-0.001	0.97	0	0.992
Economics	-0.054	0.264	-0.061	0.208
UK	0.052	0.233	0.047	0.285
India	-0.049	0.473	-0.061	0.379
Christian	-0.069*	0.071	-0.071*	0.068
Muslim	-0.188**	0.034	-0.163*	0.064
Single	0.043	0.272	0.043	0.276
Smoker	0.013	0.811	0.011	0.841
PhD	0.023	0.731	0.019	0.77
Period	-0.05	0.145	-0.054	0.121
Constant	0.022	0.886	0.03	0.847
Obs	266		266	
Prob > F	0		0	
ll	-74.675		-78.114	

Table G11: Multilevel Random Effects on Return Rates to singled out subjects

	Regression G19		Regression G20	
	b	p	b	p
Trust Rate as Trustee	0.248****	0	0.246****	0
Asterisk			-0.064*	0.077
BS	-0.024	0.7		
GS	-0.008	0.891		
PIBS	-0.080**	0.027	-0.082**	0.023
PIGS	-0.056	0.121	-0.059	0.101
RS+ABS	-0.101**	0.018		
Risk Aversion	-0.007	0.228	-0.006	0.322
SDS17 Score	0.002	0.685	0.002	0.703
RSE Score	-0.004	0.236	-0.004	0.223
Age	0.005	0.247	0.005	0.243
Gender	0.012	0.649	0.012	0.663
Economics	-0.038	0.267	-0.042	0.217
UK	0.043	0.172	0.04	0.201
India	-0.051	0.303	-0.059	0.234
Christian	-0.037	0.179	-0.037	0.177
Muslim	-0.105*	0.094	-0.085	0.167
Single	0.018	0.526	0.017	0.533
Smoker	-0.003	0.944	-0.003	0.946
PhD	0.018	0.711	0.016	0.735
Constant	0.102	0.326	0.106	0.306
Obs	266		266	
Prob > F	0		0	
ll	24.398		26.726	

Table G12: Multilevel Random Effects on Return Rates to singled out subjects controlling for decision period

	Regression G21		Regression G22	
	b	p	b	p
Trust Rate as Trustee	0.250****	0	0.249****	0
Asterisk			-0.065*	0.072
BS	-0.023	0.709		
GS	-0.014	0.815		
PIBS	-0.079**	0.029	-0.081**	0.025
PIGS	-0.056	0.118	-0.059*	0.098
RS+ABS	-0.101**	0.019		
Risk Aversion	-0.007	0.225	-0.006	0.308
SDS17 Score	0.002	0.682	0.002	0.699
RSE Score	-0.003	0.26	-0.004	0.242
Age	0.005	0.266	0.005	0.26
Gender	0.01	0.715	0.01	0.719
Economics	-0.036	0.288	-0.04	0.232
UK	0.042	0.177	0.04	0.203
India	-0.05	0.312	-0.058	0.243
Christian	-0.036	0.185	-0.036	0.183
Muslim	-0.103	0.1	-0.084	0.171
Single	0.017	0.546	0.017	0.547
Smoker	-0.005	0.897	-0.005	0.892
PhD	0.017	0.714	0.016	0.738
Period	-0.035*	0.095	-0.037*	0.084
Constant	0.165	0.136	0.169	0.123
Obs	266		266	
Prob > F	0		0	
ll	22.847		25.27	

Table G13: Tobit on Giving Rates to non-singled out controlling for the trustworthiness of the co-participant in t-1

	Regression G23		Regression G24	
	b	p	b	p
BS+PIBS	0.016	0.876		
GS+PIGS	-0.004	0.968		
RS	-0.024	0.848		
ABS	0.077	0.542		
Singled out in BS and PIBS	-0.062	0.549	-0.054	0.593
Singled out in GS and PIGS	-0.193	0.115	-0.204*	0.066
Singled out in RS	0.182	0.351	0.147	0.475
Singled out in ABS	0.049	0.702	0.117	0.317
Authority	-0.347*	0.078	-0.277	0.115
Risk Aversion	0.01	0.297	0.009	0.328
SDS17 Score	-0.021**	0.022	-0.021**	0.022
RSE Score	0.004	0.466	0.004	0.507
Age	0.005	0.556	0.006	0.485
Gender	0.052	0.293	0.052	0.293
Economics	-0.149**	0.028	-0.148**	0.027
India	-0.125	0.144	-0.121	0.141
UK	0.028	0.649	0.028	0.64
Christian	-0.156***	0.006	-0.154***	0.006
Muslim	-0.294***	0.001	-0.305****	0
Single	0.085*	0.08	0.088*	0.068
Smoker	0.04	0.588	0.043	0.547
PhD	0.079	0.354	0.08	0.354
Return rate (t-1)	0.863****	0	0.847****	0
Matched With Stigma × Return rate (t-1)	0.501	0.394	0.461	0.463
Matched With Stigma	-0.095	0.386	-0.082	0.462
Stigma			0.009	0.924
Constant	0.141	0.535	0.123	0.582
Obs	251		251	
Prob > F	0		0	
ll	-138.12		-138.61	

Chapter 2: On the Robustness of Emotions and Behavior in a Power-to-Take Game Experiment

1. Introduction

This study investigates the role played by the punishment technology in driving the results of power-to-take game (PTTG) experiments, and tests to what extent these results can be attributed to negative emotions such as anger, irritation, and contempt. In addition, it explores whether the experience gained from previous economic experiments and background of the participants affect subjects' emotions and how the latter impact on subjects' behavior in the context of the PTTG.

Over the last few decades, economists have started to pay greater attention to the complexity of emotions on economic scenarios and have been trying to capture the range of possible roles that emotions play in the decision-making process. For instance, emotions have been proposed as an explanation for important economic phenomena such as cooperation (e.g. Frank, 1988; Fehr and Gächter, 2002) and decision-making under risk (e.g. Loewenstein *et al*., 2001), and are seen to have important consequences for many other economic phenomena, such as inter-temporal choices (e.g. Rick and Lowenstein, 2008), competition (e.g. Kräkel, 2008), bidding behavior (e.g. Bosman and Riedl, 2004) and bargaining behavior (e.g. Pillutla and Murninghan, 1996). More recently, the advent of neuroeconomics has further pushed forward the interest of economists on the role played by emotions in the economic decision-making process (for a review on emotions and neuroeconomics, see Phelps, 2009).

An important branch of economic research on emotions has used experiments to study the impact of negative emotions, particularly anger, irritation and contempt, on the decision to punish (Bosman and van Winden, 2002; Bosman *et al*., 2005; Ben-Shakhar *et al*., 2007; Hopfensitz and Reuben, 2009; Joffily *et al*., 2013). This stream of research started with the seminal work of Bosman and van Winden (2002) on the PTTG. In the PTTG, there are two players, the 'take authority' (with income Y_{take}), and the 'responder' (with income Y_{resp}). The game is divided into two stages. In the first stage, the take authority selects a take rate $t \in [0,1]$, which is the proportion of the responder's income that will be transferred to the take authority at the end of the game. In the second stage, the responder chooses a destroy rate $d \in [0,1]$, which is the proportion of Y_{resp} that will be destroyed. Therefore, the payoffs of the game are $(1-t)(1-d)Y_{resp}$ for the responder, and $Y_{take} + t(1-d) Y_{resp}$ for the take authority.

If the subjects are rational profit-maximizing agents, the responder should not destroy if the take rate is less than 1, and should be indifferent between all possible destroy rates if the take rate is 1. Hence, from backward induction, the take authority should select $t = 1 - \varepsilon$, where ε is an infinitesimal positive number. The PTTG can be interpreted as an ultimatum game with continuous opportunities to punish[51] and can describe many economic situations where an agent can take away any part of the endowment of another agent (e.g. taxation, monopolistic pricing and principal-agent relationships) (Bosman and van Winden, 2002).

Most of the literature on the PTTG investigates the role played by negative emotions on responders' behavior through physiological (Ben-Shakhar *et al*., 2007) and self-report measures (Bosman and van Winden, 2002; Bosman *et al*., 2005). Both measures were found to be related to destruction decisions. In particular, participants who experienced intense anger, irritation, and contempt punished their counterparts more often and more severely. This result seems to identify these negative emotions as the main driving force of the punishing behavior in this type of context.[52]

---

[51] It is also worth noting that the endowments are allocated differently in the PTTG compared to the Ultimatum Game. From a traditional game theoretic point of view, this does not matter, but it may have important implications in terms of behavior.

[52] Other well-established findings from the experimental literature on the PTTG show that people appropriate almost 60% of responders' income, while only 20% of the responders destroy income and usually all of it (e.g. Bosman and van Winden, 2002; Reuben and van Winden, 2010). Small differences were observed between an effort treatment – where endowments were earned by doing a preliminary individual real effort task – and a no-effort treatment – where endowments were exogenously given by the experimenter (Bosman *et al*., 2005). A group version of the PTTG – where decisions were made by groups – presented the same results qualitatively as the no-group experiment (Bosman *et al*., 2006). Furthermore, in a three-player version of the PTTG with one take authority and two responders, Reuben and van Winden (2008) showed that responders who knew each other from outside the laboratory punish and coordinate more than strangers. The PTTG has also been used to study the influence of participation (Albert and Mertins, 2008), gender pairing (Sutter *et al*., 2009) and waiting time (Galeotti, 2013) on economic decision-making.

However, this finding may be confounded by the technology of the punishment adopted. In particular, the non-constant 'fine-to-fee' ratio, adopted in the PTTG and defined as "the income reduction for the targeted subject relative to the cost for the subject who requested the punishment" (Casari, 2005:107), may explain all or part of the relationship between punishing behavior and negative emotions observed in previous PTTG studies. In the PTTG, the income reduction for the authority is tdY_{resp}, whereas the cost for the responder to punish is $\mathrm{d}(1-\mathrm{t})\mathrm{Y}_{\mathrm{resp}}$. Therefore, the 'fine-to-fee' ratio is $tdY_{resp}/\mathrm{d}(1-\mathrm{t})\mathrm{Y}_{\mathrm{resp}} = t/1-t$, where t – the proportion of the responder's income that will be transferred to the take authority – is an endogenous and non-constant parameter. This implies that the 'demand' for punishment is higher when t is higher (i.e. the 'fine-to-fee' ratio is higher). In other words, for high take rates the responder has a higher incentive[53] to punish her counterpart, whereas for low take rates the incentive is lower. Hence, the role of negative emotions might be overstated; when the offence is severe subjects experience strong negative emotions, but they punish because punishing is cheaper for increasing offences and not, or not only, because they experience anger, irritation or contempt. It thus becomes important to test the robustness of the results of PTTG experiments against this possible source of confound. If a confound exists, we may need to reconsider the role played by these negative emotions in this type of context.

In order to investigate to what extent the punishment behavior observed in previous studies on the PTTG is explained by the punishment technology rather than negative emotions, we conducted a laboratory experiment using students from our university. We varied the extent to which the punishment technology embedded a variable or a constant 'fine-to-fee' ratio. Emotions were assessed through self-report measures, as in previous studies.[54] Given the variety of cultural backgrounds of the students enrolled at our university, we also ran separate sessions for UK students and

[53] Here the incentive is the cost reduction from punishing per unit of punishment.
[54] For a discussion on the reliability of self-reports in measuring emotions, see Bosman and van Winden (2002) or Hopfensitz and Reuben (2009).

non-UK students.[55] We believe this is particularly important in our experiment for two main reasons. First, we re-created, at least for half of the sessions (i.e. those with UK subjects), a lab environment analogous to the one used in previous PTTG studies where most of the participants had a similar cultural background. Second, we are able to control for and test, in a systematic way and under certain standardized conditions (e.g. location of the laboratory, university training, etc.), whether there are cultural differences in the elicitation and manifestation of emotions – as it has been observed in previous psychological and anthropological literature (for a review, see Mesquita and Frijda, 1992) – when students with a different cultural background are employed.[56] Finally, we collected, in a final questionnaire, the information about experience of our subjects in prior economic experiments, and use it as a control in the data analysis.[57] This is also extremely important as subjects with more experience of the environment and the dynamics of laboratory experiments may be more aware of what they should expect in an economic experiment and, therefore, they might experience less strong emotions and/or be better able to cope with their emotional urges than inexperienced subjects.

To give a brief overview of our results, we find that the bias caused by the punishment technology adopted in the standard PTTG is large and significant. In particular, the punishment and the role played by anger, irritation, and contempt on the decision to punish are inflated, especially for

[55] We recruited subjects using the software ORSEE (Greiner, 2004). In the non-UK sessions, we avoided recruiting students from Western societies. This was because we wanted to ensure the highest degree of cultural separation between UK and non-UK subjects. As a result, apart from one Australian participant, non-UK subjects consisted of Asian, South American, African and East European students.

[56] We cannot rule out the possibility that UK and non-UK students are different self-selected sub-samples of their native populations. However, the main purpose of this exercise is not to test whether UK individuals are in general different from non-UK individuals, but rather to establish whether it is methodologically legitimate to conduct experiments with non-native students when the aim is to study the emotional basis of punishing behavior, especially in the light of the findings of the literature in psychology.

[57] Note that there are generally two kinds of learning relevant to subjects who participate in economic experiments: 'experimental' learning and 'within game' learning (Friedman, 1969). 'Experimental' learning refers to the general experience that subjects acquire by participating in many different experiments not necessarily linked to each other. 'Within game' learning refers to a type of experience acquired in the particular experimental set-up or specific game-theoretical framework under examination. In our experiment, we controlled for and referred to the first type of learning, since none of our subjects had ever participated in a PTTG experiment before.

high take rates. When we look at the cultural background of the subjects, we do not find any difference between UK and non-UK subjects in the way in which emotions are experienced and impact on the decision to punish. Finally, we find that more experienced subjects punish less often and less severely when they experience increasing contempt compared to less experienced subjects.

The chapter is organized as follows. Section 2 presents the experimental design, Section 3 describes the theoretical background and the behavioral hypotheses, Section 4 reports the results, then Section 5 discusses the results and concludes.

2. Experimental Design

We conducted the experiment between March and September 2012 at the University of East Anglia, with 282 students participating in the experiment over many sessions.[58] Each session lasted on average 50 minutes. No subject was allowed to participate in more than one session. Subjects received a show-up fee of £5 and earned on average £9.41 (around 15 US dollars). In order to ensure the greatest comparability of our experiment with previous literature, we tried to replicate, as close as possible, the experimental procedures adopted in previous PTTG experiments. In particular, we (a) conducted a paper-and-pencil experiment, (b) employed the same instructions, exercises, examples and procedures as previous PTTG studies,[59] (c) avoided any particular or suggestive terminology during the sessions, such as ‘take authority’ or ‘take rate’, (d) adopted the same double blind procedure of Bosman and van Winden (2002) for the payments, and (e) assessed emotions on a 7-point Likert scale via self-reports after each subject learned about the decision of their counterpart. More details about the experimental procedure are provided below.

Upon arrival, each subject was randomly assigned the role of participant A (take authority) or participant B (responder) by drawing a

[58] Details of the socioeconomic characteristics of the subjects and experimental instructions are in the appendix.

[59] Minor adjustments to the original instructions were made to fit them to our laboratory routines, monetary payments and comparability of our treatments.

letter from an urn, then randomly allocated to a computer workstation which was isolated from other workstations via partitions. Then the instructions were distributed and read aloud to provide common information to the subjects. Two individual computerized exercises followed in order to check the subjects' understanding of these instructions. Clarifications were individually provided to subjects with incorrect answers. After completion of these exercises, each participant A was randomly matched with a participant B by asking participant A to randomly choose a coded envelope which was linked to a certain participant B. Each participant A was then asked to fill in the take rate, that is the proportion of participant B's endowment that would be transferred to participant A at the end of the experiment, on the form that was placed inside the envelope. Afterwards, the envelopes containing the forms were collected and given to all participants B who were asked to complete the form with the destroy rate, that is the proportion of their endowment that will be destroyed. The envelopes with the forms inside were then given back to all participants A, who could take note of the decision of their corresponding matches. Subsequently, each subject was asked to fill in a questionnaire concerning emotions, expectations about the decision of their counterpart,[60] and personal information. Meanwhile, the envelopes were collected and handed to the cashier who was outside the laboratory and, hence, not present during the experiment. Subjects were then privately paid one-by-one outside the laboratory by the cashier.

As in Bosman and van Winden (2002), we assessed a list of eleven emotions. To not direct the attention of the subjects to specific emotions, the list include both negative emotions that previous studies have found relevant for explaining the punishing behavior observed in the PTTG (i.e. anger, irritation, and contempt), other less influential negative emotions (i.e. envy, jealousy, sadness, shame, fear), positive emotions (i.e. joy, and happiness), and neutral emotions (i.e. surprise). For each emotion, subjects were asked to state how much they felt the emotion on a 7-point Likert scale when they

[60] Participant B was asked to indicate which percentage of his/her endowment he/she expected participant A would decide to transfer to himself/herself; participant A was asked to indicate which percentage of the transfer he/she expected participant B to destroy.

learned about the decision of their counterparts. The scale ranged from "no emotion at all" to "high intensity of the emotion" (Bosman and van Winden, 2002).

We employed a 2×2 factorial design crossing the nature of the 'fine-to-fee' ratio embedded in the punishment technology (constant or variable) with the cultural background of the subjects who participated in the sessions (UK or non-UK students). Note that subjects were not told about the nationality of their co-participants in the experiment. Hence, they did not know that they took part in a session with all UK or non-UK participants.[61] The main features of the design and the number of independent observations[62] per treatment are summarized in Table 1. The two treatments under a variable 'fine-to-fee' ratio are exact replications of previous PTTG experiments (i.e. Bosman and van Winden, 2002; Bosman *et al.*, 2005), in one case with only UK subjects and in the other case with only non-UK subjects (consisting of Asian, South American, African and East European subjects).[63] In these treatments, the 'fine-to-fee' ratio was increasing in the take rate and ranged from 0 to infinite. The same separation of the subjects based on cultural background occurred in the treatments under a constant 'fine-to-fee' ratio. However, here the 'fine-to-fee' ratio was constantly equal to 2 and, therefore, independent of the take rate.[64] Most of the literature on punishment behavior in economics usually employs 'fine-to-fee' ratios ranging from 1 to 4 (e.g. Nikiforakis and Normann, 2008). We opted for a

[61] Subjects also usually arrived a little at time to the experiment, and were immediately seated to their cubicles after registration to avoid them queuing at the entrance of the lab. Hence, the likelihood of subjects seeing and interacting with each other as they arrived was minimized. In addition, partitions ensured that subjects did not see each other during the experiment.

[62] An independent observation is a pair consisting of a responder and a take authority. Variations in the number of independent observations across treatments are due to different rates of attendance across sessions.

[63] The exception is one subject from Australia who played in the role of a take authority. It is also worth noting that there was a predominance of Chinese among the non-UK subjects (see the appendix).

[64] In order to employ a constant 'fine-to-fee' ratio, we simply stated in the instructions that "for each 1% of his or her endowment that participant B decides to destroy, 10 pence of the transfer to participant A will be destroyed as well". In addition, we allowed subjects to deduct the cost of punishing from their show-up fee, if needed. For this reason, the show-up fee was set at the level of £5 to ensure that, at worst, subjects (in particular, participants B) could leave the experiment with £2.50 in their pockets. This ensured that participants B could punish participants A for any possible value of the take rate without incurring losses.

value in the middle of the range, equal to 2.[65] Such a 'fine-to-fee' ratio has been extensively used in previous economic experiments to study punishment behavior (e.g. de Quervain *et al*., 2004; Cubitt *et al*., 2011). It also maximizes the comparability with the other treatment where the 'fine-to-fee' ratio is variable, and with previous PTTG experiments. Note in fact that a 'fine-to-fee' ratio equal to 2 corresponds, under a variable 'fine-to-fee' ratio, to a take rate of $2/3$, which is roughly equivalent to the mean and median take rate observed in previous PTTG experiments, and obtained in our treatment with a variable 'fine-to-fee' ratio. Hence, when we compare the two treatments, we have approximately half the observations where the incentive to punish is lower in the "constant fine-to-fee ratio" treatment, and the other half where the incentive is higher. This allows us to measure the bias, if there is any, which may occur in the standard PTTG for both sides of the distribution of the take rates: when $t < 2/3$ (i.e. when the variable 'fine-to-fee' is lower than the constant 'fine-to-fee' ratio), and when $t > 2/3$ (i.e. when the variable 'fine-to-fee' is higher than the constant 'fine-to-fee' ratio).

Table 1: Features and independent observations of the experimental treatments

	Variable ratio	**Constant ratio**	**Total**
UK subjects	37	34	71
non-UK subjects	33	37	70
Total	70	71	141

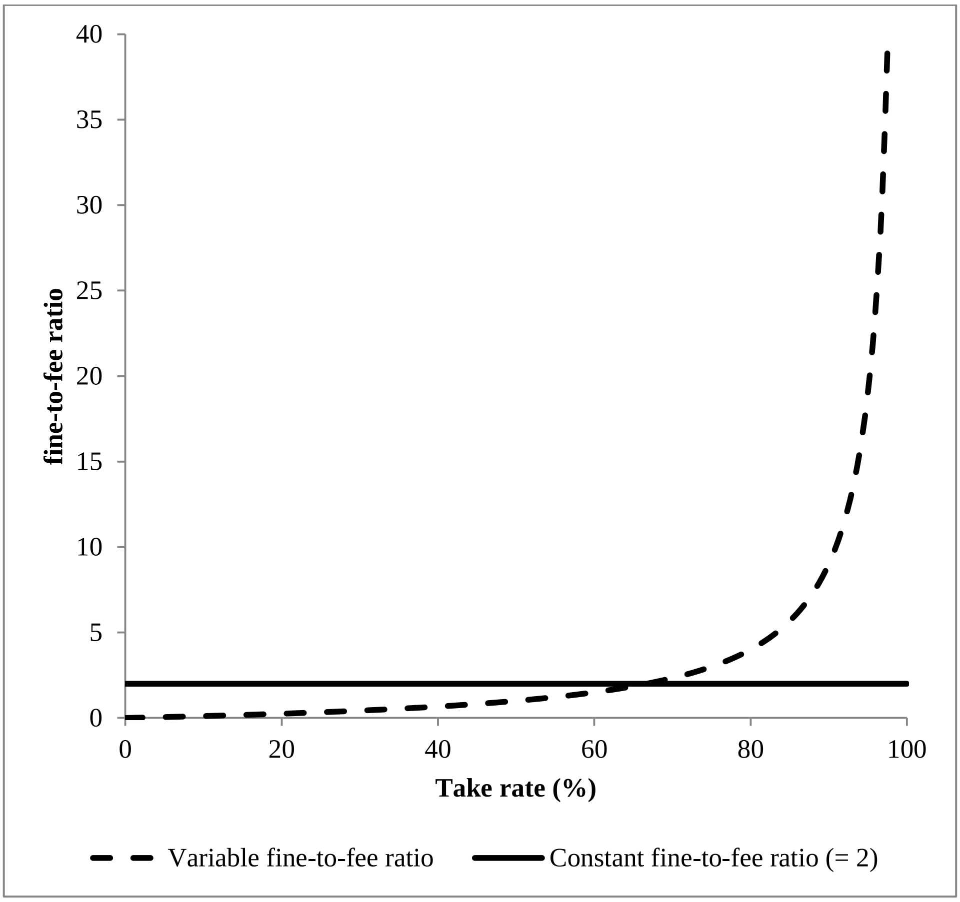
Notes: One independent observation is a pair of a take authority and responder.

Figure 1 displays how the 'fine-to-fee' ratio evolved over different values of the take rate in the treatments under a variable 'fine-to-fee' ratio and under a constant 'fine-to-fee' ratio respectively. The graph clearly shows that under a variable 'fine-to-fee' ratio the effectiveness of the punishment increases exponentially as the take rate increases. As a consequence, subjects might punish simply because it is more 'convenient' to do so and not or not only because they experience negative emotions (which is to be expected for increasing take rates). In other words, in this set-up, the idiosyncratic features of the punishment technology might induce

[65] A value of 2 means that the cost of punishing is half of how much the punishment reduces participant A's endowment.

an effect on behavior which can be confounded with that of negative emotions (for a discussion of confounds in experiments see Zizzo, 2013). This possibility is instead ruled out in the constant ‘fine-to-fee’ ratio treatments, where no confound can be attributed to the punishment technology.

Figure 1: Patterns of variable and constant ‘fine-to-fee’ ratios



3. Theoretical background

In this section, we briefly discuss the theoretical implications of having a constant versus variable ‘fine-to-fee’ ratio, embedded in the punishment technology, for the behavior and emotions of the responder in the PTTG, and present the hypotheses and conjectures that can be tested. We build on the model that Loewenstein (2000) proposed to describe the impact of visceral factors in the utility function of an agent. Let $U_i(p_i, s_i)$ be the utility function of a generic responder i, in which p_i is the consumption activity, in our case the size of the punishment, and s_i the visceral state, in our case the experience of anger, irritation or contempt. These emotions may affect behavior through two mechanisms: the “carrot” and the “stick” (Loewenstein, 2000). The “carrot” is captured by $\partial^2 U(p_i, s_i)/\partial s_i \partial p_i \geq 0$, and identifies the increase in utility that the responder experiences as a result of the mitigating effect of the punishment on the emotion. The “stick” is represented by $\partial^2 U(p_i^0, s_i)/\partial s_i < 0$, where p_i^0 is the absence of punishment, and identifies the marginal disutility that the responder experiences, if he or she does not satisfy the emotional impulse to punish. These two effects altogether induce the responder to punish when he or she experiences anger, irritation or contempt, and to punish more the higher is the intensity of the emotion. In other words, if we call p_i^* the optimal punishment, this may rise if the intensity of the negative emotion rises as well ($\partial p_i^*/\partial s_i \geq 0$). In the context of the PTTG, the responder may experience emotional distress when, for instance, the decision of the counterpart is perceived as unkind or when the distribution of the income becomes unequal, which may occur every time that t_i increases. To mitigate this distress, the responder may decide to punish. Hence, negative emotions may be one of the underlying mechanisms that explain social preferences.

Anger, irritation or contempt may not be the only explanation of the decision to punish. The responder may, for instance, punish because he or she feels the need to comply with what he or she believes it is the appropriate behavior in the lab (Zizzo, 2010) or because he or she misunderstands the instructions or the incentives in the experiment.[66] On top of these alternative explanations of punishing behavior, the decision to punish may also be sensitive to more traditional economic incentives such as the cost per unit of punishment that the responder needs to pay in order to damage the counterpart, which is a measure of the effectiveness of the punishment (Nikiforakis and Normann, 2008). To formalize all this, let assume that optimal punishment (p_i^*) depends on the intensity of the negative emotion (anger, irritation or contempt), which, in turn, depend on

[66] A significant part of the behavior observed in experiments may be attributed to confusion or mistakes, as, for instance, Andreoni (1995) found in the context of public good games. Confusion and mistakes may explain why a responder punishes in the PTTG since the only possible direction in which a responder can make a mistake is towards punishment.

the take rate experienced by the individual (t_i), the effectiveness of the punishment, which is equivalent to the 'fine-to-fee' ratio (σ_i), and a generic parameter θ_i which captures everything else, including confusion, mistakes, and experimenter demand effects.[67] We can write the optimal punishment as:

$$p_i^* = p_i^*(s_i(t_i), \sigma_i, \theta_i)$$

For the chain rule, the optimal level of punishment weakly increases in the take rate ($\partial p_i/\partial t_i \geq 0$).[68] In addition, it is weakly increasing in the fine-to-fee ratio ($\partial p_i/\partial \sigma_i \geq 0$).[69] Under a variable 'fine-to-fee' ratio, σ_i is an increasing function of t_i. Remember that $\sigma_i = t_i/(1-t_i)$. This means that the punishment may be even higher if t_i increases (i.e. t_i may have a *multiplier effect* on the decision to punish by means of the non-constant 'fine-to-fee' ratio). This multiplier effect is given by $\partial\,[t_i/(1-t_i)]/\partial t_i = 1/(1-t_i)^2$, and is exponentially increasing in t_i. In contrast, if the 'fine-to-fee' ratio is a constant ($\sigma_i = \bar{\sigma}$), its impact on the decision to punish should be the same across different level of t_i (i.e. there is no multiplier effect).[70] Hence, under a variable 'fine-to-fee' ratio, the role of negative emotions may be overstated. By comparing a situation where the punishment technology embeds a constant 'fine-to-fee' ratio with a situation where the punishment technology is characterized by a variable 'fine-to-fee' ratio, we can measure how much of the punishment is actually attributable to the

[67] For simplicity, we also assume perfect separability between θ_i and s_i, that is θ_i does not affect s_i and *vice versa*.

[68] $\partial p_i/\partial t_i = (\partial p_i/\partial s_i)\,(\partial s_i/\partial t_i) \geq 0$ since $\partial s_i/\partial t_i \geq 0$ and $\partial p_i/\partial s_i \geq 0$.

[69] If the responder displays a rational and self-interested behavior, he or she should never destroy if $t_i < 1$, and be indifferent between any levels of punishment if $t_i = 1$. This is irrespectively of the size of the 'fine-to-fee' ratio. In other words, $\partial p_i/\partial t_i \geq 0$ and $\partial p_i/\partial \sigma = 0$. If his or her behavior is driven by reciprocity (e.g. Rabin, 1993; Duwfenberg and Kirchsteiger, 2004), the punishment should be more likely as t_i increases (i.e. $\partial pi \partial ti \geq 0$), *caeteris paribus*, and, less likely as σ increases (i.e. $\partial pi \partial \sigma \geq 0$), *caeteris paribus*. If the responder cares about equality (e.g. Fehr and Schmidt, 1999), the punishment should be more likely as t_i increases (i.e. $\partial pi \partial ti \geq 0$), *caeteris paribus*. For any given level of t_i, the punishment should also be weakly increasing in σ (i.e. $\partial pi \partial \sigma \geq 0$). The proofs are in the appendix.

[70] Note that, in this study, we are not interested in how different constant levels of σ impact on the decision to punish. Our aim is to establish how much of the punishing behavior observed in PTTG experiments is actually attributable to negative emotions rather than the multiplier effect caused by the 'fine-to-fee' ratio.

multiplier effect rather than emotional distress,[71] but also whether negative emotions matter at all. Note in fact that it is possible that $\partial s_i/\partial t_i \geq 0$ and $\partial p_i/\partial s_i = 0$. If the latter is the case, punishment is better explained by the parameter θ_i rather than anger, irritation or contempt.

Based on this simple model that describes the motivations of the responder to punish, we can formulate the following hypotheses.

Hypothesis 1. In the standard PTTG, the variable 'fine-to-fee' ratio produces a multiplier effect on the decision to punish, that is $(\partial p_i/\partial \sigma_i)\,(\partial \sigma_i/\partial t_i) > 0$, and thus the role of other factors, including negative emotions, is overstated.

We can test this hypothesis by studying how the difference in punishment evolves between a constant and a variable 'fine-to-fee' ratio when the take rate increases. The null hypothesis is that there is no such multiplier effect, that is $(\partial p_i/\partial \sigma_i)\,(\partial \sigma_i/\partial t_i) = 0$.

Hypothesis 2. Once the multiplier effect is removed, the intensity of anger, irritation, or contempt does still explain the punishing behavior of the responder, that is $\partial p_i/\partial s_i > 0$.

We can test this hypothesis against the null hypothesis that punishing is due to something else (e.g. confusion, experimenter demand effects) by looking at whether these negative emotions do still explain punishing behavior when the 'fine-to-fee' ratio is constant.

Hypothesis 3. Due to the multiplier effect, anger, irritation, and contempt erroneously predict much more punishing behavior under a variable 'fine-to-fee' ratio compared to a constant 'fine-to-fee' ratio, that is $\partial p_i^V/\partial s_i > \partial p_i^C/\partial s_i$, where p_i^V is the level of punishment with a variable 'fine-to-fee' ratio, and p_i^C is the level of punishment with a constant 'fine-to-fee' ratio.

[71] Note that it is not possible to separate the multiplier effect, caused by the variable 'fine-to-fee' ratio, from the arousal effect, caused by the negative emotion, using statistical analysis on the data collected from previous PTTG studies. This is because of the very high correlation between negative emotions and the take rate.

Hypothesis 3 can be tested by comparing how much of the punishing behavior is explained by these negative emotions under a variable versus a constant 'fine-to-fee' ratio.

On top on these hypotheses, we are also able to investigate the two following conjectures.

Conjecture 1. The emotional response of UK and non-UK students is different and produces different patterns of punishing behavior.

Conjecture 1 is motivated by two streams of research: one, from psychology, suggesting that there are cultural differences in the elicitation and manifestation of emotions (e.g. Mesquita and Frijda, 1992), the other, from economics, showing that punishment may vary across societies and cultures (e.g. Hermann *et al*., 2008; Henrich *et al*., 2006; Gächter *et al*., 2005). There may be a link between these two lines of research insofar as emotions may play an important role for the decision to punish. We look at this from a methodological point of view. In particular, given the common practice of many economic experiments to employ students with different culture background, we investigate whether the emotional basis of economic behavior may be affected by the cultural background of the participants.

Conjecture 2. More experienced subjects are able to better cope with their emotional urges to punish.

This second conjecture stems from the evidence that, in many contexts, experienced participants behave differently compared to inexperienced or less experienced subjects, both in the field and in the lab (e.g. Kagel and Levine, 1999; Myagkov and Plott, 1997; Levitt and List, 2007; List and Levitt, 2005). In particular, experienced subjects seem to behave more rationally and take more "cold-blooded" decisions compared to inexpert subjects. One possible explanation, which has not so far been investigated in the literature, is that subjects with more experience are more aware of what they should expect in the economic environment where they operate, and, therefore, they might experience less strong emotions and/or be better able to cope with their emotional urges than inexperienced subjects. We investigate this possibility in the context of the PTTG with respect to the punishing behavior of the responders.

4. Results

In this section, we first check whether there is any difference in the behavior of the take authorities across the treatments. Then, we move to the main focus of this study, that is the punishing behavior of the responder, to test whether there exists any bias in the standard PTTG compared to the modified version of the PTTG where the punishment technology embeds a constant 'fine-to-fee' ratio. Afterwards, we look at emotions and the extent to which anger, irritation, and contempt explain the punishing behavior of the responders.[72] Finally, we briefly consider the expectations of the responders, and whether they help to understand behavior and emotions.

Table 2: Take rates

	Variable σ			**Constant σ**		
	n.	**Mean**	**St. dev.**	**n.**	**Mean**	**St. dev.**
UK	37	57	26.89	34	54.34	30.11
non-UK	33	72.27	20.71	37	62.70	25.35
Total	70	64.2	25.20	71	58.70	27.85

Behavior of the take authorities. Table 2 displays the take rates of UK and non-UK subjects under both constant and variable 'fine-to-fee' ratios. The results are in line with previous PTTG experiments. However, there is evidence of behavioral differences between UK and non-UK subjects. In particular, non-UK subjects display statistically significantly higher take rates than UK subjects both in aggregate (Mann-Whitney p = 0.023) and under variable 'fine-to-fee' ratios (Mann-Whitney p = 0.012).[73] If we compare the behavior of the subjects under a constant and variable 'fine-to-fee' ratios, there is no significant difference at 5% level between the take rates, both in aggregate, and within the sample of UK and non-UK

[72] In the appendix, we also look at the emotions of the take authorities.
[73] No significant difference occurs between UK and non-UK subjects under a constant 'fine-to-fee' ratio (Mann-Whitney p = 0.351).

subjects respectively.[74] These tests suggest that non-UK take authorities appropriate more resources than UK take authorities.

A Tobit regression analysis confirms this result.[75] Table 3 presents the outcomes of this analysis. The dependent variable is the take rate. In Regression 1, explanatory variables include a dummy variable for the experimental sessions under a constant 'fine-to-fee' ratio ($d_{\sigma=\bar{\sigma}}$), the nationality of the subjects (non-UK = 1 for non-UK subjects), and an interaction term between the dummy for the constant 'fine-to-fee' ratio and nationality. In Regression 2, we also include the experience of the subjects in previous experiments,[76] their gender (Male = 1 for male subjects), and their age.

Table 3: Tobit regression on take rate

	Regression 1			Regression 2		
	b	se	p	B	se	p
$d_{\sigma=\bar{\sigma}}$	-2.336	8.202	0.776	-3.232	8.084	0.69
non-UK	17.913***	6.788	0.009	22.450***	6.556	0.001
$d_{\sigma=\bar{\sigma}}$ × non-UK	-7.138	10.541	0.499	-8.483	10.92	0.439
Experience				-0.115	2.665	0.966
Male				4.673	5.489	0.396
Age				-1.719***	0.483	0.001
Constant	56.445***	5.202	0	91.046***	12.09	0
Obs	141			141		
Pseudo R-Square	0.007			0.015		
Df	138			135		
Prob > F	0.02			0.001		

Notes: Tobit regression with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01,

The coefficient of the dummy variable *non-UK* is positive and statistically significant, whereas the coefficient of the interaction term is not statistically significant. This implies that non-UK take authorities take more money than UK take authorities both under a variable and constant 'fine-to-

[74] In aggregate, and within the sample of UK subjects, the difference is not statistically significant (Mann-Whitney *p* = 0.191 and 0.732 respectively). For non-UK subjects the difference is weakly statistically significant (Mann-Whitney *p* = 0.098).

[75] 10 observations are left-censored and 16 right-censored.

[76] The data for 'experience' was collected from the final questionnaire provided to the subjects. In particular, subjects were asked to indicate whether they had previously participated in "0", "1", "2" or "3 or more than 3" experiments.

fee' ratio once covariates are controlled for.[77] We should point out that the aim of our experiment was not to study the behavior of the take authorities, but the punishing behavior and emotions of the responders. Different conjectures might explain why non-UK take authorities appropriated more resources than UK take authorities. We will briefly examine them in the next section.

Behavior of the responders. We define the *punishment rate* as the proportion of the amount taken by the take authority that was destroyed by the responder.[78] As we discussed in Section 3, the punishment rate may depend on the amount taken by the take authority because of several psychological reasons such as inequality aversion, and reciprocity. Hence, we need to control, in the analysis of the punishing behavior, for the impact of the take rate. Figure 2 displays the scatterplots, and the locally weighted smoothed regressions of the punishment rate as a function of the take rate, for each of the treatments. The behavioral pattern of UK and non-UK subjects appears to be very similar. In particular, when the 'fine-to-fee' ratio is variable, the punishment of both UK and non-UK subjects raises exponentially as the take rate increases. When the 'fine-to-fee' ratio is instead constant, both UK and non-UK subjects increase their punishment less in response to higher take rates. This is preliminary evidence in favor of Hypothesis 1. In particular, it seems that, under a variable 'fine-to-fee' ratio, there is a multiplier effect at work which induces more punishment from the responders when the take rate increases. We can test this more formally with non-parametric techniques. In particular, we can group the take rates into four classes based on their distribution: the very low take rates (the bottom 25%), the low take rates (between 25% and 50%), the high take rates (between 50% and 75%), and the very high take rates (the top 25%).[79] This distinction enables us to investigate the punishing behavior controlling for

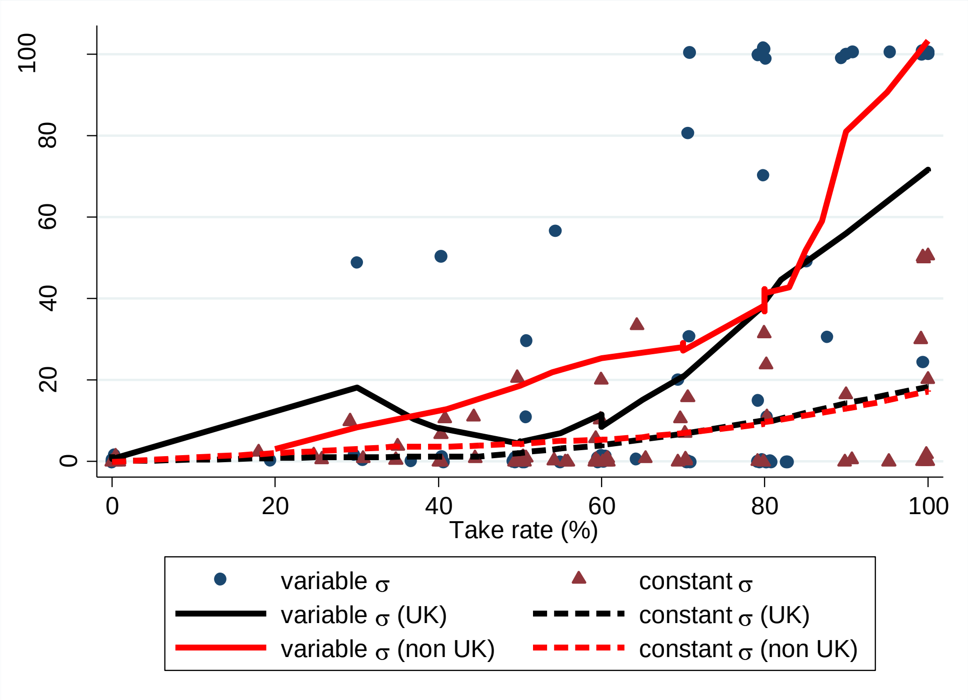
[77] There is also evidence that, other things being equal, older subjects take on average less than younger subjects.

[78] The punishment rate coincides with the destroy rate under a variable 'fine-to-fee' ratio, and is equal to $2d/t$ under a constant 'fine-to-fee' ratio, that is the amount destroyed ($10\ pence\ \times d$) over the amount taken ($5\ pounds\ \times t$).

[79] The 25th quartile corresponds to a take rate of 0.5; the 50[th] quartile (median) to a take rate of 0.6; while the 75[th] quartile to a take rate of 0.8. Hence, t_i is classified as very low take rate if $t_i \leq 0.5$, low take rate if $0.5 < t_i \leq 0.6$, high take rate if $0.6 < t_i \leq 0.8$, and very high take rate if $t_i > 0.8$.

different level of the take rate, and also discriminating between take rates for which the constant 'fine-to-fee' ratio is lower than the variable 'fine-to-fee' ratio, and *vice versa*. Note in fact that the constant 'fine-to-fee' ratio crosses the variable 'fine-to-fee' ratio at $t_i = 2/3$ (see Figure 1), which is slightly above the median of the distribution of the take rates. Hence, for the very low and low take rates, the incentive to punish is higher under a constant 'fine-to-fee' ratio. In contrast, for the high and very high take rates, the opposite is true, that is the incentive to punish is higher under the variable 'fine-to-fee' ratio.[80]

Figure 2: Relationship between punishment rate and take rate



Notes: The locally weighted regressions are computed using a bandwith of 0.8 (80% of the data).

Table 4 displays the punishment rates for UK and non-UK subjects under both constant and variable 'fine-to-fee' ratios. First of all, we can check, for each class of the take rates, whether the behavior of UK subjects is statistically different from the behavior of non-UK subjects. We do not find any statistically significant differences between the behavior of UK and non-UK subjects for each class of the take rates, either under a variable and

[80] To be precise, among the high take rates, there are 3 out of 38 observations at $t_i = 0.65$, which are slightly below $t_i = 2/3$, and where, therefore, the incentive to punish can be considered as identical between the constant and variable 'fine-to-fee' ratio.

constant ‘fine-to-fee’ ratio.[81] We can thus pool the data of UK and non-UK participants together to increase the power of our tests. If we compare the punishment rate of sessions characterized by a constant ‘fine-to-fee’ ratio with sessions where the ‘fine-to-fee’ ratio is variable, we find that, when the take rates are very high, punishment is strongly significantly more severe under a variable ‘fine-to-fee’ ratio (Mann-Whitney test, p = 0.017). In contrast, when the take rates are very low, subjects seem to punish more under a constant ‘fine-to-fee’ ratio. However, the difference is not statistically significant (p = 0.286). For low and high take rates, the punishment rates are very similar across the two treatments, and the difference is not statistically significant (Mann-Whitney test, p > 0.1).

Table 4: Punishment rates

	σ	**Very Low t**	**Low t**	**High t**	**Very High t**
UK	Constant	32.94	59	76.25	93
	Variable	26.96	58.89	75	98
non-UK	Constant	42.63	57.00	76.92	93
	Variable	36.15	59.38	71.43	97.22
All	Constant	36.17	58.43	76.60	93
	Variable	31.39	59.12	73.08	97.50

We can test the robustness of these findings in a Tobit regression analysis (see Table 5).[82] The dependent variable is the punishment rate (p_i). The independent variables are, in Regression 1, the take rate received from the take authority (t_i), a dummy variable which takes value 1 when a constant ‘fine-to-fee’ ratio was employed, the nationality of the subjects (non-UK = 1 for non-UK), and two interaction terms of the dummy used to

[81] All the differences are not statistically significant (Mann-Whitney test, p > 0.1) except for the punishment rates selected in response to take rates classified as very low, when the ‘fine-to-fee’ ratio is constant. In that case, the difference between the punishment rate of UK and non-UK subjects is weakly significant (p = 0.063). This is due the fact that more UK take authorities selected a take rate of zero, compared to non-UK take authorities. Hence, a higher proportion of UK responders did not destroy at all. If we exclude the take rates equal to zero, the difference in the punishment rate is not anymore statistically significant.

[82] There are 87 left-censored observations and 19 right-censored. We also tried a logit regression where the dependent variable was a dichotomous variable taking value 1 if the responder destroyed and 0 otherwise. The results are similar to those presented in the paper. However, this approach omits much of the information about the punishment rate and, therefore, is less preferred than the approach based on the Tobit model. The results of the logit are reported in the appendix.

identify the constant 'fine-to-fee' ratio with nationality, and the take rate respectively. In Regression 2, we also control for the experience, gender (Male = 1 for male subjects), and age of the subjects. The coefficient of the variable take rate is positive and statistically significant, meaning that the take rate from the take authority negatively affects the punishment behavior of the responder when the 'fine-to-fee' ratio is variable. Under a constant 'fine-to-fee' ratio, the relationship between the take rate and the punishment rate is also negative but markedly weaker. This brings us to the following result which supports Hypothesis 1.

Result 1. Consistently with Hypothesis 1, under a variable 'fine-to-fee' ratio, 70% (-2.727/3.885) of the punishment triggered by the take rate is attributable to the multiplier effect.

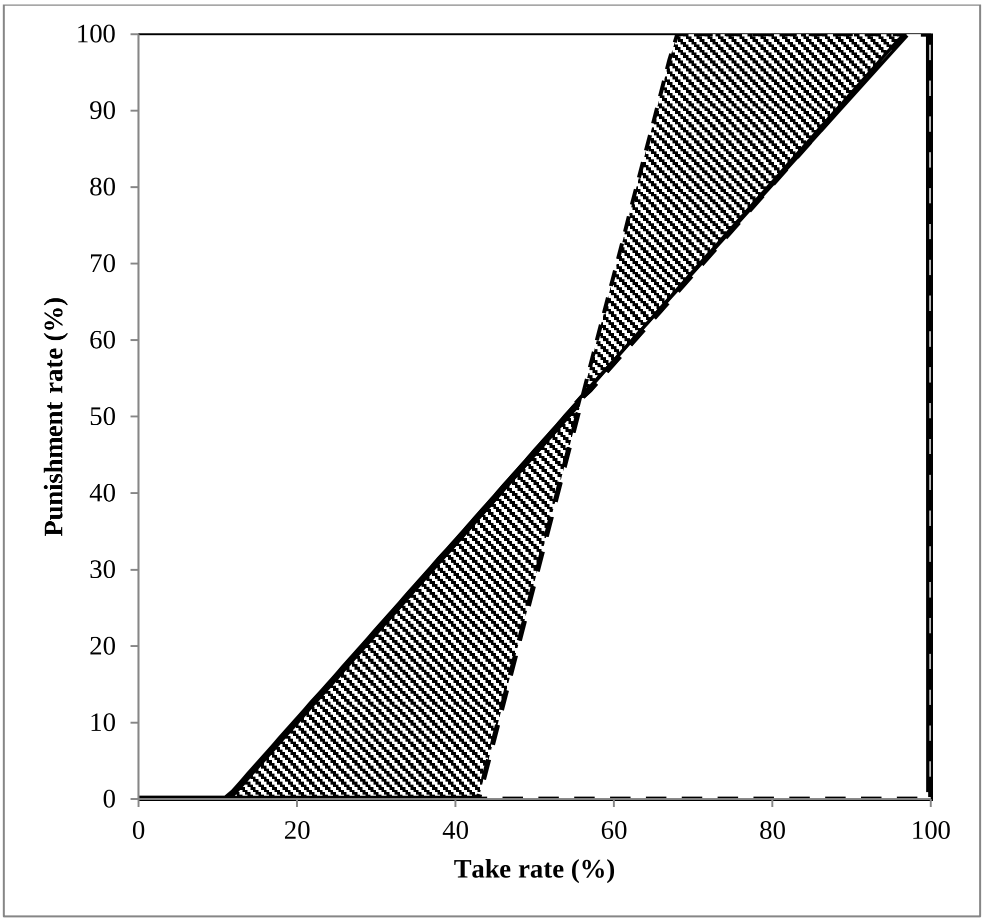
Table 5: Tobit regression on punishment rate

	Regression 1			Regression 2		
	B	se	p	b	se	p
t_i	3.885***	1.077	0	4.009***	1.005	0
$d_{\sigma=\bar{\sigma}} \times t_i$	-2.727**	1.126	0.017	-2.841***	1.063	0.008
$d_{\sigma=\bar{\sigma}}$	162.285**	79.792	0.044	158.868**	74.332	0.034
$d_{\sigma=\bar{\sigma}}$ × non-UK	-6.019	39.41	0.879	3.95	39.673	0.921
non-UK	26.465	29.38	0.369	30.978	29.276	0.292
Experience				-5.528	8.564	0.52
Male				-30.789	19.387	0.115
Age				-3.951**	1.972	0.047
Constant	-285.573***	81.39	0.001	-172*	89.096	0.056
Obs	141			141		
Pseudo R-Square	0.075			0.085		
Df	136			133		
Prob > F	0.001			0.002		

Notes: Tobit regression with robust standard errors. t_i and p_i are expressed in percentage Hence, the beta coefficients identifies percentages. * p<0.1, ** p<0.05, *** p<0.01.

Under a constant 'fine-to-fee' ratio, the punishment rates are generally higher compared to a variable 'fine-to-fee' ratio. This is captured by the positive and significant coefficient of the dummy $d_{\sigma=\bar{\sigma}}$. Hence, when the take rates are low, punishment is higher under the constant 'fine-to-fee', and, when the take rates are high, it is higher under the variable 'fine-to-fee' ratio. This is shown in Figure 3 where we plot the predicted punishment rate for the constant and variable ‘fine-to-fee’ ratio respectively against the take rate. In particular, for take rates lower than 56%, the punishment is higher under the constant ‘fine-to-fee’ ratio, whereas, for take rates higher than 56%, the punishment is higher under the variable ‘fine-to-fee’ ratio.[83]

Figure 3: Predicted punishment rate against take rate



Notes: The dashed line is the predicted punishment rate under a variable ‘fine-to-fee’ ratio; the solid line is the predicted punishment rate under a constant ‘fine-to-fee’ ratio. The crossed area measures the extent to which the punishment is inflated (t >56%) or deflated (t < 56%) under the variable ‘fine-to-fee’ ratio compared to the constant ‘fine-to-fee’ ratio.

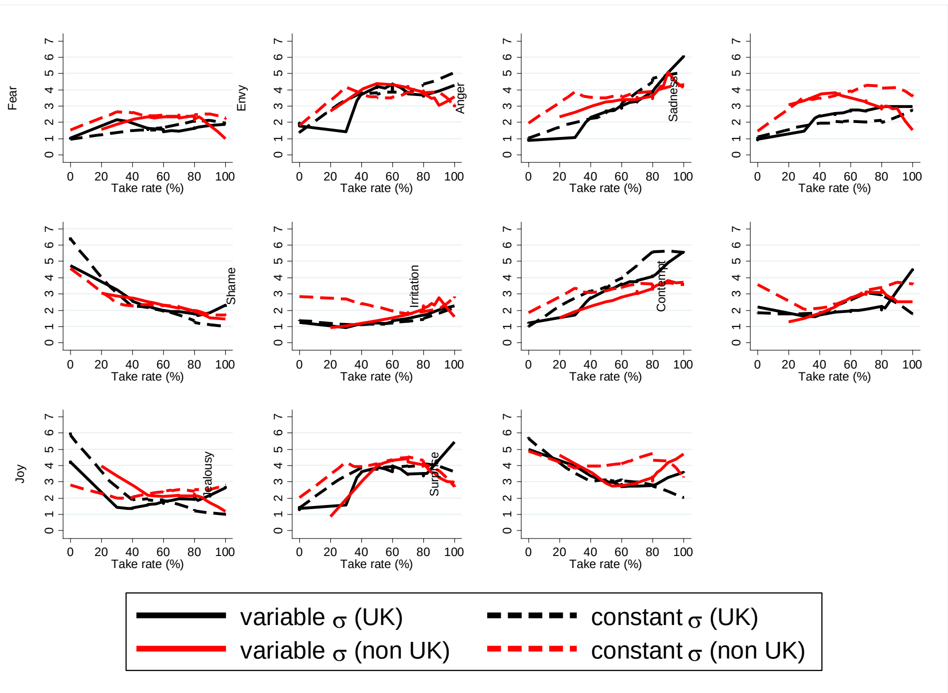
Role of emotions. We now turn to the analysis of the emotions experienced by the responders. We will initially consider all the emotions (positive, negative, and neutral) to check whether there exists any similarity between them and to investigate which emotions were driven by the take rate. We will then focus on anger, irritation, and contempt, to study the emotional basis of the punishing behavior.

[83] Among the other explanatory variables, the only coefficient statistically significant is the one for age. In particular, older subjects punish less than younger ones.

First, it is worth pointing out that, as seen in the previous literature, different emotions capture similar underlying emotional states. In particular, anger is strongly positively correlated to irritation (Spearman ρ = 0.81 and 0.74, p = 0.000), envy to jealousy (ρ = 0.84 and 0.81, p = 0.000), and happiness to joy (ρ = 0.86 and 0.72, p = 0.000) for UK and non-UK subjects respectively.

In order to study whether and which emotions are driven by the take rate, we first look at the patterns of the locally weighted smoothed regression lines between the intensity of each emotion and the take rate, for each of the treatments (Figure 4). There seems to be no differences across treatments on how the take rate impacts on each emotion. This is confirmed in non-parametric tests where, for each class of the take rates, we compare the intensity of each emotion between UK and non-UK subjects, and between the variable and constant 'fine-to-fee' ratio. None of these comparisons result statistically significant (Mann-Whitney test, p > 0.1). Figure 4 also provides some preliminary evidence of a positive relationship between take rate and negative emotions, and of a negative relationship between take rate and positive emotions.

Figure 4: Relationship between emotions and take rate



Notes: The locally weighted regressions are computed using a bandwith of 0.8 (80% of the data).

We can test for this using some ordered logit regressions,[84] one for each emotion. The dependent variable is the emotion of interest (s_i), whereas the independent variables are the take rate received from the take authority, a dummy variable which takes value 1 when a constant 'fine-to-fee' ratio was employed, experience, gender (Male = 1 for male subjects), age, nationality of the subjects (non-UK = 1 for non-UK), and two interaction terms, one between the dummy for the constant 'fine-to-fee' ratio and nationality, and another between the dummy for the constant 'fine-to-fee' ratio and the take rate. The results of these regressions are shown in Table 6.[85]

Negative emotions (in particular, anger, and irritation) are significantly positively related to the take rate. Similarly, happiness is significantly negatively related to the take rate. This evidence is consistent with previous PTTG studies, and the theory presented earlier.[86] It is also robust regardless of the background of the subjects (UK versus non-UK students) and the type of punishment technology employed.

---

[84] Robust standard errors are employed to control for heteroscedasticity. Due to some subjects failing to report all the emotions, we have 1 missing observation for sadness, shame and envy (140 observations instead of 141), and 3 missing observations for contempt (138 observations instead of 141).

[85] The qualitative results do not change if we do not include the demographic variables among the explanatory variables. The results of these regressions are reported in the appendix.

[86] The coefficients of the other explanatory variables are mostly not significant. We briefly mention here those which are significant. In particular, subjects with increasing experience in laboratory experiments experience less fear (p = 0.012) and joy (p = 0.027). Older subjects experience less fear (p = 0.044), anger (p = 0.043) , sadness (p = 0.055), and Jealousy (p = 0.049). Non-UK subjects under a constant 'fine-to-fee' ratio are on average sadder (p = 0.041). Finally, male subjects experience more sadness (p = 0.052), and jealousy (p = 0.056).

Table 6: Ordered logit regressions on emotions

	Fear			Envy			Anger			Sadness			Happiness			Shame		
	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p
t_i	0	0.01	0.63	0.01	0.01	0.49	**0.04*** **	0.01	0	0	0.01	0.66	**-0.03** **	0.01	0.01	0.02	0.02	0.31
$d_{\sigma=\overline{\sigma}}$	0.11	0.86	0.9	-0.61	0.92	0.51	0.48	0.81	0.56	-1.07	0.79	0.18	0.66	1.09	0.54	1.19	1.39	0.39
$d_{\sigma=\overline{\sigma}} \times t_i$	0	0.01	0.89	0.01	0.01	0.62	-0.01	0.01	0.52	0.01	0.01	0.58	-0.01	0.02	0.51	-0.02	0.02	0.41
non-UK	1.09*	0.57	0.06	0.17	0.51	0.74	0.34	0.48	0.48	0.6	0.46	0.19	0.29	0.52	0.57	0.63	0.63	0.32
Experience	**-0.37** **	0.15	0.01	-0.12	0.14	0.41	-0.02	0.16	0.88	-0.19	0.15	0.2	-0.07	0.19	0.72	-0.11	0.19	0.56
$d_{\sigma=\overline{\sigma}}$ × non-UK	0.05	0.77	0.95	-0.08	0.65	0.9	0.34	0.67	0.61	**1.37** **	0.67	0.04	-0.11	0.7	0.88	0.5	0.84	0.55
Male	-0.28	0.4	0.48	0.2	0.35	0.57	0.42	0.38	0.27	0.66*	0.34	0.05	0.29	0.4	0.47	0.44	0.44	0.32
Age	**-0.08** **	0.04	0.04	-0.07	0.05	0.13	**-0.07** **	0.03	0.04	-0.06*	0.03	0.05	0.01	0.04	0.89	0.04	0.04	0.38
Obs.	141			140			141			140			141			140		
Pseudo R-Square	0.05			0.02			0.07			0.06			0.08			0.05		
Prob > F	0.01			0.31			0			0			0			0.31		

	Irritation			Contempt			Joy			Jealousy			Surprise		
	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p
t_i	**0.03*** **	0.01	0	0.01	0.01	0.19	-0.01	0.01	0.24	0.01	0.01	0.2	-0.01	0.01	0.64
$d_{\sigma=\overline{\sigma}}$	0.14	0.7	0.84	0.19	0.92	0.84	1.06	1.08	0.33	0.16	0.78	0.84	0.79	1.01	0.43
$d_{\sigma=\overline{\sigma}} \times t_i$	0	0.01	0.83	0	0.01	0.97	-0.02	0.02	0.3	-0.01	0.01	0.6	-0.01	0.01	0.44
non-UK	-0.47	0.45	0.3	0.48	0.46	0.3	0.27	0.51	0.6	0.33	0.51	0.52	0.19	0.47	0.69
Experience	-0.08	0.14	0.59	0.05	0.15	0.74	**-0.35** **	0.16	0.03	-0.08	0.13	0.55	-0.16	0.13	0.22
$d_{\sigma=\overline{\sigma}}$ × non-UK	-0.08	0.67	0.91	0	0.66	0.99	0.14	0.68	0.84	0.36	0.64	0.57	0.38	0.61	0.53
Male	0.52	0.36	0.14	0.19	0.35	0.58	0.15	0.37	0.67	0.63*	0.33	0.06	-0.23	0.33	0.49
Age	-0.07	0.04	0.12	0.01	0.04	0.85	0.01	0.04	0.74	-0.09**	0.05	0.05	0.01	0.04	0.74
Obs	141			138			141			141			141		
Pseudo R-Square	0.07			0.02			0.04			0.03			0.02		
Prob > F	0			0.4			0.09			0.2			0.05		

Notes: Ordered logit regressions with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01.

We now consider the punishing behavior of the subjects and to what extent it can be explained by negative emotions. In the previous analysis of the relationship between punishing behavior and take rate, we have found that, for take rates below the median, the punishment is higher under the constant 'fine-to-fee' ratio, whereas, for take rates above the median, the punishment is higher under the variable 'fine-to-fee' ratio (see Result 3). This means that the bias in the predictive power of negative emotions that may characterize the PTTG with a variable 'fine-to-fee' ratio may be negative when the take rates are below the median, and positive, when take rates are above the median, compared to the PTTG with a constant 'fine-to-fee' ratio. It is thus important to distinguish between the role played by emotions when the take rates are low, and their role when the take rates are high. To do so, we estimate, for each emotion, the following model:

$$p_i = \beta_0 + \beta_1 \times s_i + \beta_2 \times d_{t>\tilde{t}} + \beta_3 \times s_i \times d_{t>\tilde{t}} + \beta_4 \times s_i \times d_{\sigma=\bar{\sigma}} + \beta_5 \times s_i \times d_{t>\tilde{t}} \times d_{\sigma=\bar{\sigma}} + \beta_6 \times nonUK + \beta_7 \times experience + \beta_8 \times nonUK \times d_{\sigma=\bar{\sigma}} + \beta_9 \times experience \times s_i + \beta_{10} \times age + \beta_{11} \times Male$$

where p_i is the punishment rate, s_i the intensity of the emotion of interest, $d_{t>\tilde{t}}$ a dummy which is equal to 1 when the responder experiences a take rate below the median take rate ($\tilde{t}$), $d_{\sigma=\bar{\sigma}}$ the dummy which identifies the treatment with a constant 'fine-to-fee' ratio, $nonUK$a dummy for the nationality of the subject (= 1 for non-UK subjects), $experience$ the experience of the subject in economic experiments, and age and $Male$ the age and gender (Male = 1 for male subjects) respectively of the subject. To estimate the model, we run a battery of Tobit regressions, one for each emotion.[87] Table 7 displays the results of the regressions for anger, contempt, and irritation. We focus on these emotions as they were the emotions which more likely predicted the punishing behavior in previous PTTG studies. In the appendix we report the analysis of each of the other emotions.

[87] We use robust standard errors to control for heteroscedasticity. If we exclude the demographic variables from the explanatory variables, the results remain qualitatively the same. The results of these regressions are reported in the appendix.

Table 7: Tobit regressions on punishment for anger, irritation and contempt

	Anger			Irritation			Contempt		
	b	se	p	b	se	p	b	se	p
$d_{t>\tilde{t}}$	40.29	38.31	0.29	44.76	38.78	0.25	130.84***	40.41	0
non UK	46.54	28.48	0.1	54.57*	28.17	0.05	40.99	27.56	0.14
$d_{\sigma=\bar{\sigma}}$ × nonUK	-26.22	34.97	0.45	-16.79	33.68	0.62	-3.15	35.58	0.93
Experience	1.39	16.24	0.93	19.46	19.32	0.32	33.51*	17.75	0.06
Male	-35.21*	20.07	0.08	-29.1	20.12	0.15	-22.16	19.39	0.26
Age	-2.21	1.86	0.24	-3.02	1.89	0.11	-4.27**	1.83	0.02
Anger	0.03	12.8	1						
$d_{t>\tilde{t}}$ × Anger	28.37**	11.59	0.02						
$d_{\sigma=\bar{\sigma}}$ × $d_{t>\tilde{t}}$ × Anger	-23.92**	10.15	0.02						
$d_{\sigma=\bar{\sigma}}$ × Anger	11.96	9.16	0.19						
Experience × Anger	-1.33	3.68	0.72						
Irritation				5.14	14.52	0.72			
$d_{t>\tilde{t}}$ × Irritation				32.04***	12.01	0.01			
$d_{\sigma=\bar{\sigma}}$ × $d_{t>\tilde{t}}$ × Irritation				-29.08***	10.77	0.01			
$d_{\sigma=\bar{\sigma}}$ × Irritation				14.93	9.36	0.11			
Experience × Irritation				-5.95	4.33	0.17			
Contempt							45.42**	17.55	0.01
$d_{t>\tilde{t}}$ × Contempt							1.76	13.01	0.89
$d_{\sigma=\bar{\sigma}}$ × $d_{t>\tilde{t}}$ × Contempt							-32.01***	11.66	0.01
$d_{\sigma=\bar{\sigma}}$ × Contempt							10.01	9.06	0.27
Experience × Contempt							-13.44**	5.27	0.01
Constant	85.83***	12.58	0	88.19***	12.92	0	84.38***	12.23	0
Obs	141			141			138		
Pseudo R-Square	0.09			0.09			0.09		
Df	130			130			127		
Prob > F	0			0			0		

Notes: Tobit regressions with robust standard errors. p_i is expressed in percentage Hence, the beta coefficients identifies percentages.* p<0.1, ** p<0.05, *** p<0.01.

For low take rates, an increase in the intensity of contempt induces responders to punish more (β_1= 45.42) both under a variable and constant 'fine-to-fee' ratio.[88] None of the other two negative emotions seem to explain the punishing behavior for low take rates. This brings us to the following result which, with respect to contempt, supports Hypothesis 2 and rejects Hypothesis 3.

Result 2. In line with Hypothesis 2, but in contrast to Hypothesis 3, contempt is the only negative emotion that explains the punishing behavior of the responders for low take rates, and its effect is similar under a variable and constant 'fine-to-fee' ratio.

If we now look at high take rates, in line with previous literature, subjects who experience higher anger, irritation, and contempt punish more. In particular, under a variable 'fine-to-fee' ratio, a one-unit increase in the intensity of anger, irritation, and contempt respectively produces a 28.37%, 32.04%, and 45.42% increase respectively in the punishment rate.[89] However, much of this increase is due to the multiplier effect caused by the 'fine-to-fee' ratio. Indeed the impact of irritation, anger, and contempt on the decision to punish in response to high take rates is hugely attenuated when we employ a constant 'fine-to-fee' ratio. In particular, a one-unit increase in irritation, anger, and contempt respectively raises the punishment by only 2.96%, 4.45%, and 13.41% respectively when the 'fine-to-fee' ratio is constant. This evidence supports both Hypotheses 2 and 3. We can present Results 3 and 4.

Result 3. In line with Hypothesis 2, anger, irritation, and contempt explain the punishing behavior even when the confound caused by the variable 'fine-to-fee' ratio is removed.

[88] The interaction between contempt and the dummy for the constant 'fine-to-fee' ratio is positive (as expected) but not significant.
[89] Note that the interaction between contempt and the dummy for high take rates is not significant, meaning that the effect of contempt for high take rate is the same as for low take rate.

Result 4. Consistently with Hypothesis 3, the multiplier effect causes a bias when the take rates are high, and this bias accounts for 90.76%, 87.31%, and 70.48% respectively of how much irritation, anger, and contempt respectively explain the punishing behavior in the PTTG with a variable 'fine-to-fee' ratio.[90]

Looking at the other covariates, we do not find any support to Conjecture 1. In particular, the emotional response of UK and non-UK subjects have a similar impact on the decision to punish, and is not affected by which technology punishment is employed. In contrast, we find some support to Conjecture 2 with respect to contempt. In particular, subjects with increasing experience in economic experiments are able to better cope with contempt, as they punish significantly less when they experience such emotion compared to inexperienced or less experienced subjects.[91] Note that experience in previous economic experiments does not eliminate the effect of contempt on punishing behavior, but only reduces it. We can thus present Result 5.

Result 5. Consistently with Conjecture 2, the impact of contempt on the decision to punish is lessened for subjects who have more experience in economic experiments. There is no support for Conjecture 1.

Role of expectations. As in previous PTTG studies, we also consider the role played by expectations in driving behavior and emotions.[92] First, we shall note that there is no statistically significant difference in the way in which responders reported their expectations under a constant and variable 'fine-to-fee' ratio (Mann-Whitney $p = 0.422$). Similarly, no difference in

[90] The bias is calculated as $23.92/28.37$ for anger, $29.08/32.04$ for irritation, and $32.01/45.42$ for contempt.

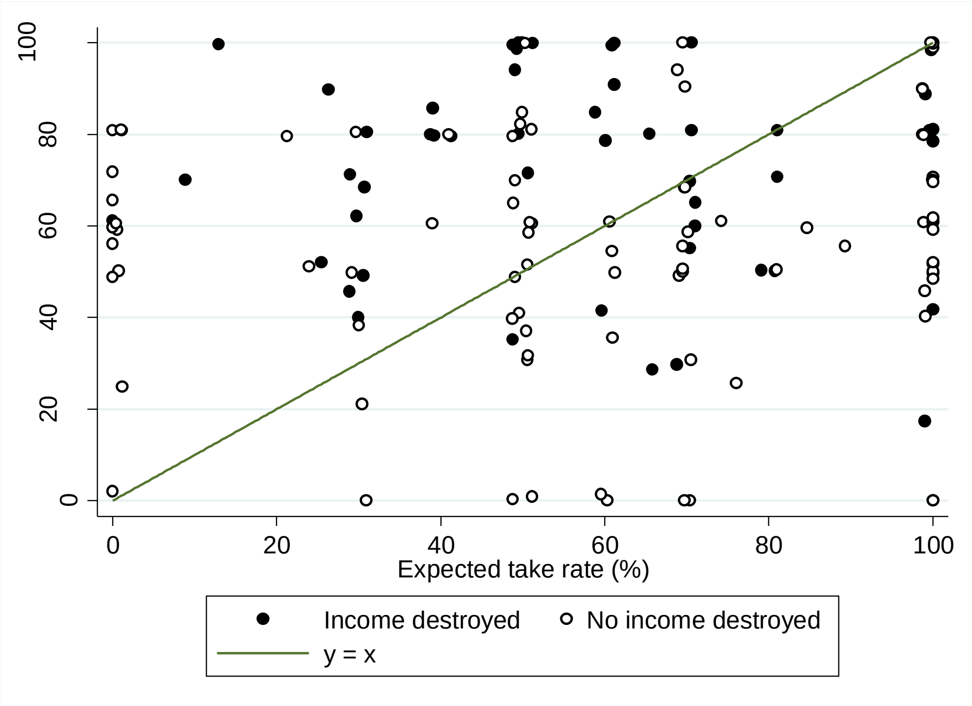
[91] The differences in contempt due to experience may be attributed to a difference in expectations concerning the take rate rather than to a difference in coping. To test for this possibility, we also conducted a tobit regression where we included the expected take rate among the explanatory variables. The result does not change. Hence, we can conclude that more experienced subjects punish less because they are able to better cope with contempt, and not because they have different expectations compared to less experienced subjects.

[92] As in previous PTTG studies, expectations were not incentivized. This is because we did not want to introduce any distortion that would have limited the comparability of our study with the previous literature. For a discussion of the reliability of measuring expectations without financial incentives, see Bosman *et al*. (2005).

expectations occur between UK and non-UK subjects (Mann-Whitney p = 0.242).

Figure 5 shows that responders who punish are generally subjects who expect lower take rates than the actual ones (dots above the 45° line). This is consistent with previous findings on the PTTG. In particular, 37 out of 77 optimistic responders (48.05%) punish the take authority, whereas only 17 out of 64 pessimistic responders (26.56%) punish. The difference achieves statistical significance ($\chi^2 = 6.83$, p = 0.009).[93]

Figure 5: Scatter plot of expected and actual take rates



Notes: optimistic responders (who expect a higher take rate than the actual take rate) are identified by dots above the 45° line; pessimistic responders (who expect a lower take rate than the actual take rate) are identified by dots below the 45° line.

We also study whether emotions and behavior are affected by expectations. In particular, we include expectations in each regression of

[93] The result holds if we conduct a separate test for the treatment with a variable 'fine-to-fee' ratio ($\chi^2 = 3.544$, $p = 0.060$) and one for the treatment with a constant 'fine-to-fee' ratio ($\chi^2 = 3.077$, $p = 0.079$). Since expectations were elicited at the end of the experiment, it might be possible that subjects wrongly reported them. In particular, too optimistic responders might have found difficult to admit that they were wrong. If such bias exists, we should observe a correlation between expected take rates and actual take rates. However, this correlation is low and not significant (Spearman $\rho = 0.025$, p = 0.772).

Table 6.[94] It turns out that expectations have a significant positive impact on envy and jealousy respectively (p = 0.031 and 0.015 respectively). In other words, responders who expect higher take rates from the take authority, especially in comparison to the actual take rate, are more envious and jealous when they learn about the decision of their counterpart. Finally, we include expectations in the regression of Tables 5 and 7. However, they do not seem to play a significant role in explaining the punishment behavior of the responders.

5. Discussion and conclusion

This study contributes to the experimental literature in economics that looks at the role of emotions on the decision to punish. In particular, we investigated whether previous findings about emotions and behavior in the PTTG were confounded by the punishment technology adopted, and to what extent punishment can be truly attributed to negative emotions. We complemented this important analysis by also testing the robustness of our findings against possible differences due to the different backgrounds of the subjects, and whether experience in previous economic experiments affects how emotions drive the punishing behavior of the subjects.

Our results provide clear-cut evidence that previous PTTG studies provided an inflated measure of the punishment. In particular, in the standard PTTG, as much as 70% of the punishment triggered by the take rate is attributable to the multiplier effect caused by the variable ‘fine-to-fee’ ratio.[95] This confirms Hypothesis 1. When we turn to the role played by negative emotions, we find that, consistently with Hypothesis 2, they are

[94] We tried different specifications to account for expectations. In one specification, we simply add among the explanatory variables the expected take rate. In another specification, we include the difference between the take rate and the expected take rate. The results are the same.

[95] If we look at the behavior of the take authorities, we do not find any significant difference between treatments under a variable or constant ‘fine-to-fee’ ratio once covariates are controlled for. In particular, the take authorities do not seem to anticipate the fact that, under a constant ‘fine-to-fee’ ratio, responders punish less when the take rates are higher. A possible explanation is that the take authorities feel guiltier to appropriate too much money when the ‘fine-to-fee’ ratio is constant because it is more costly for the responders to punish compared to a situation with a variable ‘fine-to-fee’ ratio. Alternatively, the take authorities might adopt a general norm of fairness on how to split the resources that, conditional to no destruction from the responders, should apply equally to all of the treatments.

still important predictors of the punishing behavior in the PTTG. In particular, irritation, anger, and contempt appear to be important driving forces for the punishing behavior of the responders, especially in response to high take rate, and even once we control for the punishment technology and cultural background of the subjects. However, as it is postulated by Hypothesis 3, in the PTTG with a variable 'fine-to-fee' ratio, their effect is overstated by as much as 90% for high take rates. This means that, in the previous literature, a confound exists and it affects both the punishing behavior, and the extent to which the latter is driven by negative emotions.

If we turn to the cultural background of the subjects, we do not find any support for Conjecture 1. In particular, UK and non-UK subjects experience similar emotions, and their punishment is motivated by the same underlying visceral states. The only significant effect of the subjects' cultural background is on the take authorities' behavior. In particular, we find that non-UK students in the role of take authorities appropriate more resources than UK students, particularly when the punishment technology embeds a variable 'fine-to-fee' ratio. As we have already mentioned earlier, the purpose of this study was not to study the behavior of the take authorities, and how this may vary depending on the cultural background of the subjects. Different factors may explain why non-UK students appropriated more resources than UK students. We mention here a few. First of all, non-UK students might be more sensitive to social distance since, as a minority, they are more likely to be matched with students from a different country than their own. Several experimental studies have indeed shown that as social distance increases people become more anti-social (e.g. Buchan and Croson, 2004; Charness and Gneezy, 2008). Another explanation might be related to potential cultural differences between Western societies and non-Western societies. Note that almost all the non-UK students who played in the role of take authority were from non-Western societies (in particular East Asia) and they were also those who displayed the highest take rates. This however is a rather speculative explanation, since it is based on an extreme cultural separation. In addition, we were not able to control for other unobserved variables which might instead better explain the difference in behavior. For instance, non-UK students are more likely to represent the wealthiest subset of the population of their countries of origin, and this might explain their different behavior. Similarly, the fact that non-UK students are studying abroad could indicate that they are a different self-selected subsample of their native populations in terms of risk attitude. In particular, non-UK students might be less risk averse, which could explain why they choose higher take rates.[96] Future research could investigate what are the specific differences between university subject pools that explain the differences in behavior.

Finally, we find that experience in previous economic experiments has a marginal impact on our experiment. In particular, the more experienced subjects do not appear to cope better with their anger and irritation when these drive subjects to inefficient behaviors compared to less experienced subjects. Only the impact of contempt on punishing behavior appears to be lower for the more experienced subjects, consistently with Conjecture 2.

To conclude, our findings contribute to the current state of the experimental literature in economics, particularly from a methodological point of view. First, our results suggest that a large part of the punishment behavior observed in previous PTTG studies is explained by the technology of punishment adopted, and that the role played by emotions is overstated. Second, we find that the cultural background of the subjects does not seem to be particularly relevant for experiments which use university students to investigate the emotional basis of economic punishing behavior. Finally, we find some evidence that *experimental learning* works as a potential moderator for the impact of contempt – but not for that of anger and irritation – on punishing behavior, without however nullifying its effect.

[96] Another possible explanation is that non-UK students might have found it more difficult to understand the instructions because of the language gap. However, we do not find any evidence of that. In particular, the number of mistakes in the comprehension questionnaire was the same between the UK and non-UK subjects (Mann-Whitney p = 0.627). In addition, both UK and non-UK students with incorrect answers were equally provided with individual clarification in order to ensure that everyone understood the instructions.

References

Albert, M., & Mertins, V. (2008). Participation and decision making: a three-person power-to-take experiment. *Justus-Liebig-University Giessen Discussion Paper Series in Economics 05-2008*.

Andreoni, J. (1995). “Cooperation in public-goods experiments: kindness or confusion?” *American Economic Review, 85*(4), 891-904.

Ben-Shakhar, G., Bornstein, G., Hopfensitz, A., & van Winden, F. (2007). Reciprocity and emotions in bargaining using physiological and self-report measures. *Journal of Economic Psychology, 28*(3), 314-323.

Bosman, R., & Riedl, A. (2004). Emotions and economic shocks in a first-price auction: an experimental study. *University of Amsterdam Working Paper*.

Bosman, R., & van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *Economic Journal, 112*(476), 147-169.

Bosman, R., Hennig-Schmidt, H., & van Winden, F. (2006). Exploring group decision making in a power-to-take experiment. *Experimental Economics*, *9*(1), 35-51.

Bosman, R., Sutter, M., & van Winden, F. (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology, 26*(3), 407-429.

Brown, M. B., & Forsythe, A. B. (1974). Robust Tests for Equality of Variances. *Journal of the American Statistical Association*, 69, 364–367.

Buchan, N., & Croson, R. (2004). The boundaries of trust: own and others' actions in the US and China. *Journal of Economic Behavior & Organization, 55*(4), 485-504.

Casari, M. (2005). On the design of peer punishment experiments. *Experimental Economics, 8*(2), 107-115.

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization, 68*(1), 29-35.

Cubitt, R. P., Drouvelis, M., & Gächter, S. (2011). Framing and free riding: emotional responses and punishment in social dilemma games. *Experimental Economics, 14*(2), 254-272.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science, 305*(5688), 1254-1258.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137-140.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817-868.

Frank, R. H. (1988). *Passion within reason*. New York: W.W. Norton and Company.

Friedman, J. W. (1969). On experimental research in oligopoly. *The Review of Economic Studies, 36*(4), 399-415.

Galeotti, F. (2013). An experiment of waiting time and punishing behavior. *Economics Bulletin*, 33(2), 1383-1389.

Gächter, S., Herrmann, B. and Thöni, C. 2005. Cross-cultural differences in norm enforcement. *Behavioral and Brain Sciences*, *28*(06), 822-823.

Greiner, B. (2004). The online recruitment system ORSEE 2.0 - A guide for the organization of experiments in economics. *Working Paper Series in Economics 10, University of Cologne, Department of Economics*.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*(5868), 1362-1367.

Henrich, J., McElreath, R. Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. 2006. Costly punishment across human societies. *Science, 312*(5781), 1767-1770.

Hopfensitz, A., and Reuben, E. (2009). The importance of emotions for the effectiveness of social punishment. *Economic Journal, 119*(540), 1534-1559.

Joffily, M., Masclet, D., Noussair, C. N., and Villeval, M. (2013). Emotions, sactions and cooperation. *Southern Economic Journal*.

Kagel, J. H., and Levin, D. 1999. Common value auctions with insider information. *Econometrica*, *67*(5), 1219-1238.

Kräkel, M. (2008). Emotions in tournaments. *Journal of Economic Behavior & Organization*, *67*(1), 204-214.

Levitt, S. D., and List, J. A. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 153-174.

List, J. A., and Levitt, S. D. 2005. What do laboratory experiments tell us about the real world. *NBER Working Paper*.

Loewenstein, G. F. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, *90(2*), 426-432.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127(2*), 267-286.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions - a review. *Psychological Bulletin, 112*(2), 179-204.

Myagkov, M. and Plott, C. 1997. Exchange Economies and Loss Exposure: Experiments Exploring Prospect Theory and Competitive Equilibria in Market Environments. *American Economic Review*, *87*(5), 801-828.

Nikiforakis, N., & Normann, H. T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics, 11*(4), 358-369.

Phelps, E. A. (2009). The study of emotion in neuroeconomics. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics* (pp. 233-250). London: Elsevier.

Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes, 68*(3), 208-224.

Rabin, M. (1993). Incorporating fairness into game-theory and economics. *American Economic Review, 83*(5), 1281-1302.

Reuben, E., & van Winden, F. (2008). Social ties and coordination on negative reciprocity: the role of affect. *Journal of Public Economics, 92*(1-2), 34-53.

Reuben, E., & van Winden, F. (2010). Fairness perceptions and prosocial emotions in the power to take. *Journal of Economic Psychology, 31*(6), 908-922.

Rick, S., & Loewenstein, G. F. (2008). The role of emotion in economic behavior. In M. Lewis, J. M. Haviland-Jones, & L. Feldman-Barrett (Eds.), *The handbook of emotion* (3rd ed.). New York: Guilford.

Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining - beware the same sex! *Experimental Economics*, *12*(3), 318-331.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75-98.

Zizzo, D. J. (2013). Claims and confounds in economic experiments. *Journal of Economic Behavior & Organization*, *93*(13), 186-195.

Appendix to Chapter 2: On the Robustness of Emotions and Behavior in a Power-to-Take Game Experiment

A. Experimental instructions
B. Form
C. Background information on participants
D. Theoretical predictions
E. Logit regressions on the decision to punish
F. Ordered logit regressions on emotions (without demographics)
G. Tobit regressions on punishment for all emotions
H. Tobit regressions on punishment for all emotions (without demographics)
I. Logit regressions on punishment for all emotions
J. Emotions of the take authority

A. Experimental Instructions

VARIABLE ‘FINE-TO-FEE’ RATIO TREATMENT

INSTRUCTIONS

Introduction
This is an experiment on decision making. During the experiment, you are not allowed to communicate with other participants. Please raise your hand if you have any questions at any point in the experiment.

The experiment is expected to last no more than 60 minutes. All the money that you will earn during this experiment will be paid to you in cash at the end of this experiment.

All the participants in the experiment have received the same set of instructions as you have. Each participant has been assigned randomly the role of **participant A** or **participant B**.

Each participant in this experiment receives a show up fee of 5 pounds.

Initial endowment
In this experiment each participant, participant A as well as participant B, will receive an endowment of 5 pounds.

Two phases

The experiment consists of two phases. In phase 1 only participant A must make a decision whereas in phase 2 only participant B must make a decision. Every participant thus makes one decision.

Phase 1: participant A chooses percentage

In this phase, each participant A will be paired with a participant B. This will be done by letting participant A draw a coded envelope. With the help of the code only we know which seat numbers are paired. Both participant A and B are thus anonymous. The envelope contains a form. Participant A must choose a percentage and fill this in on the form. This percentage determines how much of participant B's endowment after phase 2 will be transferred to participant A. The percentage chosen by participant A must be a number between and including 0 and 100.

When participant A has completed the form, it must be put in the envelope again. After this we will collect the envelopes and bring them to the participants B who are paired with the participants A by means of the code.

Phase 2: participant B chooses percentage

In this phase participant B has to fill in on the form which percentage of his or her *own* endowment of 5 pounds will be destroyed. The percentage chosen by participant B must be a number between and including 0 and 100.

The transfer from participant B to participant A will be based on the endowment of participant B that is left. Note that the transfer equals the percentage chosen by participant A of the endowment of participant B that is left after phase 2.

When participant B has completed the form, it must be put in the envelope again. After this we will collect the envelopes and bring them to the participants A who are paired with the participants B. Participant A will take note of the decision of participant B and, subsequently, puts the form back into the envelope. Finally, the envelopes will be collected for the payment procedure which will be clarified below.

Example how to determine one's payoffs

We will now give an example for the purpose of illustration. As you know both participant A and participant B have an endowment of 5 pounds. Suppose participant A decides that *60%* of the endowment of participant B will be transferred to him or her (participant A). In the second phase, participant B can destroy part or everything of his or her endowment. Suppose participant B decides to destroy zero percent of his or her endowment. The transfer from B to A is then equal to 3 pounds (60% of 5 pounds). The total payoff for B at the end of the experiment is equal to 7 pounds (namely, the show-up fee of 5 pounds plus the endowment of 5 pounds minus the transfer of 3 pounds). The total payoff for A at the end of the experiment is equal to 13 pounds (namely, the show up fee of 5 pounds plus the endowment of 5 pounds plus the transfer of 3 pounds).

Now suppose that in this example participant B had decided to destroy 50% of his or her own endowment. In this case the transfer from B to A is only 1 pound and 50 pence (namely, 60% of the remaining endowment of participant B after phase II, which is 60% of 2 pounds and 50 pence). The total payoff for A at the end of the experiment is equal to 11 pounds and 50 pence (namely, the show up fee of 5 pounds plus the endowment of 5 pounds plus the transfer of 1 pound and 50 pence) and for participant B 6 pounds (namely, the show-up fee of 5 pounds plus the remaining endowment of 2 pounds and 50 pence after destruction minus the transfer of 1 pounds and 50 pence).

In summary

In phase 1, each participant A will be paired with a participant B by drawing an envelope. The envelope contains a Form. Participant A fills in a percentage that indicates how much of participant B's endowment will be transferred to participant A. When participant A has completed the form, it will be brought to participant B. In phase 2, participant B decides which percentage of his or her *own* endowment will be destroyed, and fills this in on the Form. Subsequently, the Form will go to participant A who takes note of the decision of participant B. Then, the Form will be collected and the payment procedure follows. Note, that the pairing is anonymous so that nobody knows whom he or she is paired with.

Other information

Completing the Form

The decision of both participant A and B will be filled in on a Form. You have received a specimen of this Form. In phase 1, participant A completes the blue block. In phase 2, participant B completes the yellow block. The Forms must be completed with the pen that you find on your table in the laboratory. If a Form has been completed with another pen, the Form will be invalid and you will not be paid.

Finally, for making calculations you can make use of the electronic calculator that is on your table.

The payment procedure

When participant A has taken note of the decision of participant B in phase 2, the envelope containing the Form will be collected and brought to the cashier. Next, the participants will go to the reception room of the laboratory one by one. The cashier, who will not be present during the experiment, will pay the participants in the reception room. The cashier determines the payment of each participant with the help of the Form and the codes that are linked to the seats. In this way, anonymity is secured with regard to who earned what.

Exercises
We ask you to do two exercises in order to become familiar with the procedures. These exercises consist of completing the Form for an imaginary situation and determining the payoffs. You are not actually paired with another participant during these exercises. Your earnings in these exercises will not be paid out to you. When the exercises have been finished, you have the opportunity to ask questions again. After this the experiment will start.
Finally
To secure anonymity, participants A and B will be divided by partitions. The instructions on the table will be available to you during the experiment. At the end of the experiment you are asked to fill in a short questionnaire. Anonymity is again secured. After this, you are asked to leave the laboratory one by one. You must be silent and refrain from communication with others until you have left the laboratory.

COSTANT 'FINE-TO-FEE' RATIO TREATMENT

INSTRUCTIONS

Introduction
This is an experiment on decision making. During the experiment, you are not allowed to communicate with other participants. Please raise your hand if you have any questions at any point in the experiment.

The experiment is expected to last no more than 60 minutes. All the money that you will earn during this experiment will be paid to you in cash at the end of this experiment.

All the participants in the experiment have received the same set of instructions as you have. Each participant has been assigned randomly the role of **participant A** or **participant B**.

Each participant in this experiment receives a show up fee of 5 pounds.

Initial endowment
In this experiment each participant, participant A as well as participant B, will receive an endowment of 5 pounds.

Two phases
The experiment consists of two phases. In phase 1 only participant A must make a decision whereas in phase 2 only participant B must make a decision. Every participant thus makes one decision.

Phase 1: participant A chooses percentage
In this phase, each participant A will be paired with a participant B. This will be done by letting participant A draw a coded envelope. With the help of the code only we know which seat numbers are paired. Both participant

A and B are thus anonymous. The envelope contains a form. Participant A must choose a percentage and fill this in on the form. This percentage determines how much of participant B's endowment after phase 2 will be transferred to participant A. The percentage chosen by participant A must be a number between and including 0 and 100.

When participant A has completed the form, it must be put in the envelope again. After this we will collect the envelopes and bring them to the participants B who are paired with the participants A by means of the code.

Phase 2: participant B chooses percentage

In this phase participant B has to fill in on the form which percentage of his or her *own* endowment of 5 pounds will be destroyed. For each 1% of his or her endowment that participant B decides to destroy, 10 pence of the transfer to participant A will be destroyed as well. The percentage chosen by participant B must be a number between and including 0 and the maximum percentage required to destroy all the transfer to participant A.

The transfer from participant B to participant A will be based on the initial endowment of participant B (that is 5 pounds). Note that the transfer equals the percentage chosen by participant A of participant's B initial endowment of 5 pounds. However, participant A will receive 10 pence less for each 1% that participant B decides to destroy of his or her initial endowment of 5 pounds.

When participant B has completed the form, it must be put in the envelope again. After this we will collect the envelopes and bring them to the participants A who are paired with the participants B. Participant A will take note of the decision of participant B and, subsequently, puts the form back into the envelope. Finally, the envelopes will be collected for the payment procedure which will be clarified below.

Example how to determine one's payoffs

We will now give an example for the purpose of illustration. As you know both participant A and participant B have an endowment of 5 pounds. Suppose participant A decides that *60%* of the endowment of participant B will be transferred to him or her (participant A). In the second phase, participant B can destroy up to 30% of his or her endowment. Suppose participant B decides to destroy zero percent of his or her endowment. The transfer from B to A is then equal to 3 pounds (60% of 5 pounds). The total payoff for B at the end of the experiment is equal to 7 pounds (namely, the show-up fee of 5 pounds plus the endowment of 5 pounds minus the transfer of 3 pounds). The total payoff for A at the end of the experiment is equal to 13 pounds (namely, the show up fee of 5 pounds plus the endowment of 5 pounds plus the transfer of 3 pounds).

Now suppose that in this example participant B had decided to destroy 15 % of his or her own initial endowment. In this case the transfer from B to A is still equal to 3 pounds (60% of 5 pounds). However, participant A receives only 1 pound and 50 pence of this transfer (namely, 60% of the endowment of participant B, which is 60% of 5 pounds, minus 10 pence for each 1% that Participant B has destroyed of his or her initial endowment, which is 10 pence × 15 = 1 pound and 50 pence). The total payoff for A at the end of the experiment is equal to 11 pounds and 50 pence (namely, the show up fee of 5 pounds plus the endowment of 5 pounds plus the remaining transfer of 1 pound and 50 pence) and for participant B 6 pounds and 25 pence (namely, the show-up fee of 5 pounds plus the remaining endowment of 4 pounds and 25 pence after destruction minus the transfer of 3 pounds).

In summary

In phase 1, each participant A will be paired with a participant B by drawing an envelope. The envelope contains a Form. Participant A fills in a percentage that indicates how much of participant B's endowment will be transferred to participant A. When participant A has completed the form, it will be brought to participant B. In phase 2, participant B decides which percentage of his or her *own* endowment will be destroyed, and fills this in on the Form. Subsequently, the Form will go to participant A who takes note of the decision of participant B. Then, the Form will be collected and the payment procedure follows. Note, that the pairing is anonymous so that nobody knows whom he or she is paired with.

Other information

Completing the Form

The decision of both participant A and B will be filled in on a Form. You have received a specimen of this Form. In phase 1, participant A completes the blue block. In phase 2, participant B completes the yellow block. The Forms must be completed with the pen that you find on your table in the laboratory. If a Form has been completed with another pen, the Form will be invalid and you will not be paid.

Finally, for making calculations you can make use of the electronic calculator that is on your table.

The payment procedure

When participant A has taken note of the decision of participant B in phase 2, the envelope containing the Form will be collected and brought to the cashier. Next, the participants will go to the reception room of the laboratory one by one. The cashier, who will not be present during the experiment, will pay the participants in the reception room. The cashier determines the payment of each participant with the help of the Form and the codes that are linked to the seats. In this way, anonymity is secured with regard to who earned what.

Exercises

We ask you to do two exercises in order to become familiar with the procedures. These exercises consist of completing the Form for an imaginary situation and determining the payoffs. You are not actually paired with another participant during these exercises. Your earnings in these exercises will not be paid out to you. When the exercises have been finished, you have the opportunity to ask questions again. After this the experiment will start.

Finally

To secure anonymity, participants A and B will be divided by partitions. The instructions on the table will be available to you during the experiment. At the end of the experiment you are asked to fill in a short questionnaire. Anonymity is again secured. After this, you are asked to leave the laboratory one by one. You must be silent and refrain from communication with others until you have left the laboratory.

B. Form

Code: ______

FORM

Participant A fills in this block:

Endowment of participant A: 5 pounds

Endowment of participant B: 5 pounds

I (participant A) decide that ………. % of the endowment of participant B will be transferred to me.

Participant B fills in this block:

I (participant B) destroy ………. % of my endowment.

C. Background of Experimental Participants

Gender	Frequency	Percent
Female	135	47.87
Male	147	52.13
Total	282	100.00

Degree	Frequency	Percent
INTO	1	0.35
Bachelor	203	71.99
Master	49	17.38
MPhil/PhD	26	9.22
Staff	1	0.35
Other	2	0.70
Total	282	100

	Obs.	Mean	St. Dev.	Min	Max
Age	282	22.34	4.93	18	64

Economics	Frequency	Percent
No	229	81.21
Yes	53	18.79
Total	282	100.00

Religion	Frequency	Percent
No religion	165	58.51
Buddhist	14	4.96
Christian	63	22.34
Confucian	2	0.71
Hindu	6	2.13
Jain	1	0.35
Jewish	1	0.35
Muslim	13	4.61
Other	1	0.35
Prefer not to answer	16	5.67
Total	282	100

Relationship Status	Frequency	Percent
Single	144	51.06
Engaged	5	1.77
In a relationship	97	34.40
Married	17	6.03
Separated/Divorced	1	0.35
Prefer not to answer	18	6.38
Total	282	100

Experience*	Frequency	Percent
0	32	11.35
1	20	7.09
2	19	6.74
3 or more	211	74.82
Total	282	100

* n. of experiments attended in the past.

Nationality	Freq.	Percent
Australia	1	0.35
Azerbaijan	1	0.35
Bahrian	1	0.35
Botswana	1	0.35
Bulgaria	1	0.35
China**	51	18.09
Ghana	1	0.35
Hong Kong	14	4.96
Hungary	2	0.71
India	7	2.48
Indonesia	1	0.35
Japan	2	0.71
Jordania	1	0.35
Kazakistan	3	1.06
Kenya	1	0.35
Liberia	1	0.35
Lithuania	1	0.35
Malaysia	4	1.42
Mexico	2	0.71
Nigeria	7	2.48

Nationality	Freq.	Percent
Pakistan	3	1.06
Palestine	1	0.35
Peru	2	0.71
Poland	2	0.71
Romania	1	0.35
Russia	3	1.06
Singapore	3	1.06
Sri Lanka	4	1.42
Thailand	3	1.06
UK	142	50.35
UK/China*	1	0.35
UK/Hong Kong*	1	0.35
UK/Nigeria*	2	0.71
UK/Pakistan*	1	0.35
UK/Poland*	2	0.71
UK/Uganda*	1	0.35
Venezuela	1	0.35
Vietnam	5	1.77
Zimbabwe	1	0.35
Total	282	100

* Subjects who have been naturalized as British citizens later on in life. In the chapter, we treat them as non-UK subjects. We also analyzed the data by treating these few subjects as UK subjects. The results of the chapter do not change.
** China does not include Hong Kong in the table.

D. Theoretical predictions

Rational and self-interested behavior

As we have already explained in the chapter, if the responder is rational and profiting-maximizing, he or she should never punish if the take rate is less than 1, and should be indifferent between all possible punishment rates if the take rate is 1. This is irrespectively of whether the 'fine-to-fee' ratio is constant or variable.

Inequity aversion (Fehr and Schmidt, 1999)

We use the inequality aversion model of Fehr and Schmidt (1999) to describe the behavior of responders who care about equality. According to this model, the utility function of a responder is:

$$U_i(p_i, t_j) = \pi_i(p_i, t_j) - \alpha_i \max\left(\pi_j(p_i, t_j) - \pi_i(p_i, t_j), 0\right) \\ - \beta_i \max\left(\pi_i(p_i, t_j) - \pi_j(p_i, t_j), 0\right)$$

where p_i is the level of punishment of the responder i, t_j the take rate of the take authority j, $\pi_i(\cdot)$ the profit of the responder, and $\pi_j(\cdot)$ the profit of the take authority. We also assume, as in Fehr and Schmidt (1999), that $0 \leq \beta_i \leq \alpha_i$ and $\beta_i < 1$. The first term of the utility function is the material payoff of the responder; the second term the utility loss which stems from disadvantageous inequality, while the third term the utility loss which results from advantageous inequality. Note that, in the PTTG, $\pi_i(p_i, t_j) \leq \pi_j(p_i, t_j)$ for any p_i and t_j. In other words, there cannot be inequality that favors the responder. Hence, the third term of the utility is always equal to zero. We can re-write the utility function as:

$$U_i(p_i, t_j) = \pi_i(p_i, t_j) - \alpha_i\left[\pi_j(p_i, t_j) - \pi_i(p_i, t_j)\right]$$

Under a variable 'fine-to-fee' ratio, the utility is:

$$U_i(p_i, t_j) = 5(1 - p_i)(1 - t_j) - \alpha_i\left[5 + 5t_j(1 - p_i) - 5(1 - p_i)(1 - t_j)\right]$$

The responder selects p_i that maximizes his or her utility, under the constraint $0 \leq p_i \leq 1$. The optimal level of punishment depends on t_j. In particular, the responder selects either full punishment or no punishment at all. We can identify the value of t_j above which the responder select a full punishment and below which he or she does not punish at all. We call this level ${t_j}^{V*}$. This is equal to:

$$ {t_j}^{V*} = \frac{1+\alpha_i}{1+2\alpha_i} $$

For $t_j > {t_j}^{V*}$, $p_i = 1$; for $t_j < {t_j}^{V*}$, $p_i = 0$; For $t_j = {t_j}^{V*}$, the responder is indifferent between $p_i = 0$ and $p_i = 1$. Hence, the greater is t_j, the more likely is the responder to punish.

If we now turn to the case where the 'fine-to-fee' ratio is equal to 2, the utility function of the responder can be written as:[97]

$$ U_i(p_i, t_j) = 5\left(1 - \frac{t_j p_i}{2}\right)(1 - t_j) - \alpha_i\left[5 + 5t_j(1 - p_i) - 5\left(1 - \frac{t_j p_i}{2}\right)(1 - t_j)\right] $$

Let ${t_j}^{C*}$ the value of t_j above which the responder punishes the take authority and below which he or she does not punish at all:[98]

$$ {t_j}^{C*} = \frac{1-\alpha_i}{1+\alpha_i} $$

Also under a constant 'fine-to-fee' ratio, the greater is t_j, the more likely is the responder to punish.

Figure 1B displays the levels of t_j at which the responder selects $p_i = 1$, or $p_i = 0$, for different values of α_i. In the figure, we also distinguish between a variable 'fine-to-fee' ratio and a 'fine-to-fee' ratio equal to 2. The white area below the line ${t_j}^{V*}$ identifies values of t_j at which the responder destroys all the transfer to the take authority both in the case

[97] Note that, under a 'fine-to-fee' ratio equal to 2, the punishment rate is $p_i = (10d_i)/(5\,t_i)$, where d_i is the destruction rate (for every 5 pence that the responder destroys of his or her own income, 10 pence of the transfer to the take authority are destroyed as well). Solving for d_i, we obtain: $d_i = t_j p_i/2$. For a generic constant 'fine-to-fee' ratio, $p_i = \sigma\, d_i/t_i$.

[98] For a generic constant 'fine-to-fee' ratio, ${t_j}^{C*} = \frac{1+\alpha_i-\sigma\alpha_i}{1+\alpha_i}$.

of a variable and constant 'fine-to-fee' ratio. The dotted white area between the lines ${t_j}^{V*}$ and ${t_j}^{C*}$identifies values of t_j at which the responder punish the take authority only when the 'fine-to-fee' ratio is constant. The dark dotted area below the line ${t_j}^{C*}$ identifies values of t_j at which the responder does not punish at all. For any level of α_i, and both under a constant and variable 'fine-to-fee' ratio, the higher is the take rate, the more likely is the responder to punish. Also, the greater is the sensitiveness of the responder to disadvantageous inequality (α_i), the lower is the t_j above which the responder punishes the take authority.

Figure D1: Punishment of responders who care about equality as a function of t_j and α_i

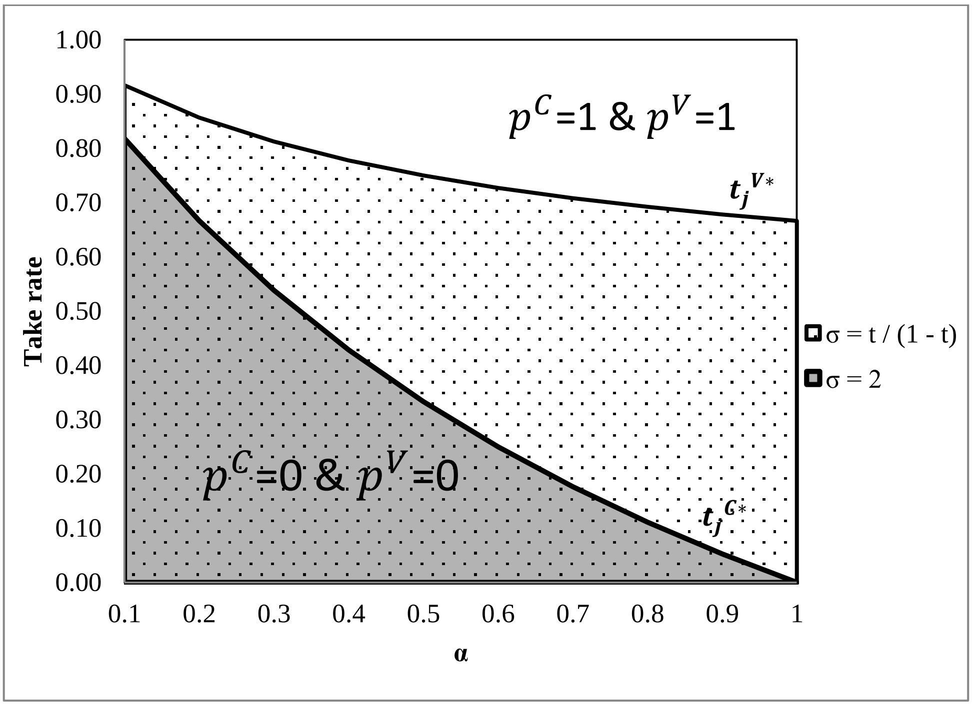
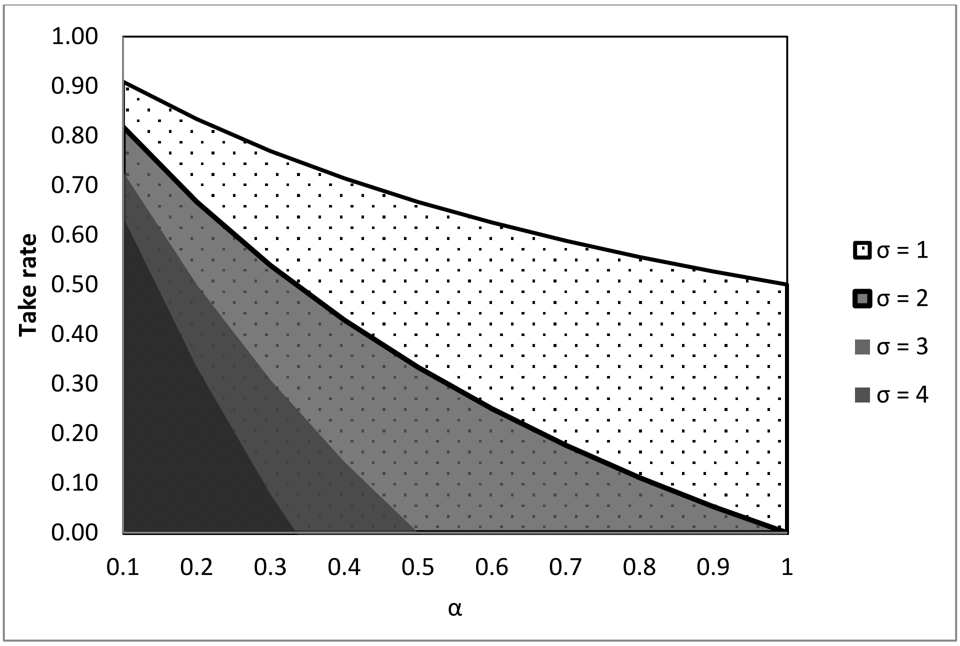


Figure 2B displays the levels of t_j at which the responder selects $p_i = 1$, or $p_i = 0$, for different values of α_i, and different constant 'fine-to-fee' ratios. The higher is the 'fine-to-fee' ratio, the more likely is the responder to punish when the take rate is high.

Figure D2: Punishment of responders who cares about equality as a function of t_j, α_i, and $\bar{\sigma}$



Reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004)

To understand the implications of reciprocity on the responder's decision to punish, we apply the model of Dufwenberg and Kirchsteiger (2004) developed to capture reciprocity in sequential games.[99] We can define the utility function of the responder *i*, evaluated at the second stage of the PTTG, as:

$$U_i\left(p_i, t_j, \left(c_{iji}\right)_{j\neq i}\right)$$
$$= \pi_i\left(p_i, t_j\right) + Y_{ij}$$
$$\cdot\left(\pi_j\left(p_i, t_j\right) - \frac{1}{2}\left(\max_{p_i\in P_i}\pi_j\left(p_i, t_j\right) + \min_{p_i\in P_i}\pi_j\left(p_i, t_j\right)\right)\right)$$
$$\cdot\left(\pi_i\left(t_j, c_{iji}\right) - \frac{1}{2}\left(\max_{t_j\in T_j}\pi_i\left(t_j, c_{iji}\right) + \min_{t_j\in T_j}\pi_i\left(t_j, c_{iji}\right)\right)\right)$$

[99] This model is an adaption of Rabin (1993)'s theory of reciprocity. The model of Rabin (1993) can be applied to normal form games with two players. The model of Dufwenberg and Kirchsteiger (2004) can be applied to a larger set of games, including games with more than 2 players, and extensive games.

where p_i is the level of punishment of the responder i, t_j the take rate of the take authority j, and c_{iji} the responder's belief of the take authority's belief about the responder's decision after the take authority has chosen t_j.[100] The first term of the equation identifies the material payoff of the responder, while the second term the *reciprocity payoff*. Y_{ij} is the a constant which captures the sensitiveness of *i* to reciprocity concerns. The term in the first parenthesis identifies the *kindness* of *i* to *j*, and is measured as the difference between the material payoff of j for a given level of punishment p_i chosen by *i*, and the so-called "equitable payoff" for *j*, that is the average between the lowest and highest material payoff of *j* that could result from the possible punishment choices of *i*. The term in the second parenthesis measures the belief of *i* about the kindness of *j*, and is given by the difference between the belief of *i* about how much material payoff *j* intends to give to *i* by choosing t_j, and the "equitable payoff" for *i*, that is the average between the lowest and highest material payoff of *i* that could have resulted from the possible choices of *j*.

To simplify the analysis, we assume that *i* believes that *j* expects no punishment from him or her, that is $c_{iji} = 0$.[101] Under a variable 'fine-to-fee' ratio, the utility function can be written as:

$$U_i(p_i, t_j) = 5(1-p_i)(1-t_j) + Y_{ij} \cdot \left(5 + 5(1-p_i)t_j - \frac{1}{2}(5 + 5t - 5)\right) \cdot \left(5(1-t) - \frac{5}{2}\right)$$

We can maximize this function with respect to p_i, subject to $0 \leq p_i \leq 1$, to obtain the optimal level of punishment for the responder. This optimal level depends on t_j. In particular, we can identify the value of

[100] Note that we do not include the belief of the responder about the behavior of the take authority. This is because we are already in the second node of the game tree, where the take authority has already chosen a level t_j, and the responder already knows what the take authority has done.

[101] This assumption is quite realistic since the take authority has no reason to induce a punishment from the responder. This is true for rational and self-interested take authorities, but also socially motivated take authorities. Note also that this simplification does not change the results in terms of the relationship between punishment and take rate. The only implication is that the optimal punishment rate can be only either 0 or 1. If we relax this assumption, the optimal punishment rate can also take values between 0 and 1.

t_j above which the responder select a full punishment and below which he or she does not punish at all. We call this level ${t_j}^{V*}$. This is equal to:

$$ {t_j}^{V*} = \frac{\frac{25}{2}Y_{ij} - 5 + \sqrt{25 + \frac{625}{4}{Y_{ij}}^2 + 375Y_{ij}}}{50Y_{ij}}. $$

For $t_j > {t_j}^{V*}$, $p_i = 1$; for $t_j < {t_j}^{V*}$, $p_i = 0$; For $t_j = {t_j}^{V*}$, the responder is indifferent between $p_i = 0$ and $p_i = 1$. Hence, the greater is t_j, the more likely is the responder to punish.

We can do the same exercise assuming now that the 'fine-to-fee' ratio is equal to 2. The utility function of the responder is:

$$ U_i(p_i, t_j) = 5(1 - \frac{t_j p_i}{2})(1 - t_j) + Y_{ij} \cdot \left(5 + 5(1 - p_i)t_j - \frac{1}{2}(5 + 5t - 5)\right) \cdot \left(5(1 - t) - \frac{5}{2}\right) $$

In this case, we call ${t_j}^{C*}$ the value of t_j above which the responder punishes the take authority, and below which he or she does not punish at all:[102]

$$ {t_j}^{C*} = \frac{\frac{1}{2} + \frac{5}{2}Y_{ij}}{\frac{1}{2} + 5Y_{ij}} $$

Similar to the case where the 'fine-to-fee' ratio is variable, the greater is t_j, the more likely is the responder to punish.

Figure 3B displays the levels of t_j at which the responder selects $p_i = 1$, or $p_i = 0$, for different values of Y_{ij}, and distinguishing between a variable 'fine-to-fee' ratio and a 'fine-to-fee' ratio equal to 2. The white area below the line ${t_j}^{V*}$ identifies values of t_j at which the responder destroys all the transfer to the take authority both in the case of a variable and constant 'fine-to-fee' ratio. The dotted white area between the lines ${t_j}^{V*}$ and ${t_j}^{C*}$identifies values of t_j at which the responder punish the take authority

[102] For a generic constant 'fine-to-fee' ratio, ${t_j}^{C*} = \frac{\frac{5}{\sigma} + \frac{25}{2}Y_{ij}}{\frac{5}{\sigma} + 25Y_{ij}}$.

only when the 'fine-to-fee' ratio is constant. The dark dotted area below the line ${t_j}^{C*}$ identifies values of t_j at which the responder does not punish at all. For any level of Y_{ij}, and both under a constant and variable 'fine-to-fee' ratio, the higher is the take rate, the more likely is the responder to punish. Also, the greater is the sensitiveness of the responder to reciprocity concerns, the lower is the t_j above which the responder punishes the take authority.

Figure D3: Punishment of reciprocal responders as a function of t_j and Y_{ij}

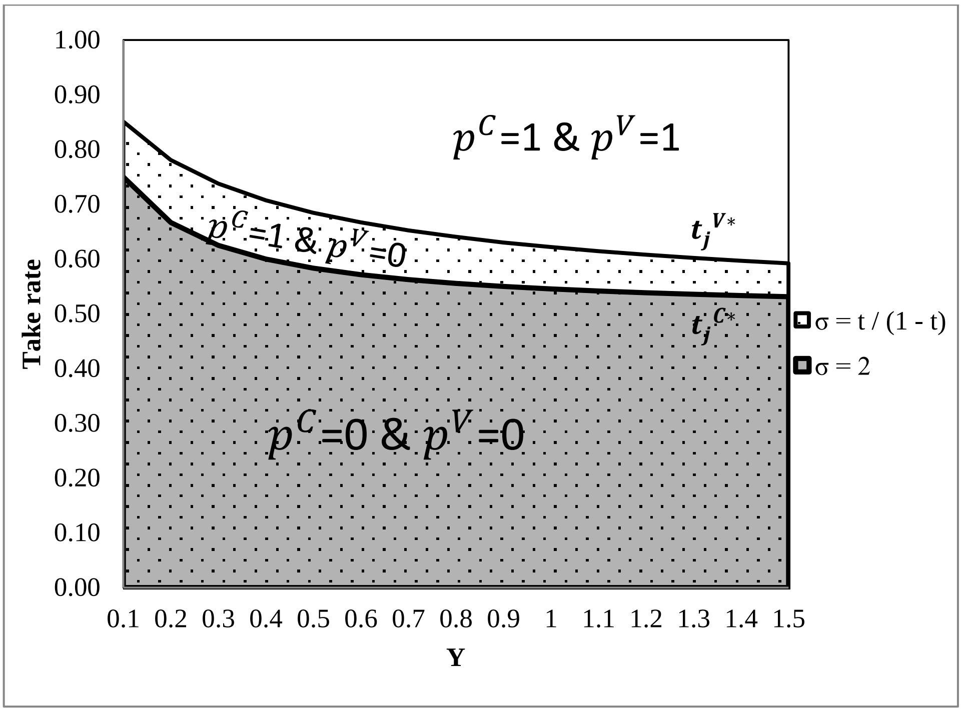
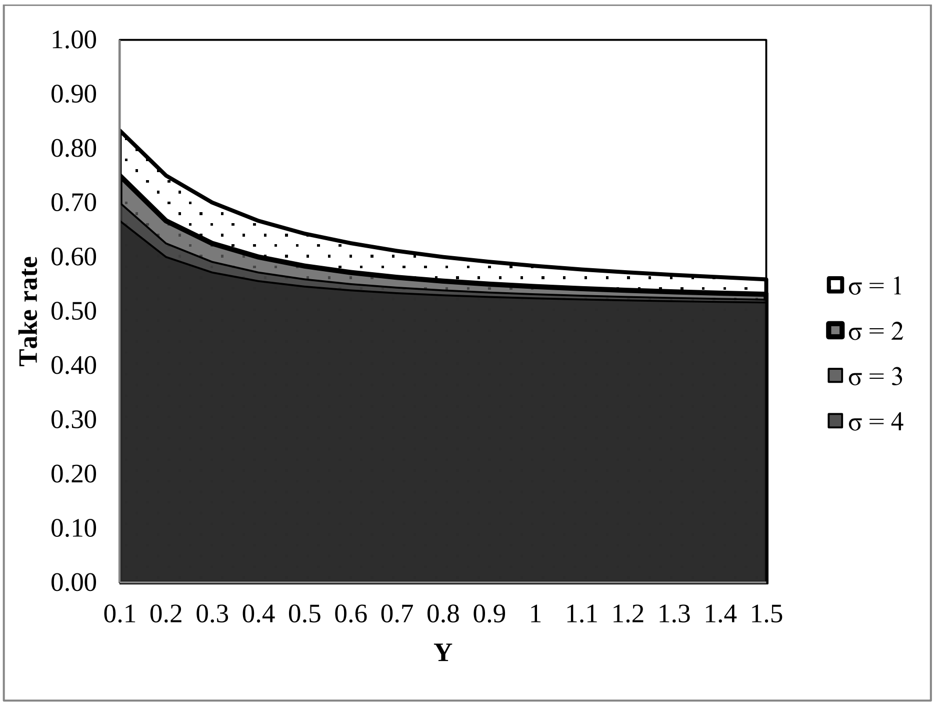


Figure 4B displays the levels of t_j at which the responder selects $p_i = 1$, or $p_i = 0$, for different values of Y_{ij}, and different constant 'fine-to-fee' ratios. The higher is the 'fine-to-fee' ratio, the more likely is the responder to punish when the take rate is high.

Figure D4: Punishment of reciprocal responders as a function of t_j, Y_{ij}, and $\bar{\sigma}$



E. Logit regressions on the decision to punish

The dependent variable is the a dichotomous variable which takes value 1 when the responder destroyed, and 0 otherwise.

	Regression 1			Regression 2		
	b	se	p	b	se	p
t_i	0.062***	0.019	0.001	0.066***	0.017	0
$d_{\sigma=\overline{\sigma}} \times t_i$	-0.047**	0.021	0.026	-0.049**	0.02	0.014
$d_{\sigma=\overline{\sigma}}$	2.771*	1.509	0.066	2.598*	1.438	0.071
$d_{\sigma=\overline{\sigma}}$ × non-UK	0.228	0.793	0.774	0.322	0.824	0.696
non-UK	0.047	0.595	0.937	0.157	0.618	0.799
Experience				-0.314	0.191	0.1
Male				-0.817*	0.469	0.081
Age				-0.072*	0.041	0.081
Constant	-4.537***	1.361	0.001	-1.854	1.668	0.266
Obs	141			141		
Pseudo R-Square	0.141			0.186		
Df	136			133		
Prob > F	0.001			0.001		

Notes: Logit regressions with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01.

F. Ordered logit regressions on emotions (without demographics)

	Fear			Envy			Anger			Sadness			Happiness			Shame		
	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p
t_i	0	0.01	0.94	0.01	0.01	0.46	0.04***	0.01	0	0.01	0.01	0.39	-0.03**	0.01	0.02	0.02	0.02	0.21
$d_{\sigma=\bar{\sigma}}$	-0.11	0.85	0.9	-0.61	0.89	0.5	0.46	0.82	0.58	-1.1	0.78	0.16	0.58	1.09	0.59	1.03	1.3	0.43
$d_{\sigma=\bar{\sigma}} \times t_i$	0.01	0.01	0.55	0.01	0.01	0.48	-0.01	0.01	0.65	0.01	0.01	0.48	-0.01	0.02	0.53	-0.02	0.02	0.39
non-UK	0.9	0.55	0.1	-0.01	0.48	0.98	0.07	0.46	0.88	0.28	0.44	0.53	0.23	0.48	0.63	0.59	0.62	0.34
$d_{\sigma=\bar{\sigma}}$ × non-UK	-0.01	0.74	0.99	-0.23	0.62	0.71	0.22	0.66	0.74	1.36**	0.66	0.04	-0.04	0.66	0.95	0.67	0.82	0.42
Obs.	141			140			141			140			141			140		
Pseudo R-Square	0.03			0.01			0.06			0.04			0.07			0.04		
Prob > F	0.1			0.32			0			0			0			0.1		

	Irritation			Contempt			Joy			Jealousy			Surprise		
	b	se	p	b	se	p	b	se	p	b	se	p	b	se	p
t_i	0.03***	0.01	0	0.01	0.01	0.15	-0.01	0.01	0.28	0.01	0.01	0.13	-0.01	0.01	0.61
$d_{\sigma=\bar{\sigma}}$	0.13	0.7	0.85	0.17	0.9	0.85	0.82	1.05	0.43	0.16	0.77	0.83	0.77	1	0.44
$d_{\sigma=\bar{\sigma}} \times t_i$	0	0.01	0.71	0	0.01	1	-0.01	0.02	0.41	0	0.01	0.78	-0.01	0.01	0.45
non-UK	-0.76*	0.44	0.08	0.44	0.42	0.3	0.24	0.49	0.63	-0.04	0.49	0.93	0.22	0.47	0.64
$d_{\sigma=\bar{\sigma}}$ × non-UK	-0.06	0.66	0.92	0.03	0.65	0.96	0.24	0.67	0.72	0.27	0.62	0.67	0.45	0.6	0.45
Obs	141			138			141			141			141		
Pseudo R-Square	0.06			0.02			0.03			0.01			0.02		
Prob > F	0			0.15			0.05			0.43			0.01		

Notes: Ordered logit regressions with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01, **** p<0.001.

G. Tobit regressions on punishment for all emotions

Notes: Tobit regressions with robust standard errors. p_i is expressed in percentage Hence, the beta coefficients identifies percentages.* p<0.1, ** p<0.05, *** p<0.01.

	Fear			Envy			Anger			Sadness		
	b	se	p	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	131.41***	43.39	0	134.95***	48.46	0.01	40.29	38.31	0.29	131.82***	46.27	0.01
non UK	65.22**	31.61	0.04	48.74	29.67	0.1	46.54	28.48	0.1	44.46	32	0.17
Constant ratio × nonUK	-58.25	41.61	0.16	-14.71	37.51	0.7	-26.22	34.97	0.45	-29.16	42.67	0.5
Experience	-22.41	18.45	0.23	-23.69	19.08	0.22	1.39	16.24	0.93	20.59	20.75	0.32
Male	-30.6	22.07	0.17	-30.64	21.4	0.15	-35.21*	20.07	0.08	-31.73	21.63	0.14
Age	-3.43	2.31	0.14	-4.61*	2.37	0.05	-2.21	1.86	0.24	-3.04	2.15	0.16
Fear	-18.55	20.59	0.37									
$t > \tilde{t}$ × Fear	-3.77	18.05	0.84									
Constant ratio × $t > \tilde{t}$ × Fear	-11.01	16.41	0.5									
Constant ratio × Fear	16.88	14.49	0.25									
Experience × Fear	7.02	5.72	0.22									
Envy				-15.75	13.86	0.26						
$t > \tilde{t}$ × Envy				4.78	10.73	0.66						
Constant ratio × $t > \tilde{t}$ × Envy				-22.07**	9.42	0.02						
Constant ratio × Envy				9.84	7.49	0.19						
Experience × Envy				4.33	4.48	0.34						
Anger							0.03	12.8	1			
$t > \tilde{t}$ × Anger							28.37**	11.59	0.02			
Constant ratio × $t > \tilde{t}$ × Anger							-23.92**	10.15	0.02			
Constant ratio × Anger							11.96	9.16	0.19			
Experience × Anger							-1.33	3.68	0.72			
Sadness										18.01	18.04	0.32
$t > \tilde{t}$ × Sadness										5.31	14.12	0.71
Constant ratio × $t > \tilde{t}$ × Sadness										-23.88*	12.28	0.05
Constant ratio × Sadness										12.08	10.77	0.26
Experience × Sadness										-7.08	5.47	0.2
Constant	94.27***	13.74	0	91.39***	12.99	0	85.83***	12.58	0	92.31***	13.62	0
Obs	141			140			141			140		
Pseudo R-Square	0.07			0.08			0.09			0.08		
Df	130			129			130			129		
Prob > F	0.01			0.01			0			0.01		

	Happiness			Shame			Irritation			Contempt		
	b	se	p	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	150.89***	42.75	0	119.98***	36	0	44.76	38.78	0.25	130.84***	40.41	0
non UK	50.12*	28.79	0.08	42.15	28.11	0.14	54.57*	28.17	0.05	40.99	27.56	0.14
Constant ratio × nonUK	-7.86	36.46	0.83	-18.14	36.37	0.62	-16.79	33.68	0.62	-3.15	35.58	0.93
Experience	9.85	14.66	0.5	15.08	14.07	0.29	19.46	19.32	0.32	33.51*	17.75	0.06
Male	-35.79*	20.78	0.09	-38.60**	19.21	0.05	-29.1	20.12	0.15	-22.16	19.39	0.26
Age	-3.92*	2.1	0.06	-4.77**	2.18	0.03	-3.02	1.89	0.11	-4.27**	1.83	0.02
Happiness	16.46	16.87	0.33									
$t > \tilde{t}$ × Happiness	-6.21	16.87	0.71									
Constant ratio × $t > \tilde{t}$ × Happiness	-37.04*	19.06	0.05									
Constant ratio × Happiness	10.38	11.31	0.36									
Experience × Happiness	-8.45	6.07	0.17									
Shame				40.67*	22.57	0.07						
$t > \tilde{t}$ × Shame				6.98	21.24	0.74						
Constant ratio × $t > \tilde{t}$ × Shame				-30.82	18.91	0.11						
Constant ratio × Shame				8.67	17.39	0.62						
Experience × Shame				-10.19*	5.23	0.05						
Irritation							5.14	14.52	0.72			
$t > \tilde{t}$ × Irritation							32.04***	12.01	0.01			
Constant ratio × $t > \tilde{t}$ × Irritation							-29.08***	10.77	0.01			
Constant ratio × Irritation							14.93	9.36	0.11			
Experience × Irritation							-5.95	4.33	0.17			
Contempt										45.42**	17.55	0.01
$t > \tilde{t}$ × Contempt										1.76	13.01	0.89
Constant ratio × $t > \tilde{t}$ × Contempt										-32.01***	11.66	0.01
Constant ratio × Contempt										10.01	9.06	0.27
Experience × Contempt										-13.44**	5.27	0.01
Constant	92.10***	13.24	0	86.62***	12.96	0	88.19***	12.92	0	84.38***	12.23	0
Obs	141			140			141			138		
Pseudo R-Square	0.07			0.09			0.09			0.09		
Df	130			129			130			127		
Prob > F	0.01			0			0			0		

	Joy			Jealousy			Surprise		
	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	157.14***	39.83	0	130.39**	47.64	0.01	97.75*	41.31	0.02
non UK	55.19*	27.63	0.05	53.25	29.74	0.08	44.97	30.14	0.14
Constant ratio × nonUK	-16.87	35.25	0.63	-25.25	37.03	0.5	-24.83	39.1	0.53
Experience	27.88	15.59	0.08	-31.6	18.32	0.09	28.61	20.76	0.17
Male	-36.32	20	0.07	-34	21.92	0.12	-25.85	19.71	0.19
Age	-3.87	2.18	0.08	-4.29	2.33	0.07	-3.51	2	0.08
Joy	27.77	18.33	0.13						
$t > \tilde{t}$ × Joy	-7.59	16.91	0.65						
Constant ratio × $t > \tilde{t}$ × Joy	-36.44*	18.4	0.05						
Constant ratio × Joy	18.12	12.77	0.16						
Experience × Joy	-16.37**	6.08	0.01						
Jealousy				-22.75	13.68	0.1			
$t > \tilde{t}$ × Jealousy				5.58	11.12	0.62			
Constant ratio × $t > \tilde{t}$ × Jealousy				-17.74	9.76	0.07			
Constant ratio × Jealousy				10.24	7.77	0.19			
Experience × Jealousy				6.62	4.35	0.13			
Surprise							16.21	15.21	0.29
$t > \tilde{t}$ × Surprise							14.39	11.96	0.23
Constant ratio × $t > \tilde{t}$ × Surprise							-22.16*	10.7	0.04
Constant ratio × Surprise							10.87	8.36	0.2
Experience × Surprise							-8.86	4.98	0.08
Constant	88.04***	12.75	0	92.75***	13.4	0	91.50***	13.41	0
Obs	141			141			141		
Pseudo R-Square	0.09			0.07			0.07		
Df	130			130			130		
Prob > F	0.01			0.01			0.01		

H. Tobit regressions on punishment for all emotions (without demographics)

Notes: Tobit regressions with robust standard errors. p_i is expressed in percentage Hence, the beta coefficients identifies percentages.* p<0.1, ** p<0.05, *** p<0.01.

	Fear			Envy			Anger			Sadness		
	b	se	p	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	129.86***	41.42	0	126.78***	47.48	0.01	27.21	38.8	0.48	109.45**	43.89	0.01
non UK	53.09*	29.06	0.07	40.57	28.34	0.15	46*	27.14	0.09	48.39	29.51	0.1
Constant ratio × nonUK	-54.59	39.68	0.17	-23.01	37.08	0.54	-32.12	34.32	0.35	-49.64	41.96	0.24
Fear	-0.3	13.51	0.98									
$t > \tilde{t}$ × Fear	-3.68	18.33	0.84									
Constant ratio × $t > \tilde{t}$ × Fear	-15.96	16.18	0.33									
Constant ratio × Fear	18.88	14.33	0.19									
Envy				-4.95	7.47	0.51						
$t > \tilde{t}$ × Envy				5.15	10.93	0.64						
Constant ratio × $t > \tilde{t}$ × Envy				-22.39**	9.72	0.02						
Constant ratio × Envy				12.29	7.52	0.1						
Anger							-4.63	10.05	0.65			
$t > \tilde{t}$ × Anger							29.85**	12.49	0.02			
Constant ratio × $t > \tilde{t}$ × Anger							-24.04**	10.66	0.03			
Constant ratio × Anger							13.35	9.59	0.17			
Sadness										-3.19	10.19	0.75
$t > \tilde{t}$ × Sadness										8.73	14.35	0.54
Constant ratio × $t > \tilde{t}$ × Sadness										-18.43	12.47	0.14
Constant ratio × Sadness										13.94	11.02	0.21
Constant	-115.49***	31.06	0	-101.83***	34.41	0	-98.41***	30.13	0	-109.19***	32.72	0
Obs	141			140			141			140		
Pseudo R-Square	0.06			0.07			0.09			0.07		
Df	134			133			134			133		
Prob > F	0			0			0			0.01		

	Happiness			Shame			Irritation			Contempt		
	b	se	p	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	123.44***	42.46	0	90.61**	35.29	0.01	31.83	40.14	0.43	119.62***	41.25	0
non UK	41.44	28.36	0.15	37.27	27.38	0.18	50.93*	27.81	0.07	27.73	27.79	0.32
Constant ratio × nonUK	-11.75	37.45	0.75	-35.7	36.46	0.33	-27.09	33.22	0.42	-1.77	35.04	0.96
Happiness	-6.16	12.04	0.61									
$t > \tilde{t}$ × Happiness	2.4	18.47	0.9									
Constant ratio × $t > \tilde{t}$ × Happiness	-31.26*	18.07	0.09									
Constant ratio × Happiness	7.81	10.85	0.47									
Shame				1.81	19.58	0.93						
$t > \tilde{t}$ × Shame				21.45	23.02	0.35						
Constant ratio × $t > \tilde{t}$ × Shame				-30.56	21.43	0.16						
Constant ratio × Shame				18.15	18.47	0.33						
Irritation							-13.45	8.41	0.11			
$t > \tilde{t}$ × Irritation							35.54***	12.07	0			
Constant ratio × $t > \tilde{t}$ × Irritation							-30.81***	10.77	0			
Constant ratio × Irritation							18.62**	8.89	0.04			
Contempt										7.82	11.2	0.49
$t > \tilde{t}$ × Contempt										7.41	14.39	0.61
Constant ratio × $t > \tilde{t}$ × Contempt										-32.21**	12.81	0.01
Constant ratio × Contempt										12.25	10.13	0.23
Constant	-96.03***	32.83	0	-109.34***	30.62	0	-92.57***	29.6	0	-125.40***	33.79	0
Obs	141			140			141			138		
Pseudo R-Square	0.06			0.07			0.08			0.07		
Df	134			133			134			131		
Prob > F	0.01			0			0			0		

	Joy			Jealousy			Surprise		
	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	126.17***	39.75	0	112.84**	45.73	0.01	78.1*	42.75	0.07
non UK	47.96*	28.25	0.09	45.89	28.37	0.11	41.16	30.77	0.18
Constant ratio × nonUK	-17.77	37.24	0.63	-32.81	36.71	0.37	-30.52	39.88	0.45
Joy	-12.58	12.85	0.33						
$t > \tilde{t}$ × Joy	2.16	18.08	0.91						
Constant ratio × $t > \tilde{t}$ × Joy	-27.1	16.69	0.11						
Constant ratio × Joy	13.23	11.74	0.26						
Jealousy				-7.21	8.47	0.4			
$t > \tilde{t}$ × Jealousy				8.01	11.65	0.49			
Constant ratio × $t > \tilde{t}$ × Jealousy				-18.63*	10.32	0.07			
Constant ratio × Jealousy				12.39	8.2	0.13			
Surprise							-8.82	9.53	0.36
$t > \tilde{t}$ × Surprise							18.27	12.54	0.15
Constant ratio × $t > \tilde{t}$ × Surprise							-18.34*	10.9	0.09
Constant ratio × Surprise							11.63	8.8	0.19
Constant	-90.61***	32.01	0.01	-97.26***	32.94	0	-89.37***	34.19	0.01
Obs	141			141			141		
Pseudo R-Square	0.07			0.06			0.06		
Df	134			134			134		
Prob > F	0			0			0.01		

I. Logit regressions on punishment for all emotions

Notes: Logit regressions with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01.

	Fear			Envy			Anger			Sadness		
	b	se	p	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	2.08***	0.71	0	2.38***	0.88	0.01	1.15	0.78	0.14	2.08***	0.81	0.01
non UK	0.86	0.61	0.16	0.51	0.61	0.4	0.51	0.63	0.42	0.44	0.66	0.5
Constant ratio × nonUK	-1.16	0.79	0.14	-0.14	0.73	0.85	-0.4	0.76	0.59	-0.47	0.86	0.58
Experience	-0.57*	0.34	0.09	-0.5	0.35	0.15	-0.19	0.33	0.57	0.01	0.36	0.99
Male	-0.77	0.48	0.11	-0.8*	0.47	0.09	-0.85*	0.48	0.08	-0.81*	0.47	0.08
Age	-0.06	0.05	0.19	-0.08*	0.05	0.09	-0.04	0.04	0.28	-0.05	0.04	0.26
Fear	-0.3	0.39	0.43									
$t > \tilde{t}$ × Fear	-0.11	0.31	0.73									
Constant ratio × $t > \tilde{t}$ × Fear	-0.03	0.34	0.92									
Constant ratio × Fear	0.35	0.28	0.21									
Experience × Fear	0.15	0.11	0.19									
Envy				-0.19	0.24	0.42						
$t > \tilde{t}$ × Envy				0.03	0.2	0.86						
Constant ratio × $t > \tilde{t}$ × Envy				-0.36*	0.19	0.06						
Constant ratio × Envy				0.18	0.15	0.24						
Experience × Envy				0.06	0.08	0.47						
Anger							0.01	0.29	0.98			
$t > \tilde{t}$ × Anger							0.35	0.25	0.16			
Constant ratio × $t > \tilde{t}$ × Anger							-0.38	0.23	0.11			
Constant ratio × Anger							0.25	0.21	0.25			
Experience × Anger							-0.03	0.08	0.76			
Sadness										0.19	0.32	0.55
$t > \tilde{t}$ × Sadness										0.09	0.26	0.73
Constant ratio × $t > \tilde{t}$ × Sadness										-0.35	0.25	0.16
Constant ratio × Sadness										0.23	0.22	0.29
Experience × Sadness										-0.07	0.09	0.43
Constant	1.15	1.49	0.44	1.61	1.48	0.28	0.08	1.42	0.95	-0.41	1.65	0.81
Obs	141			140			141			140		
Pseudo R-Square	0.17			0.17			0.18			0.16		
Prob > F	0.01			0.01			0			0		

	Happiness			Shame			Irritation			Contempt		
	b	se	p	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	2.40***	0.7	0	2.70***	0.72	0	0.93	0.77	0.23	2.32***	0.74	0
non UK	0.6	0.57	0.3	0.58	0.59	0.33	0.59	0.59	0.32	0.45	0.64	0.49
Constant ratio × nonUK	-0.14	0.73	0.85	-0.51	0.78	0.51	-0.21	0.7	0.76	0.07	0.82	0.93
Experience	-0.24	0.28	0.38	-0.07	0.28	0.8	0.29	0.41	0.48	0.34	0.34	0.32
Male	-0.8*	0.46	0.08	-0.96**	0.47	0.04	-0.76	0.46	0.1	-0.62	0.46	0.18
Age	-0.06	0.04	0.13	-0.09**	0.04	0.04	-0.05	0.04	0.17	-0.08**	0.04	0.03
Happiness	0.01	0.3	0.96									
$t > \tilde{t}$ × Happiness	-0.11	0.29	0.7									
Constant ratio × $t > \tilde{t}$ × Happiness	-0.5	0.41	0.22									
Constant ratio × Happiness	0.18	0.23	0.44									
Experience × Happiness	-0.04	0.11	0.73									
Shame				0.86*	0.5	0.09						
$t > \tilde{t}$ × Shame				-0.35	0.46	0.45						
Constant ratio × $t > \tilde{t}$ × Shame				-0.39	0.42	0.35						
Constant ratio × Shame				0.18	0.38	0.64						
Experience × Shame				-0.14	0.11	0.18						
Irritation							0.19	0.36	0.6			
$t > \tilde{t}$ × Irritation							0.50**	0.25	0.04			
Constant ratio × $t > \tilde{t}$ × Irritation							-0.47**	0.24	0.05			
Constant ratio × Irritation							0.29	0.21	0.18			
Experience × Irritation							-0.16	0.11	0.14			
Contempt										0.90**	0.4	0.02
$t > \tilde{t}$ × Contempt										-0.03	0.29	0.92
Constant ratio × $t > \tilde{t}$ × Contempt										-0.55**	0.28	0.05
Constant ratio × Contempt										0.19	0.21	0.38
Experience × Contempt										-0.26**	0.12	0.03
Constant	0.69	1.22	0.57	0.17	1.1	0.88	-0.76	1.53	0.62	-1.12	1.22	0.36
Obs	141			140			141			138		
Pseudo R-Square	0.16			0.2			0.19			0.21		
Prob > F	0.01			0			0			0		

	Joy			Jealousy			Surprise		
	b	se	p	b	se	p	b	se	p
$t > \tilde{t}$	2.39***	0.67	0	2.40***	0.87	0.01	1.60**	0.76	0.04
non UK	0.63	0.56	0.26	0.67	0.6	0.27	0.6	0.6	0.32
Constant ratio × nonUK	-0.13	0.71	0.85	-0.42	0.74	0.57	-0.54	0.78	0.49
Experience	0.01	0.31	0.98	-0.73**	0.36	0.04	0.19	0.42	0.64
Male	-0.84*	0.47	0.08	-0.88*	0.49	0.07	-0.73	0.45	0.1
Age	-0.06	0.04	0.15	-0.07*	0.04	0.09	-0.06	0.04	0.16
Joy	0.17	0.35	0.64						
$t > \tilde{t}$ × Joy	-0.03	0.33	0.93						
Constant ratio × $t > \tilde{t}$ × Joy	-0.56	0.38	0.14						
Constant ratio × Joy	0.28	0.27	0.3						
Experience × Joy	-0.15	0.11	0.17						
Jealousy				-0.38	0.26	0.14			
$t > \tilde{t}$ × Jealousy				0	0.21	0.99			
Constant ratio × $t > \tilde{t}$ × Jealousy				-0.24	0.19	0.22			
Constant ratio × Jealousy				0.19	0.16	0.24			
Experience × Jealousy				0.12	0.08	0.16			
Surprise							0.17	0.29	0.55
$t > \tilde{t}$ × Surprise							0.21	0.23	0.35
Constant ratio × $t > \tilde{t}$ × Surprise							-0.3	0.22	0.18
Constant ratio × Surprise							0.23	0.18	0.19
Experience × Surprise							-0.13	0.1	0.2
Constant	0.21	1.23	0.86	2.2	1.46	0.13	-0.37	1.61	0.82
Obs	141			141			141		
Pseudo R-Square	0.18			0.17			0.16		
Prob > F	0.01			0.01			0.01		

J. Emotions of the take authority

We also analyzed the emotional response of the take authorities to different levels of punishment of the responder. In particular, we ran some ordered logit regressions,[103] one for each emotion, where the dependent variable is the emotion of interest, whereas the independent variables are the take rate of the take authority, the punishment rate received from the responder, a dummy variable which takes value 1 when a constant fine-to-fee ratio is employed, the experience, gender (Male = 1 for male subjects), age, and nationality (non-UK = 1 for non-UK) of the subjects, and an interaction term between the dummy for the constant fine-to-fee ratio and nationality. We report the results of these regressions in Table J1 below, focusing on the role played by the punishment rate.

Negative emotions such as anger, sadness, irritation, and contempt are strongly significantly positively related to the punishment rate. Jealousy appears to be only weakly positively related to the punishment rate. Among positive emotions, happiness and joy are significantly negatively related to the punishment rate. The coefficients of the other explanatory variables are mostly not significant. We briefly mention here those which are significant. Take authorities who selected higher take rates experienced increasing surprise ($p = 0.022$). *Caeteris paribus*, non-UK subjects experienced more envy ($p = 0.000$), jealousy ($p=0.004$), and sadness ($p = 0.049$). Older subjects experienced less envy ($p = 0.031$), anger ($p = 0.079$), irritation ($p = 0.045$), contempt ($p = 0.021$), and jealousy ($p = 0.017$). Subjects under a constant 'fine-to-fee' ratio experienced on average less anger ($p = 0.029$). Finally, male subjects were more irritated ($p = 0.022$) and angry ($p = 0.088$) than female subjects.

[103] Robust standard errors are employed to control for heteroscedasticity. Due to one subject failing to report the intensity of jealousy, we have only 140 observations (instead of 141) for this emotion.

Table J1: Ordered logit regressions on emotions

	Fear		Envy		Anger		Sadness		Happiness		Shame	
	b	p	b	p	b	p	b	p	b	p	b	p
Take rate	0.006	0.411	-0.003	0.83	0.003	0.73	0.004	0.561	0.007	0.307	0.009	0.225
Destruction rate	0.004	0.499	0.007	0.389	0.019***	0.009	0.020***	0	-0.019***	0	0.002	0.641
Constant ratio	-0.336	0.571	0.889	0.456	-1.809**	0.029	0.4	0.468	-0.464	0.276	-0.058	0.903
non-UK	0.509	0.446	3.586***	0	0.908	0.148	1.105**	0.049	-0.37	0.389	-0.125	0.765
Experience	-0.038	0.859	-0.289	0.231	0.393	0.119	0.191	0.311	0.146	0.474	-0.011	0.944
Constant ratio × non-UK	0.575	0.497	-1.884	0.155	0.843	0.4	-0.932	0.216	0.838	0.231	0.073	0.913
Male	0.031	0.944	0.44	0.412	0.789*	0.088	-0.266	0.511	-0.11	0.743	-0.113	0.734
Age	-0.08	0.417	-0.224**	0.031	-0.168*	0.079	-0.067	0.372	-0.025	0.237	-0.016	0.629
Obs	141		141		141		141		141		141	
Pseudo R-Square	0.029		0.141		0.143		0.09		0.036		0.009	
Prob > F	0.602		0		0		0.002		0.023		0.835	

	Irritation		Contempt		Joy		Jealousy		Surprise	
	b	p	b	p	b	b	b	p	b	p
Take rate	0	0.957	0.004	0.534	0.012*	0.093	-0.009	0.462	0.017**	0.022
Destruction rate	0.013**	0.013	0.015***	0.009	-0.013**	0.017	0.015*	0.096	-0.004	0.412
Constant ratio	-0.324	0.54	-0.304	0.598	0.125	0.729	-1.043	0.379	0.21	0.606
non-UK	0.664	0.28	-0.121	0.834	-0.116	0.804	2.215***	0.004	0.184	0.689
Experience	0.064	0.727	0.209	0.316	0.035	0.872	0.479	0.131	-0.06	0.716
Constant ratio × non-UK	0.101	0.894	0.8	0.282	0.489	0.491	0.487	0.728	-0.572	0.37
Male	0.850**	0.022	-0.019	0.957	0.064	0.835	0.413	0.438	0.096	0.754
Age	-0.133**	0.045	-0.209**	0.021	0.005	0.816	-0.256**	0.017	-0.028	0.398
Obs	141		141		141		140		141	
Pseudo R-Square	0.062		0.064		0.02		0.154		0.017	
Prob > F	0.003		0.004		0.495		0.004		0.377	

Notes: Ordered logit regressions with robust standard errors. * p<0.1, ** p<0.05, *** p<0.01.

Chapter 3: Competence versus Trustworthiness: What Do Voters Care About?[104]

I. Introduction

We present an experiment on the preferences of voters over candidates in public elections. We are interested in two main characteristics that define the quality of a candidate: competence and trustworthiness.[105] Competence refers to the ability of a potential public official to properly perform his/her job, identifying and employing the appropriate policies that enable her to get the job done. Trustworthiness refers to the general attitude of the potential public official to fulfill the trust that the voters have placed on him or her; it usually implies a general aversion towards corrupted practices such as bribery, kickbacks, and public embezzlement which would benefit the public official to the detriment of the public.

Why may people have a preference over one of the two characteristics that define the quality of a public official? From a traditional economic point of view, a rational and purely self-interested voter should always select the candidate that ensures the highest expected return for the elector irrespectively of everything else. The underlying idea – well captured by Bill Clinton's 1992 presidential campaign strategist James Carville in his slogan "[it's] the economy, stupid" – is that people care only about the economy and want candidates who are able to improve it, and therefore their own financial position, irrespectively of everything else. The results of this study will tell us whether this is true or not based on the

---

[104] This chapter is based on a paper co-authored with Prof. Daniel Zizzo.

[105] In a previous study by Caselli and Morelli (2004), the authors used the term honesty instead of trustworthiness to identify one of the two main characteristics that define the quality of a public official. Clearly, the two concepts are related and the difference is usually subtle. A person is honest if she is sincere, truthful and reliable, and avoids cheating or lying. A person is trustworthy if she is reliable and willing to fulfil the trust that has been placed on her. In the context of electoral choices, when an elector votes for a candidate, he places trust on her, hoping that she will fulfil the interests of the public, usually avoiding any dishonest practices. Hence, trustworthiness is a more general term to use in order to define the quality of a public official, and, under certain conditions, it may be more relevant than honesty in driving the electoral choices of voters. For instance, an elector may vote for a candidate who is willing to adopt dishonourable measures or cheat in order to realize the interests of her constituency. In the context of our experiment, the two concepts are perfectly interchangeable as honesty implies trustworthiness and *vice versa*.

preferences of the voters over the characteristics of the candidates. In addition, our experiment may help us understand why democracies may at times suffer from endemic dishonesty and corruption at the public level. If voters in fact display a rational and profit-maximizing voting behavior or a preference for competence over trustworthiness, the existence of corruption and dishonesty in modern democracies might be explained by people's voting preferences.

Voters may however be reluctant to support an untrustworthy candidate if, for instance, they display what has been referred as “betrayal aversion”, that is a general dislike to “being betrayed beyond the mere payoff consequences” (Bohnet *et al*., 2008, p. 295), or if they care more about the process by which the payoffs are generated rather than the final payoffs (see, e.g., Rabin, 1993). Similarly, voters may be sensitive to a social norm that prescribes to punish a candidate who proves to be untrustworthy. As a result, voters may vote for a candidate who is more reliable but overall less worthwhile than the contender in terms of expected payoffs.

The opposite, also plausible, possibility is that voters may support the more competent candidate, quite independently of the trustworthiness of the alternative candidates and the expected returns associated to each of them. This may be the case if, for instance, voters think that the misuse of public power for personal benefit at the public level is a fact of life and, hence, justified (Peters and Welch, 1980) or if there is so much distrust in the public system that voters believe that the election of a trustworthy public official would have no impact whatsoever on the system or only a marginal one.[106]

There are other possible explanations of why voters choose a certain candidate over another which abstract from the pure preferences of the voters over trustworthiness and competence. Most notably, voters can be affected by the quality and level of information on the candidates available

[106] According to the 2012 corruption perception index published by the Transparency Organization (http://www.transparency.org/cpi2012), many countries do indeed present very high perceived levels of corruption in the public sector that could justify a total disinterest of the voters in the trustworthiness of the candidates.

to them at the moment of the vote (Peters and Welch, 1980).[107] In this study, we do not investigate the impact of these other factors that may affect the decision making of the voters in elections, but we focus solely on the voters' preferences over trustworthiness and competence, in a context where the candidates differ only over these two characteristics and where the voters are fully informed about them.

There can be found many anecdotes, or even more formal evidence, of real-world situations which could be used to support either the primacy of trust or that of competence for voting behavior. For instance, the success of the anti-establishment movement of the comedian Beppe Grillo at the general and local elections in Italy over the 2012 and 2013 might be explained by a greater weight assigned by a significant proportion of voters to trustworthiness rather than competence. Many voters might have voted for Grillo's party because of its choice to propose ordinary voters as candidates, with no experience on politics and public offices, but, as Grillo emphasized during his political campaign, much more trustworthy than conventional politicians (Bartlett, 2013). A dislike of voters for untrustworthy candidates may also explain why, in certain cases, candidates discovered or suspected to be implicated in corruption scandals fail to be elected or experience a significant drop in voters' support. For instance, in the elections of the US House of Representatives Peters and Welch (1980) and Welch and Hibbing (1997) found that incumbent candidates touched by corruption allegations lost more often their seats and received about 10 percentage points less than incumbent candidates with no corruption accusations.

A significant number of other cases seem however to support the opposite conjecture that voters are motivated by their final expected payoffs or care more about the competence of candidates rather than the trustworthiness. For instance, many of the parliamentarians who were involved in the 2009 UK parliamentary expenses scandal[108] held their seats

[107] Other important aspects that may influence the electoral choices of the voters are, for example, the electors' partisanship to a certain ideology or party or the sensitivity of certain electors to some attractive characteristics of a candidate such as beauty or charisma.

[108] In 2009, several members of the UK parliaments misused their permitted allowances and made inappropriate expenses claims for personal benefits.

in the 2010 general elections and experienced only a marginal drop in voters' support (about 1.5% on average; Eggers and Fischer, 2011). In Brazil, the former Brazilian President Luis Inacio Lula da Silva won the 2006 general elections regardless of the corruption scandals that plagued his previous administration and after a mandate characterized by steady economic growth and decrease in poverty for Brazil (Winters and Weitz-Shapiro, 2013).

Although these examples may provide important insights on how voters vote in public elections, they cannot be used to infer the preferences of voters over trustworthiness and competence as many other factors may have played a role in voting decisions. Research is needed to uncover these preferences and isolate them from other influences. Furthermore, in modern democratic elections, the vote is secret and anonymous. As a result, real-world data on voters' preferences is typically collected only in aggregate form after an election or via public opinion polls or surveys. However, aggregate data are usually difficult to interpret due to the lack of control over many unobservable variables, *in primis* the individual characteristics of the voters. In addition, the answers of voters to surveys and public opinion polls are considered to be highly affected by social pressure, especially because voters are asked about sensitive topics such as political preferences, and, therefore, not fully reliable (DeMaio, 1984).

By means of a lab experiment, we are able to bypass these limitations. We can collect data on individual voting behavior which is usually difficult to analyze with standard empirical approaches. In our experiment, we ask voters to select a public official, based on the competence and trustworthiness of which their final payoffs depend. We measure the competence of the candidates in a real effort task and their trustworthiness in a trust game. We then provide this information to the voters and ask them to select the public official. By looking at cases where there is a competence-trustworthiness trade-off, we can then measure the extent to which competence and trustworthiness matter in electoral decisions, or whether in the end only the expected financial bottom line for voters matters. We find that, in general, most voters tend to select the candidate rationally, based on who provides the highest expected profit irrespectively of trustworthiness and competence. That said, there is a bias towards caring about trustworthiness when the difference in expected profits between the two candidates is small enough.

The remaining of the chapter is organized as follows. Section 2 reviews the literature related to this study. Section 3 presents the experimental design. Section 4 describes the hypotheses to be tested and the theoretical background. Section 5 presents the results. Section 6 discusses the results, and Section 7 concludes.

2. Related literature

To the best of our knowledge, there is no study that investigates the extent to which competence and trustworthiness matter in electoral decisions. That said, one strand of related literature is about electoral delegation. In our experiment, a subject is chosen by some voters to be the public official and act for them. Several studies investigate the behavioral implications of delegating a decision about outcomes to another person (e.g. Corazzini *et al*., 2012; Hamman *et al*., 2011; Samuelson and Messick, 1986; Samuelson *et al*., 1984; Messick *et al*., 1983). These studies focus primarily on the delegate's behavior and its implications in term of welfare rather than the preferences of the people over the characteristics of the potential delegates. Similarly, voting preferences on the characteristics of the potential leaders is not a topic covered in the economic research on leadership, whereas it is the focus of our chapter.[109]

Another stream of literature related to our study is about honesty in decision making. In our experiment, voters are asked to elect a public official who can appropriate part or all of a common good by underreporting

[109] A public official can be in fact seen in many respects as a leader. The literature on leadership mostly focuses on the impact of leading-by-example (e.g. Gächter et al., 2012; Güth et al., 2007; Potters et al., 2007; Moxnes and van der Hejden, 2003). Some papers compare the implications of having randomly selected leaders with elected leaders (e.g. Levy et al., 2011; Brandts et al., 2012; Kocher et al., 2013), leaders appointed based on their past contribution (e.g. Gächter and Renner, 2005), leaders appointed based on participant's performance in a pre-task (e.g. Kumru and Vesterlund, 2010), and self-selected leaders (e.g. Rivas and Sutter, 2011; Arbak and Villeval, 2011). In our experiment, the "leader" is endogenously selected, as in some of this research. However, in contrast to this literature, our study is not about leadership-by-example, and we are not interested on the leader and followers' behavior but on subjects' preferences over the characteristics of the potential leaders.

its value. Economists have empirically investigated dishonesty mostly using experimental data. Some have studied lying and dishonesty in cheap talk games where some players can send true or false message regarding some kind of private information (e.g. future moves) to other players (e.g. Sutter, 2009; Gneezy, 2005; Croson, 2005). In these studies, deception is totally disclosed to the experimenter. Other scholars – not only in economics – have studied unobserved lying behavior by tracing its distribution from subjects' reported results of a dice roll, coin flip or matrix task (e.g. Fischbacher and Heusi, 2013; Hao and Houser, 2013; Abeler *et al*., 2012; Houser *et al*., 2012; Bucciol and Piovesan, 2011; Mazar and Ariely, 2006).

Our study is also related to some works on corruption. Barr *at al*. (2009) and Azfar and Nelson (2007) used a Public Servant's Game to study corruption in service delivery. In this game, one subject is assigned the role of service provider (or executive), a second subject the role of monitor (or attorney general), and the remaining subjects (6 subjects) are community members. The decision of the service provider, that is how many tiles (from a random distribution) to allocate to the community, is similar to the one of the public official in our experiment. A few economists and political scientists have also examined the extent to which voters may support corrupted incumbents in public elections (e.g. Peters and Welch, 1980; Welch and Hibbing, 1997; Ferraz and Finan, 2008; Winters and Weitz-Shapiro, 2013; Bågenholm, 2013). These studies are to some extent linked to ours since corruption may be a sign of untrustworthiness, particularly if the interests of the voters are aligned with those of the public. These works primarily used aggregate-level empirical approaches and focus solely on the impact of corruption on incumbents' re-election without investigating the trade-off between trustworthiness and competence.[110]

There is political science research studying the importance of the quality of the candidates, defined as a combination of integrity and competence, in electoral choices (Mondak and Huckfeldt, 2006; Mondak, 1995; Kulisheck and Mondak, 1996; McCurley and Mondak, 1995).

[110] An exception is the political study of Winters and Weitz-Shapiro (2013) who employed a non-incentivized survey experiment to investigate the attitude of respondents towards hypothetical incumbent politicians (vignettes) described in the form of qualitative sentences.

Mondak (1995) investigated the permanence of incumbents in the US House of Representatives in relation to the quality of the incumbents, measured as an index of competence and integrity constructed with content analysis. The author found that high-quality US House members remained in office longer than low-quality members. McCurley and Mondak (1995) combined the aggregate-level data on the quality of US House of Representatives' incumbents with individual-level post-election survey data to explore whether the skill and integrity of the candidates affect the voters' evaluation of the candidates and their voting choice. They found that the quality scores do affect the evaluation of the candidates. Similar findings are provided by Kulisheck and Mondak (1996) who investigated whether the information concerning the quality of hypothetical candidates influences the voting choice of subjects in a survey experiment. Mondak and Huckfeldt (2006) collected data from a series of survey experiments and a national survey to study the accessibility of the competence and integrity of hypothetical candidates in the evaluation of the contenders, and how people respond to these characteristics relative to partisanship and ideology. They found that competence and integrity are slightly more accessible than partisanship and ideology, and are perceived favorably by subjects. Altogether these studies provide evidence that the quality of candidates matter in national elections. However, they are inconclusive on which dimension of the quality matters the most. In addition, they present several features in relation to which our laboratory experiment approach based on an incentivized environment is able to provide a significant contribution.[111]

[111] First, when aggregate-level empirical approaches are used (e.g. Mondak, 1995; McCurley and Mondak, 1995), it is usually difficult to isolate and control for the effects of important unobservable variables, such as, for instance, the information available upon the candidates. In addition, one can question the subjectivity and precision of the measure used to identify the quality of a candidate, and the reliability of post-election surveys to measure the voters' support for a candidate (see, e.g., DeMaio, 1984; Lodge et al., 1990). Finally, when survey experiments are used (Kulisheck and Mondak, 1996; Mondak and Huckfeldt, 2006), the situations described to the subjects are hypothetical, there are no economic incentives associated with the choices, the focus is more on attitude and perception rather than behavior, and the quality of the candidates is identified only with qualitative statements and phrases. While these comments are not to deny the value of these studies, they suggest that an experimental approach of the kind we use would be especially useful to complement them.

3. Experimental Design

A. Outline

The experiment was conducted at the University of East Anglia between March and June 2013. 240 subjects participated in 20 experimental sessions (12 subjects per session).[112] The experiment was fully computerized and programmed with the z-Tree software (Fischbacher, 2007). Each session consisted of three stages (the *Real Effort Task*, the *Trust Game*, and the *Official's Dilemma Game*) and a final questionnaire.[113] At the beginning of each stage, subjects received both computerized and printed instructions.[114] These were context-free and written avoiding any suggestive terminology. Each set of instructions was followed by a control questionnaire which purpose was to check subjects' understanding of the instructions. Clarifications were given aloud for public knowledge. During the experiment, payoffs were calculated in *points* and converted to British pounds at the end of the experiment (1 point = 20 pence). Each subject earned on average £12.47 (around 19-20 US dollars) including £2 of show-up fee. Subjects were paid in private and in cash in a separate room outside the lab by a research assistant who was not present during the experiment and who was not aware of its content. Each session lasted around 1 hour and 15 minutes. We ran 2 treatments, described below: the Baseline treatment (14 sessions), and the CIL (*Conditional Information Lottery*) treatment (6 sessions). For each treatment, in half of the sessions we had the real effort task and the trust game second, and in the other half the reverse. Upon arrival to the lab, each subject was registered for the experiment and randomly assigned to a computer desk which was separated from the others

---

[112] Subjects were randomly recruited with the on-line software ORSEE (Greiner, 2004). Subjects were mostly students with different socio-demographic background (details are provided in the online appendix). No subject participated in more than one session.

[113] At the beginning of the experiment, subjects were informed that there were many stages in the experiment, but the details of each stage were revealed to subjects only at the beginning of each stage. This is because we want to minimize strategic behavior. We also made clear in the instructions that the information provided by the subjects in each stage may be reported to other participants at later stage of the experiment but anonymity will be preserved.

[114] Instructions are available in the online appendix.

by partitions. Afterwards, subjects received the instructions for the first stage.

B. The Baseline treatment

The Real Effort Task Stage. The task for this stage was performed individually by each subject and consisted in counting the number of 1s in a series of tables containing 0s and 1s for 10 minutes (see, e.g., Abeler *et al*., 2009; Pokorny, 2008).[115] Each subject earned 1 point for each table that he or she correctly solved on top of the first 40 tables. This number was calibrated based on the results of pilot sessions in order to obtain a similar degree of dispersion between our measure of competence and our measure of trustworthiness, and to ensure that everyone was able to pass the threshold of 40.[116] In particular, we wanted to avoid that the voting decisions of the subjects were biased towards one characteristic or the other because of the different (normalized) degree of variation of the two characteristics. We chose this particular task because it provides enough variation in performance. The task is also simple and does not require any particular knowledge. At the same time, it is tedious and, therefore, mentally costly for the subjects. Finally, the experimenter does not benefit from the output of the task. [117]

In this stage of the experiment, we obtained, for each subject, a measure of competence calculated as the number of tables correctly solved on top of the first 40 tables correctly solved.

The Trust Game Stage. The other stage 1 or 2 of the experiment, depending on the session sequence, was a modified version of the standard one-shot trust game proposed by Berg *et al*. (1995). In particular, each subject was randomly matched with another participant. For each pair of subjects, one participant was randomly assigned the role of truster, while the

[115] A table consisted of a 5 × 5 matrix of 0s and 1s. For each table, the computer randomly generated a number of 0s and 1s in a random order. In a given session, all the subjects faced the same series of randomly generated tables.

[116] Only 2 subjects out of 240 did not solve more than 40 tables in the first stage of the experiment. In particular, one subject solved 36 tables and the other one 40 tables.

[117] These are all important features of our task since we wanted to minimize the reciprocity of subjects towards the experimenter, and to ensure that our measure of competence was minimally affected by other external influences (Abeler *et al*., 2009).

other the role of trustee. The truster received an endowment of 30 points and decided whether to transfer or not the entire endowment to the trustee (it was a binary choice: transfer all/do not transfer at all). If the truster decided to transfer the 30 points to the trustee, these were multiplied by 3 and the trustee received 90 points. The trustee could then decide to give back any amount to the truster between a minimum of 9 points and a maximum of 90 points. Since the roles were revealed only at the end of the experiment, each subject made a decision in both roles[118] using a strategy method. In particular, each subject first decided how many points he or she wished to return to the truster if he or she were to be assigned the role of trustee and the truster were to transfer the 30 points to the trustee. Then, each subject decided whether he or she wanted to transfer the 30 points or not to the trustee if he or she were assigned the role of truster. This mechanism allowed us to collect a measure of trustworthiness for each participant. In particular, the proportion of points sent back to the truster by each subject in the role of trustee was our measure of trustworthiness. Note that, in order to minimize reciprocity in the following stage, the subjects could not rematch with the same person later in the experiment. In addition, we imposed a minimum amount of 9 points to be returned by the trustee in order to avoid observations at zero. This is an important aspect for the following stage as people may avoid voting for a candidate simply because he or she displays a zero in one of the characteristics rather than a minimum positive value.[119] Finally, we asked people to make a decision first in the role of trustee and, then, in the role of truster.[120]

[118] Only one of the two decisions counted for the earnings depending on the role assigned.

[119] The underlying mechanism may be similar to the one that characterized the so-called 'zero-price effect' (see, e.g., Shampanier *et al*., 2007). This effect has been studied in the marketing research. In this literature, the zero is associated with a cost and induces people to choose more often the option with the zero (other things being equal). In our context, the zero would be associated with a benefit and may induce people to choose less often the option with the zero (other things being equal), resulting in potential biased observations.

[120] This is because we wanted to minimize the possibility that the decision in the role of truster affected the decision in the role of trustee, as the latter provides our measure of trustworthiness, whereas we are not interested in the truster's decision as such in this experiment.

The Official's Dilemma Game. In stage 3 of the experiment, each participant was matched with two others to form a group of three subjects.[121] The computer assigned a *common fund* to each triad with an initial value of 0 points. The task in this stage of the experiment was to count 1s in a series of tables as in the first stage of the experiment. However, only the work of one of the three subjects counted for the earnings of this stage. This person was the public official.[122] In particular, for each table that the public official solved on top of the first 40 tables correctly solved, the value of the common fund increased by 4 points. At the end of the real effort task,[123] the common fund accumulated a certain value. Only the public official was informed of this value. He or she was then asked to report the value of the common fund to the other members of the triad, knowing that he or she could report any number between 0 and the true value of the common fund. The *reported value* of the common fund was divided equally between the three participants. The public official kept the unreported value of common fund for himself or herself. All of these rules for generating and distributing experimental earnings were common knowledge for all subjects at the point of selecting the public official: therefore, in selecting the public official, subjects knew that their earnings depended on the competence (in solving tables) and the trustworthiness (in reporting the value of the common fund) of the public official.

How was the public official selected? At the beginning of the game, each subject was informed of (a) the number of tables correctly solved by each other member of the triad in the earlier real effort task stage of the experiment; this provided a measure of the competence of each candidate; and (b) the proportion of points that each other member of the triad in the role of trustee returned to the truster in the earlier trust game stage of the experiment; this provided a measure of the trustworthiness of each candidate. Based on this information, each subject was asked to vote: that is, to choose which of the two other participants he or she wanted to appoint as

[121] In the instructions, we refrained from using any terminology (e.g. group) which could induce group identity.

[122] In the instructions, we used the neutral term *appointed co-participant* to identify the public official.

[123] Note that everyone could work on the task if they wanted to, and this was known at the time of voting. However, the task was incentive compatible only for the public official.

the public official. Then, the computer implemented the voting decision of one randomly selected subject within each triad; this mechanism ensured that voting was incentive compatible, by removing any scope for strategic voting behavior.

Final questionnaire. After stage 3, subjects had to complete a 5-parts questionnaire, reproduced in the online appendix. The first two parts were incentivized. Part 1 was a belief elicitation questionnaire.[124] In part 2, we measured the risk attitude of subjects. We employed the Eckel *et al*. (2012)'s task in the domain of gains. In this task, subjects had to choose one gamble out of six possible gambles. Each gamble was represented with a circle and involved two payoffs with 50% probability of occurrence each. Moving from gamble 1 to gamble 6, both expected return and risk increased. Part 3 was the Stöber (2001)'s 17-item Social Desirability Scale (SDS17 score) which measures how much a person desires to be perceived in a positive light. Part 4 was the Christie and Geis (1970)'s 20-item Machiavellianism scale (MACH score) which measures a person's tendency to be amoral and opportunist. In the last part of the questionnaire, we collected some demographics and elicited subjects' belief about the objective of the experiment.

Payments. At the end of the experiment, the computer randomly drew a stage. Subjects were paid the earnings of that stage plus the show-up fee of 2 pounds and any additional earnings that they obtained by answering the final questionnaire.

C. The CIL treatment

The main purpose of the CIL treatment was to collect more data, in general and by individual, on the voting behavior of the subjects, without deceiving them. In addition, the data collected in this treatment allowed us to classify the subjects based on their voting decisions. The CIL treatment differs from the Baseline treatment only in third stage of the experiment, that is in the Official's Dilemma Game. In particular, in the CIL treatment, we employed the so-called Conditional Information Lottery (Bardsley,

[124] The details are in the online appendix.

2000). This technique consists in camouflaging one true task amongst other fictional tasks, with the subjects fully aware that there is a camouflage but uninformed ex-ante of which task is the true one (Bardsley, 2000). More specifically, in the selection of the public official, each subject was presented with 7 randomly ordered situations: one real and six fictional. In the real situation, each subject was informed about the actual competence and trustworthiness of the other participants within his or her group. In the fictional situations, each subject was instead presented with fictitious information about the competence and trustworthiness of the other two participants. In particular, to make the camouflage credible and realistic, the information used in the fictional situations came from situations occurred in past sessions of this experiment (the first six Baseline sessions) and chosen at random by the computer (subjects were fully informed of this). More specifically, to generate the fictional situations, the computer randomly picked situations from past sessions using a stratification procedure which followed approximately the distribution of the cases observed until then. Two fictional situations were randomly selected from the past situations where one candidate strictly or weakly dominated the other candidate in both characteristics (competence and trustworthiness). All the other four fictional situations corresponded to cases where the characteristics of the two candidates were orthogonal and differed in the extent to which the two candidates were different in terms of expected payoffs generated for the voter. In particular, one situation was randomly selected from cases where the difference in expected payoffs between the two candidates lay in the interval [0, 5] experimental points; a second situation from cases where the difference lay in the interval (5, 10] experimental points; a third situation from cases where the difference lay in the interval (10, 20] experimental points; and a fourth situation from cases where the difference lay in the interval (20, 50] experimental points. This stratified randomization allowed us to provide to the subjects enough decoys to prevent them from spotting the true situation, and, at the same time, to collect more information on the electoral choices of subjects for different level of expected payoffs of the candidates. The order of the seven situations was randomized. For each situation, each subject was asked to choose which of the two participants he or she wanted to appoint as the public official, knowing that only the decision of one participant selected at random in the real situation was implemented if the stage was chosen for payment. All the other aspects of the experiment were identical to the Baseline treatment.

As Bardsley (2000) pointed out, the CIL procedure might induce “cold” decisions because of the hypothetical nature of the task. This might actually be desirable in our experiment as voters do usually make their electoral choices in a “cold” state, since they are typically asked to vote in polling places, anytime over a span of one or two days and after the political campaign of the candidates. The CIL procedure may also dilute the incentives of the experiment, and increase the misunderstanding of the experimental procedures. To minimize these drawbacks, we limited the fictional situations to only 6 and made sure that subjects fully understood the instructions.[125] It was also important that subjects did not spot the true situation. As we have already mentioned earlier, we adopted a procedure of stratified randomization to select the fictional situations from real situations occurred in past sessions of the experiment, making very difficult, if not impossible, for the subjects to identify the true situation. Most significantly, we can use the Baseline treatment as a control to check whether any biases were produced from using the CIL procedure.

3. Theoretical background and hypotheses

In this section, we set the theoretical background and present the hypotheses to be tested. Let us call $\pi_{n,j}$ the expected earnings of the generic voter *n* if the candidate *j* is appointed in stage 3. This can be defined as:

$$\pi_{n,j} = AT_{nj}C_{nj}$$

where *A* is the constant multiplier of voter *n*’s profit function (equal to 4/3 in our experiment, where 3 is the group size, and 4 is the value of one table correctly solved by the candidate on top of the first 40 correct tables and reported to the voter); T_{nj} captures the trustworthiness of the candidate

[125] As we have already mentioned early, subjects filled in a control questionnaire, followed by clarifications, to check their understanding of the instructions, with key questions regarding, for instance, the meaning of the fictional situations.

and is measured as the proportion of points returned by the candidate in the trust game stage; C_{nj} captures the competence of the candidate and is measured as the number of tables correctly solved on top of the first 40 tables correctly solved in the real effort task stage. The profit function is a Cobb-Douglas with profit elasticities of competence and trustworthiness equal to 1. Each elasticity measures the responsiveness of the profit to a change in competence or trustworthiness, ceteris paribus. In particular, a 1% increase in competence would lead to a 1% increase in profit. Similarly, a 1% increase in trustworthiness would lead to a 1% increase in profit.

The voter *n* must choose among two candidates (J =2). The voter obtains a certain utility if a certain candidate is elected. In particular, the utility that voter *n* gets if candidate *j* is appointed is $U_{n,j}$, j = 1, 2. Each candidate possesses two attributes (competence and trustworthiness) which are known by the voter. If the voter is rational and profit maximizing, she should choose the candidate that gives the highest utility, and her utility should be an increasing function of the expected earnings. For simplicity, let the utility be a standard Cobb-Douglas function[126] which can be defined as follows:

$$U_{n,j} = A{T_{nj}}^{\alpha_1}{C_{nj}}^{\alpha_2}e^{\epsilon_{n,j}}$$

where α_1 and α_2 are the weights (elasticities) of the trustworthiness and the competence respectively of the candidate *j* in the utility function of voter *n*. $A{T_{nj}}^{\alpha_1}{C_{nj}}^{\alpha_2}$ is the known component of the utility function, whereas $e^{\epsilon_{n,j}}$ is the stochastic component (unknown component).[127]

[126] The Cobb-Douglas function has been widely used in economics to identify the production function of a firm or the utility function of an economic agent (see, e.g., Mas-Colell *et al*., 1995). In our context, it is particularly useful as it allows us to estimate the weights that a voter places on the trustworthiness and competence of the candidates in a directly comparable way. In particular, the weights are expressed in terms of elasticities, that is how much the utility varies (in percentage) if trustworthiness or competence increases by 1%. In addition, it is logically consistent with the essential elements of our experiment. In particular, it is directly linked to the profit function used in our experiment. More precisely, it can be reduced to a function of the profit if the weights of trustworthiness and competence are identical.

[127] For simplicity, we assume that the stochastic component is non-additive.

If the two attributes have the same weight in the utility ($\alpha_1 = \alpha_2 = \alpha$), the voter cares only about his or her profit. We can rewrite the utility as a function of the profit:

$$U_{n,j} = {\pi_{n,j}}^{\alpha} e^{\epsilon_{n,j}}$$

Hypothesis 1. If voters are rational and profit maximizing, $\alpha_1 = \alpha_2 = \alpha$, with $\alpha > 0$.

If α is equal to 0, the utility does not depend on the profit. If α is less than 0, it negatively depends on the profit. If $\alpha_1 > \alpha_2$ ($\alpha_2 > \alpha_1$), it means that the voter weight more the trustworthiness (competence) of the candidate over the competence (trustworthiness), and over what would be predicted by profit maximization.

Hypothesis 2. If trustworthiness matters more than competence, α_1 will be greater than α_2.

Hypothesis 3. If competence matters more than trustworthiness, α_2 will be greater than α_1.

To test the Hypotheses 1, 2 and 3, we can take the natural logarithm of the utility to obtain a linear function in parameters:

$$\ln(U_{n,j}) = V_{n,j} = \ln(A) + \alpha_1 T_{nj} + \alpha_2 C_{nj} + \epsilon_{n,j}$$

Knowing that the probability that voter *n* chooses candidate *i* over *j* is:

$$P_{n,i} = Prob(V_{n,i} > V_{n,j}) = Prob(V_{n,i} - V_{n,j} > 0)$$

We can derive the logit choice probability assuming that the error term ($\epsilon_{n,j}$) is iid with a Type-I extreme value distribution. The equation for the logit choice probability is:

$$P_{n,i} = Prob(V_{n,i} > V_{n,j}) = \frac{\exp(\alpha_1 \ln(T_{nj}) + \alpha_2 \ln(C_{nj}))}{\sum_{j=1}^{2} \exp(\alpha_1 \ln(T_{nj}) + \alpha_2 \ln(C_{nj}))}$$

The estimation of α_1 and α_2 is relatively straightforward through maximum likelihood estimation as we observe the choices of the voters and we have measures of the trustworthiness and competence of the candidates.

4. Experimental results

A. Descriptives

Table 1 shows the average measures of competence and trustworthiness for each treatment, from the first two stages of the experiemnt.[128] In the table, *competence* is the number of tables correctly solved on top of the first 40 tables correctly solved in the real effort task; *trustworthiness* is measured by the return rate, that is the proportion of points returned to the truster in the trust game. Since the Baseline and CIL treatments were equivalent in the first two stages of the experiment (i.e. the treatment manipulation involved only stage 3), we can pool their data together. There are only weakly significant differences in competence between the sessions where the real effort task stage took place first and the sessions where the trust game stage was played first (Mann-Whitney test, $p = 0.094$).[129] We do not detect any statistically significant differences in the return rate between having first played the Real Effort Task and the Trust Game ($p = 0.419$).[130] The coefficient of variation for competence, measured as the standard deviation over the mean, is 0.466. The coefficient of variation for trustworthiness is 0.518. The normalized measure of dispersion, captured by the coefficient of variation, is similar between competence and trustworthiness. This means that the voting decision of the subjects may not be affected by a different degree of dispersion between the two variables. Also, both the competence measured in the real effort task

[128] The focus of this chapter is on the voting choice and the information that subjects had to make this choice. As a result, we do not focus on the public official's behavior in the official dilemma game, once selected. The online appendix contains an analysis of the public official's choices in the official dilemma game, as well as more analysis of stages 1 and 2.

[129] A possible interpretation of why subjects performed slightly better in the sessions where the real effort task was played first than in the sessions where the trust game was played first is that they were cognitively less tired when the real effort task was played in the first stage rather than in the second stage. All *p* values reported in this chapter are two-sided.

[130] We also ran an OLS regression (for competence) and Tobit regression (for trustworthiness) where we control for the socio-demographic and psychological characteristics of the subjects (see the online appendix).

stage and the trustworthiness measured in the trust game stage were positively correlated with the public officials' competence and trustworthiness respectively in stage 3 (Spearman's ρ = 0.89 and 0.28, and *p* < 0.001 and = 0.011, respectively).[131] This indicates that both measures were valid proxies of the public officials' behavior in stage 3.

Table 1: Competence and Trustworthiness

	Competence		Trustworthiness	
	Mean	St. Dev.	Mean	St. Dev.
Real effort task stage first	39.49	17.75	0.36	0.19
Trust game stage first	35.43	16.97	0.34	0.17
Total	37.46	17.44	0.35	0.18

To analyze the electoral choices of subjects, we first consider the number of times subjects voted for a candidate for each possible electoral situations that occurred in the experiment (Table 2). If we look at the interesting situations where there was a trade-off between trustworthiness and competence, subjects seemed to vote more often for the trustworthy candidate as opposed to the more competent one. We can test more formally whether the proportion of situations where subjects chose the more trustworthy candidate significantly differs from 50%. The result of a binomial test indicates that the proportion of situations where people voted for the more trustworthy subject (about 59%) is significantly different from 50% (*p* < 0.001).[132] It is however possible that subjects voted more often for the more trustworthy candidate simply because the latter was more often associated to higher expected payoffs. In other words, people did not vote for the more trustworthy candidate because of his/her trustworthiness, but because the combination of his/her levels of competence and trustworthiness entailed higher expected payoffs compared to the other candidate. This

131 In stage 3, competence is measured as the number of tables correctly solved on top of the first 40 tables correctly solved, while trustworthiness is measured as the proportion of tables reported.

132 In the overall sample, *p* < 0.001. However, in CIL sessions, since we have multiple observations for each subject, there may be correlation between the observations and so the test may not be valid. That said, if we restrict the test to only the observations that are totally independent (Baseline sessions), we obtain exactly the same result, that is the proportion of situations where people voted for the more trustworthiness subject (about 61% in the baseline sessions) is significantly different from 50% (*p* = 0.024).

means that we will need to take into account the expected payoffs of the two candidates if we want to correctly study the electoral preferences of the voters.

Table 2: Voting choices

Situation	**Trustworthiness**	**Competence**	**Votes for *i***	**Votes for *j***	**Tot.**
Strict dominance	Ti > Tj	Ci > Cj	160	10	170
Trade-off	Ti > Tj	Ci < Cj	238	163	401
Weak dominance	Ti = Tj	Ci ≥ Cj	86	4	90
Weak dominance	Ti ≥ Tj	Ci = Cj	8	2	10
Equality	Ti = Tj	Ci = Cj	0	1	1
Total	.	.	492	180	672

If we look at the other situations where there is no trade-off between trustworthiness and competence, in a very small proportion of cases (5.56%),[133] subjects displayed what we refer as an *inconsistent voting behavior*, that is they voted for the candidate who was strictly or weakly dominated in both characteristics (trustworthiness and competence) by the other candidate. The behavior of these subjects (from now on, we will label them as *inconsistent subjects*) is difficult to characterize and interpret. It is likely that they made random choices during the experiment or did not take the experiment seriously. Hence, we will control for their behavior in the remaining of the analysis.

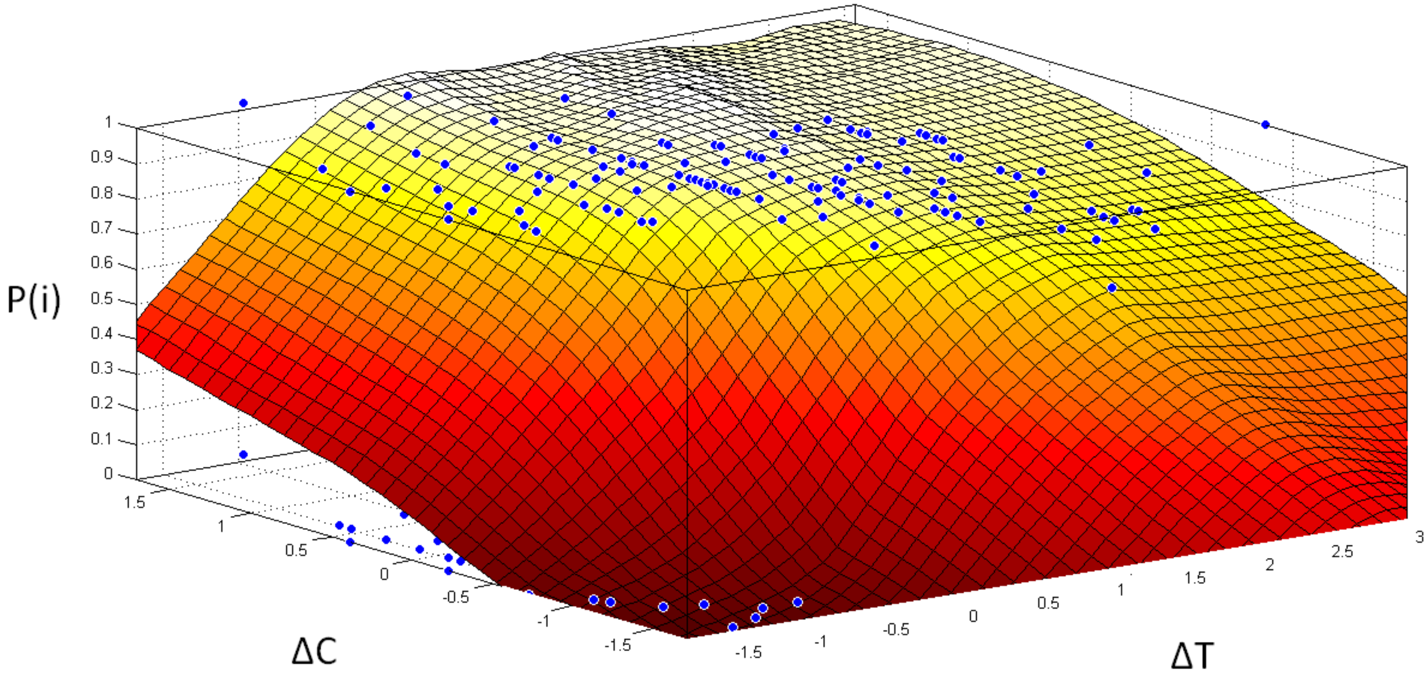
We can also look at how the probability that a candidate *i* is elected evolves as a function of the difference in *competence* and *trustworthiness* between candidate *i* and her rival, candidate *j* (Figure 1). To make competence and trustworthiness graphically comparable, we standardized them, that is we subtract the mean from each value and divide the result by the standard deviation. The graph suggests first that subjects seemed to behave quite rationally as the probability of being elected was close to 1 when the candidate was superior in both characteristics compared to the contender (upper corner of the graph), and was close to 0 when the

[133] More precisely, 14 subjects (5 in the baseline, 9 in CIL) out of 240 displayed this behavior (5.83% of subjects).

candidate was inferior in both characteristics (bottom corner of the graph). Second, the subjects seemed to slightly prefer more a trustworthy candidate to a competent one since the probability of being elected increases more steeply when the difference in trustworthiness between two candidates increases than when the difference in competence increases. We will investigate this in more detail in the regression analysis.

We now consider the expected payoffs that each candidate provided to the voters. We begin by assuming that subjects had adaptive expectations, that is they took the measures of trustworthiness and competence from the earlier stages to estimate what expected payoffs would be had by each candidate if elected public official. We shall relax this assumption later. First, we look at the probability of electing the more trustworthy candidate as a function of the difference in expected payoffs $\Delta\pi$ between the more and less trustworthy candidate (Figure 2), restricting the analysis to the observations where there was a trade-off between trustworthiness and competence. The probability is obtained by computing the weighted running means of a dichotomous variable taking value 1 when the trustworthy candidate is elected and 0 otherwise. For $\Delta\pi < 0$ (i.e. the more trustworthy candidate is also the less profitable), profit-maximizing subjects should vote for the less trustworthy candidate as he or she is associated with higher expected payoffs. Hence, the area below the smoothed means measures the extent to which subjects voted for the more trustworthy candidate when this was not the more profitable candidate. For $\Delta\pi > 0$ (i.e., the more trustworthy candidate is also the more profitable), profit-maximizing subjects should vote for the more trustworthy candidate as he or she is associated with higher expected payoffs. Hence, the area above the smoothed means measures the extent to which subjects voted for the more competent candidate when this was not the more profitable candidate. Note that the theoretical predicted probability under rational self-interest would follow a step function where the voter never chooses the more trustworthy candidate in the region where $\Delta\pi < 0$, and always chooses him or her when $\Delta\pi > 0$.

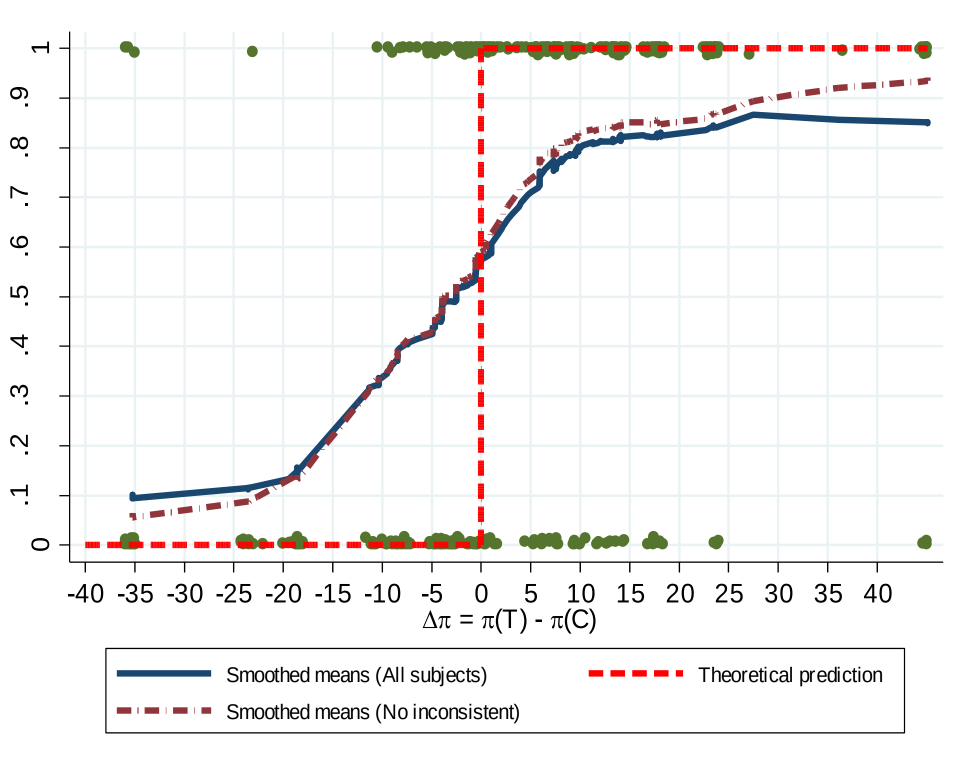
Figure 1: Probability that a candidate *i* is elected



Notes: P(i) is the probability of electing a candidate *i* in election *k*. This is computed using a locally weighted linear regression on the dichotomous variable taking value 1 when the candidate *i* is elected and 0 otherwise. ΔC is the difference in standardized *competence* between candidate *i* and candidate *j* in the situation (election) *k*. ΔT is the difference in standardized *trustworthiness* between candidate *i* and candidate *j* in the situation (election) *k*. The standardized values are obtained by subtracting the mean and dividing by the standard deviation. Each smoothed value of the locally weighted surface is computed using neighboring data points defined within the span of 0.6 (60% of the data).

As Figure 2 shows, choices follow fairly closely the rational self-interested prediction, in preliminary support of Hypothesis 1. That said, the area below the weighted running means for $\Delta\pi < 0$ is bigger than the area above the weighted running means for $\Delta\pi > 0$. This is particularly remarkable for small differences in expected payoffs ($|\Delta\pi| \leq 5$). This preliminary evidence suggests some preliminary support for a qualified version of Hypothesis 2: subjects seemed more likely to vote for the less profitable candidate when this was the more trustworthy one, particularly when the two candidates did not differ too much in terms of their contribution to the expected payoffs of the voter.

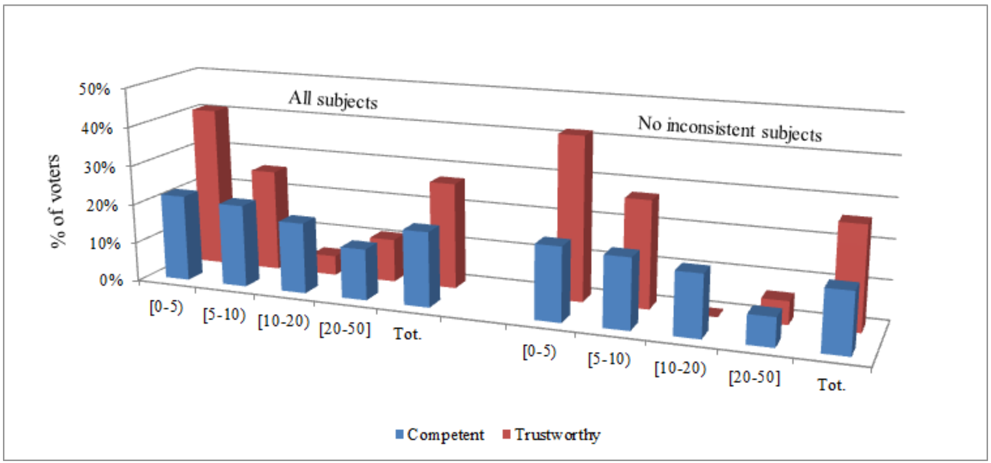
Figure 2: Probability of electing the more trustworthy candidate



Notes: P(T) is the probability of electing the more trustworthy candidate. $\Delta\pi$ is the difference in expected payoffs between the more and less trustworthy candidate. The running means are weighted to give more importance to near points than far, and computed using a bandwith of 0.6 (60% of the data). The data correspond to cases where there was a trade-off between trustworthiness and competence.

To look at this further, we consider how often subjects voted for the more trustworthy candidate when the latter was the less profitable one, and how often subjects voted for the more competent candidate when the latter was the less profitable one. Figure 3 reports the proportion of cases where the voters selected the less profitable candidate for each interval of absolute deviation in expected payoffs between the two candidates.[134]This proportion identifies the *rate of counterintuitive voting behavior* and measures the proportion of cases where subjects are willing to sacrifice their expected monetary payoffs in order to select the more trustworthy or competent candidate. In aggregate but particularly when the difference in expected payoffs between the two candidates was small (between 0 and 5 experimental points), the proportion of cases where subjects voted for the unprofitable and more trustworthy candidate was signficaintly larger than the proportion of cases where subjects voted for the unprofitable and more competenent candidate.[135]

Figure 3: Proportion of cases where the less profitable candidate was voted



Notes: The red bar identifies the proportion of cases where the voters elected the more trustworthy candidate when the latter was the less profitable one. The blue bar identifies the proportion of cases where the voters elected the more competent candidate when the latter was the less profitable one. The intervals in the x-axis are in experimental points. The data correspond to cases where there was a trade-off between trustworthiness and competence.

This is preliminary evidence of the fact that voters weighed trustworthiness more than competence, particularly when the candidates contributed a similar amount to the expected payoffs of the voters. In terms

[134] These were the same intervals that were used in the design phase of the experiment to generate the fictitious situations for the CIL treatment.

[135] If we restrict ourselves only the observations that are totally independent (Baseline sessions), we obtain χ^2 test, p = 0.070 in aggregate and 0.003 for the [0,5) interval. If we exclude the inconsistent subjects, we obtain χ^2 test, p = 0.012 in aggregate and 0.023 for the [0,5) interval. The results are if anything slightly stronger if we include the CIL treatment observations, though the validity of the tests are then questionable because of lack of independence among different observations by the same subject in the CIL treatment.

of the theoretical hypotheses presented earlier, this suggests that α_1 is greater than α_2, especially when the difference in expected payoffs between the two candidates is small enough.

B. Regression analysis

We now make our analysis more rigorous using regression analysis. We identify the candidate chosen by each voter with a dummy variable 'Vote' (= 1 if the candidate is chosen, 0 otherwise). For each situation faced by a generic voter *i*, we have two observations and for only one of the two the variable 'Vote' is equal to 1. Based on the theoretical background presented in a previous section, we estimate the probability that a subject votes for a certain candidate based on the characteristics of the alternative candidates. In particular, we estimate an alternative-specific conditional choice model. Since in the CIL sessions we have multiple observations per individual, we employ robust standard errors clustered at individual level. The dependent variable is the dummy 'Vote'. In Regression 1, the independent variables include the logs of measured trustworthiness and competence of the candidate, log(trustworthiness) and log(competence).[136] In Regression 2, we also add interaction terms of these variables with a dummy variable $|\pi| > 5$, which takes value 1 when the absolute deviation in expected payoffs between the two candidates is larger than 5 experimental points. In Regression 3, we also control for the demographic, psychological and behavioral characteristics of the voters and treatment effects by interacting them with log(trustworthiness) and log(competence).[137] In particular, we control for the nationality of the subjects (UK and China),

[136] We cannot compute the log of the competence in 12 electoral situations – where one candidate (real or fictional) did not solve more than 40 tables – out of the 401 situations characterized by a trade-off between trustworthiness and competence. This is equivalent to only the 2.99% of the relevant electoral situations. The value of competence (tables correctly solved above 40 correct tables) for these few situations is in fact zero. We thus drop these few observations from the regression analysis. Note that these 12 electoral situations come from two subjects that did not solve more than 40 tables in the Baseline sessions and that were also randomly selected as fictional candidates by the computer in few other situations of the CIL sessions.

[137] Note that since our model is alternative-specific, the characteristics of the voters do not vary over the choices of the voters, and, therefore, they would be dropped out from the model. The only way to get around this problem and account for the individual characteristics of the voter is to add interaction terms between the alternative-specific variables and the voter-specific variables as we do in our regressions.

their gender, their age, whether they study economics or not, and whether they are undergraduate students or not. In addition, we control for the risk attitude of the subjects, and their scores in the SDS17 and MACH questionnaires. We also control for the sessions where the order of the real effort task and the trust game was counterbalanced, and the CIL sessions. Finally, we control for the behavior of the voters in the real effort task stage and the trust game stage by interacting the log of competence and trustworthiness of each voter with the log of the trustworthiness and competence respectively of the candidates.[138] In Regressions 4-6, we control for the behavior of the inconsistent subjects by including an interaction of whether a subject was categorized as inconsistent with log(trustworthiness) and log(competence) respectively. Table 3 displays the results of the regressions.

In Regressions 1 and 4, both the coefficients of the log of trustworthiness (α_1) and the log of competence (α_2) are positive and significant. The coefficient of log(trustworthiness) is slightly larger than the coefficient of log(competence) but the difference is not statiscally significant (χ^2 test, p = 0.975 in Regression 1, and 0.532 in Regression 4). In Regression 4, both the interaction terms of log(trustworthiness) and log(competence) with Inconsistency are negative and statistically significant. Also, the size of the coefficients is such as to largely offset the coefficients for log(trustworthiness) and log(competence) for inconsistent subjects. We can present the first and second result.

Result 1. In line with Hypothesis 1 but in contrast to both Hypotheses 2 and 3, subjects generally displayed a rational and profit-maximizing behavior by voting for the candidate who provided the highest expected profits, irrespectvely of his or her trustworthiness and competence.

Result 2. Inconsistent subjects tended to rely less on measured competence and trustworthiness.

[138] By doing that, we lose two additional observations corresponding to the two voters that did not solve more than 40 tables in the Baseline sessions, and for which we cannot compute the log of competence.

Table 3: Alternative-specific conditional logit regressions

	Regression 1		Regression 2		Regression 3	
	b	se	b	se	b	se
log(Trustworthiness)	1.47***	0.25	3.69***	1.11	4.38***	1.31
log(Competence)	1.46***	0.23	2.82***	0.82	2.76**	1.2
log(Trustworthiness) × |π|>5			-2.29**	1.03	-2.49**	1.08
log(Competence) × |π|>5			-1.24	0.79	-1.61**	0.81
Interactions with demographic, behavioral, psychological and treatment variables	No		No		Yes	
Obs	778		778		776	
Pseudo R^2	0.25		0.26		0.31	
Df	2		4		30	
Prob > F	0		0		0	
	Regression 4		Regression 5		Regression 6	
	b	se	b	se	b	se
log(Trustworthiness)	1.90***	0.29	3.88***	1.12	5.20***	1.59
log(Competence)	1.76***	0.27	2.91***	0.82	2.98**	1.21
log(Trustworthiness) × Inconsistency	-1.94***	0.46	-1.87***	0.45	-2.89***	0.56
log(Competence) × Inconsistency	-1.17*	0.61	-1.13	0.69	-1.49**	0.68
log(Trustworthiness) × |π|>5			-2.06*	1.06	-2.37**	1.18
log(Competence) × |π|>5			-1.02	0.81	-1.46*	0.86
Interactions with demographic, behavioral, psychological and treatment variables	No		No		Yes	
Obs	778		778		776	
Pseudo R^2	0.3		0.32		0.38	
Df	4		6		32	
Prob > F	0		0		0	

Notes: Alternative-specific logit regression with robust standard errors clustered at individual level. The table reports the beta coefficients and the standard errors. The demographic variables are age, gender (= 1 for men), economics background (= 1 if applicable), nationality (UK = 1 for UK subjects, and China = 1 for Chinese subjects), and University status (= 1 for undergraduate students). The behavioral variables are the competence and trustworthiness of the voter. The psychological variables are the risk attitude, the SDS17 score and MACH score. The treatment variables are the CIL sessions, and the sessions where the trust game stage took place before the real effort task stage. The psychological and behavioral variables and age are centered at the mean in order to control for high correlation between the independent variables (see Marquardt, 1980). 'China' identifies subjects from China, Taiwan or Hong Kong. The data correspond to cases where there was a trade-off between trustworthiness and competence. The full regressions are in the online appendix. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Once we control for small and large differences in expected payoffs between the two candidates (Regressions 2 and 5), the coefficient for log(trustworthiness) becomes significantly larger than the coefficient for log(competence) for small differences (χ^2 test, p = 0.079 for Regression 2, and 0.055 for Regression 5). In other words, for small differences in expected payoffs, we observe $\alpha_1 > \alpha_2$, and, thus, we reject Hypothesis 3 in favor of some qualified support for Hypothesis 2.

Result 3. Subjects tended to weigh trustworthiness more than competence when the difference in expected profits between the two candidates was small enough, as predicted by Hypothesis 3, but, as the difference increased, people cared only about their expeted payoffs, as predicted by Hypothesis 1.

This result also holds in Regressions 3 and 6 where we control for the demographic, psychological and behavioral characteristics of the subjects, and treatment effects from using the CIL method. In particular, for small differences in expected payoffs, the coefficient for log(trustworthiness) is almost twice as large as the coefficient for log(competence) and the difference is statistically significant (χ^2 test, p = 0.070 for Regression 3, and 0.041 for Regression 6).[139]

So far we have assumed that subjects displayed adaptive expectations, that is they formed their expectations about how the potential public official will behave in the future based on the information provided to them regarding the past competence and trustworthiness of the candidates. It is possible that subjects displayed rational expectations. This means that the subjects' expectations about the future trustworthiness and competence of the public official matched exactly the true expected values of future trustworthiness and competence of the public official. In the online appendix, we replicate the analysis conducted so far by assuming that subjects display rational expectations. The results are qualitatively similar to those presented in the chapter, if anything with stronger evidence of a trustworthiness bias. It might also be possible that our results are driven by extreme cases, that is situations where the difference in expected profits between the two candidates is very large. Hence, we replicate the analysis by dropping those cases. The results are reported in the online appendix and replicate those presented in the chapter.

[139] Among the controls, the only coefficient (weakly) statistically significant is the interaction term between the CIL treatment and the log of trustworthiness (p-value = 0.070 in Regression 6). In particular, subjects in the CIL sessions put a larger weight on trustworthiness compared to Baseline subjects.

C. *Types classification*

We now classify the subjects of the CIL sessions based on their pattern of voting behavior. We can do so in the CIL sessions (and only in the CIL sessions) since in this treatment we collected multiple observations of voting behavior for each subject. We identify 6 categories of subjects, and Table 4 summarizes the results of this classification.

Table 4: Subjects' classification based on their voting behavior

Type	Utility	% (ω)	% (ω)
Unconditional competence	$\alpha_1 = 0$ $\alpha_2 > 0$	6.94% (5)	2.78% (2)
Unconditional trustworthiness	$\alpha_1 > 0$ $\alpha_2 = 0$	8.33% (6)	8.33% (6)
Conditional competence	$\alpha_2 > \alpha_1$ $\alpha_{1,2} > 0$	8.33% (6)	2.78% (2)
Conditional trustworthiness	$\alpha_1 > \alpha_2$ $\alpha_{1,2} > 0$	18.06% (13)	18.06% (13)
Profit maximizing (adaptive expectations)	$\alpha_1 = \alpha_2$ $\alpha_{1,2} > 0$	44.44% (32)	34.72% (25)
Profit maximizing (rational expectations)	$\alpha_1 = \alpha_2$ $\alpha_{1,2} > 0$	.	19.44% (14)
Profit minimizing and Inconsistent	$\alpha_{1,2} \leq 0$	13.89% (10)	13.89% (10)
Total		100% (72)	100% (72)

'Profit-maximizing' voters. These subjects always selected the more profitable candidate irrespectively of his or her competence and trustworthiness. In terms of our theoretical specification, the utility of the 'profit-maximizing' voters is characterized by $\alpha_1 = \alpha_2 > 0$. In a first classification, we only consider those subjects who were profit-maximizing based on adaptive expectations. In a second classification, we also consider those subjects who were profit-maximizing based on rational expectations.[140] Subjects that do not fall in the 'Profit-maximizing' subjects category are classified as follow.

'Unconditional competence' voters. These subjects always selected the more competent candidate irrespectively of the expected profits. The utility function of these subjects is characterized by $\alpha_1 > 0$ and $\alpha_2 = 0$.

[140] 50% of the subjects who are classified as profit-maximizing based on rational expectations, also fit in the category of the subjects who are profit-maximizing based on adaptive expectations.

'Unconditional trustworthiness' voters. These subjects always selected the more trustworthy candidate irrespectively of the expected profits. Their utility function is represented by $\alpha_1 = 0$ and $\alpha_2 > 0$.

'Conditional competence' voters. These subjects selected more often the competent candidate than the trustworthy candidate.[141] The behavior of these subjects is captured by an utility function characterized by $\alpha_2 > \alpha_1 > 0$.

'Conditional trustworthiness' voters. These subjects selected more often the trustworthy candidate than the competent candidate. The behavior of these subjects is captured by an utility function characterized by $\alpha_1 > \alpha_2 > 0$.

'Profit-minimizing and Inconsistent' voters. These subjects tended to select the less profitable candidates or displayed a random voting behavior. This category includes the inconsistent subjects (i.e. subjects who selected the less profitable subjects when the latter was strictly or weakly dominated in both characteristics by the other candidate) and subjects who displayed a negative Spearman rank correlation coefficient between their voting decision and the difference in expected profits between the more and less profitable candidate (in other words, they display qualitatively the opposite pattern of the theoretical prediction of Figure 2). In terms of the parameters of the utility functions, the behavior of the profit-minimizing subjects is captured by $\alpha_1, \alpha_2 \leq 0$.

Table 4 shows that the majority of voters displayed a profit-maximizing behavior. About a quarter of voters had a preference for the trustworthy candidate ('Unconditional and Conditional trustworthiness' voters), and possibly as little as 6% preferred a competent candidate

[141] To identify these subjects, we computed, for each subject, the average vote for the trustworthy candidates when these were the least profitable, and compared it with the average vote for the competent candidates when these were the least profitable. If the difference was positive (i.e. the subject more often voted for the less profitable and trustworthy candidate than the less profitable and competent candidate), the subject was categorized as 'Conditional trustworthiness' subject. If the difference was negative (i.e. the subject more often voted for the less profitable and competent candidate than the less profitable and trustworthy candidate), the subject was categorized as 'Conditional competence' subject.

('Unconditional and Conditional competence' voters).[142] This evidence provides additional support on what we presented earlier, that is most people tend to select the candidate rationally, based on who provides the highest expected profit irrespectively of trustworthiness and competence, but there is also a proportion of people who have a bias towards caring about trustworthiness. In terms of our theoretical specification, this means that the majority of the subjects present either a utility function characterized by $\alpha_1 = \alpha_2 > 0$ or $\alpha_1 > \alpha_2 > 0$. Obviously, this classification should be considered with caution and only as complement of the previous analysis as it is based on very few electoral situations per subject.

In this section the focus has been on the analysis of the subjects' voting behavior since this was the main objective of this chapter. In the online appendix, we also analyze the behavior of the public officials in the third stage of the experiment.

5. Discussion

We investigated how voters weigh the competence and the trustworthiness of the candidates in public elections. We did so in a controlled environment which enabled us to rule out all the other influences that may affect the electoral choices of voters. In particular, since, in our experiment, candidates differ only on their level of trustworthiness and competence, we are able to study the pure preferences of voters for trustworthiness and competence, and the extent to which they only care about what they expect to go in their pockets. By and large, we find that voters care about their expected payoffs and little else. In 84.4% of the cases, voters behaved rationally by selecting the candidate who was more expected to be profitable, based on their ex-ante trustworthiness and competence; the percentage becomes even higher (87.9%) if we also include the cases described by rational expectations. A majority of subjects always unfalteringly goes for the candidate that is expected to yield a higher payoff.

[142] Note that this does not mean that these voters did not care about expected payoffs. As shown by Table 3, sacrifices of payoffs often need to be small enough in order for the bias towards trustworthiness or competence to emerge.

That said, around 25% of voters tend to be biased towards trustworthiness, and our regression analysis confirms a bias towards trustworthy candidates when the difference in expected profits between the two candidates is small enough. This holds irrespectively of whether we assume that voters display adaptive or rational expectations.

One could argue that the bias of the voters towards caring about trustworthiness that we observe in our experiment may not be the result of their preferences but of their misunderstanding of the instructions or systematic mistakes.[143] This interpretation is not plausible for several reasons. First, we made sure that subjects understood the instructions by asking them to complete a computerized questionnaire before starting each task, where subjects had to solve some exercises and calculate the effects, in terms of payoffs, of their actions. Before the start of each task, subjects were invited to ask questions if something was not clear and clarifications were offered aloud to them if they had any doubts about the procedures and the calculation of the earnings. Second, our results show that subjects selected the more trustworthy but unprofitable candidate only under certain circumstances, that is when the difference in expected payoffs between the two candidates were small enough. If people failed to understand the instructions or the payoffs function, the bias would have characterized more generally all our data.[144] Third, a lack of understanding or incentives would have resulted in random mistakes in both directions. This however was not the case as the bias occurred systematically in one direction. Fourth, in the data analysis, we controlled for the behavior of those (few) inconsistent subjects who displayed a more random behavior and could have failed to understand the instructions or taken the experiment less seriously.

Another possible explanation of why subjects displayed a bias towards caring about trustworthiness may be related to the partially different nature of the trust game compared to the official's dilemma game. In

[143] Subjects may have voted more often for the more trustworthy candidate either because they did not understand well the implications, in terms of payoffs, of their actions or because they committed more systematic mistakes when the difference in expected profits between the two candidates was small enough and the incentives to select the profitable candidate smaller.

[144] That, is we would have also observed the bias when the difference between the expected payoffs of the candidates was larger.

particular, in the trust game, the trustee received an endowment from the truster as 'manna from heaven', whereas in the official's dilemma game the public official had to produce the endowment by counting 1s in tables. As a result, the measure of trustworthiness obtained in the trust game may have been different from the measure of trustworthiness obtained in the official's dilemma game. In particular, the trust game might have provided an inflated measure of trustworthiness compared to the official's dilemma game since, in the latter, voters might have felt more entitled to keep money for themselves as they had to work hard to produce the endowment.[145] The data analysis however showed that the two measures were highly correlated, with the measure obtained in the trust game actually lower than the measure obtained in the official's dilemma game. And, as previously noted, the trustworthiness bias is replicated (if anything, it is slightly stronger) if we assume that voters held rational expectations rather than on adaptive expectations.

Another potential relevant confound concerns the state under which the decisions in the trust game and official's dilemma game respectively were taken. In particular, in the trust game, subjects decided 'in a cold state' as they did not know yet which role the computer assigned to them and under the presumption that the truster was going to trust them. In the official's dilemma game, the decision was more in a hot state as the public officials knew that trust was placed on them. As a result, the subjects may have been more sensitive to certain psychological pressures, such as trust responsiveness (Guerra and Zizzo, 2004) and reciprocity (Falk and Fischbacher, 2006), in the official's dilemma game than in the trust game, and, therefore, they may have fulfilled trust more in the official's dilemma game. If subjects anticipated that, they might have believed that the measure of trustworthiness provided to them was understated. As a result, they might have scaled up the information about the trustworthiness of the candidates provided to them. This however does not change our results, since the

[145] Similar implications arise if subjects learned how to be untrustworthy in the trust game, and, in turn, displayed lower trustworthiness in the official's dilemma game. Note however that this type of learning could have also worked in the opposite direction: subjects learned how to be trustworthy in trust game, and, in turn, repaid trust more in the official's dilemma game.

voting choice that brings the highest utility is the same irrespectively of how utility is scaled (or the attributes are scaled).[146]

Another possible criticism of our experiment is that subjects were provided only with ex-ante measures of competence and trustworthiness which, from the standpoint of the subjects, may not necessarily capture the ex-post behavior of the public official. As a result, people may have formed certain beliefs about the ex-post competence and trustworthiness of the public official which could have not reflected the information provided to them during the voting phase. This is not however a problem as we found a very high correlation between early and later measures of trustworthiness and competence which would not justify such behavior unless subjects were extremely naïve. Second, and as previously discussed, we also analyzed the data assuming that the subjects had correct beliefs and predictions of the future behavior of the public official (rational expectations).[147]

Our experiment was conducted in the United Kingdom with subjects with a variety of different backgrounds. We do not find any differences, in our data, on how voters weigh the trustworthiness and competence of the candidates across the different nationalities of our subjects. This study is however not specifically designed to investigate cross-national differences. In particular, the number of observations that we collected for each country is relatively small, and, therefore, any definite conclusion would be too

---

[146] To illustrate this, suppose that a voter *n* believe that a unit of ex-ante trustworthiness is equivalent to τ units of ex-post trustworthiness. As long as this belief is the same for all the candidates (which is reasonable in our case since voters are only informed about the trustworthiness and competence of the candidates and nothing else), the voting choice of the voter does not change. More formally, the probability that voter *n* chooses candidate *i* is the same irrespectively of how the attributes are scaled:

$$P_{n,i} = Prob\left(U_{n,i} > U_{n,j}\right) = Prob\left(U_{n,i} - U_{n,j} > 0\right) = Prob(A^*(\tau T_{ni})^{\alpha_1} {C_{ni}}^{\alpha_2} - A^*\left(\tau T_{nj}\right)^{\alpha_1} {C_{nj}}^{\alpha_2} > 0\right) = Prob\left(A^* {T_{ni}}^{\alpha_1} {C_{ni}}^{\alpha_2} - A^* {T_{nj}}^{\alpha_1} {C_{nj}}^{\alpha_2} > 0\right)$$

In terms of our regression analysis, it means that the beliefs that change the scale of the attributes (trustworthiness and competence), do not change the estimation of the parameters α_1 and α_2.

[147] Also, as we have already mentioned earlier, the voting choice that brings the highest utility is the same irrespectively of how utility is scaled (or the attributes are scaled). This means that, as long as the beliefs of the voters change only the scale of the attributes of the candidates, our results do not change.

premature. Future research may wish to explore the extent to which our findings hold across different countries.

6. Conclusions

Our results show that voters tend mostly to care only about their final expected payoffs, irrespectively of the trustworthiness and competence of the candidates. These findings are useful to understand how voters decide in public elections. In particular, they support, in most of the cases, the idea that what ultimately matters for the voters is what they get in their pocket. Quoting James Carville's famous slogan, "[it is] the economy [that matters], stupid". As a result of this, voters may be willing to support untrustworthy candidates if the latter are perceived to contribute more to the overall welfare of the voters. This could explain why democracies may at times suffer from dishonesty and corruption at the public level.

We did identify a bias towards caring about trustworthiness, particularly when the candidates are similar in terms of their contribution to the financial welfare of the voters (that is, the difference in expected payoffs between the candidates is small enough). In these occasions, the information about the trustworthiness of the candidates can become crucial to determine which candidate will be elected, and so it is *not just* "the economy, stupid".

References

Abeler, J., A. Becker, and A. Falk. 2013. "Truth-Telling: A Representative Assessment". SSRN Scholarly Paper ID 2164648.

Abeler, J., A. Falk, L. Goette, and D. Huffman. 2011. "Reference Points and Effort Provision." *The American Economic Review*. 101 (2): 470–492.

Arbak, E., and M-C. Villeval. 2013. "Voluntary Leadership: Motivation and Influence." *Social Choice and Welfare*, 40 (3): 635–662.

Azfar, O., and W. R. Nelson Jr. 2007. "Transparency, Wages, and the Separation of Powers: An Experimental Analysis of Corruption." *Public Choice*, 130 (3-4): 471–493.

Bågenholm, A. 2013. “Throwing the Rascals Out? The Electoral Effects of Corruption Allegations and Corruption Scandals in Europe 1981–2011.” *Crime, Law and Social Change*, 60 (5): 1–15.

Bardsley, N. 2000. “Control Without Deception: Individual Behaviour in Free-Riding Experiments Revisited.” *Experimental Economics*, 3 (3): 215–240.

Barr, A., M. Lindelow, and P. Serneels. 2009. “Corruption in Public Service Delivery: An Experimental Analysis.” *Journal of Economic Behavior and Organization*. 72 (1): 225–239.

Bartlett, J. 2013.“How Beppe Grillo's social media politics took Italy by storm”. *theguardian.com*, 26 February. Available online at http://www.theguardian.com/commentisfree/2013/feb/26/beppe-grillo-politics-social-media-italy (accessed 27 February 2014).

Berg, J., J. Dickhaut, and K. McCabe. 1995. “Trust, Reciprocity, and Social History.” *Games and Economic Behavior*, 10 (1): 122–142.

Bohnet, I., F. Greig, B. Herrmann, and R. Zeckhauser. 2008. “Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States.” *The American Economic Review*, 98 (1): 294–310.

Brandts, J., D. J. Cooper, and R. Weber. 2013. “Legitimacy, Social Distance, and Leadership in the Turnaround Game.” Working Paper.

Bucciol, A., and M. Piovesan. 2011. “Luck or Cheating? A Field Experiment on Honesty with Children.” *Journal of Economic Psychology*, 32 (1): 73–78.

Caselli, F., and M. Morelli. 2004. “Bad Politicians.” *Journal of Public Economics*, 88 (3–4): 759–782.

Christie, R., F. L. Geis, and D. Berger. 1970. Studies in Machiavellianism. New York: Academic Press.

Corazzini, L., S. Kube, M. A. Maréchal, and A. Nicolo. 2013. “Elections and Deceptions: An Experimental Study on the Behavioral Effects of Democracy”. SSRN Scholarly Paper ID 1430904.

Croson, R. T. A. 2005. “Deception in Economics Experiments.” In C. Gerschlager (Ed.), Deception in Markets: An Economic Analysis, Basingstoke: Palgrave Macmillan.

DeMaio, T. J. 1984. Social Desirability and Survey Measurement: A Review. New York: Russell Sage Foundation.

Eckel, C. C., P. J. Grossman, C. A. Johnson, A. C. M. de Oliveira, C. Rojas, and R. K. Wilson. 2012. “School Environment and Risk Preferences: Experimental Evidence.” *Journal of Risk and Uncertainty*, 45 (3): 265–292.

Eggers, A., and A. C. Fisher. 2011. “Electoral Accountability and the UK Parliamentary Expenses Scandal: Did Voters Punish Corrupt MPs?” SSRN Scholarly Paper ID 1931868.

Falk, A., and U. Fischbacher. 2006. “A Theory of Reciprocity.” *Games and Economic Behavior*, 54 (2): 293–315.

Ferraz, C., and F. Finan. 2008. “Exposing Corrupt Politicians: The Effects of Brazil’s Publicly Released Audits on Electoral Outcomes.” *The Quarterly Journal of Economics*, 123 (2): 703–745.

Fischbacher, U. 2007. “z-Tree: Zurich Toolbox for Ready-made Economic Experiments.” *Experimental Economics*, 10 (2): 171–178.

Fischbacher, U., and F. Föllmi-Heusi. 2013. “Lies in Disguise-an Experimental Study on Cheating.” *Journal of the European Economic Association*, 11 (3): 525–547.

Gächter, S., and E. Renner. 2005. “Leading by Example in the Presence of Free Rider Incentives.” CeDEx Discussion Paper, University of Nottingham.

Gächter, S., D. Nosenzo, E. Renner, and M. Sefton. 2012. "Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-by-Example." *Economic Inquiry*, 50 (4): 953–967.

Gneezy, U. 2005. "Deception: The Role of Consequences." *The American Economic Review*, 95 (1): 384–394.

Greiner, B. 2004. "An Online Recruitment System for Economic Experiments". MPRA Paper.

Guerra, G., and D. J. Zizzo. 2004. "Trust Responsiveness and Beliefs." *Journal of Economic Behavior and Organization*, 55 (1): 25–30.

Güth, W., M. V. Levati, M. Sutter, and E. van der Heijden. 2007. "Leading by Example with and Without Exclusion Power in Voluntary Contribution Experiments." *Journal of Public Economics,* 91 (5–6): 1023–1042.

Hamman, J. R., R. A. Weber, and J. Woon. 2011. "An Experimental Investigation of Electoral Delegation and the Provision of Public Goods." *American Journal of Political Science*, 55 (4): 738–752.

Hao, L., and D. Houser. 2013. "Perceptions, Intentions, and Cheating". Working Paper 1039, George Mason University, Interdisciplinary Center for Economic Science.

Houser, D., S. Vetter, and J. Winter. 2012. "Fairness and Cheating." *European Economic Review,* 56 (8): 1645–1655.

Kocher, M. G., G. Pogrebna, and M. Sutter. 2013. "Other-regarding Preferences and Management Styles." *Journal of Economic Behavior and Organization,* 88: 109–132.

Kulisheck, M. R, and J. J Mondak. 1996. "Candidate Quality and the Congressional Vote: A Causal Connection?" *Electoral Studies*, 15 (2): 237–253.

Kumru, C. S., and L. Vesterlund. 2010. "The Effect of Status on Charitable Giving." *Journal of Public Economic Theory*, 12 (4): 709–735.

Levy, D. M., K. Padgitt, S. J. Peart, D. Houser, and E. Xiao. 2011. "Leadership, Cheap Talk and Really Cheap Talk." *Journal of Economic Behavior and Organization*, 77 (1): 40–52.

Lodge, M., P. Stroh, and J. Wahlke. 1990. "Black-box Models of Candidate Evaluation." *Political Behavior*, 12 (1): 5–18.

Marquardt, D.W., 1980. "You Should Standardize the Predictor Variables in Your Regression Models." *Journal of the American Statistical Association*, 75 (369): 87-91.

Mas-Colell, A., M. D. Whinston, and J. R. Green. 1995. *Microeconomic theory*. Vol. 1. New York: Oxford university press.

Mazar, N., and D. Ariely. 2006. "Dishonesty in Everyday Life and Its Policy Implications." *Journal of Public Policy and Marketing*, 25 (1): 117–126.

McCurley, C., and J. J. Mondak. 1995. "Inspected by #1184063113: The Influence of Incumbents' Competence and Integrity in U.S. House Elections." *American Journal of Political Science*, 39 (4): 864.

Messick, D. M., H. Wilke, M. B. Brewer, R. M. Kramer, P. E. Zemke, and L. Lui. 1983. "Individual Adaptations and Structural Change as Solutions to Social Dilemmas." *Journal of Personality and Social Psychology*, 44 (2): 294–309.

Mondak, J. J. 1995. "Competence, Integrity, and the Electoral Success of Congressional Incumbents." *The Journal of Politics*, 57 (04): 1043–1069.

Mondak, J. J., and R. Huckfeldt. 2006. "The Accessibility and Utility of Candidate Character in Electoral Decision Making." *Electoral Studies*, 25 (1): 20–34.

Moxnes, E., and E. van der Heijden. 2003. "The Effect of Leadership in a Public Bad Experiment." *Journal of Conflict Resolution*, 47 (6): 773–795.

Peters, J. G., and S. Welch. 1980. “The Effects of Charges of Corruption on Voting Behavior in Congressional Elections.” *The American Political Science Review,* 74 (3): 697.

Pokorny, K. 2008. “Pay—but Do Not Pay Too Much: An Experimental Study on the Impact of Incentives.” *Journal of Economic Behavior and Organization,* 66 (2): 251–264.

Potters, J., M. Sefton, and L. Vesterlund. 2007. “Leading-by-example and Signaling in Voluntary Contribution Games: An Experimental Study.” *Economic Theory,* 33 (1): 169–182.

Rabin, M. 1993. “Incorporating Fairness into Game Theory and Economics.” *The American Economic Review,* 83 (5): 1281–1302.

Rivas, M. F., and M. Sutter. 2011. “The Benefits of Voluntary Leadership in Experimental Public Goods Games.” *Economics Letters,* 112 (2): 176–178.

Samuelson, C. D, and D. M. Messick. 1986. “Alternative Structural Solutions to Resource Dilemmas.” *Organizational Behavior and Human Decision Processes,* 37 (1): 139–155.

Samuelson, C. D., D. M. Messick, C. Rutte, and H. Wilke. 1984. “Individual and Structural Solutions to Resource Dilemmas in Two Cultures.” *Journal of Personality and Social Psychology,* 47 (1): 94–104.

Shampanier, K., N. Mazar, and D. Ariely. 2007. “Zero as a Special Price: The True Value of Free Products.” *Marketing Science,* 26 (6): 742–757.

Stöber, J. 2001. “The Social Desirability Scale-17 (SDS-17): Convergent Validity, Discriminant Validity, and Relationship with Age.” *European Journal of Psychological Assessment,* 17 (3): 222–232.

Sutter, M. 2009. “Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams.” *The Economic Journal,* 119 (534): 47–60.

Welch, S., and J. R. Hibbing. 1997. “The Effects of Charges of Corruption on Voting Behavior in Congressional Elections, 1982–1990.” *The Journal of Politics*, 59 (01): 226–239.

Winters, M. S., and R. Weitz-Shapiro. 2013. “Lacking Information or Condoning Corruption: When Do Voters Support Corrupt Politicians?” *Comparative Politics*, 45 (4): 418–436.

Appendix to Chapter 3: Competence versus Trustworthiness: What do Voters Care About?

K. Experimental instructions
L. Final questionnaire
M. Background information on participants
N. Full regressions of Table 4
O. Analysis with rational expectations
P. Analysis without outliers
Q. Analysis of the public officials' behavior
R. Additional analysis of behavior in stages 1 and 2

A. Experimental instructions

- *Baseline (in half of the sessions, the order of stage 1 and stage 2 was inverted)*

Instructions

Introduction

This is an experiment on decision making. The instructions are the same for all participants. During the experiment, you are not allowed to communicate with other participants. Please raise your hand if you have any questions at any point during the experiment. If you have any questions, the experimenter will come to you and answer your questions privately. If the question is relevant to everyone, the experimenter will repeat the answer aloud.

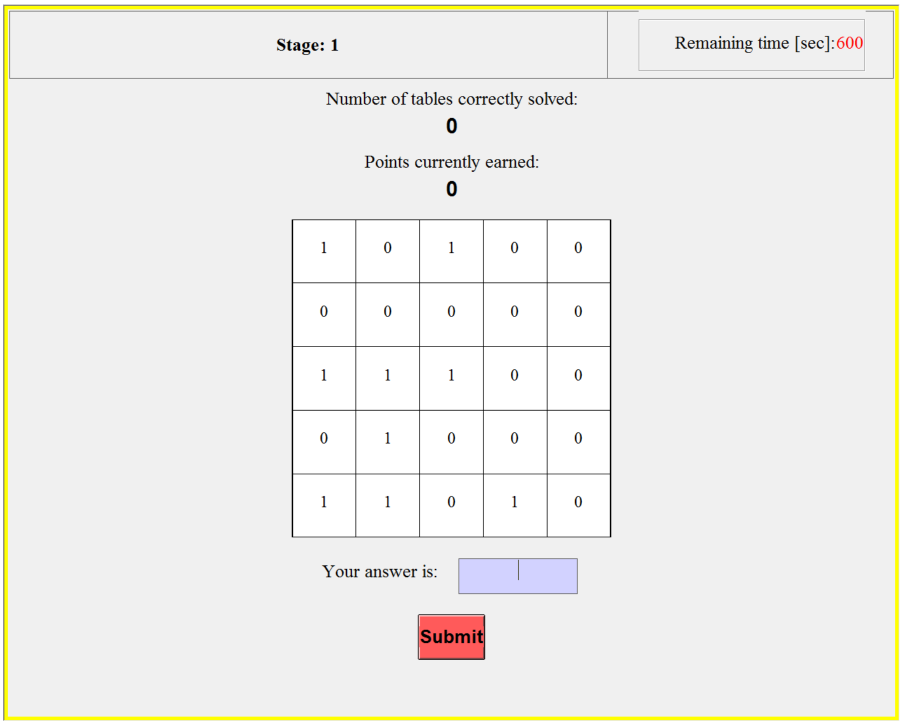
The experiment consists of **three stages**. In addition to these three stages, you will be asked some individual questions at the end of the experiment. At the beginning of each stage you will receive the corresponding instructions. The information you provide in each stage of the experiment may be reported to other participants at later stages of the experiment. However, all of your decisions and answers will remain **anonymous**.

During this experiment, your earnings depend on your decisions and the decisions of the other participants. It is therefore important that you read the instructions with care. Your earnings from the experiment will be computed in "**points**".

At the end of the experiment one stage will be chosen at random, and you will be paid the points that you earned in that stage. The points that you earn in that stage will be converted into pounds at an exchange rate of **1 point = 20 pence**. In addition to this, you will also be paid a show-up fee of 2 pounds and any additional earnings that you may obtain by answering the questions at the end of the experiment. You will be paid individually and in cash in a separate room by a person who is not present during the experiment and who is not aware of the content of this experiment.

First stage

In this stage of the experiment your task will be to count the number of 1s in a series of tables containing 0s and 1s. The figure shows the kind of screen you will see later:



You will have to enter the number of 1s into the box below the table and click the *Submit* button. After you have submitted your answer, a new table will be generated.

You will only earn money after correctly solving 40 tables. Specifically, you will receive 1 point for each table you correctly solve on top of the first 40 correctly solved tables. The greater the number of tables you solve correctly over and above the first 40 correctly solved tables, the more points you will earn.

You will have 10 minutes to complete the first stage of the experiment. Your remaining time will be displayed in the upper right hand corner of the screen.

Second stage

In this stage of the experiment, you will be matched at random with another participant. You will never interact with this person again in the remainder of the experiment. One of the two will be randomly assigned the role of **participant A**, and the other the role of **participant B**.

Participant A will receive an endowment of **30 points**. He or she will decide whether or not to transfer all the 30 points to participant B. There are two scenarios:

1. If participant A decides **not to transfer** the 30 points to participant B, participant A will earn 30 points and participant B 0 points.

2. If participant A decides **to transfer** the 30 points to participant B, these points get multiplied by **3** before they are received by participant B. Hence, participant B will receive 90 points overall. Participant B then will decide how many points to keep and how many points to return to participant A. Specifically, he/she can return to participant A any amount between a minimum of 9 points to a maximum of 90 points.

You will be informed about your role (participant A or B) only at the end of the experiment. Hence, at this stage, you will have to make decisions in the roles of both participant A and participant B:

- As participant A, you will have to decide if you want to transfer the 30 points or not to participant B.
- As participant B, you will make a decision without knowing if participant A has chosen to transfer or not the 30 points to you. Specifically, you will have to decide how many points you would wish to return to participant A if participant A were to transfer his or her 30 points to you.

At the end of this stage, if participant A has chosen **not to transfer the 30 points,** participant B's decision will be ignored and earnings will be 30 points for participant A and 0 for participant B. If participant A has chosen **to transfer the 30 points,** participant B's decision will determine the earnings of both participants.

The results and earnings for this stage will be communicated to you at the end of the experiment and will depend on the role that you have been assigned to.

Third stage

The task in this stage of the experiment is to count 1s in a series of tables as in the first stage of the experiment. However, new rules are now in effect, which did not apply in the first stage.

Specifically, you will be randomly matched with two other participants you have never been matched with before. You and these two participants will now be referred to as *co-participants*. The three of you will have a *common fund*. At the beginning of the task, the value of the common fund will be set to 0 points. Every co-participant will individually work on the task for 10 minutes (the remaining time will be displayed in the upper right hand corner of the work screen). However, only one co-participant's work will count for the earnings of this stage of the experiment. This co-participant will be referred to as the *appointed co-participant*. After the appointed co-participant correctly solves 40 tables, each additional correct answer of the appointed co-participant will increase the value of the common fund by **4 points**.

At the end of the task, the common fund will have accumulated a certain number of points equal to four times the number of tables that the appointed co-participant correctly solved on top of the first 40 correctly solved tables. Only the appointed co-participant will know the number of tables that he or she correctly solved (and so the value of the common fund). He or she will be asked to report the value of the common fund to the other co-participants. He or she can report any number between 0 and the true value of the common fund. This number corresponds to the **reported value** of the common fund. The reported value of the common fund will be split equally between the co-participants. That is, each co-participant (including the appointed co-participant) will receive 1/3 of the reported value of the common fund. If applicable, the appointed co-participant will also earn the whole of the **non-reported value** of the common fund, that is the value of the common fund not reported. Note that the appointed co-participant will be free to report or not the true value of the common fund. Moreover, the other co-participants will only be informed of the reported value of the common fund.

Example: Suppose the appointed co-participant solves 52 tables correctly (hence 12 tables over and above 40 correctly solved tables); the value of the common fund is 12 × 4 = 48. The appointed co-participant however reports a value of 33. The non-reported value of the common fund is 48 – 33 = 15. The earnings are therefore:

- 33 / 3 = 11 for each non-appointed co-participant;
- 33 / 3 + 15 = 26 for the appointed co-participant.

How is the appointed co-participant selected? Before starting the task, each co-participant will be informed of:

- the number of tables correctly solved by each other co-participant in the first stage of the experiment where the task was to count 1s in tables;

- the proportion of points that each other co-participant in the role of participant B returned to participant A in the second stage of the experiment.

Each co-participant will then be asked to choose which of the two other co-participants he or she would like to select as the appointed co-participant. The decision of one randomly selected co-participant will be implemented. Hence, it is in your best interest to choose the co-participant that you really want as the appointed co-participant.

Afterwards, the computer will inform each co-participant whether or not he or she is the appointed co-participant. Then the task of counting 1s in a series of tables will start.

In summary

- You will be randomly matched with two other participants you have never been matched with before. You and these two participants will be referred to as *co-participants*. The three of you will have a common fund with an initial value of 0 points.
- Each co-participant will choose whom he or she wants to select as *appointed co-participant* between the other two co-participants. The decision of one co-participant selected at random will be implemented.
- Everyone will work for 10 minutes on the task which consists in counting 1s in a series of tables containing 0s and 1s. However, only the work of the appointed co-participant will count for the earnings. Specifically, for each table that the appointed co-participant correctly solves on top of the first 40 correctly solved tables, the value of the common fund increases by 4 points.
- At the end of the task the appointed co-participant can report any number between 0 and the true value of the common fund to the other co-participants. This number constitutes the reported value of the common fund. The other co-participants will be informed only about the reported value of the common fund.
- The reported value of the common fund will be divided in equal parts (each worth 1/3 of the reported value) between the co-participants (including the appointed co-participant). If applicable,

the appointed co-participant will also earn the whole of the non-reported value of the common fund, that is the value of common fund not reported.

- *Stage 3 of CIL treatment*

<u>Third stage</u>

The task in this stage of the experiment is to count 1s in a series of tables as in the first stage of the experiment. However, new rules are now in effect, which did not apply in the first stage.

Specifically, you will be randomly matched with two other participants you have never been matched with before. You and these two participants will now be referred to as *co-participants*. The three of you will have a *common fund*. At the beginning of the task, the value of the common fund will be set to 0 points. Every co-participant will individually work on the task for 10 minutes (the remaining time will be displayed in the upper right hand corner of the work screen). However, only one co-participant's work will count for the earnings of this stage of the experiment. This co-participant will be referred to as the *appointed co-participant*. After the appointed co-participant correctly solves 40 tables, each additional correct answer of the appointed co-participant will increase the value of the common fund by **4 points**.

At the end of the task, the common fund will have accumulated a certain number of points equal to four times the number of tables that the appointed co-participant correctly solved on top of the first 40 correctly solved tables. Only the appointed co-participant will know the number of tables that he or she correctly solved (and so the value of the common fund). He or she will be asked to report the value of the common fund to the other co-participants. He or she can report any number between 0 and the true value of the common fund. This number corresponds to the **reported value** of the common fund. The reported value of the common fund will be split equally between the co-participants. That is, each co-participant (including the appointed co-participant) will receive 1/3 of the reported value of the common fund. If applicable, the appointed co-participant will also earn the whole of the **non-reported value** of the common fund, that is the value of the common fund not reported. Note that the appointed co-participant will be free to report or not the true value of the common fund. Moreover, the other co-participants will only be informed of the reported value of the common fund.

Example: Suppose the appointed co-participant solves 52 tables correctly (hence 12 tables over and above 40 correctly solved tables); the value of the common fund is $12 \times 4 = 48$. The appointed co-participant however reports a value of 33. The non-reported value of the common fund is $48 - 33 = 15$. The earnings are therefore:

- $33 / 3 = 11$ for each non-appointed co-participant;
- $33 / 3 + 15 = 26$ for the appointed co-participant.

How is the appointed co-participant selected? Before starting the task, each co-participant will be placed in 7 situations. Only one of these will be real, the others will be fictional.

In each situation, each co-participant will be informed of:

- the number of tables correctly solved by each other co-participant in the first stage of the experiment where the task was to count 1s in tables;
- the proportion of points that each other co-participant in the role of participant B returned to participant A in the second stage of the experiment.

However, only in the real situation, the information provided is about your actual current co-participants. In the fictional situations, the information provided is about people who participated in past sessions of this experiment.

For each situation, each co-participant will be asked to choose which of the two other co-participants he or she would like to select as the appointed co-participant. Only the decision of one randomly selected co-participant in the real situation will be implemented and count towards your earnings of this stage. Note that, for all you know, each situation could be the real one, in which case ALL information you are given about it is true, and only the real one may have any effect on who is going to be the appointed co-participant. Hence, it is in your best interest to treat each situation as if it is real and to choose, for each situation, the co-participant that you really want as the appointed co-participant.

Afterwards, the computer will inform each co-participant whether or not he or she is the appointed co-participant based on the outcome of the real situation. Then the task of counting 1s in a series of tables will start.

In summary

- You will be randomly matched with two other participants you have never been matched with before. You and these two participants will be referred to as *co-participants*. The three of you will have a common fund with an initial value of 0 points.

- Each co-participant will choose whom he or she wants to select as *appointed co-participant* between the other two co-participants in different situations. Only one of these situations is the real one. The decision of one co-participant selected at random in the real situation will be implemented.
- Everyone will work for 10 minutes on the task which consists in counting 1s in a series of tables containing 0s and 1s. However, only the work of the appointed co-participant will count for the earnings. Specifically, for each table that the appointed co-participant correctly solves on top of the first 40 correctly solved tables, the value of the common fund increases by 4 points.
- At the end of the task the appointed co-participant can report any number between 0 and the true value of the common fund to the other co-participants. This number constitutes the reported value of the common fund. The other co-participants will be informed only about the reported value of the common fund.
- The reported value of the common fund will be divided in equal parts (each worth 1/3 of the reported value) between the co-participants (including the appointed co-participant). If applicable, the appointed co-participant will also earn the whole of the non-reported value of the common fund, that is the value of common fund not reported.

B. Final questionnaire

Note: In the Part 1 of the questionnaire, appointed and non-appointed co-participants were asked different questions (see below).

Part 1 *(only for non-appointed co-participants)*

In this part of the questionnaire, we would like you to answer the following questions regarding your predictions about stage 3. You will be paid **an extra point** for each correct prediction.

1) Do you think the appointed co-participant underreported the value of the common fund?

Yes or No

2) Do you think the other co-participant who was not selected as the appointed co-participant voted for the same co-participant as you?

Yes or No

3) How do you feel the appointed co-participant was ranked among the three co-participants in terms of number of tables correctly solved in stage 1?

He/she was ranked first (i.e. he/she correctly solved the largest number of tables)

He/she was ranked second

He/she was ranked third (i.e. he/she correctly solved the smallest number of tables)

4) How do you feel the appointed co-participant was ranked among the three co-participants in terms of number of points returned to participant A in stage 2

He/she was ranked first (i.e. he/she returned the largest number of points)

He/she was ranked second

He/she was ranked third (i.e. he/she returned the smallest number of points)

- - - - - - - -

Part 1 *(only for appointed co-participants)*

1) Do you think both the other two co-participants voted for you as the appointed co-participant?

Yes or No

2) Do you think that the co-participant, selected at random, who voted for you thought that you were going to underreport the value of the common fund?

Yes or No

3) How do you feel you were ranked among the three co-participants in terms of number of tables correctly solved in stage 1?

I was ranked first (i.e. I correctly solved the largest number of tables)

I was ranked second

I was ranked third (i.e. I correctly solved the smallest number of tables)

4) How do you feel you were ranked among the three co-participants in terms of number of points returned to participant A in stage 2?

I was ranked first (i.e. I returned the largest number of points)

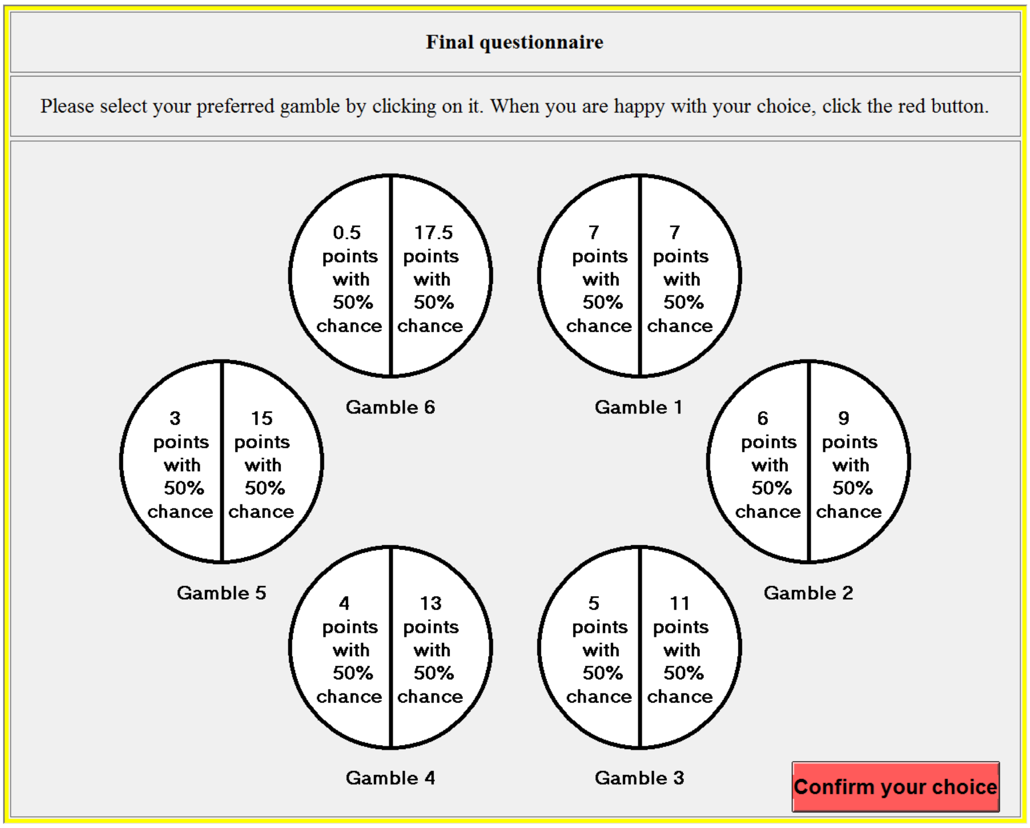
I was ranked second

I was ranked third (i.e. I returned the smallest number of points)

- - - - - - - -

Part 2

You will now be asked to select from among six different gambles the one gamble you would like to take. The figure shows the kind of screen you will use to select the gamble.



Each circle represents a different gamble. Each circle is divided in two parts. Each part is a possible outcome of the gamble. For every gamble, each outcome is equally likely, that is it has a 50% chance of happening. The number of points that the gamble will give for each possible outcome is written inside the circle.

At the end of the experiment, you will roll a six-sided die to determine which outcome of your selected gamble will occur:

- If you roll a 1, 2, or 3, you will receive the points on the left part of the circle.
- If you roll a 4, 5, or 6, you will receive the points on the right side of the circle.

Note that, no matter which gamble you pick, each outcome has a 50% chance of occurring.

To select a gamble you have to click on it with the mouse. You can revise your choice as many times as you want. When you are happy with your choice, click the "*Confirm your choice*" button to confirm.

Example: Suppose you select gamble 4 and later you roll a 1, 2, or 3, your earnings will be 4 points. If you roll 4, 5, or 6, you will earn 13 points.

- - - - - - - -

Part 3

A list of statements will be displayed. Please read each statement carefully and decide if that statement describes you or not. If it describes you click the word *true* if not, choose the word *false*. After each response a new statement will appear. There are sixteen statements.

1. I sometimes litter.

2. I always admit my mistakes openly and face the potential negative consequences.

3. In traffic I am always polite and considerate of others.

4. I have tried illegal drugs (for example, marijuana, cocaine, etc.).

5. I always accept others' opinions, even when they don't agree with my own.

6. I take out my bad moods on others now and then.

7. There has been an occasion when I took advantage of someone else.

8. In conversations I always listen attentively and let others finish their sentences.

9. I never hesitate to help someone in case of emergency.

10. When I have made a promise, I keep it--no ifs, ands or buts.

11. I occasionally speak badly of others behind their back.

12. I would never live off other people.

13. I always stay friendly and courteous with other people, even when I am stressed out.

14. During arguments I always stay objective and matter-of-fact.

15. There has been at least one occasion when I failed to return an item that I borrowed.

16. I always eat a healthy diet.

17. Sometimes I only help because I expect something in return

For each statement, subjects were asked to select True or False.

- - - - - - - -

Part 4

A list of attitude statements will be displayed. Each represents a commonly held opinion and there are no right or wrong answers. Read each statement carefully. Then indicate the extent to which you agree or disagree. First impressions are usually best in such matters. After each response a new statement will appear. There are twenty statements.

1. Never tell anyone the real reason you did something unless it is useful to do so.

2. The best way to handle people is to tell them what they want to hear.

3. One should take action only when sure it is morally right.

4. Most people are basically good and kind.

5. It is safest to assume that all people have a vicious streak and it will come out when they are given a chance.

6. Honesty is the best policy in all cases.

7. There is no excuse for lying to someone else.

8. Generally speaking, men won't work hard unless they're forced to do so.

9. All in all, it is better to be humble and honest than important and dishonest.

10. When you ask someone to do something for you, it is best to give the real reasons for wanting it rather than giving reasons that carry more weight.

11. Most people who get ahead in the world lead clean, moral lives.

12. Anyone who completely trusts anyone else is asking for trouble.

13. The biggest difference between most criminals and other people is that criminals are stupid enough to get caught.

14. Most men are brave.

15. It is wise to flatter important people.

16. It is possible to be good in all respects.

17. Barnum was very wrong when he said that there's a sucker born every minute.

18. It is hard to get ahead without cutting corners here and there.

19. People suffering from incurable diseases should have the choice of being put painlessly to death.

20. Most men forget more easily the death of their father than the loss of their property.

For each statement, subjects were asked to select one of the following options: strongly disagree, somewhat disagree, slightly disagree, no opinion, slightly agree, somewhat agree, strongly agree.

- - - - - - - -

Part 5

In this part of the questionnaire, we would like you to provide some personal information if so you wish.

What is your gender? (*Female or Male*)

What is your country of origin?

Are you a native English speaker? (*Yes or No*)

Your age?

Which course are you registered on?

Did you attend a course in Economics during your studies? (*Yes or No*)

Level of current degree? (*INTO, Undergraduate (e.g.BSc, BA, LLB, MBBS), Postgraduate Taught (e.g.MA, MSc), Postgraduate Research (e.g. MPhil, PhD) or Other*)

*If you ticked the "Other" in the question above please specify if you wish

What is your religion or belief? (*No Religion, Buddhist, Christian, Sikh, Muslim, Confucian, Hindu, Jewish, Atheist, Other or Prefer not to say*)

*If you ticked the "Other" in the question above please specify if you wish

What is your relationship status? (*Single, Engaged, In a relationship, Married, Civil Partnership, Widowed, Seperated/ Divorced or Prefer not to say*)

How many times have you participated in previous experiments? (*0, 1, 2, 3 or More than 3*)

Have you ever participated before in an experiment where the task was to count 1s in a series of table containing 0s and 1s like in this experiment? (*Yes or No*)

What do you think this experiment is about?

C. Background information on participants

Characteristics	n = 240
Gender	
Female	150 (62.50%)
Male	90 (37.50%)
Age	
Mean	23.30
St. dev.	4.43
Min.	18
Max.	65
Level of current degree	
INTO	3 (1.25%)
Erasmus	1 (0.42%)
Postgraduate Research (e.g. MPhil, PhD)	19 (7.92%)
Postgraduate Taught (e.g.MA, MSc)	95 (39.58%)
Undergraduate (e.g.BSc, BA, LLB, MBBS)	122 (50.83%)
Background in economics[148]	
No	130 (54.17%)
Yes	110 (45.83%)
Country of origin	
Bahrain	1 (0.42%)
Bangladesh	1 (0.42%)
Brazil	1 (0.42%)
Bulgaria	1 (0.42%)
China, Taiwan or Hong Kong	98 (40.83%)
Egypt	1 (0.42%)
Germany	4 (1.67%)
Greece	2 (0.83%)
Hungary	1 (0.42%)
India	2 (0.83%)
Indonesia	1 (0.42%)
Iran	2 (0.83%)
Ireland	2 (0.83%)
Italy	1 (0.42%)
Japan	1 (0.42%)
Jordan	1 (0.42%)
Kazakhstan	2 (0.83%)
Latvia	1 (0.42%)
Lithuania	3 (1.25%)
Malaysia	4 (1.67%)
Maldives	1 (0.42%)
Mauritius	4 (1.67%)
Netherlands	1 (0.42%)
Nigeria	4 (1.67%)
Norway	1 (0.42%)

[148] The question was: "Did you attend a course in Economics during your studies?" Yes or No.

Palestine	1 (0.42%)
Philippines	1 (0.42%)
Poland	5 (2.08%)
Portugal	1 (0.42%)
Romania	1 (0.42%)
Russia	2 (0.83%)
Somalia	1 (0.42%)
Sri Lanka	1 (0.42%)
Tanzania	1 (0.42%)
Thailand	2 (0.83%)
UK	67 (27.92%)
USA	4 (1.67%)
Vietnam	12 (5%)
Native English speaker	
No	160 (66.67%)
Yes	80 (33.33%)
Relationship Status	
Engaged	3 (1.25%)
In a relationship	81 (33.75%)
Married	9 (3.75%)
Prefer not to say	7 (2.92%)
Separated/ Divorced	1 (0.42%)
Single	139 (57.92%)
Religion or belief	
Atheist	20 (8.33%)
Agnostic	1 (0.42%)
Buddhist	16 (6.67%)
Christian	43 (17.92%)
Christian and Sikh	1 (0.42%)
Daoism	1 (0.42%)
Muslim	18 (7.5%)
No Religion	123 (51.25%)
Other	4 (1.67%)
Prefer not to say	13 (5.42%)
Participation in previous experiments (n.)	
Never	8 (3.33%)
1	25 (10.42%)
2	16 (6.67%)
3	15 (6.25%)
4 or more	176 (73.33%)
Participation in a similar real-effort task	
No	146 (60.83%)
Yes	94 (39.17%)

D. Full regressions of Table 4

Table D1: Conditional Logit Regressions

	Regression 3		Regression 6	
	b	se	b	se
log(Trustworthiness)	4.38***	1.31	5.20***	1.59
log(Competence)	2.76**	1.2	2.98**	1.21
log(Trustworthiness)×Inconsistency	.	.	-2.89***	0.56
log(Competence)×Inconsistency	.	.	-1.49**	0.68
log(Trustworthiness)×|p|>5	-2.49**	1.08	-2.37**	1.18
log(Competence)×|p|>5	-1.61**	0.81	-1.46*	0.86
log(Trustworthiness)×UK	1.86	1.66	1.45	1.65
log(Competence)×UK	2.1	2.13	1.65	2.1
log(Trustworthiness)×China	-0.06	0.62	-0.35	0.69
log(Competence)×China	0.14	0.66	-0.13	0.65
log(Trustworthiness)×Male	-0.19	0.65	-0.61	0.84
log(Competence)×Male	0.99	0.63	0.88	0.8
log(Trustworthiness)×Economics	-0.14	0.56	0.24	0.57
log(Competence)×Economics	-0.17	0.5	0.2	0.54
log(Trustworthiness)×Age	0.02	0.15	-0.02	0.14
log(Competence)×Age	0.05	0.17	0.02	0.16
log(Trustworthiness)×Undergraduate	0.1	0.71	-0.17	0.72
log(Competence)×Undergraduate	0.03	0.92	-0.2	0.89
log(Trustworthiness)×Trust Game first	-0.45	0.57	-0.23	0.58
log(Competence)×Trust Game first	-0.37	0.58	-0.01	0.6
log(Trustworthiness)×CIL	0.4	0.5	0.96*	0.53
log(Competence)×CIL	0.46	0.61	1	0.66
log(Trustworthiness)×Risk choice	-0.12	0.17	-0.31	0.21
log(Competence)×Risk choice	-0.02	0.18	-0.15	0.21
log(Trustworthiness)×SDS17 Score	0.02	0.09	0.08	0.07
log(Competence)×SDS17 Score	0.01	0.09	0.07	0.09
log(Trustworthiness)×MACH Score	-0.02	0.02	0.01	0.03
log(Competence)×MACH Score	0.01	0.02	0.03	0.03
log(Trustworthiness)×Voter's Return rate	-0.14	0.41	0.13	0.39
log(Competence)×Voter's Return rate	0	0.51	0.17	0.55
log(Trustworthiness)×Voter's Competence	-0.31	0.36	0.13	0.42
log(Competence)×Voter's Competence	0.26	0.41	0.59	0.44
Obs	776		776	
Pseudo R^2	0.31		0.38	
Df	30		32	
Prob > F	0		0	

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E. Analysis with rational expectations

In this section, we replicate the analysis of the chapter assuming that subjects have rational expectations. If subjects display rational expectations, they are on average able to predict the candidate that ex post will generate more profit. We can construct a measure of the expected ex post payoffs generated to the voters by the candidates by looking at how much ex ante trustworthiness and ex ante competence of the public officials explain the ex post payoffs generated to the voters by the public officials. In particular, we can estimate an OLS regression where the dependent variable is the ex post voter's payoffs generated by the public officials,[149] while the independent variables are the ex ante trustworthiness (measured as the return rate from the earlier stage) and ex ante competence (measured as number of tables correctly solved on top of the first 40 tables correctly solved in the earlier stage) of the public officials. We can then multiply the estimated coefficients with the ex ante trustworthiness and the ex ante competence respectively of all the candidates (both appointed and non-appointed participants) to obtain a statistical expected measure of the expected ex post payoffs generated to the voters by the candidates. Table E1 presents the result of this estimation.

Table E1: OLS regression on ex post voter's payoffs

	Ex-post voter's payoffs	
	b	se
Ex-ante Trustworthiness	34.24**	14.86
Ex-ante Competence	0.74***	0.12
Obs	80	
Adj. R^2	0.317	
Df	2	
Prob > F	0	

Notes: OLS regression. The table reports the beta coefficients and the standard errors. Observations are from the public officials. * p < 0.1, ** p < 0.05, *** p < 0.01. The coefficients are not normalized.

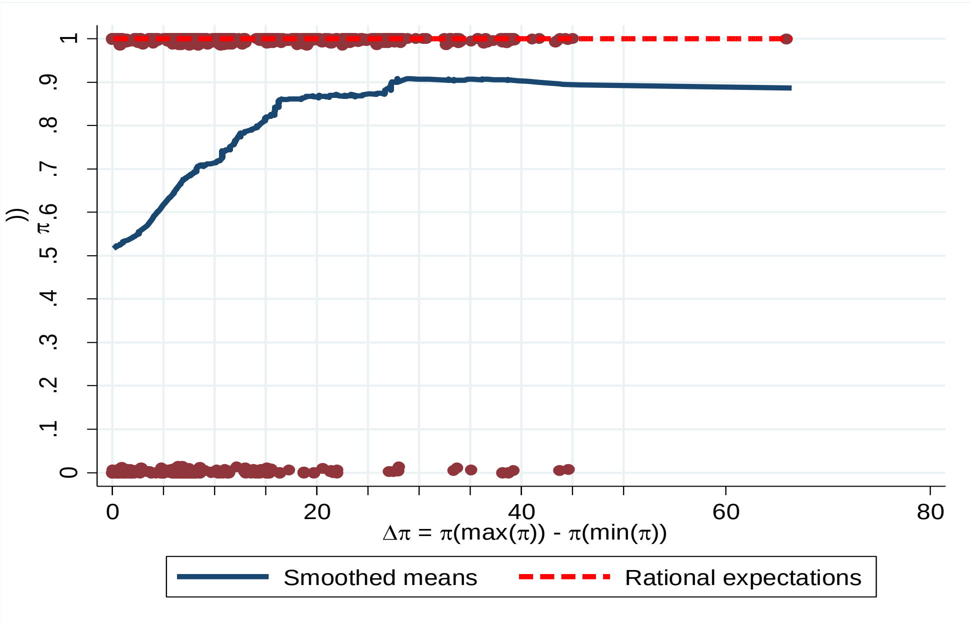
[149] These payoffs are calculated as: $\pi_{i,k}^{ex\ post} = \frac{4}{3} T_{p,k}^{ex\ post} A_{p,k}^{ex\ post}$, where i identifies the voter, k the triad, $T_{p,k}^{ex\ post}$ the ex-post rate of honesty of the public official, and $A_{p,k}^{ex\ post}$ the ex-post competence of the public official.

The expected values of the ex post profits are calculated using the following formula:

$$\pi_{i,k}^{ex\ post} = 34.24 \times T_{j,k}^{ex\ ante} + 0.74 \times A_{j,k}^{ex\ ante}$$

where $T_{j,k}^{ex\ ante}$ and $A_{j,k}^{ex\ ante}$ are the ex ante trustworthiness and ex ante competence respectively of the candidate j. Having now a measure of the ex post voter's payoffs generated by the candidates, we can study whether subjects displayed rational expectations. In particular, we can look at how the probability of voting for the more (ex post) profitable candidate evolves as the difference in ex post payoffs between the more and less profitable candidates increases (Figure E1).[150] If subjects have rational expectations, they should always select the more (ex post) profitable candidate (graphically, we should observe a straight line at P(max(π)) = 1). This seems to be the tendency when the difference in ex post payoffs between the more and less profitable candidates is large. When the difference is small, decisions are noisier.

Figure E1: Probability of electing the ex post more profitable candidate



Notes: P(max(π)) is the probability of electing the ex post more profitable candidate. Δπ is the difference in expected payoffs between the more and less ex post profitable candidate. The running means are weighted to give more importance to near points than far, and computed using a bandwith of 0.6 (60% of the data).

[150] The probability is obtained by computing the weighted running means of a dichotomous variable taking value 1 when the ex-post more profitable candidate is elected and 0 otherwise.

We can also study whether the voters favor the ex post trustworthy or competent candidate. To do so, we need to obtain a measure of ex post trustworthiness and ex post competence for all the candidates. We can estimate two OLS regressions, one for trustworthiness and one for competence, where the dependent variable is the ex post trustworthiness (competence) of the public officials, while the independent variable is the ex ante trustworthiness (competence) of the public officials. We can then multiply the estimated coefficients with the ex ante trustworthiness and the ex ante competence respectively of all the candidates (both appointed and non-appointed participants) to obtain a statistical expected measure of the ex post trustworthiness and ex post competence respectively. Table E2 presents the result of these estimations.

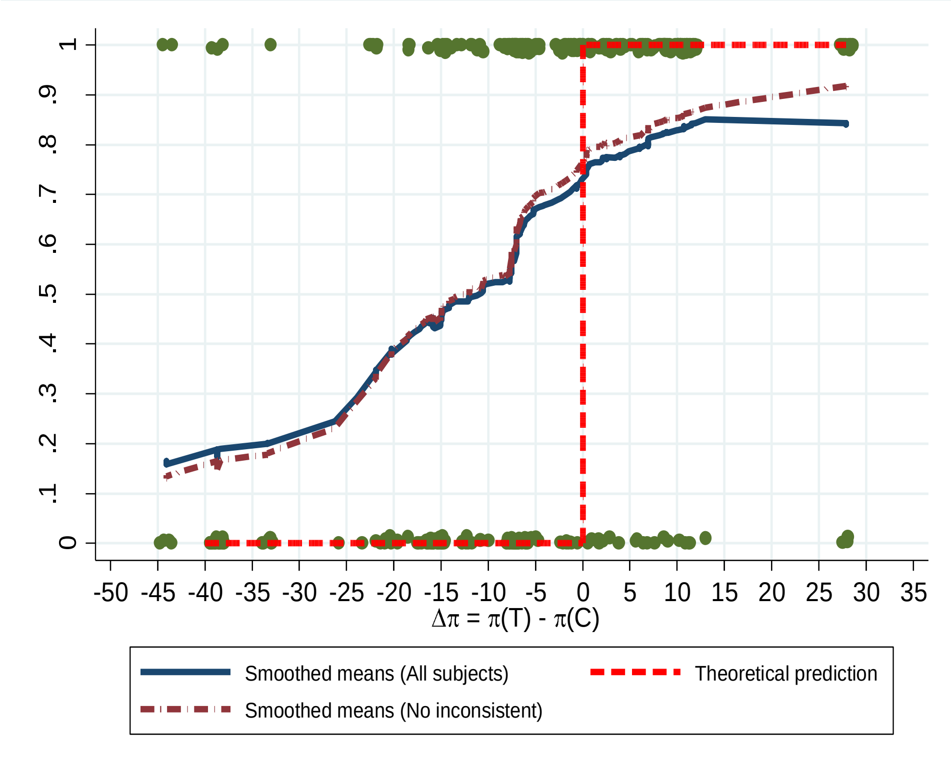
Table E2: OLS regressions on ex post trustworthiness and competence

	Ex-post trustworthiness		Ex-post competence	
	b	se	b	se
Ex-ante Trustworthiness	0.50**	0.19	.	.
Ex-ante Competence	.	.	0.89***	0.05
Obs	80		80	
Adj. R^2	0.069		0.795	
Df	1		1	
Prob > F	0.010		0	

Notes: OLS regression. The table reports the beta coefficients and the standard errors. Observations are from the public officials. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The coefficients are not normalized.

Focusing on cases where there is a trade-off between ex post trustworthiness and ex post competence, we can look at how the probability of voting for the more ex post trustworthy candidate evolves if the difference in ex post payoffs $\Delta\pi$ between the more and less trustworthy candidates increases (Figure E2). For $\Delta\pi < 0$ (i.e. the ex post more trustworthy candidate is also the ex post less profitable), profit-maximizing subjects with rational expectations should vote for the ex post less trustworthy subjects as he or she is associated with higher ex post payoffs. For $\Delta\pi > 0$ (i.e. the ex post more trustworthy candidate is also the ex post more profitable), profit-maximizing subjects with rational expectations should vote for the ex post more trustworthy subjects as he or she is associated with higher ex post payoffs. The figure shows that subjects did not seem to vote for the ex post more profitable candidate, as rational expectations would predict, but the candidate who is ex post more trustworthy, especially when the difference in ex post payoffs between the candidates is small. This pattern is similar to the one observed for adaptive expectations (see main chapter), and, perhaps, even more marked.

Figure E2: Probability of electing the ex post more trustworthy candidate



Notes: P(T) is the probability of electing the more ex post trustworthy candidate. Δπ is the difference in expected payoffs between the more and less ex post trustworthy candidate. The running means are weighted to give more importance to near points than far, and computed using a bandwith of 0.6 (60% of the data). The data corresponds to cases where there is a trade-off between ex post trustworthiness and competence.

We can also conduct a regression analysis like the one in the chapter but using the ex post measures of trustworthiness and competence. The dependent variable is the dummy 'Vote'. In regression 1, the independent variables include the log of ex post trustworthiness and ex post competence of the candidate. In regression 2, we also add interaction terms of log(ex post trustworthiness) and log(ex post competence) with a dummy variable |π|>5 which takes value 1 when the absolute deviation in ex post expected payoffs between the two candidates is bigger than 5 experimental points. In regression 3, we control as well for the demographic, psychological and behavioral characteristics of the voters and treatment effects by interacting them with log(ex post trustworthiness) and log(ex post competence).[151] In regressions 4-6, we also control for the behavior of the inconsistent subjects by including an interaction of whether a subject was categorized as

[151] As we already explained in the chapter, since the characteristics of the voters do not vary over the choices of the voters, in the regression we can only have interaction terms between the alternative-specific variables and the voter-specific variables.

inconsistent with log(ex post trustworthiness) and log(ex post competence) respectively. Table E3 displays the results of the regressions.[152]

Table E3: Alternative-specific conditional logit regressions

	Regression 1		Regression 2		Regression 3	
	b	se	b	se	b	se
log(Trustworthiness)	1.37***	0.24	1.56***	0.47	1.78**	0.81
log(Competence)	0.98***	0.22	0.49*	0.27	-0.47	0.92
log(Trustworthiness) × |π|>5			-0.24	0.47	-0.17	0.53
log(Competence) × |π|>5			0.52	0.36	0.9	0.61
Interactions with demographic, behavioral, psychological and treatment variables	No		No		Yes	
Obs	774		774		772	
Pseudo R^2	0.23		0.24		0.3	
Df	2		4		30	
Prob > F	0		0		0	
	Regression 4		Regression 5		Regression 6	
	b	se	b	se	b	se
log(Trustworthiness)	1.75***	0.28	1.81***	0.47	2.40***	0.92
log(Competence)	1.16***	0.26	0.52*	0.28	-0.59	0.96
log(Trustworthiness)×Inconsistency	-1.78***	0.45	-1.78***	0.46	-2.84***	0.58
log(Competence)×Inconsistency	-0.63	0.5	-0.64	0.53	-1.12**	0.52
log(Trustworthiness) × |π|>5			-0.11	0.5	0.12	0.57
log(Competence) × |π|>5			0.69*	0.39	1.25*	0.7
Interactions with demographic, behavioral, psychological and treatment variables	No		No		Yes	
Obs	774		774		772	
Pseudo R^2	0.28		0.29		0.37	
Df	4		6		32	
Prob > F	0		0		0	

Notes: Alternative-specific logit regression with clustered robust standard errors. The table reports the beta coefficients and the standard errors. The demographic variables are age, gender (= 1 for men), economics background (= 1 if applicable), nationality (UK = 1 for UK subjects, and China = 1 for Chinese subjects), and University status (= 1 for undergraduate students). The behavioral variables are the competence and trustworthiness of the voter. The psychological variables are the risk attitude, the SDS17 score and MACH score. The treatment variables are the CIL sessions, and the sessions where the trust game stage took place before the real effort task stage. The psychological and behavioral variables and age are centered at the mean in order to control for high correlation between the independent variables (see Marquardt, 1980). 'China' identifies subjects from China, Taiwan or Hong Kong. The data correspond to cases where there was a trade-off between trustworthiness and competence. The full regressions are available from the authors upon request. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In Regressions 1 and 4, both the coefficients of log(trustworthiness) and log(competence) are positive and strongly significantly. The coefficient of log(trustworthiness) is also significantly larger than the coefficient of

[152] The full regressions are available upon request.

log(competence) (χ^2 test, p = 0.056 in Regression 1, and 0.010 in Regression 4). This indicates that, if we assume that voters have rational expectations, the bias towards caring about trustworthiness carries through and is perhaps even stronger. If we focus on small differences in ex post expected payoffs between the two candidates (Regressions 2-3 and 5-6), the coefficient for log(trustworthiness) becomes even larger than the coefficient for log(competence). Alltogher these results support the key finding of the chapter that people care about the trustworthiness of the candidates.

Finally, we can compare how many electoral choices were consistent with rational expectations and how many electoral choices were consistent with adaptive expectations. This information is summarized in Tables E4 and E5. Both tables show that a significant proportion of choices were consistent with either adaptive or rational expectations. In addition, they suggest that more choices display adaptive (84.38% for all the subjects, 88.08% if we exclude the inconsistent subjects) rather than rational (75.30% for all the subjects, 78.31% if we exclude the inconsistent subjects) expectations (χ^2 test, $p < 0.001$).

Table E4: Electoral choices consistent with adaptive or rational expectations

	All subjects			**No inconsistent subjects**		
Adaptive expectations	**Baseline**	**CIL**	**Total**	**Baseline**	**CIL**	**Total**
NO	31 (18.45%)	74 (14.68%)	105 (15.63%)	26 (15.95%)	46 (10.43%)	72 (11.92%)
YES	137 (81.55%)	430 (85.32%)	567 (84.38%)	137 (84.05%)	395 (89.57%)	532 (88.08%)
Total	168 (100%)	504 (100%)	672 (100%)	163 (100%)	441 (100%)	604 (100%)
Rational expectations	**Baseline**	**CIL**	**Total**	**Baseline**	**CIL**	**Total**
NO	44 (26.19%)	122 (24.21%)	166 (24.7%)	39 (23.93%)	92 (20.86%)	131 (21.69%)
YES	124 (73.81%)	382 (75.79%)	506 (75.3%)	124 (76.07%)	349 (79.14%)	473 (78.31%)
Total	168 (100%)	504 (100%)	672 (100%)	163 (100%)	441 (100%)	604 (100%)

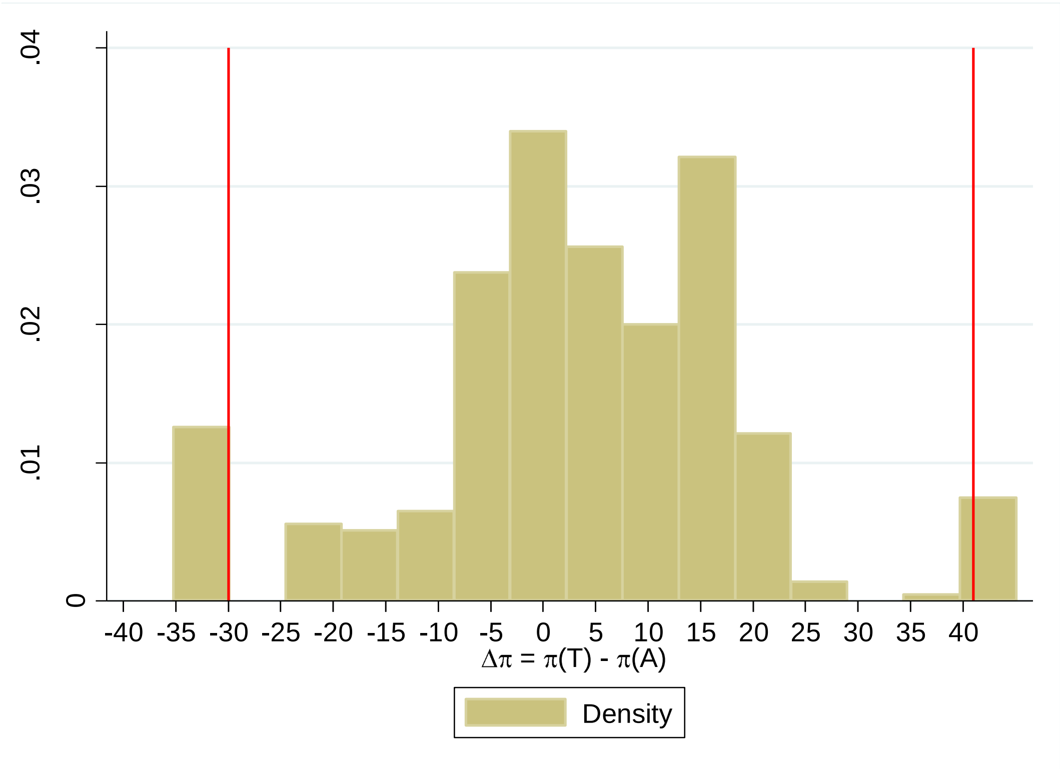
Table E5: Electoral choices consistent with adaptive and/or rational expectations

Baseline	Rational expectations		
Adaptive expectations	NO	YES	Total
NO	23 (4.56%)	8 (1.59%)	31 (6.15%)
YES	21 (4.17%)	116 (23.02%)	137 (27.18%)
Total	44 (8.73%)	124 (24.6%)	168 (33.33%)
CIL	Rational expectations		
Adaptive expectations	NO	YES	Total
NO	58 (11.51%)	16 (3.17%)	74 (14.68%)
YES	64 (12.7%)	366 (72.62%)	430 (85.32%)
Total	122 (24.21%)	382 (75.79%)	504 (100%)

F. Analysis without outliers

In this section, we test whether the results of the chapter are driven by those cases where the difference in expected profits between the two candidates was very large. We do so by replicating the analysis of the chapter with the exclusion of the extreme observations. Outliers are detected using the Carling's (2000) median rule.[153] In particular, we drop the observations where the difference in expected profits between the the more trustworthy candidate and the more competent candidate was larger than 41 experimental points and smaller than -30 experimental points (see Figure F1).[154]

Figure F1: Histogram of the difference in expected profits



Notes: The data correspond to cases where there was a trade-off between comeptence and trustworthiness. The difference in expected profits is between the profit of the more trustworthy candidate and the profit of the more competent candidate. Outliers are located on the left and on the right of the first and second vertical line respectively.

---

[153] According to the Median Rule, outliers are the observations above an upper cut-off point c^U and below an lower cut-off point c^L. These points are calculated as:

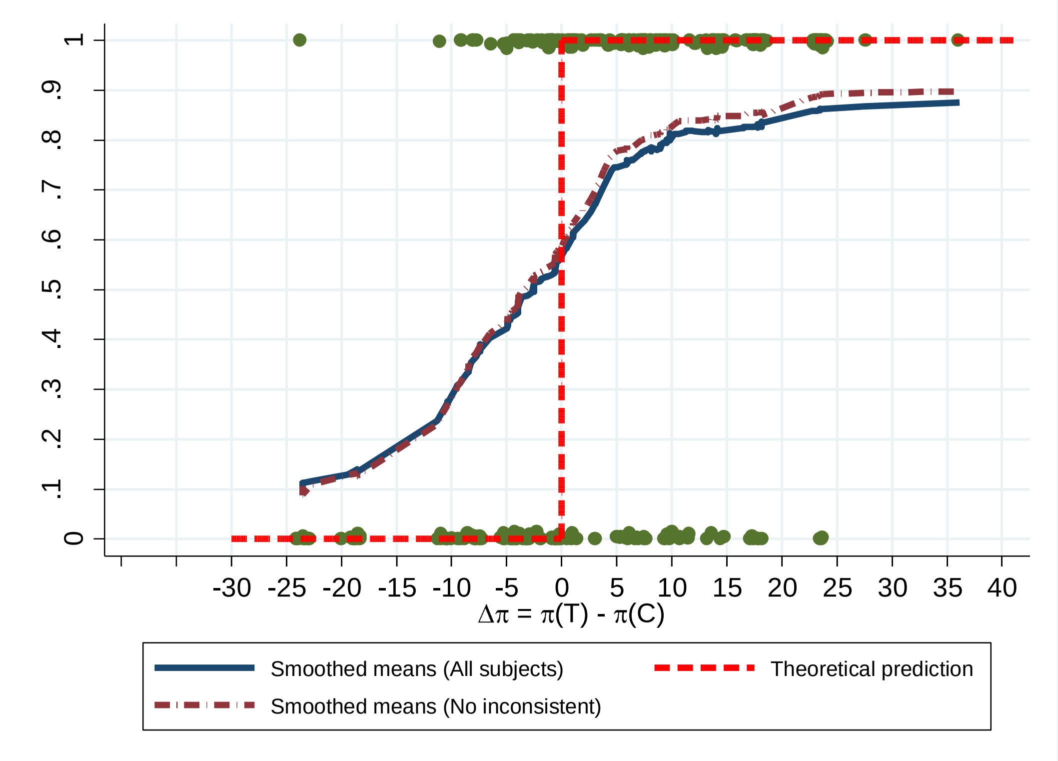
$$c^U = q_2 + k_2(q_3 - q_1)$$
$$c^L = q_2 - k_2(q_3 - q_1)$$

where q_1, q_2, and q_3 are the first, second, and third quartile respectively, while k_2 is a constant based on a pre-specified outside rate, and it is usually equal to about 2 (Carling, 2000).

[154] Similar cut-off points are obtained if we use the Tuckey's Rule (Tuckey, 1977).

Figure F2 shows the probability of electing the more trustworthy candidate as a function of the difference in expected payoffs between the more and less trustworthy candidate (as in Figure 2 of the chapter). Even once we exclude the extreme cases, the area below the weighted running means for $\Delta\pi < 0$ is bigger than the area above the weighted running means for $\Delta\pi > 0$, particularly for small differences in expected payoffs ($|\Delta\pi| \leq 5$). We also replicate the regression analysis of the chapter (Table F1). With the removal of the outliers, the results do not change. If anything, the point estimates slightly increase both for trustworthiness and competence, and their difference acquires more significance. In particular, the coefficient of log(trustworthiness) is significantly larger than the coefficient of log(competence) in Regression 2 (χ^2 test, p = 0.079), Regression 3 (p = 0.028), Regression 5 (p = 0.030) and Regression 6 (p = 0.024).

Figure F2: Probability of electing the more trustworthy candidate



Notes: P(T) is the probability of electing the more trustowrthy candidate. $\Delta\pi$ is the difference in expected payoffs between the more and less trustworthy candidate. The running means are weighted to give more importance to near points than far, and computed using a bandwith of 0.6 (60% of the data).

Table F1: Alternative-specific conditional logit regressions

	Regression 1		Regression 2		Regression 3	
	b	se	b	se	b	se
log(Trustworthiness)	1.60***	0.25	3.69***	1.11	4.27***	1.37
log(Competence)	1.56***	0.27	2.82***	0.82	2.27*	1.36
log(Trustworthiness) × |π|>5			-2.13**	1.03	-2.39**	1.11
log(Competence) × |π|>5			-1.07	0.82	-1.54*	0.89
Interactions with demographic, behavioral, psychological and treatment variables	No		No		Yes	
Obs	692		692		690	
Pseudo R2	0.24		0.25		0.30	
Df	2		4		30	
Prob > F	0		0		0	

	Regression 4		Regression 5		Regression 6	
	b	se	b	se	b	se
log(Trustworthiness)	1.81***	0.3	3.94***	1.18	4.86***	1.63
log(Competence)	1.61***	0.3	2.86***	0.84	2.28*	1.34
log(Trustworthiness)×Inconsistency	-1.38***	0.53	-1.25**	0.51	-2.38***	0.66
log(Competence)×Inconsistency	0.04	0.73	0.29	0.86	-0.26	0.89
log(Trustworthiness) × |π|>5			-2.20**	1.11	-2.45*	1.26
log(Competence) × |π|>5			-1.1	0.85	-1.53	0.93
Interactions with demographic, behavioral, psychological and treatment variables	No		No		Yes	
Obs	692		692		690	
Pseudo R^2	0.27		0.28		0.35	
Df	4		6		32	
Prob > F	0		0		0	

Notes: Alternative-specific logit regression with clustered robust standard errors. The table reports the beta coefficients and the standard errors. The demographic variables are age, gender (= 1 for men), economics background (= 1 if applicable), nationality (UK = 1 for UK subjects, and China = 1 for Chinese subjects), and University status (= 1 for undergraduate students). The behavioral variables are the competence and trustworthiness of the voter. The psychological variables are the risk attitude, the SDS17 score and MACH score. The treatment variables are the CIL sessions, and the sessions where the trust game stage took place before the real effort task stage. The psychological and behavioral variables and age are centered at the mean in order to control for high correlation between the independent variables (see Marquardt, 1980). 'China' identifies subjects from China, Taiwan or Hong Kong. The data correspond to cases where there was a trade-off between trustworthiness and competence. The full regressions are available from the authors upon request. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References Not in Chapter

Carling, K. 2000. "Resistant Outlier Rules and the Non-Gaussian Case", *Computational Statistics and Data Analysis*, 33 (3): 249-258.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

G. Analysis of the public officials' behavior

We now look at the behavior of subjects in the third stage of the expeirment (the Official's Dilemma Game). Table G1 reports the average competence in the 1st and 2nd real-effort task for each treatment. In the 2nd real effort task, we do not detect any statistically significant differences in competence across treatments both in bivariate tests (Mann-Whitney test, $p > 0.1$) and in aggregate (Kruskal-Wallis test, $p > 0.1$).[155]

Table G1: Competence in the 1st and 2nd real effort task

		Real effort Task first	Trust Game first	Total
1st real effort task				
Non-appointed[a]		38.14 (16.42)	33.66 (15.71)	35.9 (16.17)
Appointed[a]		42.2 (20.09)	38.98 (18.95)	40.59 (19.47)
2nd real effort task		**Real effort Task first**	**Trust Game first**	**Total**
Non-appointed	**Baseline**	41.27 (21.61)	40.13 (17.68)	40.7 (19.66)
	CIL	38.42 (20.86)	39.42 (19.75)	38.92 (20.11)
	Total	40.41 (21.3)	39.91 (18.21)	40.16 (19.75)
Appointed	**Baseline**	52.96 (21.07)	47.11 (19.87)	50.04 (20.51)
	CIL	49.92 (17.65)	46.5 (16.87)	48.21 (16.98)
	Total	52.05 (19.93)	46.93 (18.81)	49.49 (19.43)

Notes: Standard deviations are in parenthesis. [a] For the 1st real effort task, we only report the total average of the Baseline and CIL treatments pooled together, since the two treatments were equivalent in the first two stages of the experiment (i.e. the treatment manipulation involved only stage 3).

If we compare the performance of subjects in the 1st real-effort task (the one in stage 1 or 2) with the performance in the 2nd real-effort task (the one in the Official's Dilemma Game), we observe, not surprisingly, that, because of learning, subjects performed better in the 2nd real-effort task (Wilcoxon signed-rank test, $p < 0.001$). This result is robust across treatments and for appointed and non-appointed participants respectively. Comparing the behavior of appointed and non-appointed participants, appointed participants performed better than non-appointed participants in the second real-effort task (Mann-Whitney test, $p = 0.004$) but not in the first one ($p = 0.198$). This is not surprising since the performance of the non-appointed participants in stage 3 was not incentivized.

[155] Tests were performed for the entire sample, only for non-appointed participants, and only for appointed participants.

Appointed members had returned significantly more points in the earlier trust game than non-appointed participants (p = 0.001). This is unsurprising as trustworthy candidates were more likely to be appointed public official. Table G2 summarizes these statistics.

Table G2: Trustworthiness in the 2nd trust game

	Non-appointed participants			Appointed participants		
Variable	n.	Mean	St. dev.	n.	Mean	St. dev.
Return rate	160	0.32	0.18	80	0.41	0.16

Notes: 'Return rate' is equal to the amount returned by the trustee divided by 90.

We can also study what explains the decision of the appointed participants to truthfully report the value of the common fund. Let the *honesty rate* be the proportion of the common fund reported by the public official to the other subjects. Table G3 shows average honesty rates for each experimental treatment. The honesty rate was significantly smaller in the sessions where the Trust Game was played first, both in aggregate (Mann-Whitney test, p = 0.001) and separately for each treatment (p = 0.026 in the Baseline, and 0.011 in the CIL). A possible explanation of this finding is that, in those sessions, subjects required a higher return from the effort since the two real-effort tasks occurred one after the other, without any break or alternative task between them. In particular, subjects might have struggled more to complete the two tasks and believed to have exerted a higher effort. This potential explanation is supported by the fact that, looking at the data from the incetivized part of the final questionnaire, a higher proportion of subjects in the sessions where the trust game stage took place first felt they were ranked first among the three participants in terms of number of tables correctly solved in the first real effort task, compared to subjects who participated in the sessions where the real effort task stage took place first (χ^2 test, p = 0.020).[156]

We can regress the honesty rate against the socio-demographic, psychological and behavioral characteristics of the appointed co-participant,

[156] Specifically, in the sessions where the trust game (real-effort task) was played first, 60% (32.50%) of the subjects felt they were rank first, 37.50% (52.50%) second and 2.50% (15%) third.

and the treatments dummies.[157] Table G4 reports the results of this regression. Subjects who were more trustworthy in the trust game displayed a higher honesty rate. Subjects with higher scores in the Machiavellianism scale were more dishonest. Participants in the sessions where the trust game was played first were also generally more dishonest than other subjects.

Table G3: Rate of honesty

	Real-Effort Task first (n = 42)	**Trust Game first (n = 42)**
Baseline (n = 56)	0.67 (0.28)	0.49 (0.30)
CIL (n = 24)	0.70 (0.18)	0.47 (0.21)

Notes: Average honesty rate. Standard deviation is in parenthesis.

Table G4: Tobit regression (Rate of honesty)

	Rate of honesty	
	b	se
Return rate	0.463**	0.199
Male	-0.117*	0.07
China	-0.075	0.076
Economics	-0.031	0.069
UK	0.005	0.101
Trust Game first	-0.159**	0.063
CIL	0.011	0.07
Undergraduate	0.06	0.082
Age	0.005	0.006
Risk choice	0	0.021
MACH Score	-0.007**	0.003
SDS17 Score	-0.006	0.011
Competence (2nd real-effort task)	-0.002	0.002
Constant	0.617***	0.167
Obs	80	
R^2	0.506	
Df	13	
Prob > F	0.003	

Notes: Tobit regression. The table reports the beta coefficients and the standard errors. 4 observations are left-censored, and 6 are right-censored. The psychological variables (i.e. SDS17 Score, MACH Score) are centered at the mean in order to control for high correlation between the independent variables (see Marquardt, 1980). 'Return rate' is equal to the amount returned by the trustee divided by 90. * p < 0.1, ** p < 0.05, *** p < 0.01.

We can also look at how the honesty rate correlates with the beliefs of the subjects that we elicited in the first part of the incentivized final

[157] Given the exploratory nature of this and later analysis, we only focus on p < 0.05 (or better) significance levels in the text.

questionnaire and how these beliefs correlated each other.[158] Table G5 shows the Spearman's rank correlation coefficients of these measures. The only two statistically significant correlations at conservative significance levels are a positive one between actual honesty rate and considering oneself more competent than the others; and a somewhat paradoxical one by which public officials who *perceive* themeselves as having been relatively more trustworthy in the earlier stage are less likely to be trustworthy in the official reporting game, perhaps as a result of feeling that they already gave their share of contribution to the others early on and/or that the others had not been trustworthy enough.[159]

Table G5: Spearman's ρ (beliefs and honesty rate)

	Honesty rate	All vote	Underreporting	Rank competence
Rate of honesty	1			
All vote[a]	0.04	1		
Underreporting[b]	-0.18	-0.17	1	
Rank competence[c]	0.36***	-0.19*	-0.11	1
Rank trustworthiness[d]	-0.20*	-0.15	0.34***	-0.20*

Notes: [a] "Do you think both the other two co-participants voted for you as the appointed co-participant?" YES = 1, NO = 0; [b] "Do you think that the co-participant, selected at random, who voted for you thought that you were going to underreport the value of the common fund?" YES = 1, NO = 0; [c] "How do you feel you were ranked among the three co-participants in terms of number of tables correctly solved in stage 1/2?" 1, 2, 3; [d] "How do you feel you were ranked among the three co-participants in terms of number of points returned to participant A in stage 1/2?" 1, 2, 3. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[158] We can only look at correlations as the answers to the questionnaire could have been affected by the decision to truthfully report or underreport the value of the common fund.

[159] As noted in the main chapter, there was actually a positive relationship between actual trustworthiness in the early stage and actual trustworthiness in the official dilemma stage (Spearman's ρ = 0.28, $p = 0.011$).

H. Additional analysis of behavior in stages 1 and 2

Table H1 reports the results of an OLS regression where the dependent variable is the measured competence in the 1st real-effort task stage, and the results of a Tobit regression where the dependent variable is the measure of trustworthiness, that is the return rate, in the 1st trust game. Independent variables include the socio-demographic and psychological characteristics of the subjects, and the treatment dummies.

Table H1: Regressions on Competence (1st real-effort task) and Trustworthiness

	Competence[a]		Trustworthiness[b]	
	b	se	b	se
Male	2.766	2.376	-0.102***	0.031
China	1.85	2.651	0.007	0.034
Economics	1.94	2.351	-0.053*	0.029
UK	4.104	3.118	0.034	0.041
Trust Game first	-4.074*	2.203	-0.011	0.027
Undergraduate	-0.19	2.63	-0.070**	0.035
Age	-0.26	0.277	0.004	0.004
Risk choice	1.204	0.744	-0.020**	0.009
MACH Score	-0.252**	0.112	-0.001	0.001
SDS17 Score	-0.741*	0.417	0.005	0.005
Experience with grid	3.535	2.373	.	.
Constant	66.807***	15.914	0.395***	0.107
Obs	240		240	
R^2	0.090		0.492[c]	
Df	11		10	
Prob > F	0.03		0	

Notes: [a] OLS regression with robust standard errors; [b] Tobit regression with 51 left-censored and 2 right-censored observations. [c] This is the McFadden's pseudo R^2. The table reports the beta coefficients and the standard errors. The psychological variables (i.e. SDS17 Score, MACH Score) and Age are centered at the mean. 'Experience with grid' is a dummy taking value 1 if a subject previously participated in a similar real-effort task. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We find that competence was smaller in the sessions where the trust game stage took place first, and for participants who scored high in the MACH questionnaire. We also find that male subjects, undergraduate students, and less risk averse subjects returned less in the trust game.

If we look at the proportion of subjects who trusted the counterpart (Table H2), we do not detect any statistically significant diffferences across the sessions where the real effort task was played first and those where the trust game stage took place first (χ^2 test, $p > 0.1$).

Table H2: Trust

	Real effort task stage first (n = 120)	**Trust game stage first (n = 120)**	**Total**
Trust	0.63 (0.48)	0.7 (0.46)	0.67 (0.47)

Notes: Standard deviations are in parenthesis.

This is also confirmed in a logit regression where the dependent variable is a dichotomous variable which takes value 1 if a subject trusts the counterpart and 0 if does not trust, and where we control for the socio-demographic and psychological characteristics and treatment dummies (Table H3). Male participants were less likely to trust the counterpart, whereas UK participants were more likely to trust.

Table H3: Logit Regression on Trust

	Trust	
	b	se
Male	-0.790**	0.318
China	0.423	0.373
Economics	-0.362	0.31
UK	0.884**	0.434
Trust Game first	0.419	0.302
Undergraduate	-0.522	0.389
Age	0.088*	0.046
Risk choice	-0.062	0.089
MACH Score	-0.026*	0.014
SDS17 Score	0.002	0.052
Constant	-0.948	1.345
Obs	240	
Pseudo R^2	0.095	
Df	10	
Prob > F	0.005	

Notes: Logit regression with robust standard errors. The table reports the beta coefficients and the standard errors. The psychological variables (i.e. SDS17 Score, MACH Score) and Age are centered at the mean. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.