ESSAYS ON THE IMPACT OF SOCIAL EMBEDDEDNESS ON SOCIAL PREFERENCES, BELIEFS AND PRO-SOCIAL BEHAVIOR

Inauguraldissertation zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität zu Köln

2020

vorgelegt von

Lisa Christin Lenz
aus
Albstadt Ebingen

Erstgutachter:	Prof. Dr. Dirk Sliwka
Zweitgutachter:	Prof. Dr. Christoph Engel
Vorsitzender:	Prof. Dr. Frederik Schwerter
Datum der Promotion:	16.12.2020

Meiner Mutter

D

DANKSAGUNG

Die vorliegende Dissertation untersucht die Auswirkung der Einbettung von Individuen in soziale Systeme auf ihre Entscheidungen. Ohne meine Einbettung in ein Netzwerk bestehend aus wunderbaren Kollegen, meiner Familie und den besten Freunden, die man sich wünschen kann, wäre die Vollendung dieser Dissertation niemals möglich geworden. An dieser Stelle möchte ich all jenen danken, die mich auf meinem Weg zur Fertigstellung dieser Dissertation ermöglicht haben.

Mein erster Dank für die Zeit der Unterstützung, Förderung und Betreuung meiner Arbeit, gepaart mit verständnisvoller Geduld und einem ausgeprägten Sinn für Fairness, gilt meinem Erstbetreuer Herrn Prof. Dr. Dirk Sliwka. Seine kritischen Anmerkungen und Ideen zu meiner Arbeit waren mir stets Inspiration.

Ebenso möchte ich mich bei meinem Zweitbetreuer Herrn Prof. Dr. Christoph Engel bedanken, der mit seinen kritischen Anmerkungen und Diskussionsbeiträgen bei zahlreichen Workshops und Seminaren, entscheidend zum Gelingen dieser Arbeit beigetragen hat. Er hatte stets ein offenes Ohr für Fragen, beantwortete noch nachts Emails und lies mich so an seinem großen Erfahrungsschatz teilhaben.

Ich möchte meinem Co-Autoren Sergio Mittlaender und meinen Kollegen danken, die mir stets mit Rat und Tat zur Seite standen. Insbesondere möchte ich mich bei Frédéric Gruninger, Carina Hausladen und Sven Hoeppner für das Korrekturlesen und ihre wertvollen Hinweise und Kommentare bedanken, die dieser Arbeit mehr Substanz verliehen haben. Bei meinen Kollegen Viola Ackfeld, Anna Hartmann, Kerstin Eilermann und Daria Tisch möchte ich mich für die vielen geistreichen Gespräche und aufmunternden Worte in zahlreichen Kaffeepausen bedanken.

Ich möchte meinen Freunden Danke sagen, die mich bei allen kleineren und größeren Herausforderungen des Lebens unterstützen und es so schafften, dass ich meinen Kopf immer wieder frei bekam und meine Doktorarbeit fertigstellen konnte. Ich möchte mich daher stellvertretend, bei Anna Neumann, Alena Trauschel, Sebastian Bur und Andre Lehmann für ihre echte Freundschaft bedanken. Mein besonderer Dank gilt an dieser Stelle meinem guten Freund Philipp Fernis, der mir, insbesondere in den letzten zwei Jahren, mit einer an ein Wunder grenzenden Geduld stets eine besondere Stütze war.

Schließlich gilt mein größter Dank meiner Familie, meinen verstorbenen Großeltern, die mich in meinem Studium stets finanziell unterstützt haben und den drei Frauen, die mein Leben mit Abstand am meisten geprägt haben: Meiner Mutter, meiner Tante und meiner Großmutter. Vielen Dank für Euer Vertrauen in mich und dafür, dass ihr mich in meinem Studium und während des Schreibens meiner Dissertation bedingungslos unterstützt habt. Vor allem die jahrelange Unterstützung meiner Mutter hat mir mein Studium und dadurch meine Dissertation erst ermöglicht. Meinen Dank hierfür lässt sich nicht in Worte fassen.

A

ABSTRACT

Economic agents are no atomistic Robinson Crusoes making rational choices entirely based on stable preferences constrained exclusively by prices, budgets and alike constraints. Instead, they have social relations with one another. These relations are –other than presumed by neoclassical economics– no frictional drags impeding competitive markets. Instead, preferences, expectations and choices of others strongly influence what economic agents prefer, expect and how they decide (Granovetter, 1985). In this dissertation, I therefore discuss in three essays applications how an individual's social embeddedness in groups and networks impacts preferences, beliefs and human behavior.

Pursuing this aim, I substantiate in my first essay ("Guilt in Multi-Agent Settings") how social environmental factors impact the perception of guilt and shame (c.f. Charness & Dufwenberg, 2006) and thereby lead to a decline of pro-social behavior in multi-agent settings, such as when groups tip less generously than individual customers. In particular, I address theoretically as well as experimentally if the desire to not betray others' expectations (guilt aversion), induce lower levels of pro-social behavior in settings with more than two agents. In doing so, I distinguish between four distinct behavioral channels: first, agents may weigh the loss inflicted on a single person less in multi-agent settings. Second, agents may experience less guilt if their decisions are not attributable to them. Third, an individual agent may free ride on the pro-social behavior of others. Fourth, interaction partners may lower their expectations regarding their pro-sociality in multi-agent settings and in response agents act more selfishly.

My second essay ("A Theory of Strategic Discrimination") discusses why agents strongly consider group composition preferences of others when including new members in groups or social networks in the absence of an own taste or statistical reasons to select either candidate –such as when multi-family home proprietors discriminate against black tenants in response to prejudiced white tenants. Three potential behavioral channels explaining alike phenomena are theoretically discussed and empirically assessed: First, agents may have altruistic feelings towards their present group members and enhances their utility by living up to their group composition preferences. Second, agents anticipate that other members rest their cooperativeness upon who has been selected and adapt their inclusion decision accordingly. Third, individuals seek to trigger reciprocal behavior and positive affections towards them by signaling that they care for the preferences of present group members.

Inter-group contact has been found to increase and to decrease discrimination in field experimental studies. These conflicting results might originate from differences in addressed types of discrimination –i.e., whether discriminatory behavior arises from differences in tastes or beliefs– and from differences in contact's capacity to alter tastes and beliefs. Therefore, my third essay ("The Impact of Inter-group Contact on Economic Types of Discrimination") investigates the causal effect of inter-group contact on statistical and tasted-based discrimination as well as associated anticipation effects of the latter leading to a decrease in inter-group trust. Thereby, it studies whether the teams or network one is embedded in has an impact on ones' preferences, and on how one perceives the pro-sociality and skills of in-group and out-group members. Lessons for policymakers concerned with the reduction of discrimination involve the features that inclusive policies should strive for by changing preferences or beliefs, and thereby reducing discrimination.

C

CONTENTS

T

TABLES

F

Figures

1

INTRODUCTION

In this dissertation, I elucidate in three essays how the embeddedness of an individual in distinct environments of inter-personal relations (with whom one interacts and in what kind of social setting) shapes preferences (what one likes), beliefs (what one expects) and human behavior (what one does) –a major impact factor determining human decisions not considered by more traditional, neoclassical Economics.

Ever since the marginal revolution in Economics in the 1870s, traditional micro-economic models are based on the assumption that patterns of human behavior can be explained by atomistic Robinson Crusoes making rational choices based on stable preferences constrained by prices and one's budget. Thereby, changes in economic agents' behavior are in the standard Economic framework exclusively explained by variation in the mentioned constraints, whereas crucial questions on the emergence and determining factors of preferences are abandoned to other social scientific disciplines (Becker & Stigler, 1977). Those neoclassical models –relying upon methodological individualism (see Arrow, 1994; Menger, 1871; Schumpeter, 1908; Udehn, 2002)– do usually not consider social contexts as a behavioral constraint nor as a factor that influences preferences or beliefs (Arrow, 1994). If they do, they consider them as frictional drags impeding competitive markets (Granovetter 1985).

In light of this radical simplification of human nature, representatives of neighboring social sciences as well as (behavioral) economists raised serious concerns. The neglection of economic agents' social embeddedness comes at the price of underestimating the importance of how social structures determine information search (Granovetter, 1973, 1985), and the impact of social preferences (e.g., Andreoni, 1990; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999; Rabin, 1993; Levin, 1998) social status (Benjamin et al., 2012; Heffetz & Frank, 2011), social norms (Akerlof & Kranton, 2000; Krupka & Weber, 2013), as well as group dynamics on decision making and its underlying processes (Granovetter, 1985).

Ignoring these factors, standard Economics cannot explain phenomena such as why individuals live up to the expectations of others in the absence of any direct of strategic benefit (Charness & Dufwenberg, 2006), why discrimination at the workplace is regularly more persistent than predicted by standard economic models (c.f. Becker, 1957) as well as why, as discussed in chapter two, economic agents are more reluctant to betray individuals in comparison to groups. This dissertation accounts for this major critique by addressing the questions *how the embeddedness in distinct structures of interpersonal relations influence pro-social preferences, beliefs, as well as pro-social behavior in distinct decision-making environments*.

Pursuing this aim, I contribute in chapter 2 to the literature on social embeddedness by discussing how social environmental features impact the decline of pro-social behavior in multi-agent settings. In particular, I inquire into whether individuals perceive guilt and shame (Battigalli & Dufwenberg, 2007; Baumeister et al., 1994) differently in multi-agents settings due to variations in the transparency of actions, the ability to free-ride on the perceived moral behavior of others, as well as group discounting and anticipation effects that may vary between one-to-one, small group as well as large group settings.

Chapter 3 discusses why individuals consider others' group composition preferences selecting new group members in the absence of an own taste or statistical reasons to select either candidate, such as when a real estate owner discriminates against black clients in response to prejudiced white clients (Ondrich et al., 1999; Zhao et al., 2006). Thereby, this essay inquiries into how being socially embedded can leverage discriminatory tendencies, because individuals who would in the absence on effects on third parties refrain from any kind of unequal treatment, may discriminate once being a member of a prejudiced group. Moreover, it illustrates that with whom economic agents interact in markets and organizations is not only a question of a cost-benefit analysis but is affected by social dynamics and the composition of one's social network.

Chapter 4 examines the question which economic type of discrimination considered in the economic literature (statistical discrimination, taste-based discrimination and discriminatory tendencies arising from its anticipation) is to what extent reduced by inter-group contact, defined as short term inter-group interactions. Thereby, it analyzes the capacity of contact to alter beliefs and (social) preferences. Other than predicted by standard economic theory, the empirical analysis of our experiment reveals that a mitigation in discriminatory behavior is rather caused by a shift in relative social preferences towards out-group members arising from inter-group interactions than by alterations in beliefs. Overall, chapter reconciles insights from one of the most seminal theories of inter-group relations –the contact hypothesis (Allport, 1954)– as well as standard economic explanatory approaches for in-group favoritism and discrimination (e.g., Becker, 1957; Phelps, 1972).

The theoretical contributions and empirical findings presented in this dissertation fit into a series of distinct economic approaches which allow to analyze social interactions and systems of inter-personal relations: traditional game theory deals with strategic interaction between individuals in a rather formalistic manner. Identity economics (Akerlof & Kranton, 2000, 2005) take social categorization, identification and groups norms into consideration. Behavioral economists have studied the role of peer effects (e.g., Bursztyn et al., 2014; Card, 2013; Falk & Ichino, 2006; Mas & Moretti, 2009) or investigated social norm compliance as a rationale describing anomalous behavior (Krupka & Weber, 2013). In general, behavioral economists have developed a wide range of outcome-dependent, belief-dependent, type-based and intention-based social preference models which capture the idea that individuals have non-selfish preferences.

Assessing the different approaches, the sociologist Marc Granovetter (1985) considered traditional game theoretical models –which rest upon stable preferences that are unaffected by how a decision maker is embedded in networks of social relations– as under-socialized (Granovetter, 1985, 1992). On the other hand, models alike the identity economic approach (Akerlof & Kranton, 2000, 2005) fall into the category of over-socialized models. Those models can be criticized for assuming that individuals internalize social relation component and adhere slavishly to them (Granovetter 1985). In conclusion, under- as well as the over-socialized models are in their nature mechanic and thus ignore ongoing social processes and dynamics as well as changes in the social structure in which an individual is affect determines his or her behavior.

Both model classes lack analyses regarding how the embeddedness of individuals in a network of social relations affects their behavior. Social preference models offer –besides networks studies (e.g., Borgatti et al., 2009; Burt, 1995; Burt et al., 2004)– a way out of this dilemma. Introducing a formal framework that allows to study interpersonal relations, such as reciprocity (Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006), gift exchange behavior (Akerlof, 1982; Falk, 2007), guilt (Charness & Dufwenberg 2006, Batigalli & Dufwenberg 2007) or outcome-based social preferences (Andreoni, 1990; Fehr & Schmidt, 1999; Ockenfels et al., 2000), those models are capable to analyze how social relations affect economic decision making. Having stated this, the majority of papers on social preferences are nevertheless limited to one-to-one interactions or if they rest upon group settings, often treat group processes and dynamics as black boxes.

In contrast, this dissertation's primary focus is on the examination how changes in group structures and social relations impact social preferences (such as the perception of guilt), beliefs (e.g., about the pro-social behavior of others) and associated shifts in strategic incentives to act pro-socially. In this respect, the presented dissertation is part of a second wave of behavioral studies assessing decision making in groups which focuses on changes in the decision-making environment rather than just comparing outcomes of decisions made by either atomistic individuals or groups (see e.g., Ambrus, Mobius, & Szeidl, 2014; Gillet, Schram, & Sonnemans, 2009; He & Villeval, 2017; Zhang & Casari, 2012). Besides its theoretical contribution and practical implications, this dissertation reports results from one laboratory experiments as well as two online experiments examining the impact of embeddedness on economic decision making.

In the remainder of this chapter, I outline the three main chapters of this dissertation. Each chapter is self-contained and can be read independently. Each chapter's appendix follows after the chapter's main text. The list of the references for all three chapters is provided jointly at the end of this dissertation.

Chapter 2: Guilt in Multi-Player Games– Chapter two is a single author project. In this chapter, I elucidate how group characteristics–namely the group size, the behavior of other decision makers, how much a decision is attributable and pivotal to an individual decision maker, and potential anticipation effects–defines the perception of the moral feelings of guilt and shame. More technically, I address the question how belief-dependent other-regarding preferences (the acknowledgment of what others expect) foster selfishness in multi-agent dilemmas with more than two economic agents. Thereby, I reconcile the theory of guilt and the literature on pro-social behavior in multi-player games.

Concerning the latter, there is ample evidence that individuals regularly act less morally inclined in environments with more than two agents: the presence of bystanders reduces the likelihood that an individual intervenes in critical situations (Fischer, 2011). Isaac (1988) and Carpenter (2007) have established a negative group-size effect on cooperation in teams. Furthermore, credence good sellers provide more overpriced and unnecessary services to inflate the amount of money charged if customers' expenses are covered by an insurance, i.e., by a community of policyholders (Balafoutas 2016).

Regarding the former, embedded individuals are in general not exclusively selfish or altruistic but ask themselves "what do others expect me to do" and "what do others think of me". Put differently, they either have an intrinsic motivation to live up to the expectations of others, extrinsic image concerns or both. These kind of expectation-based social reference-point dependent preferences are captured by distinct concepts of guilt (Batigalli & Dufwenberg 2007). While predictions derived from theoretical guilt models have been widely discussed and tested in two-agent (one-to-one) dilemma settings (Abeler et al., 2019; Balafoutas & Fornwagner, 2017; Bellemare et al., 2011; Charness & Dufwenberg, 2006; Hauge, 2016; Reuben et al., 2009), this is the first study that investigates systematically and experimentally if economic agents will experience guilt in multi-agent dilemmas less severely and if so why.

In this chapter I introduce a multi-agent simple guilt and guilt from blame model which offer together four distinct behavioral explanations of why decision makers behave less pro-social in multi-agent dilemmas: first, an individual may weigh the loss inflicted on a single person less in multi-agent settings (group discounting effect). Second, deviations from other's expectations are associated with less dis-utility if individual decisions are less attributable to an individual decision maker (attributability or transparency effect). Third, individual may free ride on the pro-social behavior of others (balancing effect). Fourth, individuals might experience less guilt in multi-agent settings if they anticipate the former three effects (anticipation effect). To test the hypotheses derived from the introduced guilt models, I let subjects play a novel versions of the dictator game (see Engel, 2011 for a comprehensive review). Participants are assigned distinct settings with groups of dictators and groups of recipients of varying sizes. In the multi-agent treatment, dictators jointly determine the height of the recipients' payoff, while dictators' payoffs are solely set by themselves.

My experimental data reveal that decision makers will be more selfish if they are protected by the anonymity of the mass and the consequences of their decisions affect a collective of agents compared to a standard dictator game. This effect cannot be traced to preferences over distributions. Instead, it is caused by a joint determination of recipients' payoffs. The decline of pro-sociality is particularly in small group settings mostly affected by the lack of transparency and hence the attributability of decision to individual decision makers. Moreover, I find no support for an anticipation effect as well as the balancing effect. As a matter of fact, the data exhibit a significant positive correlation between the expected pro-sociality of other dictators and the amount allocated to recipients by an individual dictator.

While the theoretical analysis demonstrates how to utilize guilt models to explain behavior in multi-agent dilemmas, the econometric analysis facilitates the determination of distinct explanations' relative importance. The evidence on the positive impact of transparency on dictators' pro-sociality contributes to the discussion whether guilt from blame or simple guilt more precisely describes behavior (see also Batigalli & Dufwenberg 2007). The conclusion that the guilt from blame model is better suited to predict moral behavior than the simple guilt model implies that some results of previous guilt studies are not readily transferable to a wide range of real-life scenarios in which actions are opaque. Lastly, the empirical results imply that policy makers should enhance the transparency of actions and utilize the positive impact of role models in order to remedy the decline of pro-social behavior in multi-agent settings.

Chapter 3: The Theory of Strategic Discrimination– This chapter is joint work with Sergio Mittlaender and Paulo Avarte. I developed the theoretical framework presented in that paper, wrote he oTree code, developed the experimental design, conducted the experiment, analyzed the experimental data, and write the paper. Sergio Mittlaender critically supervised the experimental implementation and the statistical analysis, provided critical comments on the paper and provided funding Paulo Avarte contributed financially to the project and critically reviewed the final version of this chapter.

Chapter three discusses why and to what extent individuals live up to group composition preferences of others when deciding whom to include or exclude in a group or social network. An application in which this question is relevant are humans who find themselves wondering whether they want to invite a good friend to a party knowing that other party guests dislike this friend because of her socio-economic or ethnical background. A manager who appreciates the skill set of female candidates may refrain from hiring women, if the present workforce is heavily prejudiced against female colleagues. Finally, landlords may discriminate against homosexual clients in response to homophobic present and future clients. This essay discusses alike phenomena and its underlying behavioral channels. In particular, this is the first essay that addresses the research question *why individuals selecting new group members consider group composition preferences as well as present group members' behavioral responses in case their preferences are met or not met.*

Thereby, we investigate to what extent selectors live up to group composition preferences of their current group members, because they either have a concern for group members' well-being or, alternatively, they are motivated by the prospect of higher cooperation in case they live up to others' group composition preferences. Second, we inquire into whether third parties condition their behavior on whether the new member was deliberately included by a human or randomly selected, and, lastly, whether selectors anticipate as well as account for such procedural preferences.

The ethical assessment of the acknowledgment effect is ambiguous: while cooperation arising from the satisfaction of present team members' group composition preferences foster the success of employee referral programs, it may reinforce structural discrimination if present group members taste arises from animus towards and prejudice against minorities (discrimination spill-over effects); such as when landlords discriminate against black clients in response to prejudiced present and future white clients (Ondrich et al., 1999; Zhao et al., 2006). When managers recruit new members from their present employees' homogenous networks (e.g., including foremost white, at least middle-class males), minorities have less opportunities for social advancement. This self-re-reproduction (Luhmann, 1986) of (elite) groups may contribute to the manifestation of structural discrimination against minorities.

Our novel formal model, which rests upon the concept of reciprocal altruism (Levine, 1998) and accounts for group formation preferences of others, comprises and formally discuss three potential behavioral channels explaining the acknowledgment effect: first, individuals have altruistic feelings towards present group members and enhances present members' utility by selecting their preferred candidates. Second, they anticipate that others rest their cooperativeness upon who has been selected and adapt their selection accordingly. Third, they want to trigger reciprocal behavior by signaling that they have an interest in the well-being of others.

We test hypotheses derived from our model theoretical discussion in a stylized endogenous team formation context in which one member of a pre-existing team without any personal taste and any statistical reason (Aigner & Cain, 1977; Arrow, 1973; Phelps, 1972) to select either candidate, has the opportunity to select an additional member out of a pool in exchange for a predetermined fee. By design, the selector neither has a preference for either of the two candidates, while the other team member strongly prefers a particular candidate. The team's success depends on the contributions of its members to a public good.

Eliciting the selectors' choices and their willingness to pay to make a selection decision allows to measure whether they consider present team members' group composition preferences. Moreover, we inquire how their inclusion decision alters current team members' willingness to contribute to public good and their attitude towards the selector.

The empirical results reveal that about 60% of selectors account across all treatments for present team members' group composition preferences in the absence of any taste for or statistical reason to select either candidate. The effect persists even if present members can vary their contribution to the public good but is more pronounced if strategic incentives are present. The social-preference effect was approximately twice as large as the strategic incentive effect. In conclusion, selectors' interest in the wellbeing of present group members as well as strategic considerations together trigger the acknowledgment effect. While our empirical analysis provides first insights that endogenous selection procedures might influence team dynamics, the evidence to what extent decision environment's features, such as the transparency of selection decision, may leverage the established effects is, however, inconclusive.

Chapter 4: The Effects of Inclusive Policies and Contact on Economic Types of Discrimination – Chapter four is joint work with Sergio Mittlaender. I contributed to this chapter by providing funding, writing the oTree code, co-developing the experimental design, conducting the experiment, analyzing its outcomes, and writing the paper. Sergio Mittlaender contributed by co-developing the experimental design, critically supervising its implementation and statistical analysis, as well as providing valuable comments on various versions of this chapter.

Chapter four complements chapter three– which showed that social embeddedness may amplify structural discrimination– by discussing to what extent and under which conditions inter-group contact defined as short-term inter-group interactions impacts inter-group relations and, thereby, extenuate discriminatory tendencies and mitigate in-group favoritism. If discriminatory tendencies are reinforced by decision makers living up to the group composition preferences arising from the prejudiced tastes of others, as discussed in chapter three, this phenomenon might be mitigating by addressing the primary source of discrimination by positive inter-group interactions.

Inter-group contact (Allport, 1954, 1958) is proposed as a rationale for school desegregation policies, peacebuilding interventions (e.g., Kelman, 1998; Maoz, 2010) and the reduction of discriminatory tendencies in the society in general. Understanding its mechanisms recently became highly important, given the current wave of protest against ethnic discrimination in the U.S., migratory waves to developed countries, and the diversity of policies and proposals for the integration of immigrants in diverse societies in Europe.

However, the vast majority of previous (economic) studies did not investigate how, and to what extent, contact affect different types of economic discrimination. As a matter of fact, the contact hypothesis itself is agnostic about its underlying processes (Pettigrew, 1998). We therefore analyze in chapter four to what extent contact, implemented as short-term inter-group interactions, mitigates distinct types of discrimination considered in economics –namely taste-based (Becker, 1957), anticipated taste-based and statistical discrimination (Aigner & Cain, 1977; Arrow, 1973; Lundberg & Startz, 1983; Phelps, 1972a). Thereby, we investigate whether a mitigation in discrimination arises from alteration in social preferences or beliefs. Overall, we contribute to the discussion how optimal policy should differ contingent on whether discrimination is based on statistical, taste, or hybrid causes.

To empirically assess the impact of contact on the three economic types of discrimination, we designed a lab-in-the-field experiment in which supporters of two distinct political groups –U.S. citizens either supporting the Democratic or the Republics party– interacted. In the first stage of the experiment, participants were assigned to teams of four and had to solve a common task, namely assigning artists to their corresponding paintings. To study the effect of inclusive social policies, and some of their distinctive features, we exogenously varied across treatments whether heterogenous or homogenous teams were formed, as well as whether team members had the opportunity to communicate. To measure the extent of taste-based, anticipated tasted based, and statistical discrimination participants thereafter played an other-other allocation game similar to Chen and Li (2009), a trust game (Berg, Dickhaut, & McCabe 1995), and a novel real effort game in which they were given the opportunity to bet on the productivity of others.

Our results reveal a strong and significant attenuation effect of inter-group interactions on taste-based discrimination and confirm that changes in social preferences better predict discrimination patterns between distinct political groups than cognitive components. We show that the attenuation effect on discrimination are primarily caused by alterations in social preferences and not alterations in beliefs about in-group or out-group members. Concerning discrimination arising from the anticipation of taste-based discrimination of others, we find no significant attenuation effect. Surprisingly, contact enhanced statistical discrimination likely due to the fact that while inter-group interactions can be experienced as positive on the personal level and thereby mitigate the taste for discrimination, they can be perceived as negative on the factual level and thereby deteriorate the perception of competence.

Our empirical insights and theoretical discussion allow us to reconcile the economic and the psychological perspective on discrimination. Thereby, we derive practical implications that enable policy maker to derive more purposeful policies: inclusive policies potentially reduce taste-based discrimination and alter individuals' preferences, if they go beyond simply putting participants in contact. Therefore, it is not sufficient to include minority students, or coworkers in a group where discrimination is rampant. Instead, inclusive social policies should create an environment in which distinct social groups of equal status actively communicate.

2

GUILT IN MULTI-PLAYER GAMES[*]

Guilt aversion defined as the desire to not betray others' expectations, can substantially induce pro-social behavior. However, this is the first study that investigates systematically whether individuals will experience guilt (Charness & Dufwenberg, 2006) *in settings with more than two agents less severely and if so why. Doing so, I distinguish between four different behavioral explanations and their relative importance: first, individuals may weigh the loss inflicted on a single person less in multi-agent settings (group discounting effect). Second, deviations from other's expectations are associated with less dis-utility if individual decisions are less attributable to particular agents (transparency effect). Third, individual may free ride on the pro-social behavior of others (balancing effect). Fourth, individuals might experience less guilt in multi-agent settings if they anticipate the former three effects (anticipation effect). Overall, I find a significant decline in pro-social behavior in multi-agent settings. Determining the relative importance of different behavioral channels, I find that the relaxation of the attributability of actions is the most important channel–accounting for around 60% of the difference in pro-sociality in small group settings.*

1 Introduction

Neoclassical economic theory assumes that humans act solely in their own self-interest without paying attention to moral norms and social reference points (what others have and expect). Nonetheless, moral behavior is regularly encountered in (economic) life outside of textbooks. Individuals tend to keep their promises even if reneging on said promises would prove beneficial for them (Abeler et al., 2019; Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006, 2011), and if bureaucrats experience trust and are perceived to be persons of integrity, they are less prone to corruption (Balafoutas, 2011). Such widely observed patterns of moral behavior can be explained by concerns over others' expectations: when making decisions, individuals act not exclusively selfishly or altruistically but ask themselves "*what do others expect me to do*" and "*what do others think of me*". These kind of social reference-point dependent preferences are captured and defined by distinct concepts of guilt (Battigalli & Dufwenberg, 2007) which incorporate a social reference point that equals the perceived expectations of others regarding one's own behavior (decision maker's second order beliefs). While predictions derived from guilt models have been tested in two-agent (one-to-one) settings (e.g., Balafoutas, Kerschbamer, & Sutter, 2017; Bellemare et al., 2011; Charness & Dufwenberg, 2006; Hauge, 2016; Morell, 2017; Reuben et al., 2009), this is the first study *that investigates systematically if economic agents will experience guilt in multi-agent dilemmas less severely and if so why.*

Evidence from the lab and the field indicate that economic agents act less morally inclined in a wide range of multi-agent environments. The literature on public good provisions has established a group-size effect: the larger the number of potential contributors to a public good, the smaller is the average contribution (Carpenter, 2007; Isaac & Walker, 1988). The presence of bystanders reduces the likelihood that an

[*] I thank Dirk Sliwka, Christoph Engel, Matthias Sutter, Oliver Kirchkamp, Sven Hoeppner, Frédéric Gruninger and Carina Hausladen for helpful discussions and comments on earlier versions of the paper. Financial support by the Deutsche Forschungsgemeinschaft (DFG), grant FOR 1371, is gratefully acknowledged.

individual intervenes in critical situations (Fischer et al., 2011), i.e., the more bystanders observe an accident, the less likely one becomes a first aider. The average amount of restaurant tips decreases in the number of customers (Conlin et al., 2003). Lastly, sellers of credence goods (for instance doctors, car mechanics, lawyers and alike experts) provide more often unnecessary and overly expensive services to inflate the amount of money charged if customers' expenses are covered by an insurance, i.e., by the community of policyholders (Balafoutas et al., 2017). In the latter case, customers can often not identify whose fraudulent behavior is pivotal for an increase in insurance payments.

While some of these phenomena can be (partly) rationalized by game-theoretical or informational economic approaches,[1] the aim of this paper is to show how belief-dependent other-regarding preferences foster selfish decision making in alike multi-agent dilemmas. Therefore, I develop a multi-agent simple guilt and a guilt from blame model which offer together four behavioral explanations of why agents behave differently in multi-agent dilemmas: First, the perception of guilt might be determined by how agents weigh the harm or loss they inflict to a group compared to a single person[2] (group discounting effect). Second, to what extent individual decisions are pivotal and attributable to a particular agent may determine pro-social behavior (transparency effect). Third, agents may free ride on the moral behavior of others, i.e., if multiple agents jointly determine an outcome, and an individual agents believes that others act pro-socially, she may herself refrain from acting pro-socially (balancing effects).[3] Fourth, agents might be less likely to experience guilt in multi-agent dilemmas if affected people anticipate the previously described effects and thus expect less pro-sociality (anticipation effect). The anticipation effect reinforces the other effects and is likely to be more prevalent in the presence of learning.

To assess empirically guilt in multi-agent settings, this study introduces a novel version of the dictator game (see Engel, 2011 for a comprehensive review). Its treatments vary in the group size and the transparency of an individual decision. The number of recipients equals the number of dictators in all settings. The baseline treatment resembles a dictator game with a take framework and an endowment of $1.00. By design, the decision of a single dictator is observable by the recipient in a one-to-one dictator game. In the novel multi-agent treatment, four or respectively 20 dictators had to decide how much money they want to take out of a common pool that contained $4.00 or respectively $20.00. Dictators could take an amount up to $1.00. The amount that remained in the pool was divided equally among the recipients. This implies that in the multi-agent treatments dictators jointly determine the recipients' payoff, while the dictator's payoff is solely set by the dictator herself. To investigate to what extent the presumed behavioral channels drive the tendency to act more selfishly in multi-agent treatments, I exogenously vary dictators' second order beliefs utilizing the strategy method (Selten, 1965). In particular, I let dictators state decisions contingent on eleven possible revealed beliefs. Moreover, I varied the ex-post experimental feedback (whether participants received any information about dictators' decisions or not) and elicited dictators' first order beliefs about other dictators' behavior in one's group. Feedback variations allow not only to rationalize moral decline in multi-agent dilemmas, but additionally provides a theoretical foundation as to why changes in

[1] The bystander effect can be game-theoretically described by the volunteer's dilemma (Diekmann, 1985). However, in contrast to the theory introduced in this study, the volunteer's dilemma offers no explanation why the attributability of actions and the familiarity with bystanders mitigates the bystander effect introduced by (Latané & Darley, 1970). Holmström (1982) described how moral hazards in teams can explain the decline of effort contribution in public good scenarios. In contrast, in this paper I will show that even in the absence of changes in the incentive structure of the game decision makers act less pro-socially in multi-agent dilemmas. Sülzle and Wambach (2002) argue theoretically that customers in credence good markets are more likely to screen the market for better offers if they do not receive reimbursements, because they profit financially from difference between offers. In contrast, insured customers do hardly ever profit from finding a financially more attractive offer. In addition, sellers of credence goods anticipate the incentives in insurance markets to reduce the search effort and thus rise their prices.

[2] The first explanation approach implies that the group-size effect is driven by an increase in the number of recipients.

[3] The second and the third explanation approach indicate that the group-size effect is driven by an increase in the number of dictators as well as the associated decline in the attributability of actions.

the decision environment, like variations in the transparency of actions or attributability cues (e.g., name tags), matter.

The experimental data indicate that agents decide more selfishly if they are protected by the anonymity of the mass and the consequences of their decisions affect a collective of agents. This effect cannot be traced to mere changes in the number of subjects in one session and therefore to preferences over distributions. Instead, it is caused by a joint determination of recipients' payoffs by dictators in the multi-agent dictator game. The decline of pro-sociality is mostly driven by the lack of transparency, i.e., in multi-agent settings individuals experience less dis-utility from selfish actions if an outcome is determined by multiple decision makers in small group settings than in one-to-one settings. More precisely, a revelation of the distribution of dictators' actions can explain about 56% of the difference in recipients' payoffs between the baseline and the small group multi-agent treatment. About 37% of the difference between the one-to-one and the four dictators setting can be explained by individuals discounting the dis-utility of a single recipient differently if they interact with multiple recipients and potentially undetected behavioral channels. I find a positive but not a significant anticipation effect which makes up for 7% of the overall difference between the one-to-one and the four dictators multi-agent setting. Contrary to the presumed balancing effect, decision makers tend to conform to the behavior of other dictators within their group.

My experimental results have versatile theoretical and practical implications. First, they establish a base for improvements in formal guilt models: while the theoretical analysis demonstrates how non-linear guilt models can be used to explain behavior in multi-agent dilemmas, the econometric analysis allows me to determine the relative importance of the introduced behavioral channels.

Second, empirical evidence on the impact of transparency on dictators' decisions contribute to the discussion whether guilt from blame or simple guilt model better describes behavior (see also Battigalli & Dufwenberg, 2007). The concept of "simple guilt" covers the psychological concept of guilt, while "guilt from blame" refers to the concept of shame.[4] Shame, in everyday language is often used interchangeably with guilt (Dearing & Tangney, 2004, Chapter 2). These two terms are—in fact—closely related as psychological concepts but differ slightly in their key aspects. If betraying others is not publicly exposed, agents likely experience guilt. In contrast, if betraying others is publicly exposed, they will likely experience shame.[5] The detected transparency effect supports the hypothesis that the concept of guilt from blame more precisely describes human behavior. To the best of my knowledge, only one other study by Bracht and Regner (2013) discriminated between simple guilt and guilt from blame.[6] If economic agents make their decisions contingent on other people's expectations when said people can infer whether agents betray, meet or exceed their expectations, some results of previous studies are not readily transferable to a wide range of real-life scenarios in which actions are opaque.

Moreover, the question whether either guilt or shame prevails in moral dilemma situations is not only theoretically important, but highly relevant in practice as well: Teams with low levels of cooperation that comprises numerous shirkers (Holmström, 1982) may want to enhance the transparency of individual contribution to the team project by facilitating social control. In insurance settings, in which a significant share

[4] Henceforth, when I will use the term shame, I will refer to the psychological feeling of shame and when I talk about "guilt from blame" I will refer to its game-theoretical representation.

[5] Moreover, while with guilt people focus on their behavior and misconducts, with shame their focus is more on self-conception. This is known as the self-behavior-distinction (Cohen et al., 2011). For the sake of simplicity, in this study I rely contrary on the public-private distinction, since it can easily be modeled in a game-theoretical and tested in an experimental context.

[6] Khalmetski, Ockenfels, and Werner (2015) claim that they test to what extent pro-social behavior can be explained by guilt from blame or simple guilt. However, they do not test whether co-players can infer if the players are responsible for their monetary outcome or not, but they vary whether co-players are aware of the fact that their beliefs are revealed instead. Hence, they test whether the visibility of intentions, but not whether the visibility of actions matters. They find that players are more generous when their intentions are observable and that co-players do not strategically exploit this tendency.

of experts, such as doctors or car mechanics, defraud the collective of insureds, experts should be obliged to reveal provided services and associated costs not only to insurers, but also to clients, to reinforce feelings of shame and thereby alleviate fraudulent behavior.

Third, disentangling the different potential explanations for the moral decline in multi-agent can be utilized to elucidate why sellers of credence goods, such as medical or repair services, will more likely commit to overcharging or over-provision if the expenses of the service are covered by the community of insured customers. Sutter et al. (2013) as well as Balafoutas et al. (2016) conjecture from field experiments that—even in the absence of any adverse selection and demand side induced effects—sellers of credence goods will more likely inflate bills if they are aware that the costs are covered by a third party, either an employer or an insurance company. My experimental design allows to test in an abstract setting –mimicking the payoff of different market settings– whether differences in credence good provision persist in the absence of informational asymmetries, because providers perceive their own guilt differently in insurance settings. Thereby, I assess the impact of guilt and shame in isolation, since the experiment allows me to abstract from competing behavioral channels.[7]

Fourth, the attributability or transparency effect as well as the concept of diffusion of responsibility[8] (Darley & Latané, 1968) presume a relationship between the likelihood of being held accountable and anti-social behavior. However, the concept of diffusion of responsibility describes a behavioral effect on a phenomenon level while being agnostic about its causes (Diekmann, 1985). Although a wide range of studies has investigated various potential motives underlying the bystander effect induced by the diffusion of responsibility (see among others the meta study by Fischer et al., 2011), the theory lacks a coherent preference-based framework that conceptualizes the presumed sources of the decreased willingness to take responsibility, which the extended version of the guilt model provides.

The remainder of this paper is structured as follows. In the next section, I introduce slightly modified versions of the simple guilt and the guilt from blame models by Battigalli and Dufwenberg (2007) and derive implications about how dictators behave under two different ex-ante information schemes. Section 3 elaborates on the experimental design. Section 4 presents and discusses the empirical results. Section 5 concludes the findings of this chapter and discusses implications.

[7] Sutter et al. (2013) as well as Balafoutas et al. (2016) have conjectured that alternatively informational asymmetries between sellers, demander and insurance companies, a differing social distance between the collective of insureds and one single customer, a common history between a supplier and a customer, preferences concerning how much a demand values a good signaled by his insurance status as well as distributional preferences. These factors are kept constant between all of my experimental treatments.

[8] Diffusion of responsibility is defined as a socio-psychological phenomenon whereby an individual is in the presence of bystanders or other potential decision makers less likely to take responsibility for an action.

2 Models and Hypotheses

2.1 Simple Guilt Model

In this section, I elucidate how the proneness to experience guilt and blame theoretically determines decision in one-to-one and in multi-agent settings. In particular, I develop two different parsimonious guilt models and discuss them in a general and a dictator game context to derive testable hypotheses.[9]

In a first step, I introduce a general multi-agent guilt model. According to Battigalli and Dufwenberg (2007), an individual i (she) may not only gain utility from increasing her own endowment but will suffer if she does not fulfill the expectation of other individual(s) $j(s)$ and thus by letting them down. For instance, i may let j (he) down by making unexpected selfish decisions in dictator games, by giving no tips in a situation where it is common to do so or by selling (credence) goods, such as repair services, contrary to honest practices for an overcharged price.

The utility of an agent who experiences guilt is in the simple guilt model is given by the following utility function which persists of a monetary utility term $m(t_i)$ and a psychological utility term $P\ (\beta)$ that accounts for the expectations β of people affected by the decision:

$$U(t_i, \beta) \;=\; m(t_i) \;+\; P\,(\beta) \tag{1}$$

The key element of the simple guilt model is that making a decision *d* an individual i is not only interested in her own payoff t_i but is reluctant to betray the expectations of other js about their payoffs t_j as well. The concave monetary utility function is defined as follows:

$$m\big(t_i(d)\big) \text{ with } \frac{d\,m\big(t_i(d)\big)}{dt_i(d)} > 0 \text{ and } \frac{d\,m\big(t_i(d)\big)}{d\,t_i(d)^2} \leq 0 \textit{ e.g. } m\big(t_i(d)\big) = ln\big(t_i(d)\big) \tag{2}$$

The psychological utility function $P(\beta)$ acknowledges for negative deviations from the reference point (other agents' perceived beliefs) and is based on a guilt function $G\big(t_j(d),\, \beta\big)$, measuring whether and to what degree j is disappointed about his payoff contingent on i's decision *d*. For instance, if i shirks in a team project and does not meet the expectations of j, i will disappoint her team member and in response experiences the feeling of guilt $\big(G\big(t_j(d),\, \beta\big) < 0\big)$. The psychological utility depends on a correction term $C(\beta)$ that accounts for the effect that i does not feel responsible for the negative effects on others that she cannot prevent. That is, i only feels guilty towards j if the harm inflicted to j was self-inflicted and not caused by immutable circumstances. The psychological utility function is thus given by:

$$P \;=\; P(G((t_j(d),\, \beta),\, C(\beta)) \tag{3}$$

[9] Battigalli and Dufwenberg (2007) formalize the previously described preferences by proposing two distinct utility models with a linear income and a linear guilt term. The linear models entail having a good conscience and an increase in income are perfect substitutes. In contrast, I derive predictions from utility functions that are strictly concave in their money dimension as well as concave in their psychological dimension. For an extensive axiomatic discussion about the advantages of a logarithmic over a linear guilt model in various game theoretical settings I refer to Jensen and Kozlovskaya (2016). In short, concave functions allow for inner solutions and thus are –in contrast to linear models– apt to explain why large fraction of dictators gives less than expected in experimental studies (Ellingsen et al., 2010).

The guilt function $G\big(t_j(d),\, \beta\big)$ is based on the idea that j forms first order beliefs about his payoff t_j ($E_j\big[t_j\big] = \alpha$) and that i forms second order beliefs about j's first order beliefs ($E_i\left[E_j\big[t_j\big]\right] = E_i[\alpha] = \beta$). Subsequently, i evaluates her behavior and the consequences of her behavior with regard to her second order beliefs β. For instance, recipients expect a certain payoff in an allocation game. Hence, allocating players ask themselves what recipients expect and in response evaluate potential allocations and consider to what extent they want to live up to recipients' expectations. Moreover, I assume that the overall guilt function has the following functional form:

$$\frac{dG\big(t_j(d),\, \beta\big)}{d\, t_j} \leq 0 \text{ and } \frac{d\, G\big(t_j(d),\, \beta\big)}{d\, t_j^2} \leq 0 \tag{4}$$

These assumptions imply that i suffers from falling short of the expectation of others and the marginal disutility from falling short increases in t_j. Hence, i suffers more from violating the expectations of many to a minor extent than violating the expectations of a single person to a large extent. A guilt function $G\big(t_j(d),\, \beta\big)$ that satisfies these assumptions has the following structural form:[10]

$$G\big(t_j(d),\, \beta\big) \;=\; max\{\, 0,\, \beta - \; t_j(d)\, \} \tag{5}$$

The guilt function implies that i is sympathetic and anticipates j's utility. In particular, i feels guilty for falling short of js alleged expectations (second order beliefs).[11] However, following Battigalli and Dufwenberg (2007), I assume that i does not necessarily feel guilty if the alleged expectations of j are violated but whether she experiences the feeling of guilt depends on how much of perceived j's experience loss is due to her behavior. Thus, the psychological payoff function includes a correction term:

$$C\big(t_j(d),\, \beta\big) \;=\; min_d\{G\big(t_j(d),\, \beta\big)\} \;=\; min_d\, \{max\{\, 0,\, \beta - \; t_j(d)\, \}\} \tag{6}$$

The correction term $C\big(t_j(d),\, \beta\big)$ measures j's disappointment in case that i decides most altruistically. Not every decision maker i experiences feelings of guilt in the same intensity. Therefore, η ≥ 0 measures how prone i is towards the felling of guilt. Hence, the total psychological payoff function is defined as

$$P(G(t_j,\, \beta),\; C(t_j(d),\, \beta)) \;=\; \eta \,\cdot\, (G\big(t_j(d),\, \beta\big) \,-\, C\big(t_j(d),\, \beta\big). \tag{7}$$

[10] In the appendix B.3 I discuss an extension of the model which accounts for the fact that some decision makers may get pleasure from surprising others—as proposed by Khalmetski et al. (2015).

[11] The structural form of $G(t_j,\beta)$ implies constant marginal disutility for falling short of the expectations of others, which in turn implies that the recipients decrease in utility is independent from her wealth level. However, this might not be the case and decision makers anticipate that recipients' utility function might be non-linear in its monetary dimension instead. Nevertheless, Khalmetski et al. (2015) and Jensen and Kozlovskaya (2016) propose utility functions incorporating similar psychological payoff functions. In real life scenarios, the interdependence of wealth level and marginal utility may often not be readily intelligible by all parties concerned. Thus, I stick to a simple psychological payoff function.

Accounting for all preliminary assumptions and definitions, the overall utility function is given by

$$\begin{aligned} U(t_i, \beta) &= m(t_i) + P\left(G\big(t_j(d), \beta\big), C\big(t_j(d), \beta\big)\right) \\ &= m(t_i) + \eta \cdot \left(G\big(t_j(d), \beta\big) - C\big(t_j(d), \beta\big)\right). \end{aligned} \tag{8}$$

In multi-agent settings (such as when multiple employees cooperate in teams) in which one or multiple agents are affected by the decision of one or more decision makers, a single decision-maker (she) lets another interaction partner or a collective of interaction partners down if due to her decision the perceived expectations of one or multiple of these interaction partners are violated. Hence, the sum of all deviations from the perceived expectation *$\beta(j)$* for every individual *j* affected by decision makers i measures the extent to which i lets down his interaction partners. I assume –for the sake of simplicity– that *$\beta(j)$* is constant for all *js*.

If more than one decision maker affects the outcome of at least one interaction partner, the decision maker will have to account for the expected impact of co-decision-makers on the payoff of the affected interaction partner(s) as well. This relationship is captured by the function $t_j(d, \gamma)$, where $\gamma = E_i[t_j]$ captures i's first order beliefs about the decision of co-decision makers.

Furthermore, in settings where the decision of an individual affects more than one other agent, the utility function above is adjusted such that it captures that decision makers discount the disutility of a single individual j more if the costs of i's decisions are dispersed among many affected people by introducing a weighting function *$f(n) < 1$* ($\frac{df(n)}{dn} \leq 0$ and *$f(n) \leq n$*).[12]

I assume that $\beta_j = \beta$ for all j, i.e., the decision maker's second order beliefs are identical for all recipients, as well. This is a reasonable assumption in all settings in which the decision makers have the identical information about every interaction partner –in particular for anonymous online experiments or markets. The utility function of a single i in a multi-agent setting is therefore:

$$\begin{aligned} U(t_i, \beta) &= m\big(t_i(d)\big) + f(n) \sum_{j=1}^{n} P(G(t_j(d, \gamma), \beta), C(t_j(d, \gamma), \beta(j))) \\ &= m\big(t_i(d)\big) + f(n) \cdot \eta \sum_{j=1}^{n} G\left(t_j(d, \gamma), \beta(j)\right) - C\left(t_j(d, \gamma), \beta(j)\right) \end{aligned} \tag{9}$$

The existence of the utility function's maximum is guaranteed by the extreme value theorem by Weierstrass, whereas its uniqueness is assured by strict concavity of player B's utility function.

[12] The weighting function depicts the empirical findings of Schumacher et al. (2017) who found that less dictators in dictator games are concerned about an equal distribution of shares if the benefits of pro-social behavior are more dispersed. In line, Andreoni (2007) finds in a dictator game with multiple recipients that distributing their endowment dictators take the number of recipients with a decreasing rate into account. Moreover, there is empirical evidence that people shirk more likely on effort (Carpenter, 2007; Isaac & Walker, 1988; Isaac, Walker, & Williams, 1994) or commit social loafing if they work together in a group (Karau & Williams, 1993) but that the tendency to shirk is insensitive to changes in group size (Ingham, et al., 1974). Insensitivity to group size in form of increasing discount factors may rationalize the extension neglect cognitive bias (Kahneman, 2003). This bias arises when people tend to ignore the size of a set during an evaluation in which the size should be relevant.

A more tractable parametric specification of the overall utility function has the following form:

$$\begin{aligned} U(t_i,\beta) &= m\big(t_i(d)\big) + P\Big(G\big(t_j(d,\gamma),\beta\big),\, C\big(t_j(d,\gamma),\beta\big)\Big) \\ &= m\big(t_i(d)\big) + \eta \cdot \Big(G\big(t_j(d,\gamma),\beta\big) - C\big(t_j(d,\gamma),\beta\big)\Big) \\ &= ln\big(t_i(d)\big) - \eta \cdot (max\{\,0,\, \beta - t_j(d,\gamma)\,\} \\ &\qquad\qquad + min_d\,\{\,max\{\,0,\, \beta - t_j(d,\gamma)\,\}\,\}) \end{aligned} \tag{10}$$

In settings where more than two players are affected by i's decision the utility function of a single i in a multi-agent setting can be written as:

$$\begin{aligned} U(t_i,\beta) &= m\big(t_i(d)\big) + f(n) \cdot \sum_{j=1}^{n} P\Big(G\Big(t_j(d,\gamma),\,\beta,\, C\big(t_j(d,\gamma),\beta\big)\Big)\Big) \\ &= m\big(t_i(d)\big) - \eta \cdot f(n) \cdot \sum_{j=1}^{n} G\big(t_j(d,\gamma),\beta\big) - C\big(t_j(d,\gamma),\beta\big) \\ &= ln\big(t_i(d)\big) - \eta \cdot f(n) \cdot \sum_{j=1}^{n} max\{0,\beta - t_j(d,\gamma)\} \\ &\qquad + min_d\,\{\,max\{\,0,\, \beta - t_j(d,\gamma)\,\}\,\} \end{aligned} \tag{11}$$

2.2 Application of the Simple Guilt Model to a Dictator Game Setting

Standard Dictator Game Setting - Next, I apply the previously presented utility function to a standard one-to-one (one dictator, she, and one recipient, he) as well as a multi-agent dictator game (with *n* dictators and *n* recipients). In a standard dictator game, one dictator has to decide how to divide an initial endowment between herself and her matched passive recipients. Thereby, dictators face the moral dilemma between investing in their own wealth and investing in their good conscience by behaving pro-socially. The notation is motivated by the fact that the dictator game described in this paper has a take-framework. Clearly, considering a selfish utility maximizing agent, the dictator keeps the endowment entirely. Henceforth, I will assume that $T = 1$, i.e., the endowment is normalized to 1. It follows that the amount of money kept by the dictator is given by t_i and the amount of money assigned to the recipient is given by $t_i = T - t_i$. Furthermore, a dictator can never keep less than 0 and more than T ($0 \leq t_i \leq T = 1$). Assume otherwise that the dictator may experience guilt: in this case the dictator's utility depend on her second order beliefs regarding the recipient's belief about his income

$$E_i\left[E_j[t_j]\right] = \beta. \tag{12}$$

The optimization problem of a standard dictator game according to the simple guilt model is thus defined as follows:

$$\max U(t_i,\,\beta) \;=\; ln(t_i) - \eta \cdot \max\{\beta - (T - t_i), 0\} \;\; s.t. \;\; 0 \leq t_i \leq T = 1 \tag{13}$$

The dictator maximizes her utility with respect to her transfer payment t_i given her second order beliefs β and the budget constraint. Maximizing the utility function yields the following best response function, i.e., optimal values contingent on i's second order beliefs β:[13]

$$t_i^* = \begin{cases} 1 & \text{if } \frac{1}{\eta} > 1 \\ \frac{1}{\eta} & \text{if } 1 - \beta \leq \frac{1}{\eta} \leq 1 \\ 1 - \beta & \text{if } 1 - \beta \geq \frac{1}{\eta} \end{cases} \tag{14}$$

Dictators should state the same decisions whether recipients can observe their actions or not, because the simple guilt model implicitly assumes that dictators are intrinsically motivated. Moreover, it follows from $\eta > 0$ that $\frac{1}{\eta} > 0$. This argument implies that a dictator is not willing to give up her entire endowment under any condition. Moreover, it follows from equation 14 that the optimal dictator share $t_i{}^*$ is strictly decreasing in η if and only if $\frac{1}{\eta} \leq 1 \; - \beta$.

Multi-Agent Dictator Game with Multiple Recipients and Dictators - To study behavior in multi-agent settings, some participants played a multi-agent dictator game in my experiment. Here, t_i resembles again the dictator's payoff. In the multi-agent dictator game introduced in this paper the same number n of dictators and recipients interact.[14] Every dictator has to decide how much money of her initial endowment T = 1 she would like to keep and how much money she would like to transfer in equal shares to n recipients. Thereby, dictators commonly determine recipients' payoffs. Hence, a single dictator's payoff is again given by t_i. In the multi-agent setting $\beta = E_i[E_j[t_j]]$ captures the second order beliefs of a single dictator about the average first order belief of a recipient regarding his income. The allocating dictator takes other dictators' decisions and their consequences on recipients into consideration. Hence, the decision maker builds expectation about how much money an average other co-dictator claims for herself. Dictator i's first order beliefs are denoted by $\gamma = E_i[t_j]$. Consequently, the amount of money that all dictators leave to a single recipient j is given by:

$$t_j \; = T \; - \; \left(\frac{n-1}{n} \; \cdot \gamma \; + \; \frac{t_i}{n}\right) \tag{15}$$

A dictator in general experiences guilt if $t_j \; = \; 1 - \frac{n-1}{n} \cdot \gamma - \frac{t_i}{n} < \beta = E_i\left[E_j[t_i]\right]$, i.e., if the perceived recipients' expectations are not met. Moreover, the dictator does no longer feel guilty for the difference between the perceived expectations of the recipient's income and his actual income if the dictator has chosen the most pro-social decision. This is captured by a correction term:

$$C(\beta, \gamma) \; = \; max\,\{\beta \; - \; \left(1 \; - \; \frac{n-1}{n} \; \cdot \gamma \; - \; \frac{1}{n}\right) \; ; \; 0\,\} \tag{16}$$

[13] Proof of this proposition is provided in the appendix B.1.1.

[14] A more detailed description of the game is included in section 3.

The dictator’s utility function is therefore defined as

$$\begin{aligned} U(t_i,\beta) \; &= \; ln(t_i) \\ &-f(n) \cdot \sum_{j=1}^{n} \eta \cdot \; max\{\beta - \left(1 - \frac{n-1}{n} \cdot \gamma - \frac{t_i}{n}\right); \, 0\} \\ &+ \eta \; \cdot max\{ \; (\beta) \; - \; \left(1 - \frac{n-1}{n} \cdot \gamma - \frac{1}{n}\right) \; ; \, 0 \, \} \\ &\qquad s.t. \;\; 0 \; \leq \; t_i \; \leq \; T = 1 \end{aligned} \tag{17}$$

Overall Decision Environment Effect on the Distribution Decision- Next, to evaluate how a dictator decides in a multi-agent setting one must consider three different cases. In the first case the recipients will irrespective of i’s decision suffer from disappointment. This holds if

$$\beta \geq 1 - \frac{(n-1) \cdot \gamma}{n}. \tag{18}$$

In the second case recipients experience contingent on the dictators’ decisions, disappointment or surprise. This holds if

$$1 - \frac{(n-1) \cdot \gamma + 1}{n} \leq \beta \leq 1 - \frac{(n-1) \cdot \gamma}{n} \tag{19}$$

The third case captures the situation where the recipients will be positively surprised irrespective of the individual dictator’s decision. This holds if

$$\beta \; \leq 1 - \; \frac{(n-1) \cdot \gamma}{n} \tag{20}$$

I provide a formal proof for the derived best-reply functions under different conditions in the appendix B.1.3. Considering the first case (equation 18) the dictator’s optimal distribution choice is given by

$$\mathrm{t_i^\star} = \begin{cases} \qquad 1, & if \; \dfrac{1}{f(n) \cdot \eta} > T = 1 \\ \dfrac{1}{f(n) \cdot \eta}, & if \; \dfrac{1}{f(n) \cdot \eta} \leq \, 1 \end{cases} \tag{21}$$

Considering the second case (equation 19) the dictator’s optimal distribution choice is:

$$\mathrm{t_i^\star} = \begin{cases} \qquad 1, & if \; \dfrac{1}{f(n) \cdot \eta} > T = 1 \\ \dfrac{1}{f(n) \cdot \eta}, & if \; 1 - \beta - \dfrac{(n-1) \cdot \gamma}{n} < \dfrac{1}{f(n) \cdot \eta} \leq \, 1 \\ 1 - \beta - \dfrac{(n-1) \cdot \gamma}{n}, & if \; \dfrac{1}{f(n) \cdot \eta} > \; 1 - \beta - \dfrac{(n-1) \cdot \gamma}{n} \end{cases} \tag{22}$$

Considering the third case (equation 20) the dictator’s optimal distribution choice is given by

$$\mathrm{t_i^\star} = 1. \tag{23}$$

Overall Group Size Effect - The theoretical model leads to ambiguous results regarding whether dictators distribute less or more to recipients in multi-agent conditions. In the appendix B.1.4 I provide a detailed formal discussion under which conditions dictators distribute more or less than in the standard dictator game. In this section, I only present intuitions and predictions. Ceteris paribus, the relevant factors determining whether a dictator decides more pro-socially or more selfishly are the individual guilt parameter as well as dictators' first order beliefs about her co-dictators' decisions. If $\gamma \geq 1 - \beta$, other dictators are presumed to always distribute at least as low payoffs to themselves than in a standard one-to-one dictator game. This effect results firstly from the fact that due to the weighting term dictators attach less relative weight to the associated harm or surprise of the recipients affected by his decision. Secondly, if dictator i assumes other dictators surpass the recipients' behavior, she likely free-rides on the moral decisions of others. In other terms, if multiple actors jointly determine an outcome and in addition an individual decision maker believes that others act pro-socially, they may refrain from acting pro-socially. In contrast, if $\gamma \leq 1 - \beta$ (other dictators are considered to distribute higher payoffs to themselves than in the standard dictator games), the prediction whether dictators are more selfish is ambiguous. If $\frac{1}{f(n)\cdot(1-\beta)} \geq \eta$ dictators claim at least as high $\mathrm{t}_{\mathrm{i}}^{\star}$ than in the standard dictator game setting. If $\frac{1}{f(n)\cdot(1-\beta)} < \eta$, dictators will claim lower $t_i^{\star}$ compared to the standard dictator game.

Thus, a dictator *i* is only willing to give more in the multi-agent setting than in a standard dictator game if she expects other dictators to fall below the expectations of their recipients. Hence, it remains an empirical question whether dictators claim a higher amount of money for themselves in multi-agent dictator games–as predicted for only moderately guilt averse dictators or dictators who believe in the pro-sociality of co-dictators–or less.

Hypothesis 1: *Dictators will claim a higher amount of money for themselves in the multi-agent dictator games.*

Free Riding Effect - Next, I discuss the impact of dictator *i*'s first order beliefs of other -dictators' allocation decisions on their individual decisions. As previously shown, dictators always give less in multi-agent dictator games if the first order beliefs regarding other co-dictators' distribution decisions are higher than their second order beliefs ($\gamma \leq 1 - \beta$) or dictators are only moderately guilt averse $\left(\frac{1}{f(n)\cdot(1-\beta)} \geq \eta\right)$. Assume otherwise, i.e., $\frac{1}{f(n)\cdot(1-\beta)} < \eta$. The only condition in which $t_i^{\star}$ is linearly dependent from the dictator's first order beliefs about other dictators' allocation decision, i.e., γ, is

$$\frac{1}{f(n)\cdot\eta} \geq n\cdot\left(1-\beta-\frac{(n-1)\cdot\gamma}{n}\right) \tag{24}$$

This follows directly from equation 22 case 3. Hence,

$$\frac{\partial n\cdot\left(1-\beta-\frac{(n-1)\cdot\gamma}{n}\right)}{\partial\gamma} =- (n-1) < 0 \text{ for all n} > 1 \tag{25}$$

The more co-dictators demand for themselves the more a sufficiently guilt-averse dictator *i* will distribute to the recipients.[15]

Hypothesis 2: *There is a negative correlation between dictators' first order-beliefs about the amount taken by other dictators and the amount of money they claim for themselves.*

Group Discounting Effect - The intuition as to why an increase in the number of recipients theoretically leads to an increase in the amount of money the dictator distributes to himself is straightforward. The utility function of the multi-player game is specified as

$$U(t_i, \beta) = m(t_i(d)) - f(n) \cdot \sum_{j=1}^{n} P\left(G\left(t_j(d), \beta, C(t_j(d), \beta)\right)\right). \tag{26}$$

Furthermore, by definition $f(n) \geq f(m)$ for every $m > n$. Hence, trivially it follows that because the psychological payoff function $P\left(G\left(t_j(d), \beta, C(t_j(d), \beta)\right)\right)$ is multiplied by a smaller argument if a dictator determines the payoff of m instead of n recipients, the dictator attaches less relative importance to her psychological payoff and more relative importance to her own monetary welfare. Consequently, *i* distributes more to herself.[16]

Hypothesis 3: *In the multi-agent conditions dictators will claim a higher amount for themselves, the more dispersed the costs of their selfish actions are, i.e., the more recipients are affected by the decision.*

Moreover, to what extent an individual *i* acts more selfishly if the number of recipients increase from say four to 20 recipients depends on the structural form of $f(n)$. If the function $f(n)$ for instance is defined as $f(n) = \frac{1}{n}$ dictators will react more sensitive to changes in the number of recipients than if the function is defined as $f(n) = c + \frac{1}{n^2}$ where c is a constant. In the latter case, sufficiently guilt-averse dictators will – irrespective of the number of recipients– always show pro-social behavior but will become less sensitive to marginal changes in recipients' number the higher *n* is.[17]

Anticipation Effect - Hypothesis 4 focuses on potential anticipation effects: every case in which the average dictator's distribution decision equals the average expected distribution decision constitutes an equilibrium. In a standard dictator game according to equation 14 all beliefs between $0 \leq 1 - \beta \leq \frac{1}{\eta}$ constitute potential equilibria. In a multi-agent dictator game, according to equation 22, all values for β that satisfy

$$\frac{1}{f(n) \cdot \eta} \geq n \cdot \left(1 - \beta - \frac{(n-1) \cdot \gamma}{n}\right) \tag{27}$$

[15] The prediction of the guilt aversion model contradicts the idea that dictators may simply mimic behavior of other dictators, in order to stick to a social norm that is equal to the perceived behavior of other dictators (Krupka & Weber, 2013).

[16] Formal proof for this proposition is provided in the appendix B.1.5.

[17] Further note, Schumacher et al. (2017) provided evidence that dictators act less pro-socially if they interact with multiple recipients. However, decision makers tend to be insensitive to the group size. As a consequence, the average transfer payment in a dictator game with one dictator and 4 or 32 recipients were not significantly different from each other. In addition, Ingham et al. (1974) have shown that shirking in rope pulling does not rise significantly if an individual cooperates with 3, 4, 5 or 6 group members. This corroborates the insensitivity hypothesis.

can constitute potential equilibria. Consequently, in a standard dictator game there exists in-equilibrium beliefs that cannot constitute an equilibrium in the multi-agent setting, since the belief levels are too high. To see this, consider the case in which the dictator *i* assumes that her co-dictators live up to the beliefs of the recipients. The best response function of the multi-agent game thus simplifies to the following equation:

$$\mathrm{t}_{\mathrm{i}}^{\star} = \begin{cases} 1, & if \ \dfrac{1}{f(n)\cdot\eta} > T = 1 \\ \dfrac{1}{f(n)\cdot\eta}, & if \ \dfrac{1}{f(n)\cdot\eta} \leq 1-\beta \\ 1-\beta, & if \ \dfrac{1}{f(n)\cdot\eta} > 1-\beta \end{cases} \tag{28}$$

Now clearly, all scenarios where $\frac{1}{f(n)\cdot\eta} > 1-\beta \geq \frac{1}{\eta}$ only constitute equilibria in standard dictator games, but not in the multi-agent settings. It is likely that recipients foresee that dictators are less generous in multi-agent dictator games. Dictators' themselves may anticipate or notice these changes in the belief structure. As a consequence, dictators lower their distribution to recipients due to a decline in their second order beliefs.

Hypothesis 4: *Recipients expect that dictators claim more if they are in a multi-agent setting with multiple dictators and recipients.*

2.3 Guilt from Blame Model

In contrast to the simple guilt model, the guilt from blame model assumes that an individual *i* is not reluctant about betraying the expectations of others per se but is rather afraid of the social consequences that come with not living up to their expectations and appearing to be selfish. Yet, *j* only blames *i* in case the latter is responsible for the transfer payment (Battigalli & Dufwenberg, 2007), which implies that dictators are merely externally motivated to live up to the expectations of others and tend to act more selfishly if their actions are not (perfectly) observable.[18]

In order to integrate these preliminary thoughts, the guilt term in the guilt from blame model has to be redefined such that it is no longer contingent on the final payoffs of *j*s but instead on the extent to which *j*s blame or appraise the individual *i* for their final payoffs. Hence, it is necessary to differentiate between the disappointment about the final payoff (the relevant behavioral motivation in the simple guilt model) and the assignment of responsibility and blame for the outcome (the relevant behavioral motivation in the guilt from blame model).

The level of disappointment about the final payoff is independent from the process on how it is determined. In contrast, the individual assignment of responsibility and blame is dependent on individual levels of contribution. To assess to what extent a decision maker is considered to be responsible, every affected agent *j* defines a benchmark for every individual decision maker *i*. If the individual decision maker's (contribution) decision is less pro-social than demanded by this benchmark, the decision maker is blamed for her decision. The benchmark in a game where decision makers jointly decide upon an allocation –such as

[18] Hence, the model captures how sensitive the dictator is towards the blame or appraisal that comes with a deviation from the reference point. In contrast, the guilt from blame model introduced by Battigalli and Dufwenberg (2007) claims that a similar term measures how much an economic agent intends to let another person down. Yet, speaking about intent might be misleading, because guilt from blame models are targeted to inference about actual decision and not to whether an economic agent comes willingly to the decision to bring about a certain consequence.

the discussed dictator game– is a certain share s of the expected payoff of and by j denoted as in the simple guilt model by $\alpha = E_j[t_j]$.

In a single dictator game this share s is given by 100% of α. In a multiple dictator game in which n dictators all face the same contribution decision, this share is most likely the same for every decision maker and thus given by $s = \frac{1}{n}$. If in contrast to this setting the individual endowments differ between decision makers, the individual benchmarks may differ. This allows to capture why, for instance, service staff may expect higher tips from rich customers. Because i cannot observe $s \cdot E_j\,[t_j] = s \cdot \alpha$ she forms expectations about $E_i\,[s \cdot \alpha] = s \cdot \beta = b(\beta)$, which act as a reference point in the guilt from blame model. An individual decision maker i experiences guilt from blame if his individual contribution $t_{i \to j}$ to j is below the reference point $b(\beta)$. Taking all of these aspects into account, the guilt term in the guilt from blame model is given by

$$\frac{dG\big(t_{i\to j}(d),\, b\big)}{d\, t_{i\to j}} \leq 0 \text{ and } \frac{d\, G\big(t_{i\to j}(d),\, b\big)}{d\, t_{i\to j}^{2}} \leq 0 \text{ e.g.}, G\left(b, t_{i\to j}(d)\right) = max\{0, b(\beta) - t_{i\to j}(\tilde{d})\}. \tag{29}$$

In the guilt term above, $t_{i\to j}(d)$ measures the individual distribution of i to j. The monetary utility function $m(t_i(d))$ and the weighting function $f(n)$ are defined as in the simple guilt model.[19] A utility function that fulfills these requirements is such given by:

$$U(t_i, \beta) = m\big(t_i(d)\big) + f(n) \cdot \eta \sum_{j=1}^{n} G(d, b(\beta))) \tag{30}$$

The guilt from blame model accounts for behavioral changes due to imperfect ex-post information structure, i.e., individual actions are not perfectly observable. Consequently, affected people cannot observe (e.g., due to fact that multiple decision makers are responsible for an outcome) to what extent an individual increased their final payoffs. However, recipients[20] are not totally ignorant, but knows that with a certain probability the decision maker is responsible for a selfish or pro-social action that among others determine their payoffs. Given that j observes his payoff t_j, he draws interference about the distribution of potential decisions d given by $h(d)$, i.e., the recipient interferes from $g(t_i \to j|t_j)$ the probability that a decision $\tilde{d}$ has been made. The relationship between the j's payoff t_j and the density function $h(\tilde{d}|t_j)$ is such that for every $t_j > t_k$ it holds for the cumulative distribution function that $H_{tj}\,(t_i) \leq H_{tk}\,(t_i)$, i.e., the first distribution first order stochastic dominates the second. Given that the dictator dislikes being blamed, her overall utility function under the assumption of imperfect information can be written as:

$$U(t_i, \beta) = m(t_i(\tilde{d})) + f(n) \cdot \eta \sum_{j=1}^{n} \int_{-\infty}^{\infty} G(\tilde{d}, b(\beta)) h(d|t_j) d\tilde{d} \tag{31}$$

[19] The guilt from blame model does not include a correction term, because the changes in the perceived recipients' benchmark already account for the fact that a single decision maker is not responsible for compensating the selfish behavior of others. Secondly, the decision maker suffers from negative social consequences independently from whether she is able to live up to the expectation level behaving most pro-socially.

[20] Following Battigalli and Dufwenberg (2007) the proposed utility function does not capture that dictators experience shame if they are blamed by other co-dictators, but instead focuses on the relationship between the dictator and her direct interaction partner instead. For models that incorporate the assessment of third parties regarding the moral value of made decisions see (Krupka & Weber, 2013)

2.4 Application of the Guilt from Blame Model to a Dictator Game Setting

Two-Agent Dictator Game - Applying the previously introduced guilt from blame model to a standard two-agent dictator game setting under the assumption of perfect information, the maximization problem with respect to t_i of an individual decision maker is given by

$$max\, U\,(t_i, \beta) = ln(t_i) - \eta \cdot max\{\,(\beta) - (T - t_i), 0\} \text{ s.t. } 0 \leq t_i \leq T = 1. \tag{32}$$

Obviously, in a two-agent setting with a perfect ex-post information structure where a recipient is fully aware of the dictator's unshared responsibility for the height of the transfer payment, the simple guilt and the guilt from blame model lead to coinciding optimal strategies (c.f. Battigalli & Dufwenberg, 2007). Thus, for a deeper analysis of the dictator's behavior in a standard dictator setting I refer to the previous section.

Multi-Agent Dictator Game Setting with Perfect Information- While the simple guilt model and the guilt from blame model coincide in the standard two-player dictator game, the simple guilt model and the guilt from blame model diverge from each other in a multi-agent setting with perfect ex-post information structure. In such a setting in which recipients are able to observe the individual contributions– the maximization problem of an individual dictator with respect to t_i according to the guilt from blame model is given by:

$$max\, U\,(t_i, \beta) = ln(t_i) - f(n) \cdot \sum_{j=1}^{n} \eta \cdot max\{\left(\frac{\beta}{n}\right) - \frac{T - t_i}{n}; 0\} \text{ s.t. } 0 \leq t_i \leq T = 1 \tag{33}$$

The best response function is given the perceived beliefs β given by[21]

$$t_i^* = \begin{cases} 1 & \text{if } \frac{1}{\eta \cdot f(n)} > 1 \\ \frac{1}{\eta \cdot f(n)} & \text{if } 1 - \beta \leq \frac{1}{\eta \cdot f(n)} \leq 1 \\ 1 - \beta & \text{if } 1 - \beta \geq \frac{1}{\eta \cdot f(n)} \end{cases} \tag{34}$$

As in the simple guilt model, the group discounting effect causes the dictators to divide the endowment T more selfishly, since $f(n) < 1$ implies that $\frac{1}{\eta} < \frac{1}{\eta \cdot f(n)}$. Hypothesis 1 (overall effect), hypothesis 2 (group discounting effect), and hypothesis 4 (anticipation effect) can analogously to the simple guilt model be derived from the guilt from blame model. Regarding hypothesis 3 (balancing effect), in the guilt from blame model –in contrast to the simple guilt model– the individual decision maker, in a multi-agent setting with perfect ex-post information, has no incentive to compensate for the anti-social behavior of others, because the utility of the decision maker is independent from the expectations of the distribution decision of other dictators.

[21] Proof of this proposition is provided in appendix B.2.1.

Multi-Agent Dictator Setting with Imperfect Information– In an environment with imperfect information the dictator's utility function according to the guilt from blame model is given by:

$$U(t_i(d), \beta) = \ln\big(t_i(d)\big) + \eta \cdot f(n) \cdot \sum_{j=1}^{n} \int_0^1 \max\{\frac{\beta}{n} - \frac{T - \tilde{t_i}}{n}, 0\} \cdot h\big(\tilde{t_i} | t_j\big) d\tilde{t_i} \qquad (35)$$

In contrast to the multi-agent setting with perfect ex-post information, an individual dictator can only indirectly affect the perception of her actions due to an increase in $t_j = T - \left(\frac{n-1}{n}\gamma + \frac{1}{n}t_i\right)$ by decreasing t_i. An increase of t_j has an effect on $\tilde{t_\iota}$, because for any $t_j > t_k$ it holds for the cumulative distribution functions that $H(\tilde{t_\iota} \mid t_j) \geq H(\tilde{t_\iota} | t_k)$. Hence, intuitively, also the reverse impact of t_j on t_i ($E[t_i|t_j]$) likely decreases in *n,* since $t_j = T - \left(\frac{n-1}{n}\gamma + \frac{1}{n}t_i\right)$ and hence j can infer less about t_i by observing t_j. Therefore, in conclusion, if recipients cannot observe how much each decision maker contributes to their payoffs, a single dictator still retains the incentive to increase the amount assigned to the recipient in order to boost public image, though it decreases considerably in *n*. This implies that given other dictators demands high shares, dictator *i*'s utility of being pro-social in an imperfect ex-post information setting falls below her utility in a perfect ex-post setting, because in the former case recipients attach—due to the opaqueness of intentions—less honor to the *i*'s pro-social behavior than in the latter (transparency effect). Vice versa, in case that other dictators only claim moderate shares, *i*'s utility of being selfish in case of an imperfect ex-post information setting exceeds her utility in a perfect ex-post setting, because in the former case recipients blame—due to the opaqueness of intentions— *i* less for his selfish decision compared to the latter case. As a result, the incentive to act pro-socially decreases, while the incentive to act selfishly increases. In the appendix, I provide a formal argument proofing the prediction on which hypothesis 5 rests upon under the assumption of a degenerated distribution function.

Hypothesis 5: *Dictators demand more if dictators' decisions are opaque*

2.5 Extensions and Further Remarks

In this subsection I introduce four additional remarks. First, based on the insights of Vainapel et al. (2018), I presume that the phenomenon of diffusion of blame may leverage the conjectured attributability or transparency effect. In particular, they find experimental evidence that group members are less likely to be suspected, blamed and punished when they are judged as separate individuals compared to as a group.[22] Furthermore, I theorize that diffusion of blame might be driven by decision makers' aversion to commit errors in judgment. Apparently, the error of risk will be higher if it is unclear to what extent an individual within one group contributed to an undesired outcome.[23] Second, the guilt from blame model—as stated here—implicates that the loss in utility caused by experiencing blame for falling short from the experiences weighs as heavily as the gain in utility caused by the social appraisal associated with a positive surprise. However, this assumption has yet to be confirmed. As an alternative explanation, disapproval resulting

[22] Because immoral behavior in their cheating experiment had positive externalities on other group members in terms of final payoffs, the authors hypothesized that the reduced negative judgment of each group members is at least in parts be explained by the justifiability of lies. Since I abstract from any positive externalities this explanation is neither relevant for my model nor can it explain any of my treatment effects.

[23] This argument is supported by the finding that when both group members were brazen liars, it was no longer the case that people attribute less immoral character to each group member when judged separately compared with judging a brazen individual or group (Vainapel et al., 2018).

from falling short of other people's expectations puts high social pressure on dictators, while dictators who are willing to "go the extra mile" are more likely intrinsically motivated. In other words, dictators tend to experience loss aversion (Kahneman & Tversky, 1979; Köszegi & Rabin, 2006) in their social image dimension, i.e., dictators fear being perceived as selfish far more than they enjoy being perceived as generous.

Third, a change in the ex-post information structure allows to discriminate between the simple guilt and the guilt from blame model, and to distinguish whether it is guilt from blame or concerns over a fair outcome which motivates a dictator to act pro-socially. If giving in dictator games can exhaustively be explained by dictators' concerns over outcomes, it should make no difference whether recipients can observe dictators' actions or whether recipients cannot draw inference about the origin of their payoffs. Having stated this, it is likely that dictators' moral decisions in experimental situations are not solely driven by the internalization of the norm to not violate the expectations of others or social image concerns but most likely by internal in combination with external factors (c.f. Abeler et al., 2019).

Fourth, numerous papers attempt to find evidence in favor of guilt aversion by measuring the correlation second order beliefs and their transfer payments (e.g., Charness & Dufwenberg, 2011; Ellingsen et al., 2010; Kawagoe & Narita, 2014; Morell, 2017). The hypothesis is based on a prediction derived from the linear guilt model: it predicts that a dictator either fully matches recipients' second order beliefs or keeps the entire endowment. Considering a logarithmic instead of a linear guilt aversion model, there is only a correlation between second order beliefs and transfer payments if decision makers are highly guilt averse Jensen and Kozlovskaya (c.f. 2016). In the appendix B.1.2 I thus provide a formal discussion under which conditions there is a linear relationship between beliefs and transfer payments and show that even in the absence of a correlation guilt likely impacts behavior.

3 Experiment

3.1 Experimental Course

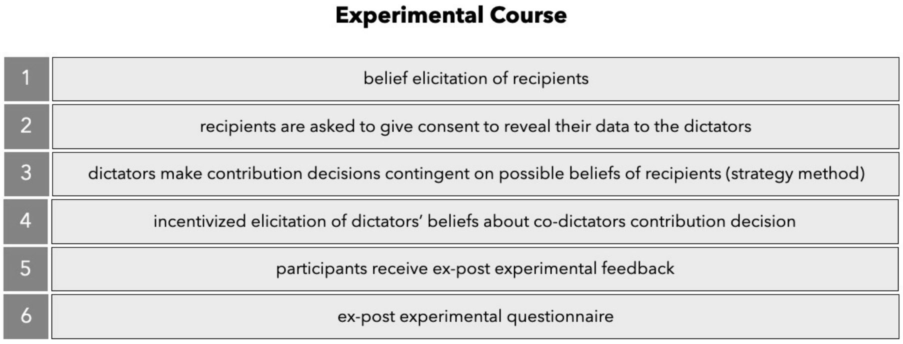
The experiment was conducted May 2018 at the online marketplace Amazon Mechanical Turk (MTurk)[24] using oTree (Chen et al., 2016). In each of the five treatments 60 dictators and 60 recipients participated. 600 subjects participated in total. Additional to their payoff from the dictator game as well as their payoff from the belief elicitation task, participants received a participation fee of \$0.25. On average participants earned \$0.77. The vast majority of workers needed between 5 - 10 minutes to complete the experiment. The average hourly payoffs were around the American minimum wage and thus significantly higher than the average hourly wage on MTurk.[25] Conducting my experiment, I inquired whether an agent acts more selfishly when the decision environment is a m:n-relationship setting compared to when it is a 1:1-relationship setting, and if so why. On an experimental level, I examined whether a dictator in a multi-agent dictator

[24] Online labor markets provide access to diverse subject pool, including low and high-skilled workers as well subjects from different age groups. Subjects in online experiments have little experience with economic experiments. Even though participants from the MTurk subject group are more naïve with respect to experiments compared to participants from standard lab populations, I excluded experienced MTurkers who finished more than 1000 Hits to sign up for the experiment. I control for sophistication effects, by asking participants whether they have already participated in a similar study and elicited the current or former subject of study. Lastly, online experiments are less sensitive for session effects than standard lab experiments.

[25] Yet, small stakes, difficulties in creating common knowledge, uncertainty about the subjects' identity and the missing opportunity in online experiments to ask and answer questions in real time raises concerns about the data quality of online experiment. Multiple studies examining these issues conclude that there exist slight differences in the experimental outcomes between classical lab experiments conducted with college students and experiments on MTurk (see e.g., Amir et al., 2012; Hauser & Schwarz, 2016; Horton et al., 2011; Raihani et al., 2013). I elicited former or current subjects of study to control for occupational effects.

game with a take-framework and more than two players acts less morally inclined. In a dictator game with a take framework, dictators can solely decide how to divide an initial endowment between themselves and recipients by taking away money from passive recipients.

Figure 2.1: *Course of the Experiment*

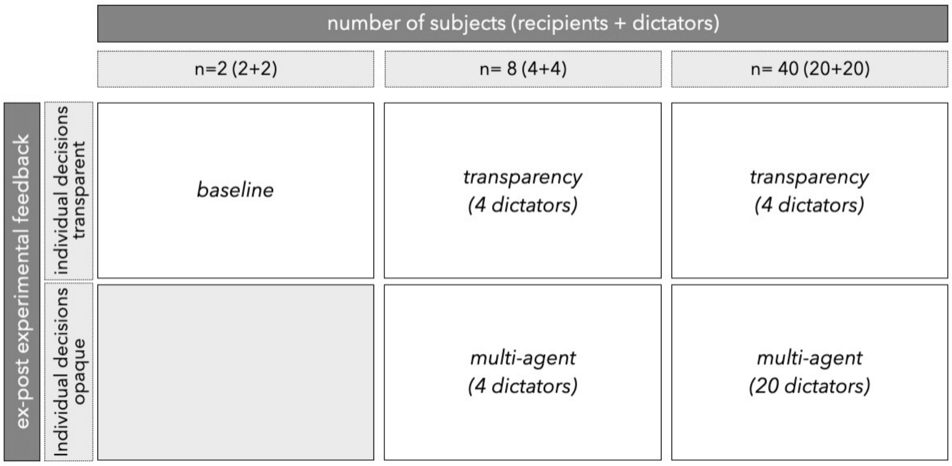


Experimental Course

1. belief elicitation of recipients
2. recipients are asked to give consent to reveal their data to the dictators
3. dictators make contribution decisions contingent on possible beliefs of recipients (strategy method)
4. incentivized elicitation of dictators' beliefs about co-dictators contribution decision
5. participants receive ex-post experimental feedback
6. ex-post experimental questionnaire

Figure 2.1 illustrates the course of the experiment. An overview about the treatments is provided in Figure 2.2. At the beginning of the experiment all subjects in each treatment were randomly assigned either "type A" (dictator) or "type B" (recipient).[26] In the baseline treatment one dictator and one recipient, while in the multi-agent as well as the transparency treatments four (respectively 20) dictators and four (respectively 20) recipients interacted. I begin with describing the dictators' basic decision in stage two before explaining how I elicited recipients' beliefs in stage one, because recipients in stage one had to build expectations about stage two.

Figure 2.2: *Treatment Overview*



		number of subjects (recipients + dictators)		
		n=2 (2+2)	n= 8 (4+4)	n= 40 (20+20)
ex-post experimental feedback	individual decisions transparent	*baseline*	*transparency (4 dictators)*	*transparency (4 dictators)*
	Individual decisions opaque		*multi-agent (4 dictators)*	*multi-agent (20 dictators)*

Stage 1 and Stage 2 Recipients and Dictators Main Decisions: The allocation decisions were structured as follows: In the baseline treatment a dictator (she) had to decide how much money she wants to take out of a pot that contains $1.00. The recipient (he) earned the remaining amount. The dictator faced a trade-off between increasing her own payoff, thus being selfish, and being kind towards the recipient by increasing his payoff. In contrast, in the multi-agent and transparency treatment four (respectively 20) dictators and four (respectively 20) recipients interacted. Dictators decided how much money they would like to take out of a common pot that contained $4.00 ($20.00). They could take any amount up to $1.00. The remaining

[26] Subjects are given instructions using neutral language to abstract from potential framing effects. the multi-agent and transparency treatment four (respectively 20) dictators and four (respectively 20).

amount was equally split between all recipients in one group. An individual dictator's decision had no impact on other dictators' monetary payoffs. However, the recipients' payoffs were determined by all dictators within one group. Thus, the dictator faces the trade-off between being selfish and being kind towards a group of recipients. Nonetheless, making selfish decisions in settings with multiple agents eventually came accompanied by moral externalities in the form of a social image loss of the entire group of dictators.

Recipients remained passive during the entire experiment but were asked to make a guess about the average dictator's allocation. They received \$0.50 if their guess did not differ more than \$0.01 from the actual dictator's decision.[27] Before recipients' first order belief were revealed to the corresponding dictators, dictators in the baseline treatment were asked to make allocation decisions conditional on all possible first order beliefs rounded to 10 cents.[28] In the multi-agent and transparency treatments dictators decide contingent on the average beliefs of 20 recipients from the particular treatment condition (c.f. Ellingsen et al., 2010; Reuben et al., 2009).

Guessing the outcome of the game, the recipients were unaware that an approximated value of their guesses or respectively the average guess of all recipients within one group will later be revealed to matched dictator(s). Hence, instead of explicitly telling recipients that their beliefs will not be transmitted to dictators (c.f. Ellingsen et al., 2010) I did not make the omission salient to avoid that recipient suspect that dictators condition their decision on recipients' beliefs, and in response try to optimize their payoff by stating the payoff maximizing instead of their actual beliefs (c.f. Reuben et al., 2009).[29] Nonetheless, to avoid deception by omission recipients were asked for consent to reveal their guesses.[30] Self-selection seems not to be an issue in my experiment, since 299 out of 300 recipients or 99.67% agreed to transmit their guesses to the dictators.[31]

Stage 3 - Elicitation of Dictators' First Order Beliefs: To elicit dictators' first order beliefs about the distribution decision of other dictators in an incentivized way, dictators had to answer the following question: "We want you to guess how much (up to \$1.00) the other A-participants [dictators], on average, take out of pot if B-participants [recipients] expected that each A-participant [dictator] takes on average \$0.00, \$0.50 or \$1.00." Each participant whose sum of the guesses was not more than \$0.03 away from the true average amount received additional \$0.50 extra.[32]

[27] While many experimenters pay subjects for the accuracy of their guesses, there is mixed empirical evidence about whether monetary incentives change both the height of the stated beliefs. In a recent study Trautmann and van de Kuilen (2015) compared different incentivized revealed preference elicitation methods with non-incentivized introspection. They find found little evidence for economically significant improved performance of more complex elicitation methods over simpler ones or introspection regarding to the accuracy of beliefs as well as their external validity. In contrast, Huck and Weizsäcker (2002) find among others that incentives increase the accuracy of the stated beliefs.

[28] To avoid that dictators vindicate selfish-behavior by telling themselves that recipients bet on dictators' selfish behavior, I do not outline the incentives associated with the elicitation to dictators.

[29] Khalmetski et al. (2015) experimentally addressed this topic. They inquired whether subjects behave differently if they know that beneficiary of her decisions is aware of the beliefs' disclosure. They found that on average decision makers are more generous when beneficiaries are informed about the revelation of their beliefs.

[30] When recipients granted permission to reveal the data and were assigned the baseline condition, the dictator made their take decision contingent on the recipients first order-beliefs. When the recipient denied permission to reveal the data, subjects play a standard dictator game and the dictator self-reports their beliefs about the recipients guess after they have made their take decision. When the recipient granted permission to reveal the data and was assigned one of the multi-agent or transparency conditions, subjects played the multi-agent dictator game, as previously described. If the recipient denied permission to reveal the data, calculating the average guess of the recipients, their guess was interchanged with a random draw from the recipients who accepted to reveal their data.

[31] The belief of one recipient in the multi-agent treatment with 4 dictators who rejected to transmit his guess was not used to calculate the average recipients' first order belief. This implies that dictators later did not conditioned their distribution decision on a potential guess of this particular recipient.

[32] The choice of the elicitation procedure was motivated by comprehensibility. I surmise that belief hedging is no serious concern in my experiment, since a singles dictator's monetary payoffs are completely certain, because she entirely determined by herself and not by a strategic interaction between her and any other subjects.

Stage 4 - Ex-post Experimental Questionnaire: At the end of the experimental session, all subjects are asked to fill out a questionnaire covering sociodemographic questions. I generate survey data to measure altruism and reciprocity based on questions included in the Global Preference Survey (Falk et al., 2018). Dictators were asked how responsible they felt for the outcome of the experiment.

Stage 5 - Ex-post Experimental Feedback: To test if the transparency of dictators' actions drives the result that dictators act more pro-socially while interacting with a single agent, I varied the level of information recipients receive at the end of the game. While in the multi-agent treatment recipients only received feedback about the amount of their final payoff, recipients in the transparency treatment received information on the distribution of individual dictators' decisions. A single dictator might be less likely to experience guilt from blame when betraying recipients' expectations concerning the height of recipients' share, since it is ambiguous as to who should be blamed for the final distribution.

Further Remarks: One major concern regarding the data quality in online experiments is that MTurkers may be less attentive, since they might get distracted by multi-tasking or available outside opportunities (Chandler et al., 2014).[33] In order to mitigate attention concerns, I restricted my MTurk samples to only high-reputation workers to mitigate potential attention concerns. Furthermore, all subjects had to answer comprehension questions about hypothetical outcomes of the experiments. Subjects could only proceed with the experiment and thus make payoff relevant decisions if they entered the correct answers in free-form fields. This particular elicitation method that humans and not bots participated in the experiment. In addition, the experiment was programmed and implemented on MTurk in such a way that every registered MTurk worker could only participate once. Therefore, to participate multiple times workers needed multiple accounts. Having stated this, MTurk provides their own financial incentives to prevent users from having multiple accounts. In particular, they use terms-of-use agreements and technical approaches to prevent multiple accounts (Horton et al., 2011). Beyond that, I checked whether two MTurkers with distinct MTurk-IDs, but equal IP-addresses participated in the experiment, since this might indicate that one individual actually participated twice. Fortunately, no duplicated IP-addresses were found.

3.2 Design Choice Rationales and Identification Strategies

In this subsection I outline the rationales behind the major design choices and discuss why they allow to (causally) identify treatment effects.

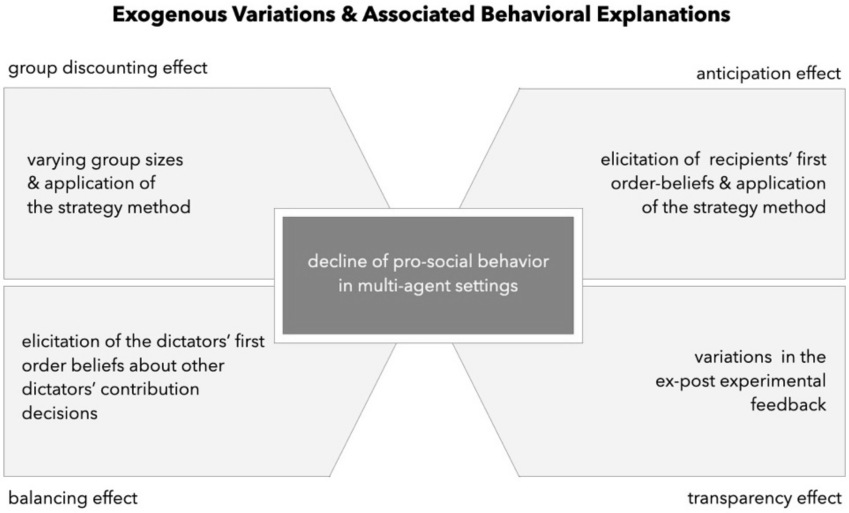
Rational for Using a Dictator Game: Numerous recent studies use dictator games to examine whether guilt aversion can explain pro-social behavior (Dufwenberg & Gneezy, 2000; Ellingsen et al., 2010; Ghidoni & Ploner, 2015; Khalmetski et al., 2015; Morell, 2017; Ockenfels & Werner, 2014). The appeal of dictator games lies in their non-strategic nature which allows to rule out kindness-based reciprocity (e.g., Dufwenberg & Kirchsteiger, 2004; Rabin, 1993), strategic risk or other confounding strategical considerations as potential explanations for participants' behavior. Traditionally, they have been used to calibrate other-regarding preference models (e.g., Blanco et al., 2011; Fehr & Schmidt, 1999). However, more recent research conjectured alternatively that pro-social behavior in dictator games is motivated by the dictators' desire to act in a socially appropriate manner in order to avoid feelings of guilt and shame (Dufwenberg &

[33] Hauser and Schwarz (2016) conducted three distinct studies to test whether MTurkers are more or less attentive to the instructions than college students. They find that MTurkers reacted more likely in response to text manipulations.

Gneezy, 2000; Ellingsen et al., 2010; Ghidoni & Ploner, 2015; Khalmetski et al., 2015; Morell, 2017; Ockenfels & Werner, 2014).

Rationale for Using a Take-Framework: The take-framework choice is substantiated by a theoretical link between the experience of guilty feelings and social norm violation. According to Baumeister et al. (1994) the prototypical causes of guilt are usually actions that are perceived as social norm violation, such as the infliction of harm, loss or distress on others. Consequently, avoiding guilt can be interpreted as a preference for not violating the social norm to fulfill the justifiable and legitimate expectations of others. To this end, feelings of guilt can be among others considered the moral consequence one draws from the anticipation of other people's loss aversion.[34] Thereby, concepts of guilt complement prospect theory (Kahneman & Tversky, 1979). The take-framework reinforces the perception that dictator inflicts loss to the recipients and incorporates the prospect theory perspective on guilt. Hence, high withdrawals are more likely perceived as a norm violation and may trigger feelings of guilt and shame.[35] **Overview of Exogenous Variations Utilized to Identify Effects:** To test the proposed group size effect, the number of interacting dictators and recipients in a dictator game with a take-framework was exogenously varied. To further disentangle the previously discussed potential reasons for the proposed decline of pro-social behavior (the discounting effect, the transparency effect, the balancing effect as well as the anticipation effect), I exogenously vary the dictators' second order beliefs, the ex-post experimental feedback and elicit the first order beliefs of recipients as well as the first order beliefs of dictators regarding the behavior of the average dictator. An overview as to which experimental approach is chosen in order to disentangle the different behavioral explanations is provided in Figure 2.3.

Figure 2.3: *Exogenous variations, elicited variables and associated behavioral explanations*


Exogenous Variations & Associated Behavioral Explanations
group discounting effect
varying group sizes & application of the strategy method
anticipation effect
elicitation of recipients' first order-beliefs & application of the strategy method
decline of pro-social behavior in multi-agent settings
elicitation of the dictators' first order beliefs about other dictators' contribution decisions
variations in the ex-post experimental feedback
balancing effect
transparency effect

[34] For an overview on the literature of loss-aversion see among others Barberis (2013) and Köszegi and Rabin (2006).

[35] Notably, while Bardsley (2008), Korenok et al. (2013, 2018), Krupka and Weber (2013) , List (2007), and Visser and Roelofs (2011) find evidence in favor of a framing sensitivity, Dreber et al. (2013) conclude from their experimental study that dictators conception of the games and the norms that govern them are not easily malleable by changes in labels.

Rationale for the Exogenous Variation of the Group Size: In order to test an overall group size effect as well as the group discounting effect, I designed small group treatments including 8 subjects and large group treatments including 40 subjects. Social psychology papers which investigate small group phenomena regularly define small groups as groups with 3 to 20 members. Hence, the size of a group within the small group treatment is in this particular range, while the size of the large group treatment is significantly larger. Small groups differ from larger groups in several aspects. Decisions made in smaller groups are usually perceived as more pivotal. Social cohesion, mutual trust and group commitment, on average, is stronger in smaller groups (Carron & Kevin, 1996; Ellemers et al., 1999). Moreover, factors such as social closeness may foster the perception of guilt (Morell, 2017).

Varying the number of recipients and dictators by the same extent allows one to study the concept of guilt in multi-agent settings while abstracting from a wide range of distributional social preferences.[36] More precisely, neither the model of Ockenfels and Bolton (Ockenfels et al., 2000) nor the model of Fehr and Schmidt (1999), or any other linear utility model that incorporates concerns about distribution would predict a treatment effect.[37] I also varied whether people in the baseline condition received the information that in total 4 or 20 subjects participated in that session of the experiment. This design choice is substantiated by the idea that if the (communicated) number of subjects within one session remain constant, even non-linear models of inequity aversion regularly predict no difference between treatments, since the number of subjects in the baseline and the multi-agent or respectively the transparency treatments are comparable.

Rationale for the Exogenous Variation of Dictators' Second Order Beliefs: I exogenously varied the dictators' second order beliefs to establish a causal effect between dictators' decisions and the recipients first order beliefs. In particular, applying the strategy method (Selten, 1965) dictators made their decision contingent on recipients' stated guesses concerning the outcome of the game (cf. Khalmetski et al. 2015).[38] Under the assumptions that the recipient had revealed his true first order beliefs as a guess and the dictator anticipated the correctness of beliefs, the dictator has correct and unbiased second order beliefs. Hence, dictators' second order beliefs and stated prosocial actions are exogenous by design.[39] Its application comes with six major advantages. An exogenous variation allows to draw causal inference on the relationship

[36] For studies that examines dictator games with multiple dictators and on recipient and with one dictator and multiple recipients see among other Andreoni (2007), Schumacher et al. (2017) and their cited papers.

[37] In distributional preference models only the outcome but not who is responsible for the outcome determines utility of an agent. Moreover, intuitively speaking, the Bolton-Ockenfels (Ockenfels et al., 2000) model states that dictators have preferences over the share of the total wealth a dictator receives. Hence, increasing the total wealth proportional to the number of participants within one group does not change dictators' optimization problems. In contrast, the underlying assumptions of the Fehr-Schmidt utility model concerning their preference parameter β causes that no dictator should be willing in a setting with subjects to give.

[38] Applying the strategy method, Balafoutas and Fornwagner (2017), Bellemare, Sebald, and Strobel (2011), Hauge (2016), Morell (2017) and Reuben et al. (2009) find evidence in favor of the guilt model. Contrary, Ellingsen et al. (2010), Chang et al. (2011) and Kawagoe and Narita (2014)found conflicting results by revealing co-players' first order beliefs. The data generated by the direct response method indicate a weaker impact on guilt on pro-social behavior than the application of the strategy method suggests. Bellemare, Sebald, and Suetens (2017) found that the direct elicitation as well as the strategy method produces similar results, while the direct response method produces higher levels of unconditional pro-social behavior. They conjectured that it might explicit communication of the belief reduces emotional distance and foster altruistic behavior. Alternatively, I argue that if only beliefs below a certain threshold value trigger an alteration of behavior, there will be no correlation between second order and dictators' decisions when recipients state beliefs above the threshold. Thus, if the strategy method is designed so that dictators have to state their decision contingent on potential beliefs beyond the threshold, it produces results in favor of guilt, while the direct response method rejects guilt as an explanation for pro-social behavior.

[39] However, in line with Ellingsen et al. (2010), applying the strategy method Khalmetski et al. (2015) do not find evidence for a significant effect between pro-social behavior and first order beliefs. Nonetheless, performing sensitivity analysis, they find evidence that a large proportion of co-players will act more pro-socially if the reference beliefs are higher.

between second order beliefs and pro-sociality.[40] Moreover, it enables me to separate anticipation effects (in form of higher first order beliefs) from primary effects. In addition, it allows to generate observations on how dictators react to rarely stated beliefs. Furthermore, it allows me to account for heterogeneity among actors regarding reactions to differences in second order beliefs in my empirical analysis. Beyond this, applying the strategy method allows me to identify a potential non-linear relationship between different levels of expectations and pro-social behavior. Finally, generating various observations per subject increases the power of the applied statistical tests.

Rational for the Exogenous Variation of Ex-Post Experimental Feedback: the ex-post experimental feedback about the dictators' behavior is modified to test whether the ability to be blamed for anti-social decision with restrictions to the anonymity of the subjects leads to more pro-social behavior.

Rational for the Elicitation of Dictators' First Order Beliefs: Finally, I measured the dictators' beliefs about other co-dictators' take decisions to assess whether dictators morally free ride on pro-social behavior or simply mimic the behavior of other dictators. All findings referring to social norms or moral free-riding effects constitute only correlational evidence, because this measure is based on a self-reported question item.

4 Results and Discussion

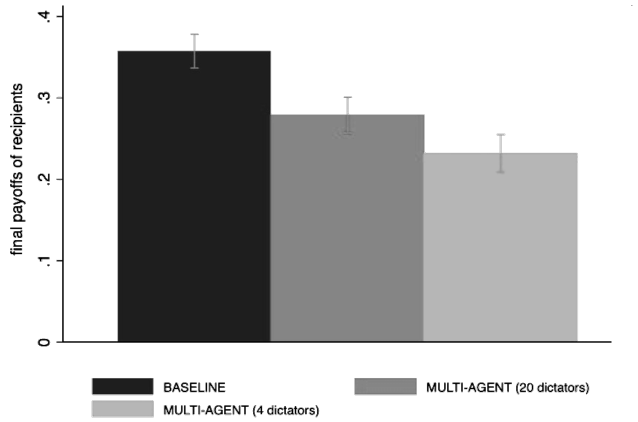
4.1 Overall Group Size Effect

This paper first tests for a significant overall group size effect (the negative impact of group size on dictators' prosocial behavior) and then continues by investigating potential underlying behavioral explanations (the group discounting, the transparency or attributability, the balancing, and the anticipation effect). Across all five treatments, data on conditional transfers of 300 dictators and 11 different first order beliefs have been collected, yielding 3300 observations on which my empirical assessments rest upon. The average age was about 32 years; approximately as many males as females participated (share of females ~ 48%.). The design choice to ask the recipients for their permission to reveal their first order beliefs leads to no pivotal self-selection bias due to this high transmission rate of over 99%.

Considering all treatments and potential recipients' beliefs, dictators on average distributed 29.81% of the original endowment of $1. In the baseline treatment, an average sum of $0.36 was distributed to recipients—57% more than in the multi-agent treatment with four dictators ($0.23) and 29% more than in the multi-agent treatment with 20 dictators ($0.28). Figure 2.4 depicts recipients' mean payoffs and the corresponding 90%-confidence intervals.

[40] The most apparent method to test guilt aversion is the direct elicitation of second order beliefs (Bacharach et al., 2007; Bracht & Regner, 2013; Charness & Dufwenberg, 2006; Dufwenberg & Gneezy, 2000; Guerra & John Zizzo, 2004) This elicitation method is related to problems with gathering unbiased data, specifically concerning to what extent the beliefs are subject to a consensus effect (Ross et al., 1977). There is empirical support that the presumed endogeneity effect leads to an overestimation of the impact of guilt on decision making (Ellingsen et al., 2010; Khalmetski et al., 2015)

Figure 2.4: *Treatment Effects (Average Recipient's Payoff with 90%-confidence Intervals by Treatments)*



The distribution between the baseline and the multi-agent treatment with four dictators was significantly different (MWU-test:[41] p < 0.001). The difference in the amount taken was between the baseline and the multi-agent treatment with 20 dictators significant (MWU-test: p<0.001). In the baseline treatment, 80% of the dictators transferred at least $0.01, while in the multi-agent treatment with four dictators 53% and in the multi-agent treatment with 20 dictators 68% did so.[42] Hence, I find ample evidence for the decline of pro-social behavior in multi-agent dictator games.

Result 1: *Dictators take an up to 59 % higher amount of money in the baseline condition than in the multi-agent condition. More dictators stated selfish decision in multi-agent treatments. Thus, I find support for Hypothesis 1.*

The recipients' shares in the baseline condition are only slightly higher than in previously conducted experiments in which average dictators kept between 70% and 90% of their initial endowment (e.g., Camerer, 2003; Engel, 2011). It is likely that the experiment's take framework triggered dictators to behave more morally inclined (see e.g., Bardsley, 2008; Engel, 2011). Overall, it does not appear that serious concerns about missing reliability of experiments with low stake sizes materializes in my setting. This result is in line with the findings of Amir et al. (2012) and Raihani et al. (2013) who find no significant stake size effect in distribution games.[43]

Robustness Checks: The treatment effect did unlikely result from imbalances in participants' characteristics in distinct treatments, because the treatment subsamples are balanced with regard to their observed prognostic variables (see appendix A.1). To further assess whether the previous results are robust, I run a random-effects linear regression[44]- including control variables for gender, age, familiarity with

[41] A MWU-test is a test of the null hypothesis that it is equally likely that a randomly selected value from the baseline treatment will be less than or greater than a randomly selected value from the multi-agent treatment with four dictators.

[42] The difference in the probability to distribute a minimum amount of $0.01 between the baseline and the multi-agent treatment with four dictators is significant at a 5%-level according to a Fisher-exact test, while the difference between the baseline and the multi-agent treatment with 20 dictators and the multi-agent treatment with four dictators and 20 dictators are not.

[43] Notably, there also exist contradicting studies, such as the study by Engel (2011).

[44] I rely on a random-effects linear model instead of a fixed-effects linear model, since it does not bias the results away from the result of the fixed-effect model. Using a fixed-effects model, some of my main effect (variation in ex-post information structure and group size) would drop out because of a lack of within-subject variance. The purpose of the random effect model is to visualize all effects in a single model. Moreover, the Hausman test is not significant for all random-effects linear regression models, i.e., all coefficients in the linear random effects model are almost identical to those of the linear fixed-effects model.

experiments, the perception of responsibility[45] as well as two survey items measuring altruism. Moreover, to account for plausible corner solutions (accumulations of observations of the dictator's monetary payoff variable at $0.00 and $1.00) as well as the panel structure of the data set, I calculate a random-effects Tobit model. Conducting both regressions, the difference between the baseline and the multi-agent treatment remained significant at a 5%- level (*see* treatment coefficients in Table 2.1 Model 3 and 4). Overall, I find further support for hypothesis 1, since dictators take significantly less in the baseline than in both multi-agent treatments.[46]

Table 2.1: *Impact of Revealed Recipients First–Order Beliefs on Dictator's Distribution Decision*

	Model (1) Random Effects	Model (2) Random Effects Tobit (AME)	Model (3) Random Effects	Model (4) Random Effects Tobit (AME)	Model (5) Random Effects	Model (6) Random Effects Tobit (AME)
			Dependent variable: ***dictator's monetary payoff***			
Belief	.080 (.009)***	.073 (.009)***	.081 (.010)***	.074 (.009)***	.081 (.010)***	.068 (.009)***
Multi–agent (4 dictators)			.131 (.041)***	.079 (.021)***		
Multi–agent (20 dictators)			.084 (.041)**	. 042 (.029)**		
Transparency (4 dictators)			.058 (.041)***	.035 (.021)*		
Transparency (20 dictators)			.076 (.041)*	.047 (.021)***		
Dictators in Session					–.0005 (.002)	–0.001 (.002)
Altruism 1			–.0004 (.0001)***	–.080 (.023)***	–.0004 (.0001)***	–.0046 (.0001)***
Altruism 2			–.018 (.0046)***	–.280 (.065)***	–0.017 (.005)***	–.023 (.006)***
Familiarity with experiments			–.015 (0.045)	–.004 (.011)	–.0255 (.046)	.037 (.058)
Age			.001 (.010)	.145 (.206)	.002 (.001)	.0022 (.001)
Female			−.006 (.027)	–0.24 (.032)	−.012 (.027)	–.032 (.033)
Responsibility			.004 (.005)	–.088 (.080)	.003 (.005)	.004 (.006)
Constant	.662 (.0147)***		.651 (.051)***		.731 (.056)***	
Number of Obs.	3300	3300	3300	3300	3300	3300

Notes: The significance level is indicated by ***p = .01, ** p = .05. * p = .1. AME = Average marginal effects. The dependent variable "dictator's monetary payoff" indicates the amount of money kept by the dictator. Random effect Tobit models are calculated to account for the share of observations with amount kept of $0 or $1 (double–censored Tobit regression) and to account for the panel structure of the data set. All reported coefficients are (average) marginal effects. (Robust) standard errors in parentheses. Altruism 1" is a survey item asking: "imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity? To elicit the "Altruism 2" we asked: "How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10".

[45] The perception of responsibility had no influence on the distribution decisions (*see* Table 2.1). This is likely the case, because the behavioral channels derived from the proposed guilt aversion model do explain the diffusion effect more precisely than the vaguer feeling of responsibility.

[46] To test the alternative theory that changes in the mere number of participants in one session of the experiment cause the treatment effect, I regress the number of subjects mentioned in the experimental instructions as well as the revealed beliefs on dictators' decision (*see* Table 2.1). It follows that not changes in the group size per se, but dictators' joint determination of recipient's payoffs induces the treatment effect. As a robustness check I also run all regressions including only observations associated with reasonable beliefs above $0.5. The described treatments remain significant at a 5%-level.

Recall that the four proposed behavioral explanations for the decline of moral behavior in multi-agent settings are derived from guilt models. To assess whether behavioral patterns emerging from the concept of guilt can be found in the experimental data, I ran a random effect and a random effect Tobit model (*see* Table 2.1, Model 1 and 2). As a matter of fact, revealed first order beliefs had a significant and direct effect on the payoff of recipients as predicted by the standard guilt model. Including control variables, the qualitative results were robust (*see* Table 2.1, Model 3 and 4). In the appendix A.2 I present a comprehensive discussion of statistical results about potentially non-linear relationships between revealed recipients' first order beliefs and dictators' choices.

4.2 Group Discounting Effect

In this subsection I inquire into whether the perception of guilt might be determined by how agents weighted the harm or loss they inflict to a group of people compared to a single interaction partner. I hypothesized in the previous section that dictators showed more selfish behavior the larger the group size is. Contrary, the tendency to act more selfishly was higher in the multi-agent setting with four dictators (recipients' average payoff = \$0.23; dictators' average payoff = \$0.77) than with 20 dictators (recipients' average payoff = \$0.28; dictators' average payoff = \$0.72). The ranks of the treatments significantly differ (MWU-test: $p = 0.002$).

However, controlling for demographics, the effects is not or only slightly significantly (t-test for the equivalence of parameters in Table 2.1, Model 3: $p = 0.24$; Model 4: $p = 0.09$, two-tailed). Controlling in addition for dictators' beliefs about other dictators' behavior, the difference between the dictator's share in the 20 dictator multi-agent treatment and the 4 multi-agent treatment becomes non-significant (*see* Table 2.3, Model 3 and 4; t-test: $p = 0.97$, respectively $p = 0.67$, two-tailed). This corroborates the empirical findings of Schumacher et al. (2017) who found that facing a trade-off between behaving pro-socially and increasing one's own payoff, decision makers attach the same weight to small as to large groups.

Result 2: Controlling for demographic variables and dictator's first order beliefs, *there is no significant difference between the amount taken by dictators in the multi-agent treatment with 4 and 20 dictators. Thus, the data do not support hypothesis 3.*

This second finding supports the experimental results of Carpenter (2007) who finds that agents in large groups contribute to a public good at rates no lower than members of small groups. Like in my study, the difference between small groups and large groups in pro-social distribution decisions was not statistically significant. However, Carpenter (2007) discovered that hindrances in monitoring players do reduce the provision to a public good, which supports that a relaxation of attributability is one of the main factors explaining the decline of pro-social behavior in the multi-agent settings.

4.3 Transparency Effect

This paper proceeds by assessing whether the extent to which individual actions are attributable impact pro-social behavior. Thereby, it investigates whether the concept of simple guilt or guilt from blame more precisely captures dictators' tendency to condition their decisions on recipients' beliefs and whether dictators exploited this moral wiggle room (Dana et al., 2006) in form of the relaxed attributability of individual decisions in multi-agent contexts. Indeed, the ex-post revelation of the decisions' distribution had a statistically and economically significant effect on dictators' decisions in the transparency treatment with four dictators (MWU-test: $p < 0.001$).

The random-effects Tobit model (*see* Table 2.1, Model 4) revealed that controlling for individual characteristics, dictators' payoffs decreased on average by 55.7% if the distribution of dictators' payoffs in a group of four dictators was revealed. The random effects model corroborates the qualitative result (*see* Table 2.1, Model 3). The result supports the findings of Dana et al. (Dana et al., 2006, 2007) who discovered that if recipients cannot determine if and to which extent an individual dictator is responsible for an unexpected low final payoff, dictators state more selfish decisions. Hence, I conclude that decision makers likely exploit moral wiggle rooms in small group settings.

Result 3: *The transparency of decisions in small-group multi-agent settings leads to a significant increase in pro-social behavior. Hence, I find strong support for hypothesis 5 in small group settings.*

Table 2.2: *Average Dictator's Distribution Decision by Dictator Types*

	Model (1)	Model (2)	Model (3)	Model (4)
	Sample A	Sample A	Sample B	Sample B
	Random	Random Effects	Random	Random Effects
	Effects	Tobit (AME)	Effects	Tobit (AME)
	Dependent variable: ***dictator's monetary payoff***			
Revealed Belief	0.212	.169	−.0486	−.0511
	(.009)***	(.010)***	(.024***)	(.027)***
Multi–agent (4 dictators)	.193	.190	−.0424	−.017
	(.047)***	(.029)***	(.065)	(.065)
Multi–agent (20 dictators)	.122	.121	−0.022	−.008
	(.046)***	(.037)***	(.063)	(.063)
Transparency (4 dictators)	.088	.093	−.070	−.058
	(.045)**	(.040)**	(.074)	(.075)
Transparency (20 dictators)	.112	.118	−.122	−.108
	(.045)**	(.038)**	(.075)	(.076)
Altruism 1	−.0005	−.0004	−.0002	−.0002
	(.0001)***	(.0001)***	(.0001)	(.0002)
Altruism 2	−.0203	−.027	−.005	−.007
	(.005)***	(.006)***	(.008)	(.008)
Familiarity with experiments	−.043	−.049	.150	.152
	(.048)	(.056)	(.104)	(.094)
Age	.003	.004	.001	.0003
	(.001)**	(002)**	(.002)	(.002)
Female	.010	−.016	-.070	-.075
	(.030)	(0.317)	(.044)	(.044)
Responsibility	.008	.009	-.006	-.005
	(.005)	(.005)*	(.080)	(.008)
Constant	.521		.937	
	(.070)***		(.086)***	
Number of Obs.	2662	2662	638	638

Notes: *The significance level is indicated by ***p = .01, ** p = .05. * p = .1. Altruism 1" is a survey item asking: "imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity? To elicit the "Altruism 2" we asked: "How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10". The dependent variable "dictator's monetary payoff" indicates the amount of money kept by the dictator. Random effect Tobit models are calculated to account for the share of observations with amount kept of $0 or $1 (double-censored Tobit regression) and to account for the panel structure of the data set. All reported coefficients are (average) marginal effects. (Robust) standard errors in parentheses.*

Eventually, I tested whether the treatment effect is also present in large group settings. While I find a significant transparency effect in small group settings (t-test based on the coefficients in Model 3 and 4, Table 2.1: $p = 0.086$; respectively $p = 0.047$; two-tailed), the multi-agent and transparency treatment with 20 dictators reveals that there is no significant similar effect in large group settings (t-test based on the coefficients presented in Model 3 and 4, Table 2.1: $p = 0.85$; respectively $p = 0.81$, two-tailed). A non-parametric MWU-test corroborates the result for the impact of transparency on distribution decisions in large group settings (MWU-test: $p = 0.30$). Hence, in large groups settings undeceiving recipients by revealing the distribution of dictators' decisions proves insufficient for dictators to assume that their decision can be attributed to them. In contrast, the attributability of decisions in small group settings are likely perceived as stronger.

Further Extensions: To examine whether dictators who show guilt-averse behavior (dictators with a positive within-correlation) react systematically different to changes in the ex-post information structure than strictly surprise-seeking dictators (dictators with a strictly negative within correlation) I split the sample into two subsamples. The first subsample (subsample A) comprises dictators having a within correlation of $r \leq 0$. The second sample (subsample B) includes all dictators having a within correlation $r > 0$. I compute random effects models as well as a random effects Tobits model comprising data from subsample A (*see* Table 2.2, Model 1 and 2) and from subsample B (*see* Table 2.2, Model 3 and 4). The dependent variable in the for models is dictators' monetary payoff. The models comprise the same control variables as the regressions in Table 2.1. Comparing the coefficients of the dummy variables "Transparency (4 dictators)" and "Transparency (20 dictators)" between both subsamples, I find that guilt-averse dictators kept with significantly less if their actions were observable (*see* Table 2.2, Model 1 and 2). Surprise-seeking dictators' payoffs were not significantly affected by the observability of their decisions (*see* Table 2.2, Model 3 and 4). Hence, dictators might be loss-averse concerning one's own social image, i.e., losses in the social image dimension loom larger than gains.

4.4 Balancing Effect

In this subsection I examine how dictators expect other dictators to behave and how co-dictators' expected decision impact dictators' individual distribution decisions. Dictators predicted that –across all treatments– co-dictators' take an average \$0.57 (actual value \$0.72) if recipients expect that dictators take on average \$ 0.00, \$0.62 (actual value \$0.69) if recipients expect on average \$ 0.50 and \$0.76 (actual value \$0.76) if recipients expect on average \$ 1.00. Comparing these numbers, I conclude that dictators are excessively confident regarding the pro-social behavior of other dictators. It follows that dictators underestimate co-dictators' selfishness if recipients' revealed first order beliefs were equal to \$ 0.00 or \$ 0.50. Figure 2.5 shows that the average share kept by dictators second order statistically dominates dictators' average first order beliefs.

The first order beliefs about other dictators' decisions differed significantly between treatments. Dictators correctly predicted that there is a decline of pro-social behavior between the baseline and pooled the multi-agent treatment (MWU-test: $p < 0.017$). Moreover, they correctly predicted that dictators' act less selfishly if the distribution of decisions in the multi-agent setting with four dictators is revealed, though only slightly significantly according to a t-test ($p = 0.08$), and not significantly according to an (MWU-test: $p = 0.11$). In addition, they predicted a transparency effect in treatments with 20 dictators (MWU-test: $p = 0.013$), which, in fact, did not exist

.

Figure 2.5: *Cumulative Distribution Functions of Dictator's Average Distribution Decision and Average Dictator's First Order Beliefs regarding other Dictators' Distribution Decisions*

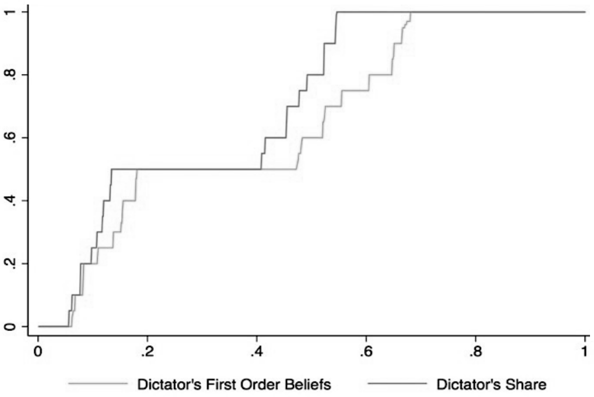


Table 2.3: *Impact of Revealed Recipients First–Order Beliefs on Dictator's Distribution Decision*

	Model (1) Random Effects	Model (2) Random Effects Tobit	Model (3) Random Effects	Model (4) Random Effects Tobit
		Dependent variable: ***dictator's monetary payoff***		
Revealed	.081	.077	−.007	−.010
Belief	(.020)***	(.019)***	(.019)	(.020)
Dictators' first order beliefs	–	–	.461***	.423***
			(.029)	(.032)
Multi–agent	.129	.148	−.072	.101
(4 dictators)	(.041)***	(.036)***	(.035)**	(.033)***
Multi–agent	.091	.104	.074	.084
(20 dictators)	(.041)**	(.039)**	(.034)**	(.034)**
Transparency	.053	.074	.032	.043
(4 dictators)	(.041)	(.041)**	(.035)	(.037)
Transparency	.081	.105	.462	.060
(20 dictators)	(.041)*	(.039)**	(.039)***	(.035)
Altruism 1	−.0003	−.0004	−.0002	−.0003
	(.0001)***	(.0001)***	(.0001)***	(.0001)***
Altruism 2	–.019	−.024	−.016	–0.021
	(0.0047)***	(.005)***	(.039)	(.005)***
Familiarity with experiments	–0.011	−.017	.017	–0.032
	(.045)	(.052)	(.038)	(.047)
Age	.002	.002	.001	.001
	(.001)	(001)**	(.001)	(.003)
Female	–.006	−.019	–.011	–.009
	(.0.27)	(.030)	(.023)	(.025)
Responsibility	.007	.008	–.010	.009
	(.005)	(.005)	(.004)*	(.004)
Constant	.613***		.420	
	(.073)		(.052)***	
Number of Obs.	900	900	900	900

Notes: *Altruism 1" is a survey item asking: "imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity? To elicit the "Altruism 2" we asked: "How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10". The dependent variable "dictator's monetary payoff" indicates the amount of money kept by the dictator. Random effect Tobit models are calculated to account for the share of observations with amount kept of $0 or $1 (double–censored Tobit regression) and to account for the panel structure of the data set. All reported coefficients are (average) marginal effects. (Robust) standard errors in parentheses. The significance level is indicated by ***p = .01, ** p = .05. * p = .1.*

In a further step I assess whether there is a negative correlation between the amount taken and dictators' first order beliefs. Therefore, I calculated a random-effects and a random-effects Tobit model that comprises in addition to the in the other models considered variables, the dictators' first order beliefs about the co-dictators' decision conditional on the assumption that recipients expected that dictators kept \$0.00, \$0.50 or \$1.00 (*see* Table 2.2, Model 3 and 4). Model 1 and 2 in Table 2.2 is based on the same subsample, but do not comprise the variable "dictator's first order beliefs. Pursuant to the presumed effect, according to a random-effects linear regression as well as a random-effects Tobit regression (*see* Table 2.3 Model 3 and 4) there is an economically significant (dictator's share increases by about \$0.47 if they expect others to take \$1.00 more) as well as statistically significant positive correlation between dictators' first order beliefs about the monetary payoffs of others and their own monetary payoffs.

Result 4: *Dictators do not free-ride on the moral behavior of others, but rather mimic the behavior of other dictators. This effect is enhanced by overly confident first order beliefs about co-dictators' behavior. Hence, I find no support for hypothesis 2.*

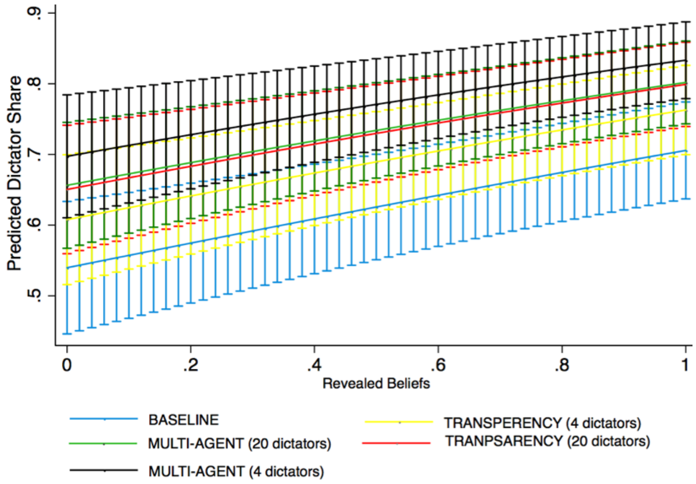
Consequently, it seems that in my experiment the preference for norm compliance (Krupka and Weber, 2013) seem to better capture the behavioral patterns than the moral free-riding explanation and dictators have a desire to not stand out of the crowd. Nonetheless, I only find a correlational and not necessarily a causal relationship between the dictators' take decision and their expectation about other dictators' decisions. It is possible that the correlation between the amount taken and the first order beliefs would have been lower, if I could abstract from a consensus effect by exogenously manipulating dictators first order beliefs. Thus, it remains to future research to study this topic and establish a causal relationship between dictators' actual behavior and their beliefs about other dictators' decisions.

4.5 Anticipation Effect

The theoretical guilt models derived in this paper predict that the higher the revealed reasonable beliefs of recipients' the higher are dictators' second order beliefs and, consequently, the higher are dictators' shares. In addition, recipients likely foresee the decline in pro-sociality in settings with multiple dictators. Hence, it may hold that their anticipation fosters the already established group size effect. While applying the strategy method allowed me to abstract from potential anticipation effects in the previous analysis steps, I now explicitly predict the height of potential anticipation effects. Therefore, I first test whether the recipients' first order beliefs differ between treatments. The histograms of the recipients' first order beliefs collected in Figure 2.6 provides a first graphical indication that recipients expect dictators to be least selfish in the standard dictator setting and most selfish in the multi-agent settings. Indeed, recipients guessed that in the baseline treatment dictators take on average \$0.66, \$0.74 (MWU-test: p = 0.04) in the multi-agent treatment with four dictators as well as \$0.74 in the multi-agent treatment with respectively 20 dictators (MWU-test: p = 0.089).

Result 5: *Recipients expect that dictators claim more in the transparency, as well as in the multi-agent treatments. Hence, hypothesis 4 can be confirmed.*

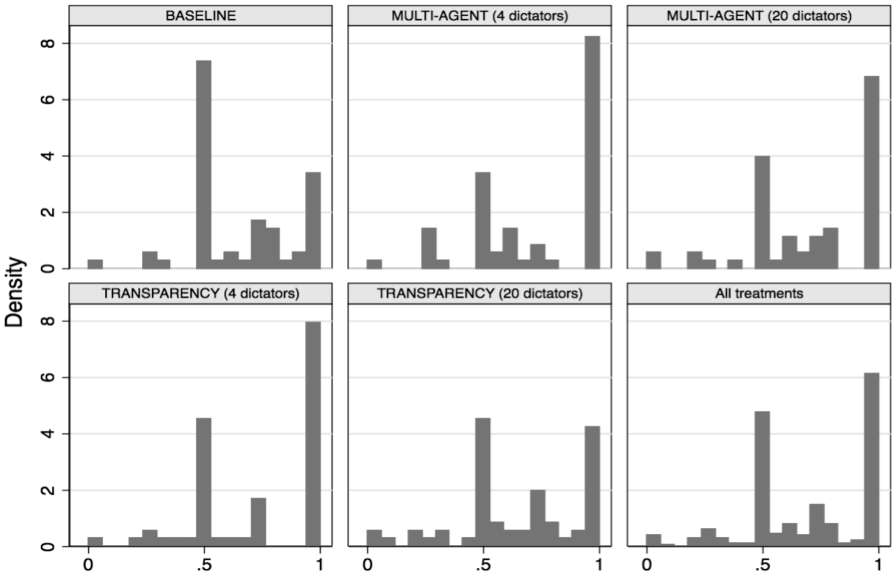
Figure 2.6: *Recipients' First Order Beliefs about the Average Share of a Single Dictator*



Next, I test whether the differences in beliefs actually cause an overall increase in selfish behavior in the multi-agent treatment. Therefore, I predicted the amount taken in the multi-agent treatments given the actual average beliefs as well as the average beliefs in the baseline condition and compare the prediction results. Figure 2.7 shows all ceteris paribus predictions of dictators' average share if beliefs are varied. According to the random-effects linear model, the difference in predicted dictator share in the multi-agent treatment with four dictators based on the average beliefs in the baseline condition and the predicted dictator share based on the average beliefs in the multi-agent condition is with a difference of $0.02 statistically significant at any conventional significance level (*see* Figure 2.7). Similar conclusions can be derived from predictions of the Tobit regression model.

Result 6: *There exists no significant anticipation effect that explains the decline of pro-sociality in the multi-agent and the transparency conditions.*

Figure 2.7*: Predictions of Dictator's Share Conditioned on Different First Order Beliefs*



4.6 Concluding Remarks on the Experimental Results

It remains to analyze to what extent the single as relevant identified behavioral explanations contribute to the overall treatment effect–the difference in the amount taken by dictators between the baseline condition and the 4 dictator multi-agent condition. The statistics presented in this subsection rest upon the regressions presented in Table 2.3. Taking into account that the average revealed belief in the baseline treatment is \$0.65 and in the multiagent treatment with four dictators is \$0.73, the Tobit model presented in column 2 (*see* Table 2.1) predicts that dictators share in the baseline treatment is \$0.65 (95%-confidence interval = [0.58,0.71]) and in the multi-agent treatment \$0.81 (95%-confidence interval = [0.76,0.86], thus the overall difference adds up to \$0.16.

Relative impact of the anticipation effect: Now, assume otherwise. If recipients' in the multi-agent setting with four dictators would have revealed the same beliefs as in the baseline treatment, the dictators' share would on average only decrease by \$0.01 to \$0.80 (95%-confidence interval = [0.75,0.85]). The anticipation effect is not significant and only accounts for 7% of the difference between the two treatments.

Relative impact of the transparency effect: Furthermore, the multi-agent treatment is not only different from the baseline treatment in how payoffs are determined, but individual decisions can also not be attributed to individual dictators precisely. It is for this reason that given recipients' revealed first order beliefs from the baseline treatment the ex-post revelation of the distribution of dictators' distribution decision lead according to a point prediction derived from the Tobit model to \$0.06 higher recipients' shares in the transparency condition than in the multi-agent with four dictators (average dictator share = \$0.75; 95%-confidence interval = [0.72,0.77]). The transparency effect thus accounts for 56% of the difference between two-agent and multi-agent setting with four dictators.

Relative impact of other factors: Overall, I conclude that controlling for dictators' first order beliefs 56% of the difference between two-agent and multi-agent settings with four dictators and 4 recipients can be put down to changes in the attributability of decisions and about 7% might be due to anticipation effects. The remaining difference (about 37%) is likely explained by the structural form of the dictators' utility function–dictators weigh the harm of an individual recipient less if they interact with multiple individuals– and changes in the dictators' first order beliefs regarding other dictators' decisions.

5 Conclusion and Policy Implications

In this paper I introduced two multi-agent guilt models and presented a novel experimental design that allows for the evaluation of the impact of guilt and blame on decisions in both small group as well as large group settings. It enabled me to not only inquire whether a group size effect per se exists, but also to discriminate between four different behavioral explanations as to why economic agents experience guilt and shame in multi-agent settings less severely.

Overall, I find a significant decline of pro-social behavior in multi-agent settings. About 56% of the difference between the multi-agent setting with 4 decision makers and the baseline treatment can be explained by changes in the attributability of actions in multi-agent settings. 7% of the difference are related to anticipation effects. Hence, the anticipation effect is neither economically nor statistically significant. The remaining 37% could likely be explained by agents who account the dis-utility of an individual in a multi-agent setting to a lesser extent and further disturbance factors. I find no evidence that decision makers free ride on others moral and try to offset their immoral decisions. Contrary, I established the effect that individuals comply with their expectations about the behavior of others.

My empirical results have versatile theoretical implications: First, the inquiry into how the inference about decision makers' responsibility –i.e., to what extent the transparency of dictators' actions has an impact on pro-social behavior and guilt avoidance in my experiment– revealed that the desire to avoid feelings of shame prevails when stating pro-social decisions, in particular in small group settings. This implies that in environments with an imperfect ex-post information structure, such as multi-agent settings, the presented guilt from blame model best (c.f. Battigalli & Dufwenberg, 2007) captures the underlying cause of pro-social behavior. In addition, I found that guilt-averse dictators give significantly more if their actions are observable, but that surprise-seeking dictators transfer payments are not significantly affected by the observability of actions. This indicates that dictators might be loss-averse in their image dimension. Eventually, in larger group settings it requires more effort to generate an attributability effect, likely because here a single decision maker is less likely to be the center of attention and therefore blame is (perceived as) less severe. Moreover, a simple disclosure of the distribution of decisions, as in this study, does not suffice to generate transparency.

Second, the established transparency or attributability effect entails that numerous results generated in previous studies might be not generalizable to situations where others cannot detect to what extent an individual is responsible for a deviation from expectations. For instance, the high levels of guilt aversion in numerous games (e.g., Charness et al., 2007; Dufwenberg & Gneezy, 2000) might be attenuated if decisions are not perfectly observable.

Third, the concept of guilt from blame constitutes a formal, behavioral explanation of the bystander effect (Latané & Darley, 1970). The finding that foremost feelings of shame and social image loss causes a decrease of moral behavior in multi-agent settings trigger the bystander effect is in line with a study by van Bommel et al. (2012). They found that cues which eliminate feelings of anonymity, like the presence of a camera or wearing a name tag, trigger people to become aware of the attributability of their actions. As a consequence, subjects assigned to attributability cues more likely help individuals in the presence of other people in a critical situation. Fischer et al. (2011) find in a meta-study that another attributability cue, namely the familiarity within other people, reduces the bystander effect.

Fourth, based on the idea of Schumacher et al. (2017) I also investigated whether dictators discount the disutility of a single recipients more if the costs or their actions are more dispersed. Astonishingly, if decisions are opaque, pro-sociality does not further decrease in large groups compared to small groups. This finding that decision makers are insensitive to changes in the group size is highly relevant for experimental economists who are concerned about external validity. Having stated this, I am confident that a different perception of behavior that potentially causes guilt and shame in multi-agent settings is able to explain not only small group phenomena, such as tipping behavior, but large-scale phenomena, such as insurance fraud by credence good providers. With respect to (partly) transparent decision, future research should continue to examine why the observability of decisions is a weaker predictor for pro-social behavior in large group settings.

Fifth, the experiment was not only able to test the concept of guilt in multi-agent settings in general but mimicked in particular the insurance market for credence goods in an abstract manner.[47] Thereby, I offer a

[47] The baseline treatment described a situation where a customer (recipient) is willing to acquire a credence good for a price up to his reservation price (dictators' endowment). The actual expenses (price charged by his matched dictator) were covered by the customer. The credence good market's feature that sellers have discretionary price setting power due to informational asymmetries was captured by the dictator game-setting. The take framework reinforced the perception that sellers in credence good markets inflict losses to the

novel theoretical explanation why credence good sellers show more fraudulent behavior in insurance market is valuable because the most prominent alternative theory is inconsistent (Balafoutas et al., 2017; Kerschbamer et al., 2016). In credence good markets sellers have the opportunity to exploit informational asymmetries between them and their consumers (Darby & Karni, 1973; Emons, 1997), because the latter are unaware whether sellers commit over-treatment (providing a higher quality than actually needed), overcharging (charging for a higher quality than has been provided), under-treatment (choosing a quality that is insufficient to satisfy the consumer's needs) or a combination of the former (Kerschbamer et al., 2016). It has been argued that being insured lowers the costs for customers to take measures in order to decrease expenses or monitor sellers, because they profit only to a negligible extent from a lower bill (c.f. Balafoutas, Kerschbamer, et al., 2017; Kerschbamer et al., 2016). As a consequence, the detection risk will decrease if customers are insured. Hence, sellers may react by committing more fraud (Sülzle & Wambach, 2002).

Nevertheless, insurance companies have more sound industry knowledge and access to aggregated data of multiple customers and sellers and thus obtain an advantage in detecting systematic fraud and threaten sellers with more severe penalties. It follows that the detection risk cannot solely explain expert or supply-side induced insurance fraud. Instead, behavioral explanations –such as how guilt[48] is experienced in various settings – should be taken into greater account when explaining insurance fraud. The experimental results imply that suppliers of credence goods commit fraudulent behavior more often in multiagent settings even in the absence of informational economic incentives, (at least partly) because the experience less shame in an insurance setting.[49]

Sixth, the experimental design allows to derive managerial implications: in particular, the transparency effect has important implications for practitioners. The existence of the transparency imply that insurance companies should instead of just covering the expenses of their customers, reveal to them the services and their costs provided by e.g., their doctor or a car mechanic that outlines the single items comprehensibly and transparently. Teams with low levels of cooperation that comprises numerous shirkers (Holmström, 1982) may want to enhance the transparency of individual contribution to the team project and facilitate social control. This enhancement might be achieved by a shared instead of an individual office policy and regular meetings in which individuals communicate their progress. Service staff may increase their earned tips by charging customers individually and thus benefit from the established transparency effect (c.f. Conlin et al., 2003). In addition, the established norm compliance effect indicates that tips are higher if the first customer paying is benevolent, since customers tend to mimic pro-social behavior of others.

recipients' current wealth levels by practicing overcharging. Customers (recipients) were restricted in their possibility to actively take part in the market, because the focus of this study lies exclusively on the behavior of credence good sellers. The multi-agent treatment described an insurance market for credence goods in which recipients in one group form the collective of insured customers. All customers received a reimbursement for the bought credence good (price equals the amount taken by the dictator) but had to pay an insurance fee (the average amount taken by all dictators in one group). The final share of recipients after reimbursement and payment of the insurance fee therefore equaled the average amount of money that remains in the common pot. Such a design of the full coverage insurance contract equaled the equilibrium contract in a market with risk-neutral insurance companies, risk-averse customers and perfect competition (c.f. Rothschild & Stiglitz, 1976).

[48] Beck et al. (2013) found some evidence concerning the influence of experts' guilt aversion on fraudulent behavior in credence good markets using an indirect test procedure. They found that if an expert uses the opportunity to make a non-binding promise and thereby influencing buyers' belief, the expert will more likely to behave morally inclined. Contrary, buyers' opportunity to burn money and thus to induce guilt aversion has no significant effect on experts' pro-sociality. A potential explanation for the missing significance is that the underlying theory relies on iterated forward induction Beck et al. (2013). Hence, due to the complexity of the mechanism customers may not take advantage of it. By using a more simplistic design and accounting for non-linear utility functions, I found clear evidence in favor of the guilt aversion in credence good markets.

[49] My online experiment abstracted from other potential confiding behavioral explanation such as initial wealth status, unequal social distance, informational economic and strategic considerations as well as a common interaction history and signals concerning customers' willingness to pay that regularly differs between self-paying customers and the collective of insured customers.

Appendix

A Further Statistical Analysis

A.1 Randomization Checks

My experimental setting rests upon random assignment of subjects to different treatments and exogenous variations in group-size, recipients' first order beliefs, as well as the experimental ex-post information structure. Randomization and exogenous variations of independent variables secured that the data is collected in a ceteris paribus fashion– a necessary prerequisite in order to estimate a causal average treatment effects (ATE) of different group sizes as well as institutions on the willingness to act pro-socially. Nevertheless, to establish a precise cause-and-effect relationship, it is necessary that the treatment groups are balanced with respect to observed and unobserved prognostic variables. If this is not the case, it cannot be decided whether behavioral differences across treatments are caused by the ATE or by differences in (unobservable) predictive variables.[50] Therefore, I test whether the randomization process balanced the groups with respect to observable variables as intended.

Above all, because this study focuses on investigating distribution decisions in dictator games in the context of guilt aversion, characteristics that are assumed to influence giving in dictator games should be balanced across treatments. While it is impossible to determine all characteristic features that influence transfer decisions ex-ante, I test whether characteristics, which previous studies found to be correlated with the transfer decision, are balanced across treatments.

A range of experimental studies seeking to investigate potential gender effects find that on average, female dictators transfer a higher share of their initial endowment (see among others Croson & Gneezy, 2009; Eckel & Grossman, 1998; Engel, 2011) or behave more altruistic in general (Falk et al., 2018). The share of female dictators is between 38% in the transparency treatment with 20 dictators and 60% in my transparency treatment with 4 dictators. This effect is according to a two-sided Fisher exact test significant at a 1%-level. To mitigate a potential bias caused by such an imbalance, I include an age coefficient in all of my regressions.

In an empirical investigation, Bracht and Regner (2013) find that in dictator games economics students are more likely to behave pro-socially. By contrast, Engelmann and Strobel (2006) as well as Fehr, Naef, et al. (2006) find that economic students appear to be more selfish and concerned about the efficiency of the results.[51] The share of economic students was between 34% and 40%. Applying two-sided Fisher exact-tests the differences between the treatments are not significant at any conventional significant level.

In a meta study, Engel (2011) find a significant statistical relationship between dictators' age and the height of their transfer payment: The older a dictator is, the more prone she is to transfer a positive amount of money to the recipient. In my experiment, participating dictators are between 18 and 76 years old—with an average age of 32.9 years. The average age in the different treatment is between 32 and 34. Hence, it is unlikely that an age effect drives the overall treatment effect.

In addition, I use a qualitative and a quantitative survey item from the Global Preference Survey (Falk et al., 2018) to measure to what extent dictators are altruistic. The quantitative item described a situation

[50] For a technical discussion of the ATE and the corresponding statistical framework of counter-factual causality, see Angrist and Pischke (2008, chapter 2), Heckman and Vytlacil (2007), and Holland (1986).

[51] The influences of taking economic classes on giving behavior can be classified as either an education or a selection effect. That is, either the action of taking economic classes directly encourages students to make more selfish decisions or that vice versa students who are more prone to be selfish tend to take economic classes.

in which the respondent unexpectedly received 1,000 and asked them to state how much of this amount they would donate. The qualitative question asked respondents how willing they are to give to charity without expecting anything in return on an 11-point Likert-scale. The average score stated in the former question are between \$77 and \$100 in the different treatments. The average score stated in the latter question is between 4.8 and 6. According to MWU-test the difference between the answers to neither question one nor question two between two treatments are significant at a 10-%-level. Overall, the investigated characteristics are except of the proportion of female dictators, for which I control in my regressions, equally balanced across treatments. Consequently, I expect that selection biases unlikely distort my empirical results.

A.2 Impact of Dictator’s Second Order Beliefs on Her Distribution Decision

In this appendix A.2. I examine whether dictators condition their distribution choice on recipients’ guesses. Therefore, I calculate the correlation as well as the rank correlation between the amount taken by the dictator and the recipients’ guesses. In line with the empirical findings of Khalmetski et al. (2015) the effect size of Pearson’s $r = .0862$ and Spearman’s $\tau = .0857$ is rather low, even though the correlation coefficients are significant on a 5%-level. A random-effects regression as well as a random-effects panel Tobit regression corroborate the results (*see* Table A2.1, Model 1 and 2 in the result section).

The linear random-effects model implies that a change in revealed beliefs by $1.00 leads to $0.08 lower amounts kept by dictators. The Tobit regression model predicts an average marginal effect on the actual dictator’s payoff of $0.08, as well. While the reported coefficients are statistically significant, the average effect size is low. However, the within-subject data tell a different story: overall 53 % of all subjects condition their choice on recipients’ first order beliefs, though the number of conditional distributors is slightly lower than in the study of Khalmetski et al. (2015). In particular, the distribution decision of 31,6% of all dictators have a positive-within correlation and the distribution if 19,3% had a negative correlation (show surprise-seeking behavior):

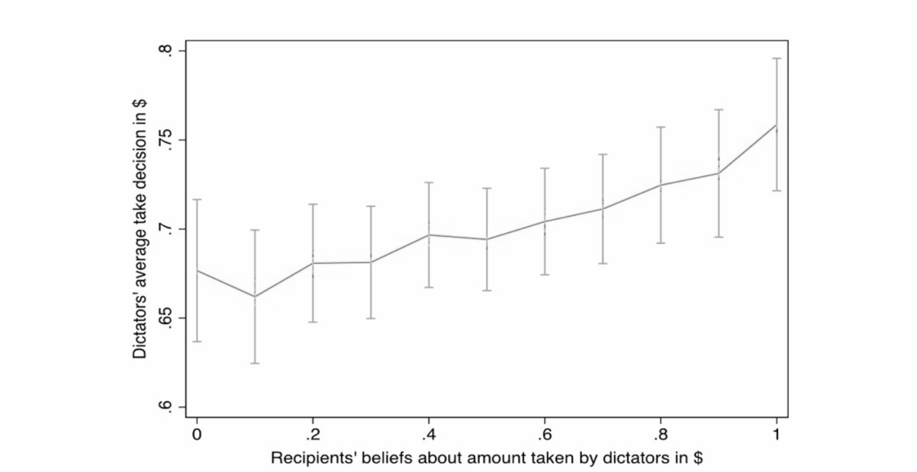
Table A2.1: *Relationship between Dictator’s Allocation Decision and Recipients’ Revealed First Order Beliefs*

Between-subject correlation coefficient of transfers with guesses	$r = .09$
Share of dictators who vary transfers conditional on beliefs	53%
Share of dictators with a positive correlation	31.6%
Share of dictators with a negative correlation	19.3%
Share of completely selfish dictators	32.0%
Share of altruistic dictators	17.1%
Share of dictators with a positive correlation > .3	29.7%
Share of dictators with a negative correlation > .3	16.0%
Average correlation of dictators with a positive correlation	$r = .80$
Average correlation of dictators with a negative correlation	$r = -.77$

A fine-grained analysis of the different distribution types revealed that of those dictators with a positive correlation, almost 94% have a within correlation above 0.3. Over 82% of dictators with a negative correlation have a coefficient below −0.3. 32% of the dictators show completely selfish behavior.

In contrast to Khalmetski et al. (2015), Balafoutas and Fornwagner (2017) proposed that there is a u-shaped relationship between beliefs and dictators’ share. This reflects that dictators only appear to be guilt averse up to a certain threshold. Beyond that threshold dictators tend to punish recipients when the latter ask for too much. A graphical plot of the average amount kept conditional on different beliefs does not exhibit such patters (*see* Figure A2.1).

Figure A2.1: *Impact of Revealed Recipients' First orders Beliefs on Average Co-Dictators' Distribution Decision on the Final Payoffs*



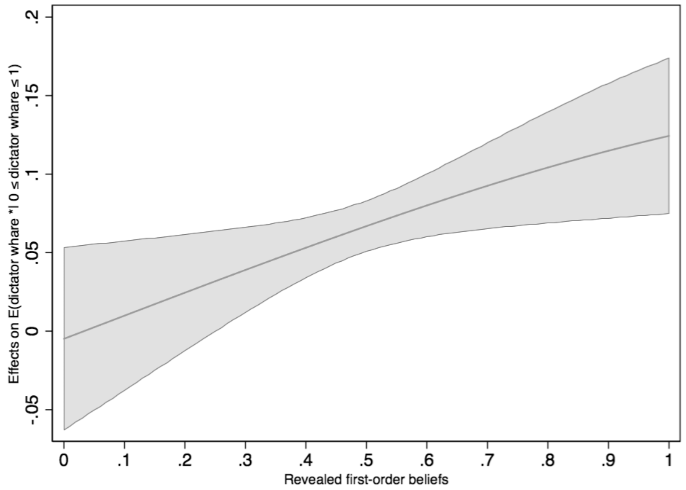
Moreover, I tested whether I could confirm this effect by including a quadratic term into the random-effects linear model as well as the random-effects Tobit model (*see* Table A2.2, Model 1 and 2). While the quadratic term in the random-effects linear regression was significant, the vertex was according to a t-test not significantly different from zero on a 10-% level. Consequently, I find no support for the hypothesis proposed by Balafoutas and Fornwagner (2017). The intuition from linear regression considering the interpretation of the quadratic form does not extend to the interpretation of the actual variable in the Tobit model. The statistical significance cannot be determined from the reported z-statistic (Norton et al., 2004). To correctly interpret the effect, I plot the respective marginal effects on the dictator's share conditional on different potential belief values (*see* Figure A2.2). Figure A2.2 revealed that the overall marginal effect is always significantly positively different from zero or not significantly different from zero I therefore conclude that my experimental design allows me to discriminate between the assumptions about the functional from proposed by Balafoutas and Fornwagner (2017) and Khalmetski et al. (2015). I find that the model proposed by Khalmetski et al. (2015) more precisely captures distribution patterns in dictator games.

Table A2.2: *Impact of Revealed First Order Beliefs on Different Dictators' Distribution Decision Patterns*

	Model (1) Random Effects	Model (2) Random Effects Tobit
	Dependent variable: ***dictator's monetary payoff***	
Revealed Beliefs	.013(.036)	.064(.008)***
Squared Revealed Beliefs	.067(.035)**	
multi-agent treatment (4 dictators)	.131(.041**)	.175(.047)***
multi-agent treatment (20 dictators)	.082(.041)**	.094(.039)**
transparency treatment (4 dictators)	.058(.041)	.078(.041)
transparency treatment (20 dictators)	.076(.041)*	.105(.039)**
Altruism 1	–.0004 (0.001)***	–.0004 (0.001)***
Altruism 2	–.018(.005)***	–.024(.005)***
Study Known	–.015(.045)	–.017(.052)
Age	.001(.001)	.002(.001)
Female	–0.006(.027)	.023(.030)
Responsibility	.004(.005)	.006(.005)
Constant	.66(.060)***	
Number of Obs.	3300	3300

***Notes**: The dependent variable kept indicates the amount of money kept by the dictator. Random effect Tobit models are calculated to account for the share of observations with amount kept of $0 or $1 (double-censored Tobit regression) and to account for the panel structure of the data set. All reported coefficients are (average) marginal effects. Standard errors in parentheses. The significance level is indicated by ***p = .01, ** p = .05. * p = .10*

Figure A2.2: Conditional Marginal Effect of the Variable Revealed Beliefs with 95%-Confidence Intervals of the Random Effects Tobit Model



A.3 Impact of Dictator’s First Order Beliefs regarding Other Dictators’ Decision on Dictators’ Allocation Decisions

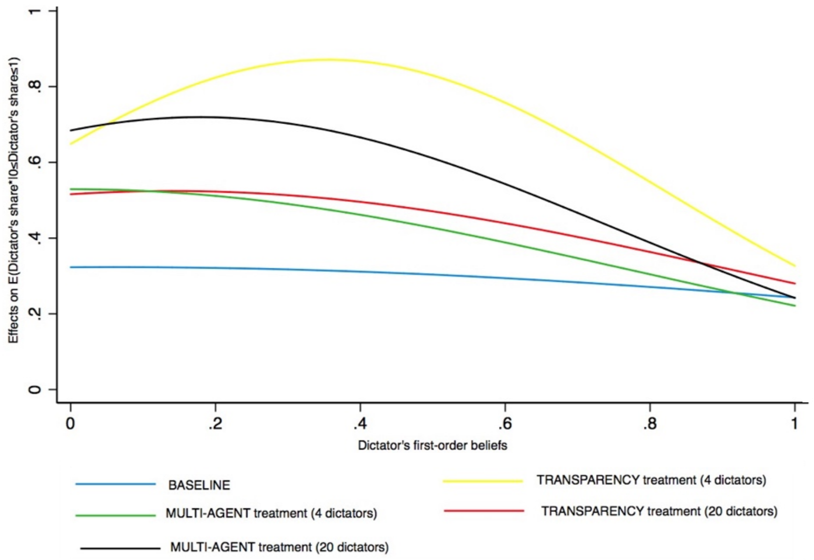
In appendix A.3 I investigated whether the impact of beliefs about other dictator’s behavior is stronger in multi-agent settings as proposed by both guilt models. Therefore, I included the interaction terms between the treatments and the height of the revealed beliefs in a random-effects linear model as well as a random effects Tobit model. According to the results of Table A2.3 there is mixed evidence whether dictators care more about the behavior of others in multiagent settings: The linear model imply that dictators responded more strongly to the other dictators’ assumed behavior in all multi-agents. However, this effect is only significant at a 5%-level in the multi-agent setting with 4 dictators and the transparency treatment with 20 dictators. According to the Tobit model, all average marginal effects are significant at a 5%-level, but not for all different treatments values of beliefs considering the actual variable. The conditional marginal effects of an average dictator’s first order beliefs on his share by treatment are illustrated in Figure A2.3). In conclusion, the evidence on a potential relationship between other dictators’ first order beliefs and dictators’ payoffs is inconclusive.

Table A2.3: *Impact of Dictator’s First Order Beliefs regarding other Dictators’ Decision on Average Dictator’s Distribution Decision Including Interaction Effects*

	Model (1) Random Effects	Model (2) Random Effects Tobit
Dependent variable: ***dictator’s monetary payoff***		
Revealed Beliefs	–.006 (.019)	–.018 (.018)
Multi-agent treatment (4 dictators)	0.24 (.067)	.026 (.063)
Multi-agent treatment (20 dictators)	–.052 (.065)	−.072 (.061)
Transparency treatment (4 dictators)	–.221 (.063)***	−.19(.060)***
Tansparency treatment (20 dictators)	–0.26(.063)	.016 (.060)
Dictator’s first order beliefs	.300 (.060)	.423 (.029)***
Dictator’s first order beliefs X multi-agent treatment (4 dictators)	.096 (.087)	.098 (.029) ***
Dictator’s first order beliefs X multi-agent treatment (20 dictators)	.21 (.089) **	.074(.033) ***
Dictator’s first order beliefs X transparency treatment (4 dictators)	.38(.083) ***	.026(.034)
Dictator’s first order beliefs X transparency treatment (20 dictators)	.10(.086)	.059(.031) *
Altruism 1	–.0003(.0001)***	.–.0002(.0001)***
Altruism 2	–.016(.004)***	–.019(.003)***
Study Known	–.016(.038)	–.023(.041)
Age	.001 (.001)	.001(.001)
Female	.017(.023)	–.001 (.025)
Responsibility	.007(.003)*	.008(.004)*
Constant	.509(.062)***	
Number of Obs.	900	900

Notes*: The dependent variable “dictator’s monetary payoff” indicates the amount of money kept by the dictator. Random effect Tobit models are calculated to account for the share of observations with amount kept of $0 or $1 (double-censored Tobit regression) and to account for the panel structure of the data set. All reported coefficients are (average) marginal effects. Standard errors in parentheses. The significance level is indicated by ***p = .01, ** p = .05. * p = .1.*

Figure A2.3: Conditional Marginal Effects of Dictator’s First Order beliefs by Treatment



B Proofs and Formal Discussions

B.1 Simple Guilt Model

B1.1. The Dictator's Optimal Transfer Payment in a One-to-one Setting

Consider a utility function of the following form:

$$\max U\,(t_i, \beta) = ln(t_i) - \eta \cdot \max\{\,(\beta) - (T - t_i), 0\} \text{ s.t. } 0 \leq t_i \leq T \qquad (1)$$

where $m(t_i)$ denotes the monetary payoff function and $\eta \cdot \max\,\{\beta - (T - t_i), 0\}$ the psychological payoff function. The normalized endowment is given by $T = 1$ and the dictator's payoff by t_i. Moreover, assume that the dictator cannot keep a sum that exceeds the initial endowment $T=1$ (either assigned to the dictator or the recipient) or is less than 0 ($0 \leq t_i \leq 1$) and the dictator's second order belief is consequently s.t $0 \leq \beta \leq 1$. Furthermore, $\eta \geq 0$ depicts how much a dictator dislikes "letting the recipient" down.

The dictator maximizes her utility with respect to her payoff t_i given her belief β. The maximum operator included in $U(t_i,\beta)$ causes the function to be not differentiable anymore. This operator separates the function into two parts:

$$U_1(t_i, \beta) = ln(t_i) - \eta \cdot (\beta - (T - t_i)) \qquad \text{if } 1 - \beta \leq t_i \leq 1 \qquad (2)$$

$$U_2(t_i, \beta) = ln(t_i) \qquad \text{if } 1 - \beta > t_i \qquad (3)$$

Solving for $t^{**}_{i,un}$ yields the optimal transfer payment of the unconstrained maximization problem. It follows that the optimal transfer payment of the unconstrained maximization problem is given by

$$\frac{\partial U_1(t_i,\beta)}{\partial t_j} = \frac{1}{t^{**}_{i,un}} - \eta = 0 \text{ or } \frac{1}{\eta} = t^{**}_{i,un} \qquad (5)$$

Next, take into account that the function $U_1(t_i, \beta)$ is constrained by $0 \leq t_i$ and $t_i \leq 1 - \beta$ and additionally note that the maximizing argument of

$$U_2(t_i, \beta) = ln(t_i) \quad \text{if} \quad 1 - \beta > t_i \qquad (6)$$

is constraint by $0 \leq t_i \leq T = 1$ is given by $t_i^{***} = 1$.

It follows that the utility maximizing dictator payoff of $t_i^{*}$the constrained function $U(t_i, \beta)$ is given by

$$t_i^* = \begin{cases} 1 & \text{if } \frac{1}{\eta} > T = 1 \\ \frac{1}{\eta} & \text{if } 1 - \beta \leq \frac{1}{\eta} \leq T = 1 \\ 1 - \beta & \text{if } \frac{1}{\eta} \leq 1 - \beta \end{cases} \qquad (7)$$

or alternatively

$$t_i^{**} = \max\{1 - \beta; \min\{1; \frac{1}{\eta}\}\}. \qquad (8)$$

B.1.2 Guilt Aversion Does Not Necessarily Imply Correlation Between Second Order Beliefs and Dictators' Decisions

In this section, I discuss that guilt aversion does not necessarily imply a correlation between beliefs and allocation decisions in dictator game. Therefore, consider and arbitrary β s.t. $0 \leq \beta \leq 1$, an arbitrary $\eta \geq 0$, as well as the best response function of the introduced simple guilt model applied to a standard dictator game (see Appendix B.1.1 for the derivation) of the following form:

$$t_i^* = \begin{cases} 1 & \text{if } \frac{1}{\eta} > 1 \\ \frac{1}{\eta} & \text{if } 1 - \beta \leq \frac{1}{\eta} \leq 1 \\ 1 - \beta & \text{if } 1 - \beta \geq \frac{1}{\eta} \end{cases} \tag{1}$$

I am going to consider two different cases:

- First, the beliefs and the dictators' payoffs will be perfectly correlated if and only if $1 - \beta \geq \frac{1}{\eta}$, because under this assumption dictators' payoffs equal their beliefs.
- Second, consider that $1 - \beta < \frac{1}{\eta}$ for all βs. This implies that there is no correlation between second order beliefs and dictators' payoffs, since dictators allocate to themselves a share of $t_i^* =$ if $\frac{1}{\eta} > 1$ or $t_i^* = \frac{1}{\eta}$ if $1 - \beta \leq \frac{1}{\eta} \leq 1$.

In conclusion, if the utility function entails moderate guilt aversion parameters and if beliefs are sufficiently high, then the simple guilt model predicts the absence of a correlation between second order beliefs and the height of the dictator share. This holds, even though the dictator shares are determined by the guilt aversion coefficient η. A similar reasoning applies to a setting with an imperfect ex-post information structure and a dictator who is hurt by blame.

B.1.3 The Dictator's Optimal Transfer Payment in a Multi-Agent Setting

Consider a simple guilt utility function in a multi-agent dictator game of the following form:

$$\begin{aligned} U(t_i, \beta) &= ln(t_i) \\ &- f(n) \cdot \sum_{i=1}^{n} \eta \cdot \max\{(\beta) - (1 - \frac{n-1}{n} \cdot \gamma - \frac{t_i}{n}); 0\} \\ &+ \eta \cdot max\{\beta - (1 - \frac{n-1}{n} \cdot \gamma); 0\} \text{ s.t. } 0 \leq t_i \leq T = 1 \end{aligned} \tag{1}$$

The dictator maximizes her utility with respect to her payoff t_i given her belief β. Depending on different parameter values of the dictator's first order belief about other dictators' decisions γ, three different cases have to be considered:

- In the first case the perceived recipients' expectations will irrespective of the individual dictator's decision always be violated, since $1 - \frac{(n-1)\cdot\gamma}{n} \leq \beta$.
- In the second case the perceived recipients' expectations will depending on the dictators' decision either be violated, met or surpassed, since $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$.
- In the third case the perceived recipients' expectations will be surpassed irrespective of the individual dictator's decision, since $\beta \leq 1 - \frac{1+(n-1)\cdot\gamma}{n}$.

Case 1: First, consider the case where $1 - \frac{(n-1)\cdot\gamma}{n} \leq \beta$. In this case the utility function simplifies to:

$$U(t_i, \beta) = ln(t_i) - f(n) \cdot \eta \sum_{i=1}^{n} \left(\beta - \left(1 - \left(\frac{n-1}{n} \cdot \gamma + \frac{t_i}{n}\right)\right)\right) + \left(\beta - \left(1\text{-}\left(\frac{n-1}{n} \cdot \gamma\right)\right)\right) \text{ if } 0 \leq t_i \leq 1 \tag{2}$$

The first order condition of the function yields

$$\frac{\partial U(t_i,\beta)}{\partial t_i} = \frac{1}{t_i} - f(n) \cdot \eta = 0 \tag{3}$$

Furthermore, note that $U(t_i, \beta)$ is strictly concave. Thus, setting $\frac{\partial U(t_i,\beta)}{\partial t_i} = 0$ and solving for $t_{i,un}^{\dagger}$ yields the optimal transfer payment of the maximization problem. It follows that the optimal dictator payoff of the unconstrained maximization problem is given by $t_{i,un}^{\dagger} = \frac{1}{f(n)\cdot\eta}$. Taking into account that the function $U(t_i, \beta)$ is constrained by $0 \leq t_i \leq 1$, the utility maximizing recipient's payoff $t^{\dagger}_i$ of the constrained function $U(t_i, \beta)$is given by:

$$t_i^{\dagger} = \begin{cases} 1, & if \ \frac{1}{f(n) \cdot \eta} > 1 \\ \frac{1}{f(n) \cdot \eta}, & if \ \frac{1}{f(n) \cdot \eta} \leq 1 \end{cases} \tag{4}$$

Alternatively, $t_i^{\dagger}$ can be denoted by $t_i^{\dagger}$= $\min\{1; \frac{1}{f(n)\cdot\eta}\}$.

Case 2: Second, consider the case where $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$, i.e., it is up to the dictator *i* whether the expectations of the recipients are violated or surpassed. It follows that

$$max\,\{\,(\beta - \,(1\,-\,\frac{(n-1)}{n}\cdot\gamma);\;0\,\}\;=\;0. \tag{5}$$

Therefore, the utility function simplifies to

$$U(t_i, \beta) = ln(t_i) - f(n)\cdot\eta\cdot\sum_{i=1}^{n}\max\{\beta - \left(1 - \frac{n-1}{n}\cdot\gamma - \frac{t_i}{n}\right); 0\}\text{ s.t. } 0 \leq t_i \leq 1. \tag{6}$$

The dictator maximizes her utility with respect to her payoff t_i given her belief $0 \leq \beta \leq 1$. First, consider an unconstrained version of $U(t_i, \beta) =$. The first order condition of the function yields

$$\frac{\partial U(\mathrm{t_i}, \beta)}{\partial \mathrm{t_i}} = \frac{1}{\mathrm{t_i}} - \mathrm{f(n)}\cdot\eta\;= 0. \tag{7}$$

$\mathrm{U}(t_\mathrm{i}, \beta)$is strictly concave. Thus, setting

$$\frac{\partial U(\mathrm{t_i}, \beta)}{\partial \mathrm{t_i}} = 0 \tag{8}$$

and solving for $t_{i,un}^{**}$ yields the optimal transfer payment of the unconstrained maximization problem. It follows that the optimal transfer payment of the unconstrained maximization problem is given by

$$t_{i,un}^{**} = \frac{1}{f(n)\cdot\eta}. \tag{9}$$

Now, taking into account that the function $U_1(t_j, \beta)$ is constrained by $t_i \leq 1$ and $t_i \geq (1-\beta)n - (n-1)\cdot\gamma$, the utility maximizing argument (recipient's payoff t_i^{**}) of the constrained function $\mathrm{U}(t_j, \beta)$ is given by

$$\mathrm{t_i^{**}} = \begin{cases} 1,\; if\; \frac{1}{f(n)\cdot\eta} > T = 1 \\ \frac{1}{f(n)\cdot\eta},\; if\; (1-\beta-\frac{(n-1)\cdot\gamma}{n}) \leq \frac{1}{f(n)\cdot\eta} \leq 1 \\ 1-\beta-\frac{(n-1)\cdot\gamma}{n},\; if\,\frac{1}{f(n)\cdot\eta} > \;(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \end{cases} \tag{10}$$

Alternatively, t_i^{**} can be denoted by $t_i^{**} =\; \max\,\{n\cdot\left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right); \min\{1; \frac{1}{f(n)\cdot\eta}\}\}$.

Case 3: Third, consider the case where $\beta \leq 1 - \frac{(n-1)\cdot\gamma+1}{n}$. In this case, the utility function simplifies to $U(t_j, \beta) = ln(t_i)$. Taking into account that the function $U(t_i, \beta)$ is constrained by $0 \geq t_i \geq 1$, the utility maximizing transfer payment $t_i^{**}$of the constrained function $U(t_i, \beta)$ is given by $\mathrm{t_i^{*}} = 1$.

B.1.4 Comparison of Dictators' Distribution Decision between the Multi-Agent and the Standard Dictator Game

In this section, I compare dictators' distribution in the standard dictator games as well as in an n:n-setting. I consider the same utility function and best-reply function in the multi-agent game as provided in B.1.3. Furthermore, I consider a weighting function $f(n) < 1$, an arbitrary η and arbitrary β s.t. $\eta \geq 0$ and $0 \leq \beta \leq 1$. Moreover, I consider a best reply function of a dictator in a standard dictator game that is given by

$$t_i^* = \begin{cases} 1 & \text{if } \frac{1}{\eta} > T = 1 \\ \frac{1}{\eta} & \text{if } 1-\beta \leq \frac{1}{\eta} \leq T = 1 \\ 1-\beta & \text{if } \frac{1}{\eta} \leq 1-\beta \end{cases} \tag{1}$$

The dictator's best-reply function of a dictator in the multi-agent setting derived in appendix B.1.3. is dependent on γ. In particular, we considered three different cases:

- First, consider that $\beta > 1 - \frac{(n-1)\cdot\gamma}{n}$. Hence the maximizing argument in the multi-agent setting is given by

$$t_i^\dagger = \begin{cases} 1, & if \ \frac{1}{f(n)\cdot \eta} > 1 \\ \frac{1}{f(n)\cdot\eta}, & if \ \frac{1}{f(n)\cdot\eta} \leq 1 \end{cases} \tag{2}$$

- Second, consider that $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$. Hence, the maximizing argument in the multi-agent setting is given by

$$t_i^* * = \begin{cases} 1 & \text{if } \frac{1}{f(n)\cdot\eta} > 1 \\ \frac{1}{f(n)\cdot\eta} & \text{if } n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \leq \frac{1}{f(n)\cdot\eta} \leq 1 \\ n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) & \text{if } \frac{1}{f(n)\cdot\eta} \leq n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \end{cases} \tag{3}$$

- Third, consider the case where $\beta \leq 1 - \frac{(n-1)\cdot\gamma+1}{n}$. Hence, the maximizing argument in the multi-agent setting is given by

$$t_i^{**} = 1 \tag{4}$$

Case 1: First, we consider the case in which $\beta > 1 - \frac{(n-1)\cdot\gamma}{n}$. Furthermore, assume that that $\frac{1}{\eta} \geq 1$. It follows from equation 1 that the utility maximizing dictator payoff is given by $t_i^* = 1$ in the standard dictator game. Moreover, it holds that by definition $\frac{1}{f(n)} > 1$ which implies that $\frac{1}{f(n)\cdot\eta} \geq \frac{1}{\eta} \geq 1$. Hence, it follows equation 2, 3, 4 that the maximizing argument in the multi-agent setting is in all cases given by $t_i^{**} = 1$. Consequently, $t_i^{**} = t_i^* = 1$. Consequently, under the assumption that $\frac{1}{\eta} \geq 1$ the dictator in the standard dictator games as well as in the multi-agent game always acts selfishly.

Case 2: Consider that $1 - \beta \leq \frac{1}{\eta} \leq 1$. It follows from equation 1 that the utility maximizing dictator payoff is given by $t_i^* = \frac{1}{\eta}$ in the standard dictator game. We next discuss three different cases in which the dictator in the multi-agent setting acts more or less selfishly than in the dictator setting:

- If $\frac{1}{f(n)\cdot\eta} \geq 1$ it follows from equation 2, 3, 4 that the maximizing argument of the dictator in the multi-agent dictator game is given by $t_i^{**} = 1$. Hence, the dictator acts more selfishly in the multi-agent compared to the standard dictator game, since $t_i^* = \frac{1}{\eta} < 1 = t_i^{**}$.

- Consider that either $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$ and $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) < \frac{1}{f(n)\cdot\eta}$<1 or alternatively that $\beta > 1 - \frac{(n-1)\cdot\gamma}{n}$ and $\frac{1}{f(n)\cdot\eta} < 1$.

  In both of these cases, the dictator's utility maximizing payoff is given by $\frac{1}{f(n)\cdot\eta} = t_i^{**}$. Hence, the dictator acts more selfishly in the multi-agent setting than in the standard dictator game, since $t_i^* = \frac{1}{\eta} < \frac{1}{f(n)\cdot\eta} = t_i^{**}$.

- Assume that $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$ and $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) > \frac{1}{f(n)\cdot\eta}$.

  Hence, it follows from equation 3 that $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) = t_i^{**}$. It follows that $t_i^* = \frac{1}{\eta} \leq n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) = t_i^{**}$. In order to see this, note that it follows from equation 1 that $1 - \beta \leq \frac{1}{\eta}$ *and* consider the following two cases:

  - It follows that if $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) < 1 - \beta < \frac{1}{\eta}$, $t_i^* > t_i^{**}$ if $\gamma < 1 - \beta$, because $1 - \beta > n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right)$ if $\gamma < 1 - \beta$.
  - Assume otherwise, that $\gamma > 1 - \beta$. Hence, it follows that $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) < 1 - \beta \leq \frac{1}{\eta} < \frac{1}{f(n)\cdot\eta}$. It directly follows from equation 3 that $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) \neq t_i^{**}$, which contradict the initial assumption. Hence,
  - $t_i^* = \frac{1}{\eta} \leq n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) = t_i^{**}$ *if* $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$ and $n \cdot \left(1 - \beta - \frac{(n-1)\cdot\gamma}{n}\right) > \frac{1}{f(n)\cdot\eta}$.

Case 3: Consider that $\frac{1}{\eta} \leq 1 - \beta$. It follows from equation 1 that the utility maximizing dictator payoff is given by $t_i^* = 1 - \beta$ in the standard dictator game.

- If in addition $\frac{1}{f(n) \cdot \eta} \geq 1$ it follows from equation 2, 3, 4 that the maximizing argument of the dictator in the multi-agent dictator game is given by $t_i^{**} = 1$. Hence, the dictator acts more selfish in the multi-agent compared to the standard dictator game, since $t_i^* = 1 - \beta < 1 = t_i^{**}$.
- Contrary, if in addition either $1 - \frac{(n-1) \cdot \gamma + 1}{n} \leq \beta \leq 1 - \frac{(n-1) \cdot \gamma}{n}$ and $n \cdot \left(1 - \beta - \frac{(n-1) \cdot \gamma}{n}\right) < \frac{1}{f(n) \cdot \eta}$ <1 or alternatively $\beta > 1 - \frac{(n-1) \cdot \gamma}{n}$ *and* $\frac{1}{f(n) \cdot \eta} < 1$, it follows from equation 2 and 3 that the dictator's utility maximizing payoff in the multi-agent dictator game is given by $\frac{1}{f(n) \cdot \eta} = t_i^{**}$. Whether the dictator acts more or less selfishly depends on whether $\frac{1}{f(n) \cdot \eta} < 1 - \beta$. Trivially, this can only be true if $\gamma < 1 - \beta$, because $1 - \beta > n \cdot \left(1 - \beta - \frac{(n-1) \cdot \gamma}{n}\right)$ if $\gamma < 1 - \beta$.
- If in addition $1 - \frac{(n-1) \cdot \gamma + 1}{n} \leq \beta \leq 1 - \frac{(n-1) \cdot \gamma}{n}$ and $n \cdot \left(1 - \beta - \frac{(n-1) \cdot \gamma}{n}\right) > \frac{1}{f(n) \cdot \eta}$. Hence, it follows from equation 3 that $n \cdot \left(1 - \beta - \frac{(n-1) \cdot \gamma}{n}\right) = t_i^{**}$. It holds that if $t_i^* = 1 - \beta > n \cdot \left(1 - \beta - \frac{(n-1) \cdot \gamma}{n}\right) = t_i^{**}$, dictators act more selfishly in the multi-agent treatment. This holds if and only if $\frac{(n-1) \cdot \gamma}{n} > 1 - \beta$.

In summary, the only case in which a dictator in a standard dictator games acts more selfishly than a dictator in a multi-agent setting, i.e., the only situation where $t_i^* \geq t_i^{**}$ can hold under the above specified conditions, is if $\gamma < 1 - \beta$. If $\gamma > 1 - \beta$ it holds that $t_i^{**} \geq t_{i.}^*$

B.1.5. Influence of the Number of Recipients on Dictator's Distribution Decision

In this section I derive predictions regarding the impact of the number of recipients on dictators' distribution decision. I consider the same utility function as discussed in B.1.3. Recall that *f*(*n*) is defined such that $\frac{df(n)}{dn} \leq 0$. Hence, if $m \leq n$ it follows that $f(m) \geq f(n)$.

Furthermore, recall that under the assumption that $\beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$ the dictator's optimal distribution choice is given by

$$t_i^{**} = \begin{cases} \frac{1}{f(n)\cdot\eta} & \text{if } \frac{1}{f(n)\cdot\eta} \leq 1 \\ 1 & \text{if } \frac{1}{f(n)\cdot\eta} \geq 1 \end{cases} \tag{1}$$

If $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$ the dictator's optimal distribution choice is:

$$t_i^* * = \begin{cases} 1 & \text{if } \frac{1}{f(n)\cdot\eta} > 1 \\ \frac{1}{f(n)\cdot\eta} & \text{if } n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \leq \frac{1}{f(n)\cdot\eta} \leq 1 \\ n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) & \text{if } \frac{1}{f(n)\cdot\eta} \leq n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \end{cases} \tag{2}$$

Finally, if $\beta \geq \frac{(n-1)\cdot\gamma+1}{n}$ the dictator's optimal distribution choice is given by $t_i^\star = 1$.

Case 1: Consider that $\beta \;\leq 1 - \frac{(n-1)\cdot\gamma}{n}$. It follows equation 1 that a dictator claims

$$t^* = 1 \;\; \text{if} \; \frac{1}{f(n)\cdot\eta} > 1 \tag{3}$$

irrespective of the number of recipients, though the threshold value is more likely surpassed since $\frac{1}{f(n)\cdot\eta} > \frac{1}{f(m)\cdot\eta}$. Assume otherwise that $\frac{1}{f(n)\cdot\eta} \leq 1$, it follows from $m \leq n$, $\frac{df(n)}{dn} \leq 0$ and equation 1 that

$$\mathrm{t}^*(\mathrm{n}) = \frac{1}{\mathrm{f(n)}\cdot\eta} \geq \mathrm{t}^*(\mathrm{m}) = \frac{1}{\mathrm{f(m)}\cdot\eta} \tag{4}$$

i.e., a dictator distributes greater or equal amount to herself the more dispersed the costs of her decisions are.

Case 2: Consider that 1–$\frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq$ 1–$\frac{(n-1)\cdot\gamma}{n}$. And $m \leq n$ It follows equation 2 that a dictator claims an amount

$$t^* = 1 \text{ if } \frac{1}{f(n)\cdot \eta} > 1 \qquad (5)$$

irrespective of the number of recipients. It follows from *f*(m) ≥ *f*(n) and equation 2 that

$$\mathrm{t}^*(\mathrm{n}) = \frac{1}{\mathrm{f(n)}\cdot\eta} \geq \mathrm{t}^*(\mathrm{m}) = \frac{1}{\mathrm{f(m)}\cdot\eta} \qquad (6)$$

Assume now that $\frac{1}{f(n)\cdot\eta} \leq 1-\beta$,

$$t^* = 1 \text{ if } \frac{1}{f(n)\cdot\eta} > 1\ -\ \beta \qquad (7)$$

Next, assume that $\frac{1}{f(n)\cdot\eta} \geq n\cdot\left(1-\beta-\frac{(n-1)\cdot\gamma}{n}\right)$. It follows from m ≤ n and *f(m)* > *f(n)* that

$$t^*(n) = n\cdot\left(1-\beta-\frac{(n-1)\cdot\gamma}{n}\right) \geq t^*(m) = m\cdot\left(1-\beta-\frac{(m-1)\cdot\gamma}{m}\right), \qquad (8)$$

In conclusion, given that 1–$\frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq$ 1–$\frac{(n-1)\cdot\gamma}{n}$ a dictator distributes an amount greater or equal to themselves the more dispersed the costs of their decisions are.

Case 3: Consider case three. If $\beta \geq \frac{(n-1)\cdot\gamma+1}{n}$. Trivially, the dictator claims irrespective of the number of recipients $t^* =\ 1$. Hence, n has no direct impact on t^*.

B.2 Guilt from Blame Model

B.2.1 The Dictator's Optimal Transfer Payment in a Multi-Agent Dictator Game with Perfect Knowledge

Consider a utility function of the following form:

$$\max U(t_i, \beta) = ln(t_i) - \eta \cdot f(n) \cdot max(\beta) - (1 - t_i), 0 s.t. 0 \leq t_i \leq 1 \tag{1}$$

The maximum operator included in $U(t_i,\beta)$ causes the function to be not differentiable anymore. This operator separates the function into two parts $U(t_i, \beta) = ln(t_i)$ if $\beta \leq 1 - t_i$ and $U(t_i, \beta) = ln(t_i) - \eta \cdot f(n) \cdot \big(\beta - (1 - t_i)\big)$ otherwise.

First, consider an unconstrained version of $U(t_i, \beta)$. The first order condition is given by

$$\frac{\partial U(t_i, \beta)}{\partial \mathrm{t_i}} = \frac{1}{\mathrm{t_i}} - \mathrm{f(n)} \cdot \eta = 0. \tag{2}$$

Furthermore, note that $U(t_i, \beta)$ is strictly concave.

Thus, setting $\frac{\partial U(t_i,\beta)}{\partial t_j} = 0$ and solving for $t_{i,un}^{**}$ yields the optimal transfer payment of the unconstrained maximization problem. $t_{i,un}^{**}$ is given by

$$\mathrm{t_{i,un}^{**}} = \frac{1}{\mathrm{f(n)} \cdot \eta} \tag{3}$$

Next, take into account that the function $U(\mathrm{t_i},\beta)$ is constrained by $1 - \beta \leq \mathrm{t_i} \leq T = 1$, since the utility function simplifies to $\mathrm{U}(\mathrm{t_i},\beta) = ln(\mathrm{t_i})$ if $1 - \beta > \mathrm{t_i}$. In this case $\mathrm{t_i^{***}} = 1$. Hence, it follows that the utility maximizing dictator payoff $\mathrm{t_i^*}$ of the constrained function $U(t_i,\beta)$ is given by

$$t_i^* = \begin{cases} 1 & \text{if } \frac{1}{f(n)\cdot\eta} > T = 1 \\ \frac{1}{f(n)\cdot\eta} & \text{if } 1 - \beta \leq \frac{1}{f(n)\cdot\eta} \leq T = 1 \\ 1 - \beta & \text{if } \frac{1}{f(n)\cdot\eta} \leq 1 - \beta \end{cases} \tag{4}$$

B.2.2 The Dictator’s Optimal Transfer Payment in a Multi-Agent Dictator Game with Imperfect Knowledge

Consider a guilt from blame utility function applied to a multi-agent dictator game with imperfect knowledge of the following form:

$$U(t_i(d), \beta) = \ln\big(t_i(d)\big) + \eta \cdot f(n) \cdot \sum_{j=1}^{n} \int_0^1 \max\{\frac{\beta}{n} - \frac{T - \tilde{t}_\iota}{n}, 0\} \cdot h\big(\tilde{t}_\iota \big| t_j\big) d\tilde{t}_\iota \qquad (1)$$

To illustrate that dictators demand more if their actions are opaque, I assume a particular and simple specification of $h\big(\tilde{t}_\iota \big| t_j\big)$ denoted by $g(\tilde{t}_\iota|t_j)$. In the following discussion, $g(\tilde{t}_\iota|t_j)$ is by assumption a degenerated function so that its cumulative distribution function is given by

$$G(t_j) = \begin{cases} 1, & \text{if } x \geq t_j \\ 0, & \text{if } x < t_j \end{cases} \qquad (2)$$

Under this assumption that the utility function of a single dictator depicted in the utility function (see function 1) reduces to:

$$\max U\,(t_i(d), \beta) = ln\big(t_i(d)\big) + \eta \cdot f(n) \cdot \max\{\,\beta - T - (\frac{n-1}{n}\,\gamma - \frac{t_i}{n})),0\} \text{ s.t. } 0 \leq t_i \leq T = 1 \qquad (3)$$

The dictator maximizes her utility with respect to her payoff t_i given her second order beliefs β. Depending on different parameter values of the dictator’s first order beliefs about co-dictators’ decisions γ, three different cases have to be considered:

- In the first case the recipients will irrespective of the individual dictator’s decision always blame the dictator, since $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta$.
- In the second case whether the dictator will be blamed or not relies in her decision, because $1 - \frac{(n-1)\cdot\gamma}{n} \leq \beta \leq 1 - \frac{(n-1)\cdot\gamma+1}{n}$.
- The third case captures the situation where the dictator will not be blamed by the recipients irrespective of the individual dictator’s decision, since $\beta \leq 1 - \frac{(n-1)\cdot\gamma}{n}$.

Case 1: First consider the case where $1 - \frac{(n-1)\cdot\gamma+1}{n} \leq \beta$. In this case the utility function simplifies to:

$$\max U\ (t_i(d),\beta) = ln\big(t_i(d)\big) + \eta \cdot \mathrm{f(n)} \cdot \left[\beta - \mathrm{T} - \frac{\mathrm{n}-1}{\mathrm{n}}\gamma + \frac{1}{\mathrm{n}}\mathrm{t_i}\right] \qquad (4)$$
$$\text{s.t. } 0 \le \mathrm{t_i} \le \mathrm{T} = 1$$

The first order condition of the function yields

$$\frac{\partial U(t_i,\beta)}{\partial t_i} = \frac{n}{t_i} - f(n) \cdot \eta = 0 \qquad (5)$$

Furthermore, note that $U(t_i,\beta)$ is strictly concave. Thus, setting $\frac{\partial U(t_i,\beta)}{\partial t_i} = \frac{n}{t_i} - f(n) \cdot \eta = 0$ and solving for $t^{\dagger}_{i,un}$yields the optimal transfer payment of the maximization problem. It follows that the optimal dictator payoff of the unconstrained maximization problem is given by $t_i^{\dagger} = \frac{n}{f(n)\cdot\eta}$. Now, taking into account that the function $U(t_i,\beta)$ is constrained by $0 \le t_i \le 1$, the utility maximizing recipient's payoff $t^{\dagger}_i$ of the constrained function $U(t_i,\beta)$ is given by.

$$t_i^{\dagger} = \begin{cases} \frac{n}{f(n)\cdot\eta} & \text{if } \frac{n}{f(n)\cdot\eta} \le 1 \\ 1 & \text{if } \frac{n}{f(n)\cdot\eta} \ge 1 \end{cases} \qquad (6)$$

Case 2: Second, consider the case where $1 - \frac{(n-1)\cdot\gamma}{n} \le \beta \le 1 - \frac{(n-1)\cdot\gamma+1}{n}$, i.e., it is up to the individual dictator *i* whether she will be blamed by the recipients or not. In this case the utility function is given by

$$U(t_i,\beta) = ln(t_i) - f(n) \cdot \eta \cdot \sum_{i=1}^{n} \max\{\beta - \left(T - \frac{n-1}{n} \cdot \gamma - \frac{t_i}{n}\right); 0\} \text{ s.t. } 0 \ge t_{bi} \ge T = 1 \qquad (7)$$

First, consider an unconstrained version of $U(t_i,\beta)$. The first order condition of the function yields

$$\frac{\partial U_1(t_i,\beta)}{\partial t_i} = \frac{n}{t_i} - f(n)\ \cdot\ \eta \qquad (8)$$

Furthermore, note that $U_1(t_i,\beta)$ is strictly concave. Thus, setting $\frac{\partial U_1(t_i,\beta\ )}{\partial t_i} = \frac{n}{t_i} - f(n)\ \cdot\ \eta = 0$ and solving for $t^{**}_{i,un}$ yields the optimal transfer payment of the unconstrained maximization problem. It follows that the optimal transfer payment of the unconstrained maximization problem is given by

$$t^{**}_{i,un} = \frac{n}{f(n)\cdot\eta}. \qquad (9)$$

Now, taking into account that the function $U_1(t_i,\beta)$ is constrained by $t_i \le T = 1$ and $t_i \ge n - \beta - (n-1) \cdot \gamma$, the utility maximizing recipient's payoff $t_i^{*}$of the constrained function $U_1(t_i,\beta)$ is given by

$$t_i^* = \begin{cases} 1 & \text{if } \frac{n}{f(n)\cdot\eta} > T = 1 \\ \frac{n}{f(n)\cdot\eta} & \text{if } \frac{n}{f(n)\cdot\eta} < n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \leq 1 \\ n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) & \text{if } n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \geq \frac{n}{f(n)\cdot\eta} \end{cases} \tag{10}$$

Case 3: Third, consider the case where $\beta \leq 1-\frac{(n-1)\cdot\gamma}{n}$ In this case, the utility function simplifies to $U(t_j,\beta)$ = $ln(t_i)$ Now, taking into account that the function $U(t_i,\beta)$ is constrained by $0 \geq t_i \geq 1$, the utility maximizing transfer payment t_i^* of the constrained function $U(t_i,\beta)$ is given by $t_i^* = 1$.

Finally, I compare the optimal allocation in a game with perfect and with imperfect information under the assumption that the guilt from blame model applied. First conder case three ($\beta \leq 1-\frac{(n-1)\cdot\gamma}{n}$)
Hence, it follows that $t^*{}_i = 1$. Consequently, dictators who are according to equation (34) in the main text willing to act pro-socially under the condition of perfect knowledge[52], act completely selfish under the condition of imperfect knowledge.

Next, I analyze the situation in which independent from the contribution decision the recipient will always interfere that the decision maker will be responsible for the violation of his expectations (case 1). Such a situation is given if $\frac{n-1}{n}\gamma + \frac{1}{n} < \beta$. In this scenario the maximizing argument is as previously derived given by:

$$t_i^* = \begin{cases} 1 & \text{if } \frac{n}{\eta\cdot f(n)} > 1 \\ \frac{n}{\eta\cdot f(n)} & \text{if } 1-\beta \leq \frac{n}{\eta\cdot f(n)} \leq 1 \end{cases} \tag{11}$$

Comparing the best response function in case of perfect knowledge (see equation 34, in the main text) with the best response function given an imperfect knowledge (see equation 11) it holds that while dictators would live up to the expectations of the recipients under perfect knowledge, they would violate the expectations or the recipients in case their actions are no longer directly observable. Contrary, the share of dictators who act selfishly increases as they are less rewarded for pro-social decision. The share of dictators who completely act selfishly under imperfect knowledge but show pro-social behavior under perfect knowledge is given by the share of dictators for which the following two inequalities hold as $\frac{n}{\eta\cdot f(n)} > 1$ well as $\frac{1}{\eta\cdot f(n)} < 1$. Moreover, those who are willing to live up to the expectations of others in both models still behave n times more selfishly if individual decisions are non-observable. Decision makers for which the inequality 1-β < $\frac{n}{\eta\cdot f(n)}$ holds, will act more pro-socially if the decisions are opaque. Overall, I still find strong theoretical support for the decline in pro-social behavior in settings with imperfect knowledge.

[52] This is true if for an individual *i* the following inequality holds $\frac{1}{\eta\cdot f(n)} < 1$.

Third, I discuss the intermediate case in which $t_j = \frac{n-1}{n}\gamma \leq \beta \leq \frac{n-1}{n}$. In this scenario the maximizing argument is given by:[53]

$$t_i^* = \begin{cases} 1 & \text{if } \frac{n}{\eta \cdot f(n)} > 1 \\ \frac{n}{\eta \cdot f(n)} & \text{if } 1 - (\beta - \frac{n-1}{n} \cdot \gamma) \leq \frac{1}{\eta \cdot f(n)} \leq 1 \\ 1 - (\beta - \frac{n-1}{n} \cdot \gamma) & \text{if } 1 - \frac{n}{\eta \cdot f(n)} \leq 1 - (\beta - \frac{n-1}{n} \cdot \gamma) \end{cases} \tag{12}$$

More precisely, for all dictators that fulfill the requirement that $1 - \beta \leq \frac{n}{\eta \cdot f(n)}$ but $1 - \left(\beta - \frac{n-1}{\mathrm{n}} \cdot \gamma\right)$, it holds that while those dictators would live up to the expectations of the recipients under perfect knowledge, they would violate the expectations of the recipients in case that their actions are no longer directly observable. Contrary, the share of people who act selfishly increase, since people are less rewarded for pro-social decision. The share of dictators who completely act selfishly under imperfect knowledge but show pro-social behavior under perfect knowledge is given by the share of dictators for which the following to inequalities hold $\frac{n}{\eta \cdot f(n)} > 1$ as well as $\frac{1}{\eta \cdot f(n)} < 1$. Again, those who are willing to live up to the expectations of others in both models still behave n times more selfishly than if actions are attributable. Thus, also the comparison of case three with the benchmark model leads to the prediction that under imperfect information dictators act less pro-socially: in summary, having discussed all three cases, the theory supports a decline in pro-social behavior in instance of imperfect knowledge.

[53] Proof of this proposition is provided in appendix B.2.2.

B.3 Surprise Model

B.3.1 Introduction of the Surprise Model

In this section I introduce an extension of the simple guilt model: a surprise model that incorporates that some dictators get pleasure from surprising others—as proposed by Khalmetski et al. (2015).[54]

As in the simple guilt model, making a decision d an individual i is not only interested in his own payoff t_i, but is reluctant to betray the expectations of other js as well. However, in contrast to the simple guilt model, some is may even want to surprise others by exceeding their expectations (surprise-seeking dictators), e.g., by making generous tips or demanding unexpected-low prices for high-quality credence goods. These preferences are represented by a utility function, which is as in the simple guilt model split up into two functions, namely the monetary utility function $m(t_i(d_i))$ and a psychological utility function $P(S(\beta,),C(\beta))$. The concave monetary utility function is defined as follows:

$$m\big(t_i(d)\big) \text{ with } \frac{d\,m\big(t_i(d)\big)}{dt_i(d)} > 0 \text{ and } \frac{d\,m\big(t_i(d)\big)}{d\,t_i(d)^2} \leq 0 \text{ } \textit{e.g.} \text{ } m\big(t_i(d)\big) = ln\big(t_i(d)\big) \qquad (1)$$

The psychological utility function incorporates positive and negative deviations from the reference point (other agents' perceived beliefs) and is based upon a surprise function $S\big(t_j(d),\ \beta\big)$, measuring to what degree j is disappointed or in contrast to the simple guilt model, also to what extent j surprised by i's decision's outcome. This functional form allows me to incorporate the key findings of (Khalmetski et al., 2015). Moreover, as in the simple guilt model, it includes a term specifying how an agent takes external shocks (e.g., pro-social behavior by a third party) into account $C\big(t_j(d), \beta\big)$.

As in the simple guilt model, $S\big(t_j(d),\ \beta\big)$ relies on the idea that j forms first order beliefs about his payoff t_j ($E_j[t_j] = \alpha$) and i forms seconder beliefs about j's first order beliefs ($E_i[E_j[t_j]] = E_i[\alpha] = \beta$) and evaluates her behavior and the consequences of her behavior with regard to her second order beliefs. Now, note that $\eta \geq 0$ measures how prone i is towards guilt and $\mu \geq 0$ how much i seeks to surprise other js. Further, assume $\eta \geq \mu$, i.e., i suffers more from falling short of the recipients' expectations by one Dollar than she gains by exceeding their second order beliefs by the same amount.[55] Overall, I assume that

$$\frac{dS\big(t_j(d_i), \beta\big)}{dt_i} \geq 0 \text{ and. } \frac{dS\big(t_j(d_i), \beta\big)}{dt_i^2} \leq 0 \qquad (2)$$

[54] However, in contrast to Khalmetski, Ockenfels, and Werner's (2015) model I consider utility functions of a certain functional form with a deterministic, instead of a stochastic reference point for mainly three reasons. First, empirically testable predictions can more easily be derived from a utility function with a determined functional form, because these functions have less degrees of freedom. Secondly, these functions are more vivid and illustrative. Thirdly, I consider deterministic reference points, because stochastic reference points entail a level of complexity in the decision-making process that a dictator will likely not consider. Note that if I substitute the concave and increasing utility function of money in the model developed by Khalmetski et al. (2015) with a logarithmic function and if I substitute the distribution in their psychological payoff function with a degenerated distribution function, I will arrive at the utility function introduced in this section. Consequently, the derived propositions of my proposed models are qualitatively in line with the predictions of Khalmetski et al. (2015).

[55] This assumption is consistent with other reference-dependent utility models (e.g., Kahneman and Tversky, 1979; Köszegi and Rabin, 2006a). Moreover, Khalmetski et al. (2015) corroborate this assumption with their empirical findings.

A surprise function $S\big(t_j(d_i), \beta\big)$ that satisfies these assumptions has the following form:

$$S\big(t_j(d_i), \beta\big) = -\eta \cdot max\{0, \beta - t_j(d)\} + \mu \cdot max\{0, t_j(d) - \beta\}. \tag{3}$$

As in the simple guilt model, *i* does not necessarily feel guilty if the alleged expectations of *j* are violated, but whether he experiences the feeling of guilt depends on how much of perceived *j*'s experience loss is due to his behavior. The same reasoning applies if *j* experiences surprise. Thus, the psychological payoff function includes a correction term

$$C\big(t_j(d), \beta\big) = min_d\big(\eta \cdot max\{0, \beta - t_j(d)\}\big) - min_{d_i}\big(0, \mu \cdot max\{0, t_j(d) - \beta\}\big) \tag{4}$$

It measures *j*'s disappointment in case that *i* decides most altruistically or surprise in case she decides most selfishly. Hence, the psychological payoff function is defined as

$$P_i\left(S\big(t_j, \beta\big), C\big(t_j(d), \beta\big)\right) = S\big(t_j(d), \beta\big) - C\big(t_j(d), \beta\big) \tag{5}$$

The overall utility function thus can be written as

$$U(t_i, \beta) = m(t_i) + P\left(S\big(t_j(d), \beta\big)\right) = m(t_i) + S\big(t_j(d), \beta\big) - C\big(t_j(d_i), \beta\big) \tag{6}$$

Similar to both other introduced models, the surprise model entails a weighting function $f(n)$ (s.t. $\frac{df(n)}{dn} \leq 0$ and $f(n) \leq n$) that captures the hypothesized group discounting effect. The utility function of a single *i* in a multi-agent setting is therefore:

$$\begin{aligned} U(t_i, \beta) &= m(t_i(d_i)) + f(n) \cdot \sum_{j=1}^{n} P(S(t_j(d), \beta, C(t_j(d), \beta))) \\ &= m(t_i(d_i)) + f(n) \cdot \sum_{j=1}^{n} S(t_j(d), \beta) - C(t_j(d), \beta) \end{aligned} \tag{7}$$

A more tractable parametric specification of the overall utility function has the following form:

$$\begin{aligned} U(t_i, \beta) &= m(t_i(d)) + f(n) \cdot \sum_{j=1}^{n} P(S(t_j(d), \beta, C(t_j(d), \beta))) \\ &= m(t_i(d)) + f(n) \cdot \sum_{j=1}^{n} S(t_j(d_i), \beta) - C(t_j(d), \beta) \\ &= ln(t_j(d)) + f(n) \cdot \sum_{j=1}^{n} -\eta \cdot max\{0, \beta - t_j(d)\} + \mu \cdot max\{0, t_j(d) - \beta\} \\ &+ min_d(\eta \cdot max\{0, \beta - t_j(d)\}) - min_d(\mu \cdot max\{0, t_j(d) - \beta\}). \end{aligned} \tag{8}$$

B.3.2 The Dictator's Optimal Transfer Payment

Consider a utility function of the following form:

$$\begin{aligned} U(t_i, \beta) &= ln(t_i) \\ &- f(n) \cdot \sum_{i=1}^{n} \eta \cdot \max\{(\beta) - (1 - \frac{n-1}{n} \cdot \gamma - \frac{t_i}{n}); 0\} + \mu \cdot max\{(1 - \frac{n-1}{n} \cdot \gamma - \frac{t_i}{n}) - (\beta); 0\} \\ &+ \eta \cdot max\{\beta - (1 - \frac{n-1}{n} \cdot \gamma); 0\} - \mu \cdot max\{(1 - \frac{n-1}{n} \cdot \gamma - \frac{1}{n}) - \beta; 0\} \text{ s.t. } 0 \leq t_i \leq T = 1 \end{aligned} \tag{1}$$

In addition, it holds that $0 \leq \mu \leq \eta$. The dictator maximizes her utility with respect to her payoff t_i given her second order beliefs $0 \leq \beta \leq 1$. Depending on different parameter values of the dictator's first order beliefs about co-dictators' decisions γ, three different cases have to be considered:

- In the first case the perceived recipients' expectations will irrespective of the individual dictator's decision always be violated, since $1 - \frac{(n-1) \cdot \gamma}{n} \leq \beta$
- In the second case the perceived recipients' expectations will depending on the dictators' decision either be violated, met or surpassed, since $1 - \frac{(n-1) \cdot \gamma + 1}{n} \leq \beta \leq 1 - \frac{(n-1) \cdot \gamma}{n}$.
- The third case captures the situation where the perceived recipients' expectations will be surpassed irrespective of the individual dictator's decision, since $\beta \leq 1 - \frac{1 + (n-1) \cdot \gamma}{n}$.

Case 1: First consider the case where $1 - \frac{(n-1) \cdot \gamma}{n} \leq \beta$. In this case the utility function simplifies to

$$\begin{aligned} U(t_i, \beta) = ln(t_i) - f(n) \cdot \sum_{i=1}^{n} \eta \cdot (\beta - (1 - (\frac{n-1}{n} \cdot \gamma + \frac{t_i}{n}))) \\ + \eta \cdot (\beta - (1 - (\frac{n-1}{n} \cdot \gamma))) \text{ if } 0 \leq t_i \leq 1 \end{aligned} \tag{2}$$

The first order condition of the function yields

$$\frac{\partial U(t_i, \beta)}{\partial t_i} = \frac{1}{t_i} - f(n) \cdot \eta \tag{3}$$

Furthermore, note that $U(t_i, \beta)$ is strictly concave. Thus, setting $\frac{\partial U(t_i, \beta)}{\partial t_i} = 0$ and solving for $t^{\dagger}{}_i$ yields the optimal transfer payment of the maximization problem. It follows that the optimal dictator payoff of the unconstrained maximization problem is given by

$$t_i^{\dagger} = \frac{1}{f(n) \cdot \eta}. \tag{4}$$

Now, taking into account that the function $U(t_i, \beta)$ is constrained by $0 \leq t_i \leq 1$, the utility maximizing recipient's payoff $t_i^{\dagger}{}_i$ of the constrained function $U(t_i, \beta)$ is given by

$$t_i^{\dagger} = \begin{cases} \frac{1}{f(n) \cdot \eta} & \text{if } \frac{1}{f(n) \cdot \eta} \leq 1 \\ 1 & \text{if } \frac{1}{f(n) \cdot \eta} \geq 1 \end{cases} \tag{5}$$

Case 2: Now consider the case where $1-\frac{(n-1)\cdot\gamma+1}{n} \leq \beta \leq 1-\frac{(n-1)\cdot\gamma}{n}$, i.e., it is up to the individual dictator *i* whether the expectations of the recipients are violated or surpassed. Trivially,

$$\max\{(\beta - (1-\frac{(n-1)\cdot\gamma}{n}); 0\} = 0 \text{ and } \max\{(1-\frac{(n-1)\cdot\gamma+1}{n} - \beta; 0\} = 0 \tag{6}$$

Therefore, the utility function can be simplified to

$$\begin{aligned} U(t_i,\beta) = ln(t_i) + f(n)\sum_{i=1}^{n} -\eta\cdot\max\{\beta - \left(q - \frac{n-1}{n}\cdot\gamma - \frac{t_i}{n}\right); 0\} \\ + \mu\cdot\max\{\left(1 - \frac{n-1}{n}\cdot\gamma - \frac{t_i}{n}\right) - \beta; 0\} \text{ s.t. } 0 \leq t_i \leq 1 \end{aligned} \tag{7}$$

The dictator maximizes her utility with respect to her payoff t_i given her second order beliefs β. Themaximum operators included in $U(t_i,\beta)$ causes the function to be not differentiable anymore. These operators separate the function into two parts:

$$U_1(t_i,\beta) = ln(t_i) - f(n)\cdot\eta\cdot\sum_{j=1}^{n}\beta - (T - (\frac{n-1}{n}\cdot\gamma + \frac{t_i}{n})) \qquad \text{if } t_i \geq n - n\cdot\beta - (n-1)\cdot\gamma \tag{8}$$

$$U_2(t_i,\beta) = ln(t_i) + f(n)\cdot\mu\cdot\sum_{j=1}^{n}((T - (\frac{n-1}{n}\cdot\gamma + \frac{t_i}{n})) - \beta) \qquad \text{if } t_i < n - n\cdot\beta - (n-1)\cdot\gamma \tag{9}$$

First, consider an unconstrained version of $U(t_i,\beta)$. Its first order condition is given by

$$\frac{\partial U_1(t_i,\beta)}{\partial t_i} = \frac{1}{t_i} - f(n)\cdot\eta = 0 \tag{10}$$

Furthermore, note that $U_1(t_i,\beta)$ is strictly concave. Thus, setting $\frac{\partial U_1(t_i,\beta)}{\partial t_i} = 0$ and solving for $t_{i,un}^{**}$ yields the optimal transfer payment of the unconstrained maximization problem. It follows that the optimal transfer payment of the unconstrained maximization problem is given by $t_{i,un}^{**} = \frac{1}{t_i}$. Consider that the function $U_1(t_i,\beta)$ is constrained by $t_i \leq T = 1$ and $t_i \geq n - \beta - (n-1)\cdot\gamma$, the utility maximizing payoff $t^{**}{}_I$ of the constrained function $U_1(t_i,\beta)$ is given by

$$t_i^{**} = \begin{cases} 1 & \text{if } \frac{1}{f(n)\cdot\eta} > T = 1 \\ \frac{1}{f(n)\cdot\eta} & \text{if } \frac{1}{f(n)\cdot\eta} < n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \leq T = 1 \\ n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) & \text{if } n\cdot(1-\beta-\frac{(n-1)\cdot\gamma}{n}) \geq \frac{1}{f(n)\cdot\eta} \end{cases} \tag{11}$$

Second, consider an unconstrained version of

$$U_2(t_i, \beta) = ln(t_i) + f(n) \cdot \mu \cdot \sum_{j=1}^{n} \left(\left(T - \left(\frac{n-1}{n} \cdot \gamma + \frac{t_i}{n} \right) \right) - \beta \right) \tag{12}$$

The first order condition is given by:

$$\frac{\partial U_2(t_i, \beta)}{\partial t_j} = \frac{1}{t_i} + f(n) \cdot \mu = 0. \tag{13}$$

Furthermore, note that $U_2(t_j,\beta)$ is strictly concave. Thus, setting $\frac{\partial U_2(t_i,\beta)}{\partial t_j} = 0$ and solving for $t_{i,un}^{***}$ yields the optimal transfer payment of the unconstrained maximization problem. It follows that the optimal transfer payment of the unrestricted maximization problem is given by

$$t_{i,}^{***} = \frac{1}{f(n) \cdot \mu}. \tag{14}$$

Taking into account that the function $U_2(t_j,\beta)$ is constrained by $T \geq t_j \geq 1 - \beta$, the utility maximizing transfer payment $t_i^{***} = \frac{1}{f(n) \cdot \mu}$ of the constrained function $U_2(t_i,\beta)$ is as follows:

$$t_i^{***} = \begin{cases} \frac{1}{f(n) \cdot \mu} & \text{if } \frac{1}{f(n) \cdot \mu} \geq n \cdot (\beta - \frac{(n-1) \cdot \gamma}{n}) \\ n \cdot (1 - \beta - \frac{(n-1) \cdot \gamma}{n}) & \text{if } n \cdot (1 - \beta - \frac{(n-1) \cdot \gamma}{n}) \leq \frac{1}{f(n) \cdot \mu} \end{cases} \tag{15}$$

Overall, the maximizing argument is thus given by

$$t_j^* = \begin{cases} 1 & \text{if } \frac{1}{f(n) \cdot \eta} > T = 1 \\ \frac{1}{f(n) \cdot \eta} & \text{if } \frac{1}{f(n) \cdot \eta} \leq n \cdot (\beta - \frac{(n-1) \cdot \gamma}{n}) \\ n \cdot (1 - \beta - \frac{(n-1) \cdot \gamma}{n}) & \text{if } \frac{1}{f(n) \cdot \eta} > n \cdot (\beta - \frac{(n-1) \cdot \gamma}{n}) \text{ and } \frac{1}{f(n) \cdot \mu} < n \cdot (\beta - \frac{(n-1) \cdot \gamma}{n}) \\ \frac{1}{f(n) \cdot \mu} & \text{if } \frac{1}{f(n) \cdot \mu} \geq n \cdot (\beta - \frac{(n-1) \cdot \gamma}{n}) \end{cases} \tag{16}$$

Case 3: Third, consider $\beta \leq 1 - \frac{1+(n-1)\cdot\gamma}{n}$. Hence, the utility function simplifies to

$$U(t_j, \beta) = ln(t_i) + f(n) \cdot \sum_{j=1}^{n} \mu \cdot \left(\left(T - \frac{(n-1)\cdot\gamma + t_i}{n}\right) - \beta\right) - \mu\left(\left(T - \frac{n-1}{n}\cdot\gamma\right) - \beta\right) \tag{17}$$

The first order condition of the function yields

$$\frac{\partial U(t_i, \beta)}{\partial t_i} = \frac{1}{t_i} - f(n)\cdot\mu = 0 \tag{18}$$

Furthermore, $U(t_j,\beta)$ is strictly concave. Thus, setting $\frac{\partial U(t_i,\beta)}{\partial t_i} = \frac{1}{t_i} - f(n)\cdot\mu = 0$ and solving for t_i^{**} yields the optimal transfer payment of the unconstrained maximization problem. It follows that the optimal transfer payment of the unrestricted maximization problem is given by

$$t_i^{**} = \frac{1}{f(n)\cdot\mu} \tag{19}$$

Now, taking into account that the function $U(t_i,\beta)$ is constrained by $0 \geq t_i \geq 1$, the utility maximizing transfer t_i^{*}

$$t_i^{\star} = \begin{cases} 1 & \text{if } \frac{1}{f(n)\cdot\mu} > 1 \\ \frac{1}{f(n)\cdot\mu} & \text{if } \frac{1}{f(n)\cdot\mu} \leq 1 \end{cases} \tag{20}$$

C Instructions

Introduction

[All participants]

Dear participant,

you are taking part in an economic experiment of the University of Cologne. Experiments such as today's help us to collect reliable data about human decision making that is needed for scientific publications.

This economic experiment is anonymous. The data is collected in a way that we cannot link individual responses to participants' identities. Moreover, participants will receive no information about the identity of other participants.

Information Concerning the Course of the Experiment and the Bonus Payments

Please read the following instructions carefully. A clear understanding of the instructions will help you make better decisions and increase your payoff. All statements made in these instructions are true. In particular, all actions will be implemented exactly in the way they are described.

You will receive a fixed amount of $0.25 for completing this experiment. During this economic experiment you will be paid additional money depending on your stated decisions. We will explain you in detail how you can earn a significant bonus.

In addition, some of your decisions have real consequences on other participants. To every decision you stated or question you have answered it can be related to in later parts of the experiment. The experiment consists of two parts. Each part will be introduced on a screen with the header "instructions". These instructions will explain in detail what the respective part of the experiment is about.
Click on "next" if you have read the instructions.

Instructions Part 1

[Player A (Dictator) - BASELINE TREATMENT]

In this session 8 (respectively 40) participants participate in total.
In the first part 2 participants interact each. One of these two participants is assigned type A and the other participant is assigned type B.

You are assigned **type A**.

Every participant A is randomly matched with one participant B. Every participant A can decide how much she / he wants to take out of a pot that contains $1.00. Every amount between $0.00 and $1.00 in steps of $0.01 can be taken. Participant B earns whatever is left in the pot.

This means:
Earnings of participant A = amount taken by participant A
Earnings of participant B = $1.00 - amount taken by participant A

Participant B cannot act. Before participant A makes his/her decision, participant B is asked about his/her guess of the average amount of money taken by an A-participant. Participant A is informed at the end of the experiment what the guess of his/her matched B-participant was. However, participant A can condition his/her taking decision on different possible estimates.

In other words:

Participant A is telling us by the use of the attached table which transfer he/she likes to give participant B for each level of participant B's guess. The computer will then put into effect the answer that was conditional on the actual stated guess closest.

Decision Table

If participant B expects that the average participant A takes $0.00,	I will take $???
If participant B expects that the average participant A takes $0.10,	I will take $???
If participant B expects that the average participant A takes $0.20,	I will take $???
If participant B expects that the average participant A takes $0.30,	I will take $???
If participant B expects that the average participant A takes $0.40,	I will take $???
If participant B expects that the average participant A takes $0.50,	I will take $???
If participant B expects that the average participant A takes $0.60,	I will take $???
If participant B expects that the average participant A takes $0.70,	I will take $???
If participant B expects that the average participant A takes $0.80,	I will take $???
If participant B expects that the average participant A takes $0.90,	I will take $???
If participant B expects that the average participant A takes $1.00,	I will take $???

Comprehension Test Part 1

[Player A (Dictator) - BASELINE TREATMENT]

We now ask you the following questions to check whether you have correctly understood the instructions. You can only continue with the actual experiment (i.e., make your own payoff relevant decisions) if you correctly answer these control questions. If you answer the questions incorrectly, you will receive an error message.

Imagine the following hypothetical scenario and note that we randomly come up with the numbers used here. In other words, the decisions described here do not represent actual decisions made by other participants. A hypothetical participant A stated the following decisions:

Participant A's Decisions

If participant B expects that the average participant A takes	participant A will take
$0.10	$ *random number*.
$0.20	$ *random number*.
$0.30	$ *random number*.
$0.40	$ *random number*.
$0.50	$ *random number*.
$0.60	$ *random number*.
$0.70	$ *random number*.
$0.80	$ *random number*.
$0.90	$ *random number*.
$1.00	$ *random number*.

Participant B guessed that participant A takes an amount of $ *random number*.

How much would this hypothetical participant A receive from this part of the experiment? How much would this hypothetical participant B receive from this part of the experiment?

Part 1

[Player A (Dictator) - BASELINE TREATMENT]

Now the real implementation of the payoffs starts. Which amount (up to $1.00) do you want to take out of the pot, if participant B expects that $0.00, $0.10, $0.20, $0.30, $0.40, $0.50, $0.60, $0.70, $0.80, $0.90 or $1.00 will be taken? You have to state a decision for every possible guess.

Additional Questions Part 1

[Player A (Dictator) - BASELINE TREATMENT]

We want you to guess how much (up to $1.00) the other A-participants on average take out of pot in case that their matched B-participants expected that an A-participant takes on average $0.00, $0.50 or $1.00?

Each participant whose sum of the guesses differs not more than $0.03 from the true average amount will win $0.50 extra. Please enter your guesses in the box below.

Instructions Part 1

[Player B (Recipient) - BASELINE TREATMENT]

In this session 8 (respectively 40) participants participate. In the first part 2 participants interact each. One of these two participants is assigned type A and the other participant is assigned type B.

You are assigned **type B**.

Every participant A is randomly matched with one participant B. Every participant A can decide how much he/she wants to take out of a pot that contains $1.00. Every amount between $0.00 and $1.00 in steps of $0.01 can be taken. Participant B earns whatever is left in the pot.

This means:

Earnings of participant A = amount taken by participant A
Earnings of participant B = $1.00 - amount taken by participant A

Further note that participant B cannot act.

Comprehension Test

[Player B (Recipient) - BASELINE TREATMENT]

We now ask you the following questions to check whether you have understood the instructions correctly. You can only continue with the actual experiment if you correctly answer these control questions. If you answer the questions incorrectly, you will receive an error message. Imagine the following hypothetical scenario and note that we randomly come up with the numbers used here. In other words, the decisions described here do not represent actual decisions made by other participants.
A hypothetical other participant type A stated the following decisions: I will take $ *random number* out of the pot.

How much would this hypothetical participant A receive from this part of the experiment?
How much would his/her matched participant B receive from this part of the experiment?

Declaration of Consent

[Player B (Recipient) - BASELINE TREATMENT]

Considering a fair choice by participant A, what amount of money should he/she take out of the pot? Please enter your number in the box below.

Please give us your permission to reveal your guess to your matched A-participant. If you give us your permission, your matched A-participant will be informed about your guess at the end of the experiment (i.e., after his/her decision). However, he/she might relate his/her taking decision to different possible guesses. Your matched participant will receive no personal information about you other than your guess.

Do you agree with this procedure?

Instructions Part 1

[Player A (Dictator) - MULTI-AGENT TREATMENT]

In this session 40 (8) participants participate in total.
In this part the 40 (8) participants will be assigned either type A or type B.

You are assigned type A

The 20 (4) A-participants and the 20 (4) B-participants interact.

There is a common pot that contains $20.00 ($4.00). Every A-participant can decide how much he/she wants to take out of the common pot. Every amount between $0.00 and $1.00 in steps of $0.01 can be taken. The amount of money that remains in the pot after every of the 20 (4) A-participants has made his/her decision will be split equally among the B-participants.

This means:

Earnings of participant A = amount taken by participant A
Earnings of participant B = ($20 ($4.00) - sum of amount taken by 20 (4) A-participants) / 20 (4)

B-Participants cannot act. Before A-participants make their decision, every B-participant is asked about his/her guess of the average amount of money taken by an A-participant. The A-participants are only informed at the end of the experiment what the guess of the average B-participants in such a situation was, but they can condition their transfer on different possible estimates.

Participant A tells us by the use of the attached table which transfer he/she would like to give the B-participants for each possible guess. Then participant B's guess is rounded to the nearest 10 cents. Depending on this rounded guess, the computer will then put into effect the related choice of participant A from the table.

We will inform all participants after the end of the experiment about their individual payoff from this stage. We will not inform participants about the individual take decisions of A-participants.

Decision Table

If B-participants expect that the average participant A takes $0.00,	I will take $???
If B-participants expect that the average participant A takes $0.10,	I will take $???
If B-participants expect that the average participant A takes $0.20,	I will take $???
If B-participants expect that the average participant A takes $0.30,	I will take $???
If B-participants expect that the average participant A takes $0.40,	I will take $???
If B-participants expect that the average participant A takes $0.50,	I will take $???
If B-participants expect that the average participant A takes $0.60,	I will take $???
If B-participants expect that the average participant A takes $0.70,	I will take $???
If B-participants expect that the average participant A takes $0.80,	I will take $???
If B-participants expect that the average participant A takes $0.90,	I will take $???
If B-participants expect that the average participant A takes $1.00,	I will take $???

Comprehension Test

[Player A (Dictator) - MULTI-AGENT TREATMENT]

We now ask you the following questions to check whether you have understood the instructions correctly. You can only continue with the actual experiment if you correctly answer these control questions. If you answer the questions incorrectly, you will receive an error message.
Imagine the following hypothetical scenario and note that we randomly come up with the numbers used here. In other words, the decisions described here do not represent actual decisions made by other participants.
Assume for reasons of simplicity that there are only 2 participants of type A (A1 and A2) and 2 participants of type B (B1 and B2). That is payoffs are for the moment calculated as follows.

This means:

Earnings participant A = amount taken by participant A
Earnings participant B = ($2.00 - sum of amount taken by 2 A-participants) / 2

The common pot contains an amount of $2.00. Participant A1 and participant A2 state the following decisions.

Participants type B guessed that an average participant type A takes an amount of $ *random number*. A-participants made the following decisions:

A–Participants' Decisions

If B-participants expects on average that the average participant A takes	participant A1 will take	participant A2 will take
$0.10	$ *random number*	$ *random number*.
$0.20	$ *random number*	$ *random number*.
$0.30	$ *random number*	$ *random number*.
$0.40	$ *random number*	$ *random number*.
$0.50	$ *random number*	$ *random number*.
$0.60	$ *random number*	$ *random number*.
$0.70	$ *random number*	$ *random number*.
$0.80	$ *random number*	$ *random number*.
$0.90	$ *random number*	$ *random number*.
$1.00	$ *random number*	$ *random number*.

How much would the hypothetical participant type A1 receive from this part of the experiment?

How much would the hypothetical participant type A2 receive from this part of the experiment?

How much would the hypothetical participant type B1 receive from this part of the experiment?

How much would the hypothetical participant type B2 receive from this part of the experiment?

Dictator’s Decision

[Player A (Dictator) - MULTI-AGENT TREATMENT]

Now the real implementation of the payoffs starts. Please remember that you are in a group with 20 (4) A-participants and 20 (4) B-participants. Also note that we will not inform participants about the individual take decisions.

Which amount (up to $1.00) do to you want to take, if the 20 (4) B-participants expect that an average participant A takes $0.00, $0.10, $0.20, $0.30, $0.40, $0.50, $0.60, $0.70, $0.80, $0.90 or $1.00?
You have to state a decision for every possible guess.

Additional Questions Part 1

[Player A (Dictator) - MUlTI-AGENT TREATMENT]

We want you to guess how much (up to $1.00) the other A-participants on average take out of pot in case that B-participants expected that each A-participant takes on average $0.00, $0.50 or $1.00

Each participant whose sum of the guesses differs not more than $0.03 from the true average amount will win $0.50 extra.

Please enter your guesses in the box below.

Instructions Part 1

[Player B (Recipient) - MULTI-AGENT TREATMENT]

In this session 40 (8) participants participate in total.
In this part the 40 (8) participants will be assigned either type A or type B.

You are assigned **type B.**

The 20 (4) A-participants and the 20 (4) B-participants interact.

There is a common pot that contains $20 ($4.00). Every A-participant can decide how much he/she wants to take out of the common pot. Every amount between $0.00 and $1.00 in steps of $0.01 can be taken. The amount of money that remains in the pot after every of the 4 type A participants have made his/her decision will be split equally among the B-participants.

This means:
Earnings of participant A = amount taken by participant A
Earnings of participant B = ($20 ($4.00). - sum of amount taken by 20 (4) A-participants) / 20 (4)

Participants of type B cannot act.

Comprehension Test

[Player B (Recipient) - MULTI-AGENT TREATMENT]

We now ask you the following questions to check whether you have understood the instructions correctly. You can only continue with the actual experiment if you correctly answer these control questions. If you answer the questions incorrectly, you will receive an error message.

Imagine the following hypothetical scenario and note that we randomly come up with the numbers used here. In other words, the decisions described here do not represent actual decisions made by other participants. Assume for reasons of simplicity that there are only 2 participants of type A and 2 participants of type B. The common pot thus contains an amount of $2.00.

This means:

Earnings of participant A = amount taken by participant A
Earnings of participant B = ($2.00 - sum of the amount of money taken by the 2 A-participants) / 2

A hypothetical participant A1 takes an amount of $ *random number* out of the pot,
while a hypothetical participant A2 takes an amount of $ *random number* out of the pot.

How much would the hypothetical participant A1 receive from this part of the experiment?
How much would the hypothetical participant A2 receive from this part of the experiment?
How much would the hypothetical participant B1 receive from this part of the experiment?
How much would the hypothetical participant B2 receive from this part of the experiment?

Declaration of Consent

[Player B (Recipient) - MULTI-AGENT TREATMENT]

Considering a fair decision made by an A-participant, what amount of money should he/she take out of the pot? Please enter your number in the box below.

Please give us your permission to reveal your guess to the A-participants.

If you give us your permission, A–participants will be informed at the end of the experiment and thus after their decisions what the average guess of all B-participants was. However, they might condition their taking decisions to different possible guesses. Other participants will receive no personal information about you other than your guesses.

Do you agree with this procedure?

Instructions Part 1

[Player A (Dictator) - TRANSPARENCY TREATMENT]

In this session 40 (8) participants participate in total.
In this part the 40 (8) participants will be assigned either type A or type B.

You are assigned **type A**

The 20 (4) A-participants and the 20 (4) B-participants interact.

There is a common pot that contains $20.00 ($4.00). Every A-participant can decide how much he/she wants to take out of the common pot. Every amount between $0.00 and $1.00 in steps of $0.01 can be taken. The amount of money that remains in the pot after every of the 20 (4) A-participants has made his/her decision will be split equally among the B-participants.

This means:
Earnings of participant A = amount taken by participant A
Earnings of participant B = ($20 ($4.00) - sum of amount taken by 20 (4) A-participants) / 20 (4)

B-Participants cannot act. Before A-participants make their decision, every B-participant is asked about his/her guess of the average amount of money taken by an A-participant. The A-participants are only informed at the end of the experiment what the guess of the average B-participants in such a situation was, but they can condition their transfer on different possible estimates.

Participant A tells us by the use of the attached table which transfer he/she would like to give the B-participants for each possible guess. Then participant B's guess is rounded to the nearest 10 cents. Depending on this rounded guess, the computer will then put into effect the related choice of participant A from the table. We will inform all participants after the end of the experiment about their individual payoff from this stage.

We will also inform B-participants via email about the distribution of the individual take decisions. In particular we will tell all B-participants how many A-participants took what amount of money.

Decision Table

If B-participants expect that the average participant A takes $0.00,	I will take $???
If B-participants expect that the average participant A takes $0.10,	I will take $???
If B-participants expect that the average participant A takes $0.20,	I will take $???
If B-participants expect that the average participant A takes $0.30,	I will take $???
If B-participants expect that the average participant A takes $0.40,	I will take $???
If B-participants expect that the average participant A takes $0.50,	I will take $???
If B-participants expect that the average participant A takes $0.60,	I will take $???
If B-participants expect that the average participant A takes $0.70,	I will take $???
If B-participants expect that the average participant A takes $0.80,	I will take $???
If B-participants expect that the average participant A takes $0.90,	I will take $???
If B-participants expect that the average participant A takes $1.00,	I will take $???

Comprehension Test

[Player A (Dictator) - TRANSPARENCY TREATMENT]

We now ask you the following questions to check whether you have understood the instructions correctly. You can only continue with the actual experiment if you correctly answer these control questions. If you answer the questions incorrectly, you will receive an error message.

Imagine the following hypothetical scenario and note that we randomly come up with the numbers used here. In other words, the decisions described here do not represent actual decisions made by other participants.

Assume for reasons of simplicity that there are only 2 participants of type A (A1 and A2) and 2 participants of type B (B1 and B2). That is payoffs are for the moment calculated as follows.

This means:

Earnings of participant A = amount taken by participant A
Earnings of participant B = ($2.00 - sum of amount taken by 2 A-participants) / 2

The common pot contains an amount of $2.00. Participant A1 and participant A2 state the following decisions.

Participants type B guessed that an average participant type A takes an amount of $ *random number*.

A-participants made the following decisions:

A–Participants' Decisions

If B-participants expects on average that the average participant A takes	participant A1 will take	participant A2 will take
$0.10	$ *random number*	$ *random number*.
$0.20	$ *random number*	$ *random number*.
$0.30	$ *random number*	$ *random number*.
$0.40	$ *random number*	$ *random number*.
$0.50	$ *random number*	$ *random number*.
$0.60	$ *random number*	$ *random number*.
$0.70	$ *random number*	$ *random number*.
$0.80	$ *random number*	$ *random number*.
$0.90	$ *random number*	$ *random number*.
$1.00	$ *random number*	$ *random number*.

How much would the hypothetical participant type A1 receive from this part of the experiment?
How much would the hypothetical participant type A2 receive from this part of the experiment?
How much would the hypothetical participant type B1 receive from this part of the experiment?
How much would the hypothetical participant type B2 receive from this part of the experiment?

Dictator's Decision

[Player A (Dictator) - TRANSPARENCY TREATMENT]

Now the real implementation of the payoffs starts. Please remember that you are in a group with 20 (4) A-participants and 20 (4) B-participants. Also note that we will inform participants via email about the distribution of the individual take decisions. Which amount (up to $1.00) do to you want to take, if the 20 (4) B-participants expect that an average participant A takes $0.00, $0.10, $0.20, $0.30, $0.40, $0.50, $0.60, $0.70, $0.80, $0.90 or $1.00? You have to state a decision for every possible guess.

Which amount (up to $1.00) do to you want to take, if the 20 (4) B-participants expect that an average participant A takes $0.00, $0.10, $0.20, $0.30, $0.40, $0.50, $0.60, $0.70, $0.80, $0.90 or $1.00?

You have to state a decision for every possible guess.

Additional Questions Part 1

[Player A (Dictator) - TRANSPARENCY TREATMENT]

We want you to guess how much (up to $1.00) the other A-participants on average take out of pot in case that B-participants expected that each A-participant takes on average $0.00, $0.50 or $1.00

Each participant whose sum of the guesses differs not more than $0.03 from the true average amount will win $0.50 extra.

Please enter your guesses in the box below.

Instructions Part 1

[Player B (Recipient) - MULTI-AGENT TREATMENT]

In this session 40 (8) participants participate in total.
In this part the 40 (8) participants will be assigned either type A or type B.

You are assigned **type B.**

The 20 (4) A-participants and the 20 (4) B-participants interact.

There is a common pot that contains $20 ($4.00). Every A-participant can decide how much he/she wants to take out of the common pot. Every amount between $0.00 and $1.00 in steps of $0.01 can be taken. The amount of money that remains in the pot after every of the 4 type A participants have made his/her decision will be split equally among the B-participants.

This means:

Earnings of participant A = amount taken by participant A
Earnings of participant B = ($20.00 (4.00) - sum of amount taken by 20 (4) A-participants) / 20 (4)

Participants of type B cannot act. We will inform B-participants via email about the distribution of the individual take decisions. In particular we will tell all B-participants how many A-participants took what amount of money.

Comprehension Test

[Player B (Recipient) - MULTI-AGENT TREATMENT]

We now ask you the following questions to check whether you have understood the instructions correctly. You can only continue with the actual experiment if you correctly answer these control questions. If you answer the questions incorrectly, you will receive an error message.

Imagine the following hypothetical scenario and note that we randomly come up with the numbers used here. In other words, the decisions described here do not represent actual decisions made by other participants. Assume for reasons of simplicity that there are only 2 participants of type A and 2 participants of type B. The common pot thus contains an amount of $2.00.

This means:

Earnings of participant A = amount taken by participant A
Earnings of participant B = ($2.00 - sum of the amount of money taken by the 2 A-participants) / 2

A hypothetical participant A1 takes an amount of $ *random number* out of the pot,
while a hypothetical participant A2 takes an amount of $ *random number* out of the pot.

How much would the hypothetical participant A1 receive from this part of the experiment?
How much would the hypothetical participant A2 receive from this part of the experiment?
How much would the hypothetical participant B1 receive from this part of the experiment?
How much would the hypothetical participant B2 receive from this part of the experiment?

Declaration of Consent

[Player B (Recipient) - MULTI-AGENT TREATMENT]

Considering a fair decision made by an A-participant, what amount of money should he/she take out of the pot? Please enter your number in the box below.

Please give us your permission to reveal your guess to the A-participants.

If you give us your permission, A–participants will be informed at the end of the experiment and thus after their decisions what the average guess of all B-participants was. However, they might condition their taking decisions to different possible guesses. Other participants will receive no personal information about you other than your guesses.

Do you agree with this procedure?

Part 2

[All participants]

Finally, we ask you to answer questions concerning your assessment of four hypothetical situations as well as a few demographic questions. After you have answered all questions, you will receive your validation code.

1. Imagine the following situation: You won 1,000 dollars in a lottery. Considering your current situation, how much would you donate to charity? (Values between 0 and 1000 are allowed)

2. How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10 means you are "very willing to share". You can also use the values in-between to indicate where you fall on the scale.

□ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □7 □ 8 □ 9 □ 10

3. Imagine the following situation: You are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 dollars in total. The stranger does not want money for it. You carry six bottles of wine with you.
The cheapest bottle costs 5 dollars, the most expensive one 30 dollars.
You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you give?

□ $5 □ $10 □ $15 □ $20 □ $25

4. IHow do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly? Please use a scale from 0 to 10, where 0 means you are "not willing at all to incur costs to punish unfair behavior and a 10 means you are "very willing to incur costs to punish unfair behavior". You can also use the values in-between to indicate where you fall on the scale'

□ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □7 □ 8 □ 9 □ 10

3. Next, please indicate how responsible you feel for the outcome of part 1 of the experiment (outcome of your take decision). Please use a scale from 0 to 10, where 0 means you feel "not responsible for the outcome of part 1 of the experiment at all" and a 10 means you are "very responsible for the outcome of part 1 of the experiment". You can also use the values inbetween to indicate where you fall on the scale"

□ 0 □ 1 □ 2 □ 3 □ 4 □ 5 □ 6 □7 □ 8 □ 9 □ 10

Before you will receive your validation code, we ask you to answer a few final demographic questions.

6. How old are you?

7. What is your gender?

8. What is your highest education degree (in case you are still studying, pick the highest one you have already achieved)?

9. Please indicate the filed that best described your subject of study

10. Have you ever participated in a similar study?

11. Is there something you want to tell us about this study

3

A THEORY OF STRATEGIC DISCRIMINATION*

This is the first paper that studies how individuals account for other's group composition preferences when deciding whom to include in a group or network, in the absence of any personal taste or monetary benefits associated with the inclusion of a particular person, such as when landlords discriminate against black clients in response to current and future prejudiced white tenants. We study three different potential causes why individuals live up to the group composition preferences of their group members. First, individuals have altruistic feelings towards current group members and enhances current members' utility by selecting their preferred candidates. Second, they anticipate that other group members' cooperativeness dependent upon who has been selected and adapt their selection decision accordingly. Third, they want to trigger reciprocal behavior by signaling that they care for group composition preferences of others. We test our theory in a public good game in which we allow for endogenous team formation. Thereby, we show that discriminatory behavior in embedded context may appear even if the individual has no taste for either candidate or any reason to discriminate statistically, for both altruistic and strategic reasons and thereby identify a new major source for structural discrimination.

No man is an island, entire of itself; every man is a piece of the continent, a part of the main
—John Donne Devotions on Emergent Occasions Meditation XVII

1 Introduction

People engaging in markets are no metaphorical islands, instead they cooperate, reciprocate and interact with each other. These interactions regularly impose a social or economic effect on third parties due to individuals' embeddedness in social environments of interpersonal relations: for instance, if owners of apartment buildings rent out apartments, they will not only establish new relationships with the selected clients but will also alter the composition of the apartment buildings' communities and thereby the social dynamics in it. These re-compositions trigger positive or negative effects on the relationships between landlords and current residents, because they blame or appraise landlords for their selection decisions. When landlords select new clients, they consider these behavioral responses and likely select candidates favored by current residents. In a corporate environment, the inclusion of a new employee in a team influences its dynamics – such as employees' willingness to cooperate in the team and their personal attitudes towards the selecting manager– contingent on whether the manager satisfies current employees' group composition preferences or not when deciding whom to include. Thus, the inclusion of a single member can change an entire group or network structure, its dynamics, and operating principles. The acknowledgment of current network members' composition preferences is a pressing societal and political issue when their tastes arises from animus towards and prejudice against minorities (discrimination spill-over effects), such as when landlords discriminate against black clients in response to prejudiced current and future white clients (Ondrich et al., 1999; Zhao et al., 2006).

* I thank Dirk Sliwka, Christoph Engel, Oliver Kirchkamp, Frédéric Gruninger, Ulrich Becker and participants of various workshops for helpful discussions and comments on earlier versions of this paper. Financial support of the Max Planck Institute for Social Law and Social Policy is gratefully acknowledged.

In this paper we address *why and to what extent individuals selecting new group members consider current group members' group composition preferences as well as their behavioral responses in case their preferences are met or not met.* Thereby, we focus on two questions: first, we investigate to what extent a potential consideration of group composition preferences is caused by either selector's social preferences regarding other group members' utility or, alternatively, by strategic motives arising from the prospect of higher cooperation. Selectors might select other team members' preferred candidate to enhance their utility out of altruistic motives. Contrary, selectors might be motivated by (material) benefits resulting from an increase in team members' willingness to cooperate if their preferred candidate is selected. Thus, we investigate to what extent, and under which conditions individuals potentially discriminate –in the absence of any taste, or statistical reason against potential candidates– to attend current group members' preference. Thereby, they maintain or foster cooperation and homogeneity in groups with the aim to maximize their own profits. We denote this phenomenon as *strategic discrimination*. Second, we inquire if third parties condition their behavior and willingness to cooperate on whether the new member was deliberately included by a human or randomly selected by a computer, as well as whether selectors anticipate and account for such potential reciprocity effects.

To address these two questions, we study selection decisions in a stylized endogenous team formation context in which one out of two team members has the opportunity to select one additional member out of two candidates in exchange for a predetermined fee. The selector has no personal taste (Becker, 1957) or any statistical reason (Aigner & Cain, 1977; Arrow, 1973; Phelps, 1972) to pick either candidate by design, while the other team member prefers a particular candidate. If the fee exceeds the selector's reservation price, the computer will randomly select one candidate. The team's success depends on the contributions of its members to a team project –a public good (Holmström, 1982).

Eliciting selectors' reservation prices, and their selection choices with the Becker-deGroot-Marschack mechanism (1964) allows to identify if they consider current team members' group composition preferences in embedded settings. We assess to what extent the inclusion decision alters current team members' willingness to contribute to the public good and their attitude towards selectors contingent on whether selectors satisfy their preferences or not as well. To disentangle whether strategic considerations or social preference towards current team members explain the acknowledgment of group composition preferences, we exogenously vary to what extent current team members are enforced to cooperate. To study potential reciprocity effects, we study whether current team members show reciprocal behavior if selectors live up to their preferences by varying whether the new member is randomly selected by the computer or deliberately by the selector. To evaluate whether selectors anticipate this potential reciprocal behavior, we vary whether current members can observe who selected the candidate and elicit selectors' reactions to this variation in the transparency of actions.

Our results reveal that 60% of selectors consider across all treatments current team members' group composition preferences in the absence of an own taste for or in the absence of any statistical reason to select either candidate. This effect persists even if current team members cannot vary their contributions to the public good but is more pronounced if strategic incentives are present. Therefore, selectors' social preferences as well as strategic considerations together trigger that selectors account for others' group composition preferences. On average, the social preference dominates the strategic incentive effect, though heterogenous treatment effects are likely present: while the selection decisions of some selectors are mostly driven by strategic incentives, the social preference channel prevails for others. While our analysis provides some inconclusive indication that endogenous selection procedures trigger reciprocal behavior, we find no evidence that selectors accounted for environmental features, such as the transparency of the decision, which may leverage the established consideration effect.

The results' ethical implications are diverse. In some situations, the consideration of group composition preferences is from an ethical point of view irrelevant or even desirable: for instance, if selectors have incentives to include open-minded candidates into a team, because current team members refuse to cooperate with prejudiced people, then considerations of group composition preferences reinforce the groups' moral standards. In other situations, social embeddedness leads to reinforcement of discrimination, because agents accounting for prejudiced group composition preferences likely discriminate, though they have no own taste (Becker, 1957) or reasons for statistical discrimination (Phelps, 1972). Herein, discrimination is considered as an unjust and prejudiced distinction in the treatment of people based on associated status categories functionally irrelevant for the outcome in question (Merton, 1972) –i.e., discrimination is the unequal treatment of equals.[56] This kind of discrimination may be collateral or even unconscious: if high-status groups account for their members' group composition preferences by recruiting new members from their current members' homophile networks (e.g., including foremost white, middle-class males), minorities have less opportunities for social advancement. These dynamics provide a rationale why homophily often goes beyond direct links and individuals instead often prefer a homogenous set of indirect friends (Mele, 2020). On a broader level, self-reproduction (Luhmann 1986) of (elite) groups may contribute to the manifestation of structural discrimination against minorities– a societal issue which leaded, besides unjust police violence against black citizens in the U.S., to the emergence of the "Black Lives Matter" movement.

The contributions of this paper are two-fold: first, in respect to the economics of discrimination, it is one of the scarce studies investigating behavioral channels underlying discrimination in the absence of any taste-based (Becker, 1957), statistical (e.g., Arrow 1973, Phelps 1972, Guryan & Charles, 2013, Aigner & Cain, 1977) or combined reasons (Neilson & Ying, 2016)[57] for the selection of a particular candidate by studying *strategic discrimination*. While individuals are usually embedded in social environments of interpersonal relations, taste-based and statistical discrimination models abstract from the above-described strategic incentives arising from others' social preferences and study discriminatory behavior of atomized actors not affected by any social relation. However, in the world outside Economic textbooks, individuals will perform a variety of discriminatory behavior if they presume that, for instance, colleagues or friends gain utility from homophile groups or if they anticipate punishment and rewards as a response to their inclusion decisions. This paper aims to close this gap by studying whether individuals consider others' group composition preferences and strategically discriminate. These discrimination spill-over effects are a potential leverage for discrimination in embedded contexts and may exist irrespective of whether others' group composition preferences originate from prejudice, statistical reasons or the desire to act with a member of their own homophile networks.

Consequently, strategic considerations and altruistic preferences likely leads discrimination to be more widespread in group than in individual decision making[58] (Daskalova, 2018), and does not necessarily diminish in competitive markets. In fact, taste-based discrimination models (Becker, 1957) only predict

[56] Whether the unequal treatment of otherwise equals is considered discrimination depends on the basis of, the circumstances and the cultural context associated with the (discriminatory) act. Distinctions based on inborn characteristic such as race or sex are regularly considered clear norm violations, while distinctions based on talent are justifiable in job selection processes, there are considered unjust in organ donations assignment process.

[57] Numerous studies indicate that both taste-based as well as statistical discrimination can explain many patterns of discriminatory behavior in the field (Chaim Fershtman et al., 2005; Gneezy et al., 2012; List, 2004b)

[58] The experiment of Daskalova (2018) on the role of social identity in individual and joint assignment decisions revealed that in joint decision treatments decision makers hire significantly more in-group than other out-group candidates. While her paper was not designed to disentangle different potential factors, our theory as well as our experimental design allows to do so.

persistent discrimination if a large share of employees has strong preferences to be associated and interact with particular persons instead of others. However, there is empirical evidence that –in the absence of statistical reasons for discrimination– discriminatory patterns seem to be stable even if most employees are mildly prejudiced (Lang & Lehmann 2012). We offer an explanation for such patterns: contrary to previous market discrimination models (Becker, 1957) –which rest upon fixed-cost arguments and argue that due to arbitrage opportunities (taste-based) discrimination is inefficient and pervasive– we state that spill-over effects might explain the persistence of discrimination in markets, because shifts in the willingness to cooperate have a direct effect on employees' productivity. These spill-over effects imply that not only inborn characteristics and human capital but also to what extent the work environment is encouraging, cooperative, and non-hostile determines individual performance, discrimination and its actual costs (Bergmann & Darity, 1981). In a recent paper Lang & Spitzer (2020) promoted that economists should consider discrimination as a system and study how discriminatory acts ae reinforced by institutions and decision environments. We account for this proposal by studying how taste-based discrimination might be reinforced by spill-over effect arising from its anticipation.

Second, this paper addresses whether selection decisions' transparency in endogenous group formation settings may backfire. Thereby, it contributes to the question whether organizations struggling with structural discrimination and demanding a higher share of minority employees should use externally imposed selection procedures –such as quotas or outsourcing of recruiting processes– to overcome heterogeneity induced by endogenous group formation (e.g., Azmat, 2019; Charness et al., 2011). If the increase in performance in homophile teams could be inter alia explained by positive reciprocity triggered by a favorable selection, exogenously imposed selection criteria, such as quotas, might mitigate the productivity decline driven by negative reciprocity, since prejudiced team members will no longer punish selectors for actively selecting a minority candidate. If the acknowledgment of group composition preferences stem from selector's social preferences towards current group members, a limitation of the selectors' choice sets due to external selection criteria, may mitigate discrimination.

Finally, this paper offers an additional explanation why employee referral programs do not only increase the efficiency as well as the quality of the employer-employee matching process by reducing search friction (Cappellari & Tatsiramos, 2015; Ioannides & Loury, 2004; Topa, 2011) but also may positively affect the motivation, both, of the hired employee as well as of the referrer. The referrer may reciprocate for the acknowledgment of her team composition preferences by increasing her effort level.

The remainder of this paper is organized as follows. The next section will present the formal framework of our endogenous group formation model. Section 3 will present the experimental design and discuss how our derived hypotheses are tested in an experimental context. Section 4 presents the empirical results and entails a discussion of our findings. Section 5 concludes.

2 Model

2.1 Outline of the Endogenous Group Formation Model

In this section, we model the decision of a selector *A* concerning whom to include in a group or network contingent on its current members' composition preferences. Her inclusion decision affects other members' utility and behavior, such as when a manager hires an employee and thereby alters how pleasant the present staff perceives the work environment and in response how cooperative employees behave. We show how changes in selector *A's* social environment (i.e., how other members' preferences and cooperativeness contingent on the chosen candidate) affects *A's* decision. Thereafter, we apply our model to a public good game (Andreoni, 1988; Isaac & Walker, 1988) in which *A* includes a new group member. The public good game resembles cooperative interactions in the group or network.

When selecting whom to include into a group, or network, *A* assesses and compares the utility she and current group members will gain from relationships with each other and with the potential new member. Thereby, *A* considers the monetary, taste, and social preferences dimension from each relationship in each individual's utility function. The relative weight of the dimensions and their determining coefficients might be assessed differently by each individual. For instance, some employees strive to profit from a cooperation with ambitious colleagues who generate high revenues. Other priories harmonious work environment when searching for a job.

By assessing a relationship on the basis of distinct dimensions, we account for White's (2008) critique that the nature of an interaction cannot be characterized by a one-dimensional scale, such as the strength of a single tie (Granovetter, 1973) but should rather be assessed on multiple dimensions. A multi-dimensional assessment enables to explain why one evaluates a single relationship differently contingent on distinct settings, e.g., whether one is situated in private or professional contexts (in which potential monetary benefits from a relationship are usually higher). Lastly, our model allows for variations in individuals' preferences and assumes that affections can be non-mutual. Non-mutuality comprises that for instance *B* likes *C*, while *C* dislikes *B*.

The monetary dimension covers direct pecuniary benefits associated with a relationship. For instance, employers want to hirer high-skilled and cooperative employees to increase their profits. Participants in public good games profit monetarily from being in a group with reciprocal and cooperative players, and thus likely strive to interact with cooperators and socially exclude free riders, if possible (Dannenberg et al., 2020; Gürerk et al., 2014; Njozela et al., 2018).The taste dimension covers to what degree an individual has a preference for an interaction with another person, e.g., how much one appreciates to work with women. Finally, the social preferences dimension includes outcome-based social preferences (briefly referred to as *social preferences*), in particular altruism and its counterpart, envy– i.e., how much one cares for the well-being of others. Reciprocity arises endogenously in the model. In particular, individuals can be reciprocal altruists who gain utility if other altruists benefit from their relationship, and vice versa suffer from disutility if an envious person does so.

We assume that the taste and the social preference dimensions are interdependent, because empirical studies showed that strong affections manifest themselves in positive social preferences (e.g., positive reciprocity or altruism) towards candidates for which *As* have a taste and negative social preferences (e.g., spite, envy and negative reciprocity) towards candidates for which they have a distaste (e.g., Bohnet & Frey, 1999; Buchan, Johnson, & Croson, 2006; Cardella, 2015; Charness, Haruvy & Sonsino, 2007).

The following model specifies how *A* assesses selection decisions. We base our model on an adapted version of the type-based reciprocity model of Levine (1998) –a model of reciprocal altruism.[59] A utility maximizing individual i assesses the utility she receives at the terminal nodes of an extensive form game.

[59] Alternatively, we could also have incorporated a multi-player version of an intention-based reciprocity model and derive similar quantitative results (Cardella, 2016; Cox et al., 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Ockenfels & Bolton, 2000; Rabin, 1993). We choose the Levine (1998) model, because it is a hybrid model between an outcome and a reciprocity-based model and at the same time much simpler and relies on less assumptions than the often- cited intention-based reciprocity model by Falk and Fischbacher (2006)

The utility function of every individual i has the same structural form, although parameters and constraints vary between individuals and contexts. The utility of an individual i that is part of a group $I = 1, \ldots, n$ contingent on the payoffs of the other group members $j \in I$ is given by:

$$u_i = x_i^{\alpha} + \frac{1}{n-1} \sum_{j \in I \setminus i} \left[\sigma_{i \to j} + \left(x_j + \frac{1}{n-1} \sum_{k \in I \setminus j} \sigma_{j \to k} \right) \left((1 - \lambda_i) a_{i \to j} + \lambda_i a_{j \to i} \right) \right] \tag{1}$$

where

x_i^{α} denotes the monetary gains of individual i, with $0 \leq \alpha \leq 1$; and α measures the relative sensitivity of the monetary dimension while allowing for risk aversion
$\sigma_{i \to j}$ denotes the utility that individual *i* derives from interacting with individual j
x_j denotes the monetary gains of individual j
$a_{i \to j}$ $\in [-1,1]$ captures to what extent *i* experiences altruistic feelings or envy towards *j*
$a_{j \to i}$ $\in [-1,1]$ captures to what extent j experiences altruistic feelings or envy towards *i*
λ_i $\in [0,1]$ measures the relative importance of outcome-based and reciprocity-based social preferences of individual i

There are four components in each individual's social preferences dimension. First, if *i* has altruistic preferences towards *j* ($a_{i \to j} > 0$), then ceteris paribus the more *j* earns, the higher is the utility of *i*. Vice versa, if *i* experiences envy ($a_{i \to j} < 0$), *i* suffers from an increase in *j's* earnings.

Second, if ceteris paribus *j* is an altruistic type $a_{j \to i} > 0$, *i.e.*, *j* has altruistic preferences towards *i*, then the more *j* earns, the higher is the utility of *i*. This captures *i*'s reciprocal altruism, i.e., *i* acts reciprocal and experiences a feeling of warm glow while being benevolent towards an altruist. The model captures, both, positive and negative reciprocity: while i may be reciprocal in that sense that i derives utility from an increase in the welfare of an arbitrary altruistic j $(a_{j \to i} > 0)$, i suffers from an increase in the utility of an envious *j* ($a_{j \to i} < 0$).

Third, if *i* has altruistic preferences towards *j*, and *j* derives utility from interacting with *k* ($\sigma_{j \to k} > 0$), then ceteris paribus *i* derives utility if *j* in effect interacts with *k*. That is, *i* derives utility not only from *j*'s monetary well-being, but also for *j*'s utility from interacting with *j*'s preferred partner. Furthermore, this paper makes the intuitive assumption that *i* does not only care about the monetary payoffs of his interaction partner *j*, but also about the direct psychological payoffs of *j*. Altruists derive utility from the fact that their interaction partners feel happier and not only that they are richer. This implies that i derives more utility from *j*, the more *j* enjoys staying with decision maker i.

Fourth, if ceteris paribus *j* has altruistic preferences towards *i*, and *j* derives utility from interacting with *k*, then ceteris paribus *i* also derives utility if *j* in effect interacts with *k*. This captures that *i* 's reciprocal feelings also extend to non-monetary utility dimensions.

Furthermore, we assume that if i has a stronger taste for j than for k ($\sigma_{i\to j} > \sigma_{i\to k}$), then i's altruistic affections towards j are stronger than those towards k, i.e., $a_{i\to j} \geq a_{i\to k}$, as well.[60] This captures the proposed interdependence of taste and social preference dimensions.[61] Lastly, the term $\lambda_i \in [0,1]$ measures the relative importance of i's outcome-based and reciprocity-based social preferences.[62]

2.2 Application of the Endogenous Group Formation Model to a Public Good Game

Next, we apply our model to a multi-agent setting in which one person includes a candidate in a pre-existing team or network. The setting is reflected–except for minor alterations– in our experimental design. Consider a selector *A* and an additional current group member *B* who are both in the same group. The selector *A* can either select candidate *C* or *D* to become a new group member. *A*'s utility function conditional on the selection of *C* is given by:

$$U_A = (x_A^{\alpha} + 0.5\,(\sigma_{A\to B} + \sigma_{A\to C})) + 0.5\ (x_B + 0.5\,(\sigma_{B\to A} + \sigma_{B\to C}))((1-\lambda_A)a_{A\to B} + \lambda_A a_{B\to A}) \\ +\ 0.5\,(x_C\ +\ 0.5\,(\sigma_{C\to A} + \sigma_{C\to B}))((1-\lambda_A)a_{A\to C} + \lambda_A a_{C\to A}). \tag{2}$$

Respectively, the utility function of *B* conditional on the selection of *C* is given by:

$$U_B = (x_B^{\alpha} + 0.5\,(\sigma_{B\to A} + \sigma_{B\to C})) + 0.5\ (x_A + 0.5\,(\sigma_{A\to B} + \sigma_{A\to C}))((1-\lambda_B)a_{B\to A} + \lambda a_{A\to B}) \\ +\ 0.5\,(x_C\ +\ 0.5\,(\sigma_{C\to A} + \sigma_{C\to B}))((1-\lambda_B)a_{B\to C} + \lambda_B a_{C\to B}). \tag{3}$$

By interchanging *C* with *D,* one can derive the utility function of *A* or, respectively, *B* conditional on the selection of D. In a public good game, subjects choose what share t_i of their private endowment T they want to invest in a public good. The invested money is multiplied by a factor greater than one and smaller than the number of players. Finally, all group members receive an equal share of the public good. In a three-person public good game where all group members earn an endowment of T = 1 (we normalize the endowment to one) and the multiplication factor is without the loss of generality set to 2 < n, a standard value in public good games.

[60] In a seminal paper Tajfel (1970) finds that subjects favor in-group members in other-other allocation games and that the social distance towards an in-group level is smaller, which justifies our assumption. By comparing pro-social behavior in a class room as well as in an otherwise identical internet experiment, Charness et al. (2007) find that subjects behave significantly more pro-social on average in the class room experiment, in which people often have a close emotional proximity.

[61] Numerous experimental studies (Chen and Li, 2009; Falk and Zehnder 2013; Fershtman and Gneezy 2001; Ben-Ner et al. 2009; Ockenfels and Werner 2014) implicitly rely on the assumption that differences in with whom wants to be associated comes along with differences in outcome dependent social preferences towards different persons.

[62] If $\lambda_i = 0$ our model is a model of pure altruism, as among other proposed by Ledyard (1995). A reciprocal decision gains utility from his interaction partners psychological payoff, the more altruistic the interaction partner is. This also implies, that decision suffers from the fact, the more candidate j enjoys interacting with him while being spiteful towards i.

A's utility conditional on the selection of *C* is given by:

$$\begin{gathered}
U_A = \left(1 - \frac{1}{3}\, t_A + \frac{2}{3}\, (t_C + t_B)\right)^{\alpha} + \, 0.5\, (\sigma_{A \to B} + \, \sigma_{A \to C})) \\
+ 0.5 \left(1 - \frac{1}{3} t_B + \frac{2}{3} (t_A + t_C) + 0.5(\sigma_{B \to A} + \sigma_{B \to C})\right) \big((1 - \lambda_A) a_{A \to B} + \lambda_A a_{B \to A}\big) \\
+0.5 \left(1 - \frac{1}{3} t_C + \frac{2}{3} (t_A + t_B) + 0.5(\sigma_{C \to A} + \sigma_{C \to B})\right) \big((1 - \lambda_A) a_{A \to C} + \lambda_A a_{C \to A}\big) \\
\text{s.t.}\, 0 \leq t_i \leq T = 1
\end{gathered} \tag{4}$$

Analogously, one can derive *A*'s utility function conditional on the selection of *D*. *B*'s utility function conditional on the selection of *C* and *D* can be derived accordingly as well. We are interested in assessing reasons why *A* consider *B*'s group composition preferences. To simplify the discussion, we assume that *A* is enforced to cooperate, i.e., $t_A = 1$. To abstract from any statistical reasons to select either candidate, we assume that the contributions of *C* and *D* are deterministic, i.e., $t_C = t_D = 1$. Furthermore, *C* and *D* equally enjoy interacting with *A*, i.e., $\sigma_{C \to A} = \sigma_{D \to A}$; and have identical social preferences towards *A*, i.e., $a_{C \to A} = a_{D \to C}$. We consider that *A* is selfish towards C (and *D)*, i.e., $a_{A \to C} = a_{C \to D} = 0$ and is indifferent between interacting with either *C* or *D*, i.e., $\sigma_{A \to C} = \sigma_{A \to D} = 0$. Qualitative results remain the same if we depart from this paragraph's assumptions. Overall, *A*'s utility function reduces to:

$$\begin{gathered}
U_A = \left(1 - \tfrac{1}{3}\, t_A + \tfrac{2}{3}\, (t_C + t_B)\right)^{\alpha} \; +(1 - \tfrac{1}{3}\, t_B + \tfrac{2}{3}\, (t_A + t_C) \, + \, 0.5\, (\sigma_{B \to A} + \, \sigma_{B \to C}))(1 - \lambda_A) a_{A \to B} + \lambda_A a_{B \to A}) \\
U_A = \left(\tfrac{2}{3}\, (2 + t_B)\right)^{\alpha} \; +(1 - \tfrac{1}{3}\, t_B + \tfrac{4}{3} \, + \, 0.5\, (\sigma_{B \to A} + \, \sigma_{B \to C}))(1 - \lambda_A) a_{A \to B} + \lambda_A a_{B \to A}).\text{[63]}
\end{gathered} \tag{5}$$

Before making a selection decision, *A* assesses how her relationship with *B*, *B*'s behavior, and *A*'s gained utility changes contingent on different inclusion. We assume, without the loss of generality, that *B* gains more utility from a relationship with C than from a relationship with *D ($\sigma_{B \to C} \geq \sigma_{B \to D}$)*. Consequently, *B* has more pronounced social preference towards *C* ($a_{B \to C} \geq a_{B \to D}$) as well. *A* might be affected by *B*'s willingness to cooperate with team member *C*, because all three team members contribute to the same public good.[64] In what follows, we discuss that even in the absence of a personal taste, *A* might select *C*, even if the selection of *C* is costly, because *A* directly gains utility from acknowledging *B's* preferences or because *A* correctly anticipates that *B* is more willing to cooperate if *C* is selected. We also consider that by selecting *C*, *A* may signal her altruistic preferences towards *B* and thus seeks to trigger reciprocal behavior. To make the discussion tractable, we discuss all channels in isolation.

In a first step, we provide a formal argument that describes that if *A* has altruistic feelings towards *B* *($a_{A \to B} > 0$)*, then *A* considers *B*'s group composition preferences, because an altruistic *A* gains utility from the fact that *B* gains higher utility from a relationship with *C* ($\sigma_{B \to C} > \sigma_{B \to D}$). To begin, we abstract from reciprocal motives ($\lambda_A = 0$ and $\lambda_B = 0$) and strategic incentives (t_B is constant).

[63] Alternatively, we could also assume that *A* is not indifferent between interacting with *C* and *D*. Qualitatively, the results would not change but the theoretical discussion would be overly complex

[64] The difference in *B*'s utility from having a relationship with *C* compared to having a relationship with *D* emerges from differences in taste but it could in general also result from statistical discrimination practiced by *B*.

Proposition 1a: we assume that neither taste-based ($\sigma_{A\to C} = \sigma_{A\to D}$) nor statistical motives ($a_{C\to A} = a_{D\to C} = 0$; and $t_A = t_C = t_D = 1$) for the selection of a particular candidate exists. When additionally abstracting from reciprocal motives ($\lambda_A = \lambda_B = 0$), any strategic advantage of higher levels of cooperation arising from higher levels of altruism from B towards C ($a_{B\to C} = a_{B\to D}$), and strategic incentives ($t_B$is constant), it holds that an *A* with altruistic preferences towards *B* ($a_{A\to B} > 0$), chooses *C* if *B* experiences stronger social preferences towards *C* than towards *D* *($\sigma_{B\to C} > \sigma_{B\to D}$),*

Proof: see appendix A2.

When A has a taste or incentive for being associated with *C* instead of *D,* she acts as if she is willing to pay something to interact with C (Becker 1957, p.14). Her reservation price *p*, defined as her maximum willingness to be associated with C instead of D, thus constitutes a quantitative measure for *A's* taste. *A's* altruism towards *B* affects *A's* reservation price to select *C* and not *D*, i.e., the reservation price of an altruistic A increases in the difference in the utility *B* gains from a relationship with *C* in contrast to a relationship with *D* ($\sigma_{B\to C} - \sigma_{B\to D}$). This can be expressed formally by proposition 1b:

Proposition 1b: : under the assumptions that neither taste-based ($\sigma_{A\to C} = \sigma_{A\to D}$) nor statistical motive ($a_{C\to A} = a_{D\to C} = 0$; and $t_A = t_C = t_D = 1$) for the selection of a particular candidate exists, and abstracting from reciprocal motives ($\lambda_A = \lambda_B = 0$) as well as strategic incentives ($t_B$is constant), it holds that an *A* with altruistic preferences towards *B* ($a_{A\to B} > 0$), , *As' are willing to pay a higher price p to select C the stronger the difference in the taste of B towards C in comparisons to the taste of B towards D is, i.e., the larger* $\sigma_{B\to C} - \sigma_{B\to D}$ *is.*

Proof: see appendix A2.

Next, we provide a formal argument that even in the absence of any social preferences towards *B* (i.e., $a_{A\to B} = 0$ *, and* $\lambda_A = 0$) and reciprocal motives of *B ($\lambda_B = 0$), A* selects *C* if *B* has more pronounced social preferences towards C ($a_{B\to C} \geq a_{B\to D}$), i.e., *A* discriminates in expectancy of a higher level of cooperation of others, and thereby of a higher payoff, and hence exclusively motivated by strategic considerations.

Proposition 2a: *under the assumptions that neither taste-based ($\sigma_{A\to C} = \sigma_{A\to D}$) nor statistical motives ($a_{C\to A} = a_{D\to C} = 0$; and $t_A = t_C = t_D = 1$) for the selection of a particular candidate exists, abstracting from reciprocal motives ($\lambda_A = \lambda_B = 0$) and social preferences towards B ($a_{A\to B} = 0$), A prefers to select C if B has stronger social preferences towards C than towards D ($a_{B\to C} \geq a_{B\to D}$).*

Proof: see appendix A2.

Intuitively, the proof provided in the appendix rests upon the fact that increase in the monetary dimension based on *B's* more pronounced altruistic feelings towards *C* increases B's willingness to cooperate and hence *A*'s monetary income. The established form of strategic discrimination can again be expressed in terms of a reservation price: *A*'s reservation price is given the previously outlined assumptions increasing in $a_{B\to C} - a_{B\to D}$, i.e., the difference in the strength of *B*'s social preferences towards C in contrast to *B*'s social preferences towards *D*.

Proposition 2b: under the assumptions that neither taste-based ($\sigma_{A\to C} = \sigma_{A\to D}$) nor statistical motive ($a_{C\to A} = a_{D\to C} = 0$; and $t_A = t_C = t_D = 1$) for the selection of a particular candidate exists, and abstracting from reciprocal motives ($\lambda_A = \lambda_B = 0$), for any arbitrary $a_{B\to C} \geq a_{B\to D} \geq 0$, it holds that the price *A* is willing to pay to be able to include *C* instead of D increases in the difference between $a_{B\to C} - a_{B\to D}$.

Proof: see appendix A2.

Next, we assess our third behavioral channel, the reciprocity channel, in isolation. We inquire into the impact of *B's* reciprocal behavior on *A*'s selection decision. In contrast to the scenarios considered before, we now assume in addition that neither *A* nor *B,* enjoys or suffers from interacting with each other, i.e., ($\sigma_{A\to B} = \sigma_{B\to A} = 0$).[65] Moreover, consider that B experiences no social preferences towards A ($a_{B\to A} = 0$)[66] and B acts reciprocally ($\lambda_B > 0$). Our reciprocity model assumes that *A's* social preferences towards *B* cannot be directly observed by *B*. However, *B* can infer the types of *A* (a type is a set of individuals who experience equal levels of envy or altruism towards, say, *B*) by observing *A's* selection choice. Thus, *A* can signal her type (see Spence, 1974) by selecting B's prefered team member to provoke reciprocal behavior. Cosequently, there exists some type distribution such that some types of *A* that are spiteful towards *B* but select *B* to trigger reciprocal behavior. Moreover, the more *B* conditions his or her behavior on the type of *A* (λ_B), the more spiteful selection decision makers are willing to select *C*.

Proposition 3: Assume that neither taste-based ($\sigma_{A\to C} = \sigma_{A\to D}$) nor statistical motives ($a_{C\to A} = a_{D\to C} = 0$; and $t_A = t_C = t_D = 1$) for the selection of a particular candidate exists. Furthermore, abstract from direct social preferences of *B* ($0\,, \mathrm{a}_{\mathrm{B}\to\mathrm{C}} = \mathrm{a}_{\mathrm{B}\to\mathrm{D}}$) and assume that only *B* but not *A* acts reciprocally ($\lambda_{\mathrm{A}} = 0$). Moreover, consider that *A* has no taste to interact with *B* and vice versa $\sigma_{A\to B} = \sigma_{B\to A} = 0.$ The types of A are uniformly distributed such that $\mathrm{a}_{\mathrm{A}\to\mathrm{B}} \sim \mathrm{U}(-1,1)$ and are private knowledge of *A*. Under these assumptions, there exists –for some reciprocity values of λ_{B} – some types of *A* ($\mathrm{a}_{\mathrm{A}\to\mathrm{B}} < 0$) who will in the presence of informational asymmetries select *C* but not in its absence, if *B* experiences a stronger taste for *C* than for *D* ($\sigma_{B\to C} > \sigma_{B\to D}$).

Proof: see appendix A2.

In summary, we have introduced three different explanations why individuals account for others' group composition preferences. First, the more developed the social preferences of *A* towards *B* ($a_{A\to B}$) are, the more likely *A* lives up to the team preferences of *B*. Second, the more pronounced the social preferences of *B* towards *C* in comparison to her preferences towards *D* are *(*$\sigma_{B\to C} - \sigma_{B\to D}$), the more likely is it that *A* selects *B*'s preferred candidate. Third, under the assumption of imperfect information about the type of *A*s, *A*s incentive to signal that she cares for the well-being of *B* in order to provoke reciprocal behavior increases with the relative weight *B* puts on the type of his interaction partner *(*λ_B).

[65] Allowing for $\sigma_{A\to B} \neq 0$, $\sigma_{B\to A} \neq 0$ would not change the result derived in proposition 3.
[66] Deviation from this assumption would not change the qualitative results of our discussion but would enhance its complexity.

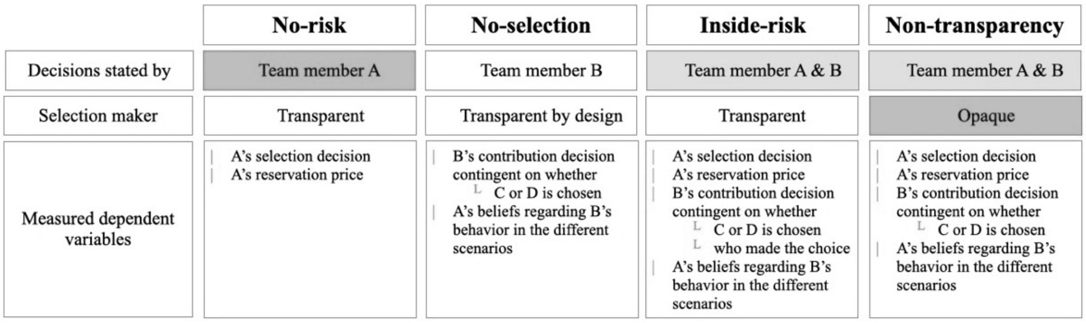
3 Experimental Setting

3.1 Treatment Overview and Hypotheses

To study the impact of current team members' group composition preferences on *A*'s selection decision, we designed an experiment comprising four treatments. Each of them consists of two stages, a team formation and a payoff determination stage. In the team formation stage, one additional team member has to be selected out of two candidates *(C* or *D*) to join a team consisting of two participants (*A* and *B*). The excluded participant receives no additional payoff. The present team member *B* has a preference for candidate *C*, while selector *A* has no preference for either candidate by design. After the team formation stage, all team members play a deconstructed version of a linear public good game (see Zelmer, 2003 for an overview) from which they monetarily benefit. We assess the impact of selectors' decisions on *Bs'* contribution to the public good, thereby investigating how *B*s reciprocate to *A*s' decision.

Contingent on the treatment, there is at maximum one team member *A* who can make a selection decision, and one team member *B* who can make a deliberate contribution decision. The other group members are enforced to cooperate. Hence, *B* determines her, *A*'s and the third passive member's (either *C*'s or *D*'s) payoff. The main dependent variables of interest are the selection choice of *A*, her reservation price (willingness to pay) to make a selection decision, and *B*'s contribution to the public good contingent on who has been selected. Furthermore, we elicit *A*'s beliefs about *B*'s contributions as well as control variables about all participants.

Figure 3.1: *Treatment Overview*



	No-risk	**No-selection**	**Inside-risk**	**Non-transparency**
Decisions stated by	Team member A	Team member B	Team member A & B	Team member A & B
Selection maker	Transparent	Transparent by design	Transparent	Opaque
Measured dependent variables	A's selection decision A's reservation price	B's contribution decision contingent on whether C or D is chosen A's beliefs regarding B's behavior in the different scenarios	A's selection decision A's reservation price B's contribution decision contingent on whether C or D is chosen who made the choice A's beliefs regarding B's behavior in the different scenarios	A's selection decision A's reservation price B's contribution decision contingent on whether C or D is chosen A's beliefs regarding B's behavior in the different scenarios

Our four treatments (*see* Figure 3.1) differ in three dimensions: we exogenously vary (i) whether the payoff from the public good is contingent on the decision of the present team member *B*, (ii) whether the team formation process is endogenous or exogenous, and (iii) whether it is observable whose decision (either *A's* or the computer's) is implemented. The treatments can be summarized as follows:

- In the *No-risk Treatment, A* has the opportunity to select one of the team members (*C* or *D*) for a randomly set fee. All players (including *B*) are enforced to cooperate in the public good game. If *A* is not willing to pay for the opportunity to select the third team member, the computer randomly selects one of the two candidates (*C* or *D*). The current team member *B* will be informed whether *A* or the computer has selected the team member.
- In the *Inside-risk Treatment, A* can again select in exchange for a randomly determined fee one of two candidates as team members. *B* can vary her contribution to the public good. All team members will be informed whether *A* or the computer has selected the additional team member.
- In the *No-selection Treatment*, the third candidate is randomly chosen by the computer, but *B* can decide what amount of her endowment she wants to invest into the public good.

– The *Non-transparency Treatment* is similar to the inside-risk treatment besides that team members other than *A* do not receive any information about whether *A* or the computer has selected the third team member.

Next, we explain why the experimental treatments allows to shed light on the introduced propositions.

Question 1: To what extent do strategic considerations and social preferences explain why selectors account for others' group composition preferences?

In a first step we investigate to what extent the consideration of other's group composition preferences in endogenous group formation processes rests upon social preferences or upon strategic considerations. Therefore, we investigate whether selectors, even in the absence of any strategic advantage, gain utility from living up to the preferences of current team members. If so, we can establish a causal effect between having social preferences towards other team members and the consideration of their preferences. A theoretical justification of a social preference effect on *A*s' selection choice is given by proposition 1a and 1b. To establish a causal effect, we assess *A*s' selection decisions in the no-risk treatment in which strategic considerations are not present by design, since *B* remains passive.

Hypothesis 1: *Selectors tendency to consider current team members team preferences are partly driven by social preferences towards them.*

Hypothesis 1 implies that *A* is willing to pay a fee to make a selection decision in the no-risk treatment and is in addition more likely to select candidate *C* out of altruistic motives.[67] Contrary, it may hold that individuals account for the current group members' tase for candidate *C*, because satisfying their group composition preferences may induce higher levels of altruism between current group members and selected candidates from which *A*s expect to profit monetarily. We test the strategic incentive effect by exogenously varying the decision environment such that in the inside-risk and the non-transparency treatment strategic incentives are potentially present, while in the no-risk treatment strategic incentives are absent by design. A comparison between *A*s' reservation prices and selection choices in the no-risk condition and the conditions associated with social risk thus enables us to measure to what extent the incorporation of preferences of others in *A*'s own selection decision is based on strategic considerations.

Hypothesis 2: *Selectors tendency to consider current group members preferences are partly driven by strategic considerations arising from the prospect of higher cooperation of current group members.*

Hypothesis 2 implies that *A*s' reservation price is higher in the inside-risk and the non-transparency than in the no-risk treatment. Moreover, *A*s are more likely to choose *C* in the inside-risk and the non-transparency treatment compared to the no-risk treatment.

[67] However, in contrast to the trade-off depicted in the formal model, participant ***A*** has to make a selection between getting his preferred candidate with certainty in exchange for a price or only getting his preferred candidate with probability of 50%. All predictions regarding the latter scenario are in line with the propositions derived from the former scenario, though the proposed effect size is larger in the former scenario. This relationship holds for all four discussed hypotheses in this section.

Question 2: **Do team members punish or reward selection decision makers for favorable or unfavorable selection choices and do selection makers account for this?**

Second, we investigate whether there is a causal effect on *B*s' contributions contingent on whether preferred or disfavored group members were randomly included or selected deliberately by a human selector. Therefore, we exogenously varied the selection procedure: in the no selection treatment the included candidate was randomly chosen by the computer. In the inside-risk treatment it is transparent whose choice is implemented. This information is opaque to *B*s in the non-transparency treatment.

Hypothesis 3: Current group members will reciprocate for kind behavior and thus contribute more to the public good if selectors include their preferred candidate than when the former one was exogenously included.

Finally, we study whether selectors account for procedural effects. Comparing whether *B*s contribute more if their preferred candidates are randomly or deliberately chosen by *As* reveals whether *B*s reciprocate for favorable decisions. In proposition 3, we formally derived this prediction. Exogenously varying the transparency of the selection decision does not only allow us to establish a causal link between whether reciprocal motives are important for behavioral responses of the candidate directly determining the final payoffs but also whether the selectors anticipate and react to such kind of reciprocal behavior.

Hypothesis 4: Selectors consider current group member group formation processes because they anticipate that current group members show reciprocal behavior in response to a favorable selection decision.

Hypotheses 4 implies that *A*s believe that *B*s' contributions are, first, more sensitive in the inside-risk in comparison to the other two treatments. Second, *A*s likely pay a higher reservation price in the inside-risk treatment than in the non-transparency and the no-selection treatment to make a selection decision.

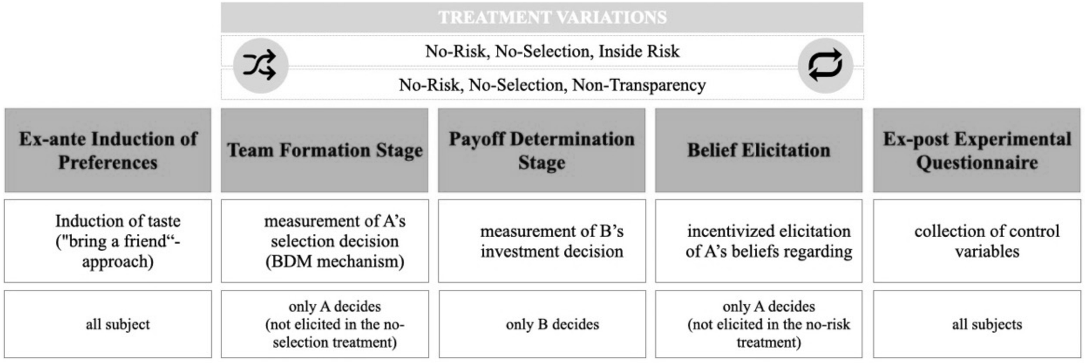
Randomization of Treatments– Exogenous variations and the random treatment assignment allow for the identification of causal effects. We use a within-subject design to address our research questions: every subject is assigned to three distinct treatments which she plays once: the no-risk, the no-selection and either the inside-risk or the non-transparency treatment. The design allows us to identify effects on an individual level, classify participants, and thereby study heterogenous treatment effects. To reduce experimenter demand effects (Zizzo, 2010) on the other hand, we collect between data between the no-selection and the non-transparency treatment.

Hence, participant *A* and *B* make each in total two selection or distribution decisions. The participants keep their role *(A, B, C* or *D*) over the entire course of the experiment. We apply a random strange matching in all of our treatments –with the exception that *B* and *C* are always in the same group– to secure the independence of observations. To mitigate order effects, we counterbalance the implementation of the treatments' order. To eliminate hedging concerns, participants only receive payments from one randomly selected treatment. To assure that participants understood the instructions well, we ask control questions at the end of the treatments' instructions which needed to be answered correctly to be able to continue.

3.2 Experimental Design and Procedural Details

Course of the Experiment – The experiment was run in November 2019 in the CLER in Cologne, Germany, with a total of 264 subjects (66 *As, Bs, Cs*, and *D*s) that participated in 12 sessions. It was programmed in oTree (Chen et al., 2016). Subjects were recruited using the software ORSEE (Greiner, 2004) and received a show-up fee of 4 €. All subjects play for the experimental currency unit "Taler", which equaled €0.05. The overall course of the experiment is depicted in Figure 3.2.

Figure 3.2: *Course of the Experiment*



TREATMENT VARIATIONS

No-Risk, No-Selection, Inside Risk

No-Risk, No-Selection, Non-Transparency

Ex-ante Induction of Preferences	Team Formation Stage	Payoff Determination Stage	Belief Elicitation	Ex-post Experimental Questionnaire
Induction of taste ("bring a friend"-approach)	measurement of A's selection decision (BDM mechanism)	measurement of B's investment decision	incentivized elicitation of A's beliefs regarding	collection of control variables
all subject	only A decides (not elicited in the no-selection treatment)	only B decides	only A decides (not elicited in the no-risk treatment)	all subjects

Preference Manipulation– We use an indirect approach to induce *B*s' group composition preferences while assuring that *A*s themselves are in a non-embedded setting indifferent between *C* and *D*.[68] In particular, *A* has to decide whom to include –a friend of *B* or a person neither *A*, nor *B* know. *B* likely has a strong preference for the inclusion of his or her friend *C*, while *A* has no preferences for either of the two candidates.[69] We consider the possibility that *B* and *C* arrange an agreement to split their payoffs evenly after the experiment. However, under the assumption that some pairs of friends agree on such an arrangement, an inequity averse *A* has an additional incentive to select *D*, because including *D* increases the number of participants benefiting from the experimental payoffs and the overall inequity between all participants. This opposes our presumed acknowledgment effect and therefore does not impede an assessment of qualitative treatment effects.

Team Formation Stage – In this stage, A selects a particular candidate if the utility gained form a selecting this candidate is larger than its costs.[70] To measure to what extent *As* consider *B*s' group composition process in the no-risk, inside-risk and the non-transparency treatment we pin down this threshold value: following Becker (1957), the willingness to discriminate can be measured by eliciting the difference between the

[68] Alternatively, one may directly induce preferences by changing the parametrization of the experiment such that Bs monetarily profit from discrimination in the public good game holding the efficiency parameters for A constant. However, changing tastes by altering the incentive structure do not only change the *B*s' group composition preferences, but also the economic efficiency of the choices as well as distributions and thus do not allow to identify the proposed effect.

[69] Similar procedures have been introduced to test the impact of social distance on social preferences (Binzel & Fehr, 2013; Brañas-Garza et al., 2010; Candelo et al., 2018; Goeree et al., 2010; Leider et al., 2009). A minor downside of that approach is that they may underestimate the treatment effect: Attending an experiment together with a friend is usually more entertaining than attending an experiment as a single individual. Hence, the effort costs of attending an experiment are likely higher for a single participant. As a consequence, selection makers have an antagonistic incentive to select player *D*, while our introduced theory would predict that *A* should prefer *C*.

[70] This design fulfills various purposes: first, in the no-risk, inside-risk as well as the non-transparency treatment we establish a precise measure which assesses to what extent *A* incorporates the preferences of *B* into his selection process. Second, by varying whether *A* knowingly and deliberately selects one of the candidates or whether the computer has implemented the choice changes the willingness to contribute the public good. That is, we test whether the circumstances of the selection influences *B*'s contribution. Thereby, we assess whether *B* punishes or rewards A for making a favorable selection. Third, to explore whether *B* conditions his contribution on who made the decision and whether *A* anticipates this we vary the ex-post experimental feedback

willingness to pay to interact either with candidate *C* or candidate *D* (Becker, 1957). To elicit ta proxy for this reference price, we implement the Becker-DeGroot-Marschak mechanism (Becker, Degroot, & Marschak, 1964).[71] In particular, we ask *A*s to decide between two options: if option 1 is chosen, *A* acquires the opportunity to select a team member in exchange for a determined selection fee. If option 2 is chosen, the computer will randomly select a candidate. *As* have to choose 10 times between delegating the selection decision to the computer or paying a potential fee of 1, 5, 10, 15, 20, 25, 30, 35, 40, or 45 and actively select on candidate. Then, the computer will randomly select one scenario and put the option chosen into effect. While in the no-risk, inside risk or non-transparency treatment the group formation process in endogenous, it is exogenous in the no-selection treatment, i.e., the computer always selects a candidate.

Payoff Determination Stage – After the formation of the team, the three team members play a modified version of a public good game. In a standard public good game, participants have to decide how to use their initial endowments. The task of each player is to decide how many Taler she wants to contribute to a team project and how many she wants to keep for herself. Contrary, in our experiment, *A* and the new group member are by design enforced to cooperate. Contrary, the current group member (*B)* can in three out of four treatments freely decide how to use her endowment. In particular, *B* has to decide how many out of 100 Taler she wants to contribute to a team project and how many she wants to keep for herself. Her income comprises the amount that she keeps for herself and her share of the team project. All team members' contribution to the team project are multiplied by two and evenly divided. The income of each group member from the project is calculated similarly, even though *A* and *C* are enforced to contribute their entire endowment.

To assess if *B* conditions her cooperation behavior on whether her likely favored candidate is chosen or not, we apply the strategy method (Selten, 1965) and let *B* state to what extent she wants to contribute to the public good contingent on whether *C* or *D* was selected.[72] To assess whether *As* considers *B*s' behavioral reactions in response to the selection decision, we vary whether *Bs* are enforced to cooperate or have the opportunity to shirk. To further study this question, we also varied the feedback *B*s receive: while in the non-transparency treatment *B*s receive no information whose selection choice was implemented, we announced prominently in the inside-risk treatment whether the computer or *A* selected the third candidate. Varying the amount of information enables us to assess whether *B* will react reciprocally if *A* satisfies *B*'s group composition preferences. To ensure the comparability of treatments, we decided to provide no information about the reservation price to participant *B*.

Belief Elicitation – To test whether the changes in *A*s' selection decisions are based on anticipated changes in *Bs'* behavior contingent on the selection procedure as well as its outcome, we elicit *A*s' beliefs regarding *B*s' contribution to the public good in an incentivized manner (*A*s receive 25 additional Taler when their actual guess is no more than 3 Taler away from the true guess). We elicit those beliefs in the no-selection

[71] Note that the BDM mechanism allows to measure the *A*'s preferences more precisely than a simple choice option between selecting one candidate or randomizing the options for a fixed alternative payment that is larger or equal to zero. A precise measurement is important for the identification of the effect, because the result of the experiment is sensitive to the alternative payment's parametrization. Moreover, for a wide range of preference structures, it can be formally shown that the BDM mechanism incentivizes selection decision makers to truthfully reveal their valuation for a private good. In particular, incentive compatibility implies that the BDM mechanism incentivizes the bidder to truthfully reveal her cut-off price.

[72] Since computations of final payoffs are considered complicated, we equipped participants with a slider tool to calculate the final payoffs of all participants contingent on Bs' contribution. Our data revealed that most participants made use of the slider tool.

and non-transparency treatment by asking about the average *B*'s contribution decision to the public good contingent on whether *A* or *B* has been chosen. In the inside-risk treatment, we ask *A*s' to answer the same question for four different scenarios: if *A* has chosen *C* or has chosen *D* and if the computer has chosen *C* or *D*.

Questionnaire– At the end of the experimental session, subjects are asked to fill out a questionnaire on personal data as well as a number of control questions. We also generate survey data to measure altruism, reciprocity and social preferences based on the questions included in the "Global Preference Survey" (GPS) introduced by (Falk et al., 2018). In addition, we elicit the social value orientation of participants using a slider measure –which is more fine-grained refinement of nine-item triple dominance measure (Au & Kwong, 2004)– introduced by Murphy, Ackermann, and Handgraaf (2011).

4 Results and Discussion

4.1 Baseline Effects

In total, 264 subjects took part in our experiment, with each 66 *A*-, *B*-, *C*- and *D*-participants, accounting for 132 observations on selectors' (*As*') reservation prices, and selection decisions, as well as for 264 observations on current group members' (*Bs*') contribution decisions. The average age was 26 years. Slightly more females (59%) than males participated.

The aim of the experiment is to study selectors' (*A*s') motives to consider other team members' (*B*s') group composition preferences in the absence of any taste-based or statistical reasons to select either candidate. Its empirical assessment requires, first, that *B*s have a taste for interacting with their friend (*C*s) and, second, that *As* in general anticipate and account for *Bs'* group composition preferences.

Figure 3.3: *Histogram of B's Contribution Decision by Selected Candidate (including averages)*

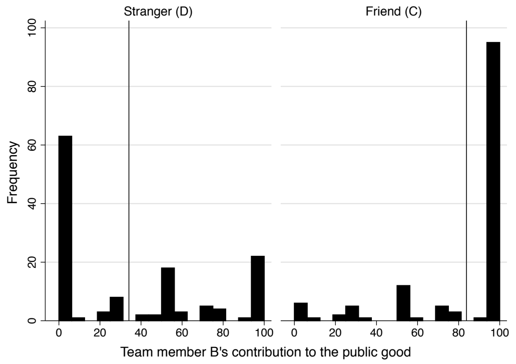
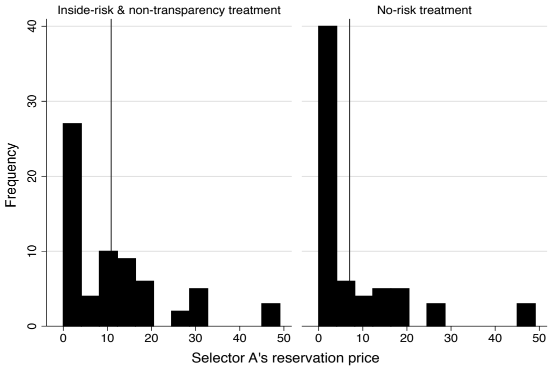


Figure 3.4: *Histogram of Average A's Reservation Price by Treatments (including averages)*



First, when current group members (*B*s) were matched with their friends (*C*s) across all three treatments in which *B*s were not enforced to cooperate, 96 out of 132 *B*s (or 72%) contributed a higher share of their endowment to the public good (Fisher-exact test: $p < 0.01$). Being matched with a friend, *B*s contributed on average considerably 50 Taler more (out of an endowment of 100 Taler) than being matched with a stranger (a MWU-test depicts a significant difference in rank orders: $p < 0.001$) and only 6 out of 132 *B*s (~ 4,5%) contribute nothing to the public good (*see* Figure 3.3). Contrary, when being matched with a stranger 50% of *B*s acted completely selfishly. In Table 3.1, we report *B*s' average contributions by treatments which corroborate our finding that *B*s had more pronounced social preferences towards their friends and consequently group composition preferences across all treatments. In conclusion, our treatment manipulation worked as intended.

Regarding the latter prerequisite (*A*s consider *B*s' group composition preferences) we find that, pooling all treatments associated with risk, selectors (*A*s) were willing to pay to select a candidate in about 60% (79 out of 132)[73] of all cases. If *B*s could deliberately vary their contributions, the share of *A*s willing to pay to make a selection decision increased to 70% (46 out of 66 *A*s). Contrary, it dropped to 50% (33 out of 66 *A*s) in the no-risk condition. 18% of all *A*s do not discriminate on the basis of social relations in any scenario (*see* Table 3.1).[74]

Table 3.1: *Descriptive Statistics*

	Treatment					
	No Risk	No Selection	Inside-Risk	Non-Transparency	Pooled Treatments (Inside Risk & Non-Transparency)	All Treatments
	Selection maker A's decisions					
Mean A's reservation price (in Taler)	7.00 (1.40)	–	10.67 (2.20)	11.10 (1.99)	10.86 (1.49)	8.93 (1.03)
Share of A's with a reservation price >0 Taler	50% (0.06)	–	67% (0.08)	73% (0.08)	70% (0.06)	60% (0.04)
Share of A's who select C (conditional on a reservation price >0 Taler)	73% (0.08)	–	91 % (0.06)	86% (0.08)	89% (0.05)	82% (0.04)
	Current team member B's decisions					
B's contribution conditional on C's (friend) selection	–	35.76 (4.87)	29.83 (6.18)	35.97 (6.98)	32.37 (4.61)	34.1 (3.35)
B's contribution conditional on D's (stranger) selection	–	84.61 (3.60)	83.88 (5.31)	82.67 (5.17)	83.03 (3.70)	83.81 (2.57)
Number of Obs.	66	66	36	30	66	132

Notes: *standard errors in parentheses*

We pool the data associated with social risks (the non-transparency and the inside-risk treatment) to enhance the statistical power of our analyses by increasing the number of observations. Before pooling the data, we tested for the equality in ranks (MWU: p = 0.54), means (t-test: p = 0.88, two-tailed), variances between treatments (variance comparison test: p = 0.29) and distributions (Kolmogorov-Smirnov test: corrected p = 0.73). We find no significant differences and thus no statistical arguments opposing pooling. Both treatments associated with social risk are –apart from variations in the information *B*s receive– similar. We assess the treatments separately in section 4.3.

Furthermore, the average reservation price to make a selection across all treatments equaled 8.9 Taler (*see* Table 3.1), which is considerable, because even if *B*s can decide on their contribution, risk-neutral *A*s should never be willing to pay more than 33 Taler.[75] Moreover, even in the no-risk treatment *As* had an

[73] We reason with confidence that *As* are not indifferent between the two choices, because *As* are willing to pay a price in order to make a selection decision while randomization is costless.

[74] In experiments that impose costs to selection decisions the disposition to discriminate in exchange for a price is not incisive (Neumark, 2018), plausibly as a consequence of social desirability. Thus, the number of *As* who is in the absence of an own taste for either of the two candidates willing to state a choice is remarkably high.

[75] Given these assumptions *A* should pay 33 Taler in case that *B* contributes everything to the team project if *C* was chosen and nothing to it otherwise as well as that the computer randomizes with 50%, because in this case the expected loss of not making choice is $\frac{2}{3} \times 100\ Taler - \left(\frac{1}{2} \times \frac{2}{3} \times 100\ Taler - \frac{1}{2} \times \frac{2}{3} \times 0\ Taler\right) \approx 33\ Taler$

average reservation price of 7 Taler. Thus, *A*s had a considerable interest in actively selecting a candidate in the presence and the absence of strategic considerations. Figure 3.4 depicts a histogram of *A*s' reservation prices that comprises information about both the frequency of selectors who were willing to make a selection choice for a minimum fee of one Taler, and their actual reservation price. Among those *A*s that strived to make a selection decision, around 80% selected *Cs* (*see* Figure 3.5). Thus, there is evidence that individuals account for current group members' (*B*s') group composition preferences.

4.2 The Impact of Strategic Considerations and Social Preferences on Selection Decisions

Question 1: *To what extent do strategic considerations and social preferences explain why selectors account for others' group composition preferences?*

In the following, we disentangle to what extent selection makers' social preferences explain the consideration of group composition preferences, and to what extent this effect can be traced back to strategic considerations.

Non-parametric Test of Hypothesis 1 (Reservation Price): To address the social preference channel, we analyze *A*s' selection decision in the no-risk treatment in which by design no strategic incentive to select either candidate was present. In the no-risk treatment a share of 50% of *A*s had a reservation price above 0. The average reservation price equaled 7 Taler (t-test with H0: reservation price = 0; $p < 0.001$).

Non-parametric Test of Hypothesis 1 (Selection Decision): Moreover, 73% of those *A*s who had a reservation price exceeding 0 lived up to *B*s' preferences by selecting *C* in the no risk-treatment (proportion test, H0= 50%: p < 0.001). These two results imply that a large share of individuals cared about their current group members' utility if being embedded in social groups. They were willing to pay for the opportunity to make a selection decision, and, in addition, select the candidate preferred by present team members. Consequently, the concept of social preferences, in particular the concept of altruism, is not limited to monetary outcomes but extends to group formation processes.

Finding 1: *Selectors tendency to account for current group members' group composition preferences are partly driven by social preferences towards current group members. Thus, we find empirical support for hypothesis 1.*

Notably, the subset of *A*s who selected *Bs'* friend *C* had a mean reservation price of 15.2 Taler. Contrary, the mean reservation price of *A*s who selected *D* was 10.6 Taler, though a MWU-test postulates no significant difference in distributions ($p = 0.17$)[76]. This indicates that selectors who were willing to live up to the preferences of other team members value the option of making a selection decision more, though not significantly.

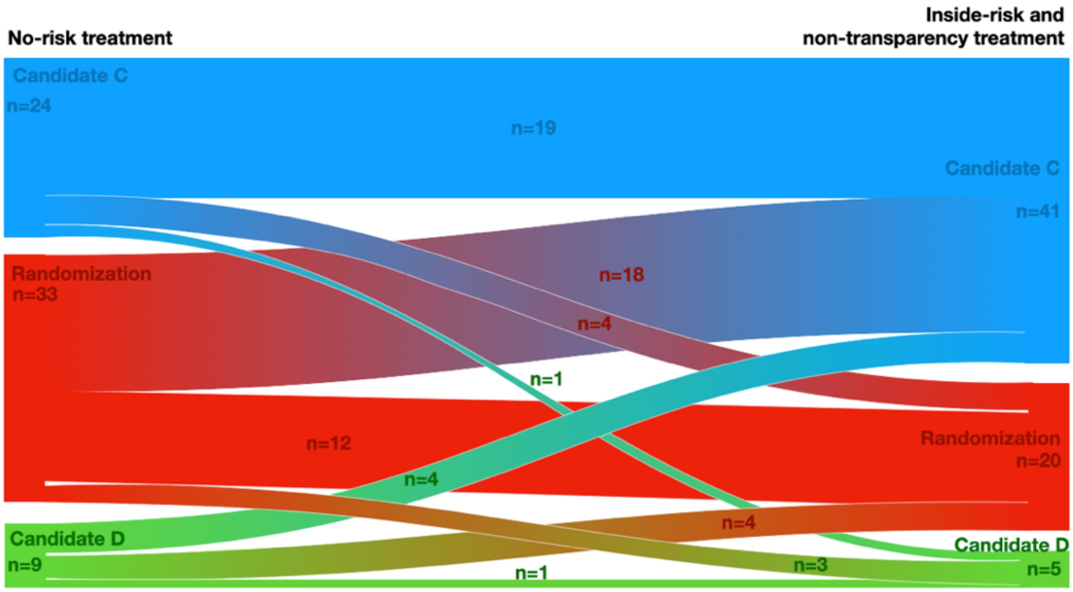
[76] The Mann-Whitney-U test is considered to be the non-parametric equivalent of the mean comparison t-test. However, note that the MWU-test does in contrast to the t-test not rely on the normal distribution assumption and evaluates whether two samples are drawn from the same population or more specifically, it tests whether the probability is 50% that a randomly drawn member of the first population will exceed a member of the second population or vice versa.

Non-parametric Test of Hypothesis 2 (Reservation Price): To assess the extent to which the tendency to account for *B*s' group composition preferences is driven by strategic considerations (hypothesis 2), we analyze whether selectors have a significantly higher reservation price in the pooled inside-risk and the non-transparency compared to the no-risk treatment. The reservation price was significantly higher in the pooled conditions associated with social risks compared to the no-risk condition (MWU-test: p =0.0164). We found that about 70% and thus a 20% higher share of *A*s had a reservation price above 0 in the presence of social risks (Fisher-exact: p = 0.033, two-tailed).

Notably, the reservation price in the no-risk treatment (7 Taler) was equivalent to roughly 66% of the reservation price in the social risk treatments (10.86 Taler) in which social preferences and strategic considerations together explain the consideration of others' group composition preferences (MWU-test: p = 0.016). Consequently, the social preference effect on the reservation price prevails and is about twice as strong as the strategic incentive effect. However, practical implications from this finding should only be drawn cautiously, since our experiment does not allow to assess strategic incentives in isolation.

Non-parametric Tests of Hypothesis 2 (Selection Decisions): Having studied treatment effects on reservation prices, we continue by analyzing *A*s' selection decisions in the treatments associated with social risks (inside-risk & non-transparency treatment) in comparison to selections in the no-risk treatment: Figure 3.5 depicts a Sankey diagram illustrating *A*s' decisions in the inside-risk and the non-transparency treatment (pooled conditions) conditional on their selection choices in the baseline treatment contingent on a BDM-price that equals 1 Taler. If *B*s could decide whether to cooperate or defect, *A*s were significantly more likely to make a decision and select candidate *C*.

Figure 3.5: *Sankey Diagram of As' Selection Decision at a Reference Price of 1 Taler by Treatment*



First, in the presence of social risks about 90% of *A*s with a reservation above 0 Taler selected candidate *C,* compared to 73% in its absence (Fisher exact test: $p = 0.077$, two-tailed). Second, 63% of *A*s who were not willing to pay any price to make a selection in the baseline scenario had a reservation price exceeding 0 Taler in treatments associated with social risks (MWU-test: $p < 0.01$). Third, while 24 *A*s selected *C* in the baseline condition, 41 did so in the pooled conditions (Fisher exact test: $p < 0.01$, two-tailed). Fourth, about 45% of *A*s who selected candidate *D* decided to select *C* if *B*s could condition their decision on *A*s' choices.

These patterns reflect that *B*s' discretionary power add a strategic incentive to select *C*. *A*s with either weak preferences for *D* or who are indifferent between either candidate, had an incentive to select candidate *C* in treatments associated with social risk or at least prefer a random selection over a deliberate selection of candidate *D*.[77]

Finding 2: *Selectors' are more willing to account for current group members' group composition preferences if current group members can vary their level of cooperation. First, in the presence of strategic incentives, selectors reservation prices are 66% higher. Second, selectors are more likely to choose current group members preferred candidate in the presence of strategic incentives. Overall, we find support for hypothesis 2.*

Additional Discussion – Classifying Types: Figure 3.5 and Table 3.2 reveal that roughly 87% of selectors can be assigned either the non-discriminatory type (~18%), the strategic type (~40%) or the pro-social type (~29%). The non-discriminatory type comprises *A*s who always refrain to make a selection decision, even in the presence of strategic incentives. The strategic type comprises *A*s who refrain from making a decision or select *D* in the absence of a strategic incentive arising from the prospect of higher cooperation but adhere to it by selecting *C* in its presence or refrain to make a selection decision when they preferred to select D in the absence of social risks. The pro-social type comprises *A*s who select *C* in the presence and the absence of social risks.

We continue by studying strategic and pro-social types' reservation price patterns (*see* Table 3.2). Trivially, the non-discriminatory type has a reservation price of zero in the presence and the absence of strategic incentives. First, *pro-social types* had an average reservation price of 16.30 Taler in the baseline and of 19.05 Taler in treatments associated with social risks (MWU-test; $p = 0.47$). Hence, social preferences account for about 85% the overall reasons underlying the consideration of group composition preferences in this subsample. Strategic incentives played a minor, even non-significant role in predicting the behavior of pro-social types.

Second, concerning participants of the *strategic type*, the average reservation price of those 18 participants who refused to make a selection decision in the no-risk treatment but selected *C* in the presence of social risks was 13 Taler (MWU-test; $p < 0.001$). Moreover, the average reservation price of those decision makers who selected *D* in the absence and *C* in the presence of strategic incentives increased from 11.5 to 13.75 Taler (MWU-test; $p = 0.77$). Finally, the reservation price of those selectors who selected *D* in the presence and refrained to pay to make a selection decision decreased from 6.5 Taler to 0 Taler. These patterns are in line with our theory, since it predicts that the utility from selecting *D* shrinks and the utility from selecting *C* increases in the presence of strategic incentives.

[77] Our model is not able to explain the decision of one *A* who switched between the baseline treatment and one of the treatments in which B had discretion from *C* to *D*. Neither can it explain the behavior of four *A*s who had no willingness to pay to make a selection decision, but were willing to pay at least a small fee to make a selection decision in one of the other two treatments (inside-risk treatment and the non-transparency treatment) or the decisions of the four that selected *C* in the baseline treatment, but were no more willing to pay in order to make a selection decision in the latter treatments. By and large, our model is able to explain the selection patterns of about 86% of our 66 *A*s.

Table 3.2: *Selection Decision and Reservation Price Patterns*

			Selection Decisions in Treatments Associated with Risk (pooled inside-risk and non-transparency treatment)		
			No Selection	**C**	**D**
Selection Decision in No-Risk Treatment	**No selection**	Reservation Price (no-risk)	0 (0)	0 (0)	0 (0)
		Reservation Price (pooled)	0 (0)	13 (2.1)	7 (3.0)
		Share of obs.	12/66 (18%)	18/66 (27%)	3/66 (2%)
	C	Reservation Price (no-risk)	10 (3.5)	16.3 (3.4)	15 (–)
		Reservation Price (pooled)	0 (0)	19.1 (3.3)	15 (–)
		Share of obs.	4/66 (7%)	19/66 (29%)	1/66 (2%)
	D	Reservation Price (no-risk)	6.5 (2.9)	11.5 (4.1)	25 (–)
		Reservation Price (pooled)	0 (0)	13.8 (4.3)	30 (–)
		Share of obs.	4/66 (7%)	4/66 (7%)	1/66 (2%)

Notes: *standard errors or percentage points in parentheses; black = pro-social type; dark grey = non-discriminatory; grey = strategic type*

Additional Discussion – Impact of Beliefs of Selection Makers Decisions: Eventually, it remains to inquire whether the treatment effect is driven by *A*s' beliefs about *B*s' contribution decision. We find that *A*s assume that *B*s contribute in the pooled treatments on average 80.1 Taler and thus about 50 Taler more if *C* has been selected (*see* also Table 3.5). The rank differences of the expected contributions are significant (MWU test $p < 0.001$). Second, *B*s actually contributed with 83.81 Taler[78] on average about 145% more to public good in case they were matched with their friends (MWU-test: $p < 0.001$).

Figure 3.6: *Kernel Density Estimation (Epanechnikov Kernel) of A's Strategic Price and Reservation Price (pooled treatments)*

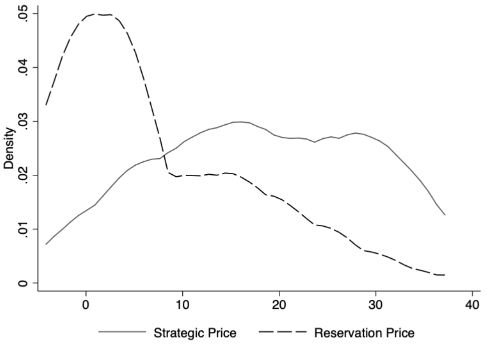
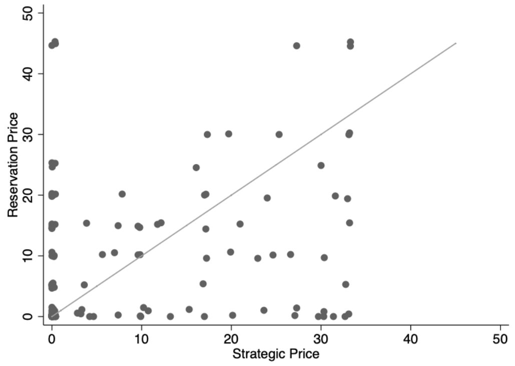


Figure 3.7: *Scatter Plot of A's Strategic Price and Reservation Price (pooled treatments; 45°-line included)*



To study ***A***s' strategic considerations more comprehensively, we calculated a strategic price which is defined as the price a risk neutral and profit maximizing ***A*** should be willing to pay based on her stated beliefs. The average strategic price is with 18.1 Taler significantly higher than the average reservation price in the pooled risk treatments (10.89 Taler), possibly also due to participants risk-aversion. The difference in ranks is significant (MWU-test: $p < 0.001$) as well.

[78] The average contribution to strangers is with 38% slightly lower in our experiment than in an average public good game (Zelmer, 2003), plausibly because the enforced cooperation is considered as a less kind act compared to a voluntary contribution and hence triggers less reciprocal behavior.

A kernel density estimation of the distribution of the strategic price and the reservation price (Figure 3.6) and a scatter plot of the strategic and the reservation price (Figure 3.7) reveals that a large share of ***A***s practice overbidding (potentially because ***A***s experience strong social preferences towards ***B*** and are less motivated by strategic incentives) and underbidding (potentially due to risk-aversion.[79]

The scatter plot reveals that a substantial share of decision makers (mostly non-discriminators) knowingly forgoes the benefit of current team members' higher willingness to cooperate by refraining to make a selection decision, while others (mostly pro-social types) were willing to pay to make a selection decision without expecting any strategic benefit from it. Overall, the Spearman rank order correlation between the reservation price and the strategic price is given by ρ = 0.3166 (p = 0.0096) which corroborates that selectors' consideration of group composition preferences of others is –inter alia– driven by strategic considerations, though the relationship is not perfect. The positive correlation between selectors' strategic prices and their reservation price provides additional support for hypothesis 2 and shows that strategic considerations determine selectors willingness to make a selection decision.

Regression Analyses: To corroborate the treatment effects, we estimated the effects of different decision environments on A's selection decisions utilizing different econometric approaches (see Table 3.3): ordinary least square (OLS) models with robust standard errors (*see* Table 3.3, Model 1 and 2), random effect models[80] (*see* Table 3.3, Model 3 and 4) to account for dependence in error term caused by the data's panel structure as well as random effect Tobit models[81] (*see* Table 3.3, Model 5 and 6) to account in addition for censoring effects and their average marginal effects. Model 2, 4 and 6 comprise –in addition to the treatment dummy depicting whether ***B***s could vary their contribution– an age, a gender as well as two additional variables measuring ***A***s' altruism[82] based on two item from the study by Falk et al. (2018). Establishing a correlational effect between altruism and reservation prices would generate further support for our introduced behavioral model which rests upon reciprocal altruism.

[79] About 17% of *A*s had a reservation price that is more than 3 Taler higher than the strategic price. 30% of *A*s had a reservation price is in the range of +/- 3 Taler of the strategic price and the reservation price of 53% of *A*s was lower than the strategic price.

[80]We selected random effects over fixed-effects models, because first, the aim of our experiment is to be generalizable to other settings and to derive policy implications. Thus, we are interested in examining correlational relations between time-invariant independent variables (such as age, altruism and gender effects) and the reservation price. Second, we are interested in testing whether the intercept is equal to zero in order to establish a potential social preference effect. Since fixed-effect models have fixed-effects (individual intercepts) instead of a common intercept, they do not allow for such an analysis. Third, random effect regression coefficients are more efficient than those for fixed effects; and ruining a Hausmann test, we find that, the random effects model is preferable. Fourth, the fact that we have experimental data based on exogenous variations and random assignment to treatments the random effect assumption is most likely satisfied and hence, unlike with happenstance data, we can calculate average treatment effects using random effect models.

[81] As an alternative to a random effect model we considered a multi-level mixed effect model. However, comparing the three level mixed effects models with the two level (random effect) model using a LR test, we find that its null hypothesis cannot be rejected at any convenient significance level. Hence, a standard two level (random effect) model is sufficient to deal with the dependence of error terms. The same reasoning applies comparing random effect and a three-level mixed effect Tobit model.

[82] To elicit the variable *Altruism 1* we asked: imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity? To elicit the variable *Altruism 2* we asked: How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10.

Table 3.3: *The Impact of Social Preferences and Strategic Incentives on A's Reservation Price*

	All Treatments (inside-risk & non-transparency treatments are pooled)[83]							
	Model 1: Linear regression	**Model 2:** Linear regression	**Model 3:** Random effects regression	**Model 4:** Random effects regression	**Model 5:** Panel Tobit regression	**Model 6:** Panel Tobit regression	**Model 5:** Panel Tobit regression (AME)	**Model 6:** Panel Tobit regression (AME)
	Dependent variable: ***reservation price***							
Deliberate Decision of B	3.864*	3.864*	3.863***	3.964***	6.921***	7.114***	6.921***	7.114**
	(2.046)	(2.018)	(1.339)	(1.339)	(2.305)	(2.326)	(2.305)	(2.326)
Altruism 1		-0.007		-0.007		-0.012		-0.012
		(0.009)		(0.130)		(0.0186)		(0.019)
Altruism 2		.980**		0.964 *		1.61**		1.61**
		(0.821)		(0.581)		(0.822)		(0.822)
Age		-0.179*		-0.174		-.759		-0.759
		(0.094)		(0.215)		(0.412)		(0.416)
Female		-3.678*		0.130		-6.637		-6.637
		(2.194)		(4.004)		(3.756)		(3.755)
Constant	7.000***	8.791***	7.000***	8.791 ***	-0.76	15.54		
	(1.399)	(3.175)	(1.447)		(2.333)	(10.948)		
# of obs.	132	132	132	132	132	132	132	132
Left-cens. obs.	–	–	–	–	53	53	53	53
Uncens. obs.	–	–	–	–	79	79	79	79
# of groups	132	132	66	66	66	66	66	66

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01; robust standard errors in parenthesis; AME = average marginal effects. The dummy variable "Deliberate Decision of B" indicates whether participants were assigned the no-risk treatment (=0) or either the inside-risk or the non-transparency treatment (=1.) "Altruism 1" is a survey item asking: "imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity? To elicit the "Altruism 2" we asked: "How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10"*

The ordinary least square regression as well as the random effect panel model in column 1–4 confirm our previous findings that the reservation price in the baseline treatment (the models' intercept) is significantly different from 0 at a 1%- level which corroborates hypothesis 1 or respectively finding 1. Furthermore, in all of our specifications the treatment (*see* Table 3.3, Model 1–6) the effect of strategic incentives *B*s could decide on their contribution) on the average *A*s' reservation price is significant at a 10%-level in the no-risk treatment, and at a 1%-level in the random effects and Tobit models. The null hypothesis of a t-test testing for the equivalence of the slope parameter and the dummy variable of model 1 capturing whether *B* could actively decide, could not be rejected (*p* = 0.2439). In addition, model 2, 4, and 6 reveal –as predicted by our behavioral model– that *A*s with more pronounced altruistic preferences had according to the significant and positive coefficient of the variable Altruism 2, a higher willingness to pay in order to be able to make a selection decision (in line with *B*s' group composition preferences).

To assess whether *As*' beliefs justify the acknowledgment of *Bs*' group composition preferences, we calculated two ordinary least square regression (*see* Table 3.4, Model 7), and two Tobit regressions presented in Table 3.4 in which we regress *As*' beliefs conditional on the implemented selection on the height of their reservation payment. Throughout all specifications, we find that, as predicted, *As*' willingness to make a selection decision increases with their beliefs about *B*s' contribution if candidate *C* is selected. Vice

[83] Here we present both, the raw Tobit data that can be interpreted as the effect of independent variables on the latent variable censoring as well as the average marginal effect, describing the average marginal effect on the actual censored variable. Furthermore, accounting for left-censoring the average marginal treatment effect of the pooled inside-risk treatments is about twice as large than in the linear random effect model, i.e., the linear model likely underestimates the actual treatment effect.

versa, *As*' reservation price decreases with their beliefs about *B*s' contribution if candidate *D*, i.e., the stranger, is chosen. Notably, only the latter effect is significant throughout all specifications. Hence, we find additional support for finding 2 or respectively hypothesis 2.

Table 3.4: *The Influence of Beliefs on As' Reservation Prices*

	Pooled Inside-risk & non-transparency treatment					
	Model 7: Linear regression	Model 8: Linear regression	Model 9: Tobit regression	Model 10: Tobit regression	Model 9: Tobit regression (AME)	Model 10: Tobit regression (AME)
	Dependent variable: ***reservation price***					
Belief Friend	0.100 (.065)	0.0812 (0.066)	0.100 (0.065)	0.114 (0.078)	0.142* (0.074)	0.114 (0.078)
Belief Stranger	-0.112** (0.050)	-0.119** (0.056)	-0.112** (0.050)	-0.150* (0.076)	0.141* (0.074)	-0.149* (0.076)
Constant	6.142 (4.72)	9.834 (6.619)	6.143 (4.717)	12.964 (12.683)		
Control Variables	no	yes	no	yes	no	yes
Number of observations	66	66	66	66	66	66
Left-censorded obs.	_	_	20	20	20	20
Uncensorded obs.	_	_	46	46	46	46

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01; robust standard errors in parenthesis; AME = average marginal effects. To elicit the variable "Beliefs Fried" we asked: How many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C)?; To elicit the variable "Beliefs Stranger" we asked: How many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D)?*

4.4 The Impact of Selection Processes and Transparency on Selection Decisions

Question 2: *Do team members punish or reward selection decision makers for favorable or unfavorable selection choices and do selection makers account for this?*

Non-parametric Treatment Tests of Contribution Decisions in the Inside-Risk Treatment: Next, we consider the inside-risk and the non-transparency treatment separately. In doing so, we analyze whether *B*s rewarded or punished *As* if *As* did or did not satisfy their team composition preferences, as well as if *As* consider alike reciprocal behavior in their selection decisions. To this end, we analyze in a first step whether *Bs* in the inside-risk treatment conditioned their contribution on whose selection choice –the computer's or *A's*– has been implemented (hypothesis 3). Table 3.5 reveals that *Bs* contributed on average about 10 Taler more to their friends if *As* –instead of the computer– deliberately chose *Cs* (MWU-test: *p* = 0.29). Contrary, *Bs* contributed about 10 Taler less if *As* deliberately chose *Ds* (MWU-test: *p* = 0.34), summing up to a total difference of about 20 Taler. However, due to the low number of observations (9 observations), the applied MWU-tests are underpowered. Regressions corroborating the results of the non-parametric analyses are provided in the appendix A1 (Table A3.1).

Table 3.5: *Overview of Current Team Member's Contribution Decisions & Selectors' Beliefs*

	Treatment			
	No-selection	Non-transparency	Inside-risk (DM: Computer)	Inside-risk (DM: A)
	Current team member B			
B's contribution conditional on ***C***'s	84.61	82.66	80.55	91.67
(friend) selection in Taler	(3.6)	(5.19)	(6.51)	(8.33)
B's contribution conditional on ***D***'s	35.76	35.96	31.96	21.67
(stranger) selection in Taler	(4.88)	(6.98)	(7.17)	(12.53)
	Selection maker A			
A's beliefs	–	77.86	76.11	81.88
conditional on ***C***'s (friend) is selected		(4.71)	(5.02)	(4.79)
A's beliefs conditional on ***D***'s	–	21.53	37.66	36.47
(stranger) selection		(3.59)	(5.04)	(5.22)
Number of Obs. (***B***s / ***A***s)	66	30	27/36	9/36

Notes: *standard errors in parenthesis. To elicit the variable "Beliefs Fried" we asked: how many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C)? To elicit the variable "Beliefs Stranger" we asked: how many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D)? In the inside-risk treatment we additionally varied whether A's or the computer's choice has been implemented. DM = (selection) decision maker.*

Non-parametric Treatment Tests on Comparing Contribution Decisions between treatments: In the second step, we compare *B*s' contributions to the public good in the inside-risk treatments with *B*s' decisions in other treatments. Our formal model introduced in section 2 implies that *B*s' contributions should be the highest in the inside-risk treatment, the second highest in the non-transparency treatment and the lowest in the no-selection treatment if *Cs* have been included and *A*s' choices were implemented.[84] Vice versa, theory predicts that *Bs* contribute the least in the inside-risk treatment, the second lowest in the non-transparency treatment and the highest in the no-selection treatment if *A*s' choices have been implemented *As* had selected *Ds*.

When *Cs* were chosen, the contributions were with on average 92 Taler in the inside-risk higher than in the no selection treatment with 85 Taler (MWU-test: $p = 0.65$) and in the non-transparency treatment with 83 Taler (MWU-test: $p = 0.94$). The effect size of the average contributions in the no-selection treatment as well as the non-transparency treatment are indistinguishable. It connotes that individuals are reluctant to punish or reward others if their responsibilities are opaque. Contrary, we find that when *Ds* has been chosen, *Bs'* average contributions was with 35 Taler in the non-transparency almost identical to average contributions in the no-selection decision (36 Taler; MWU-test: $p = 0.97$).[85] In contrast, if *A*s' decisions have been implemented in the inside risk treatment, the average *B* contributed 22 Taler or almost 40% less than in the other two treatments (MWU-test: $p = 0.26$).

Finding 3: *Current group members punish and reward selectors contingent on their respective selection decision. This effect is, however, not significant, plausibly due to the low number of observations. Thus, we find no support for hypothesis 3.*

[84] If A choses *C* in the inside-risk treatment in exchange for a fee, *B* will be informed about A's selection choice and thus may reciprocate. Contrary, in the non-transparency treatment *B* only knows that *A's* choice has been implemented with some probability. B thus may reciprocate for the selection choice, but probably to a smaller extent. In the no-selection treatment B does not reciprocate, since the computer's choice was implemented.

[85] These results confirm a wide range of empirical studies (Binzel &Fehr 2013; Brañas-Garza et al. 2010; Candelo et al. 2018; Goeree et al. 2010; Leider et al., 2009). Vainapel et al. (2018) find in an experimental study that group members are less likely to blamed punished, and reported on, when they are judged as separate individuals compared with as a group. The established bias in judgment of group members is plausibly caused by punishers' blame aversion. Punisher likely shy away from punishing individuals in settings with shared responsibility, if the responsibility and the pivotality of a decision cannot be pin downed easily.

Non-parametric Treatment Tests of Selectors Anticipation: Next, we investigate whether *As* anticipate that *Bs* positively or negatively reciprocate contingent on whether *A*s live up to *Bs'* group composition preferences. Indeed, *As* believe that on average *Bs* tend to punish or reward them for their deliberate favorable or unfavorable selection by varying their contributions: in the in-side risk treatment *A*s believes that Bs contribute about 5 Taler more to the public good if the friend was selected by the themselves (MWU-test: *p* = 0.003). Vice versa, *A*s believes that *B*s contribute about 1 Taler less to the public good if strangers were selected by *A*s (MWU-test: *p* = 0.913).

Finally, we evaluate if selectors' reservation prices differ between the inside-risk and the non-transparency treatments, i.e., if *A*s are more willing to pay to make a selection system in the presence of the opportunity to signal that they care *B*s' group composition preferences and thus their well-being. The average reservation price in the non-transparency condition was, unexpectedly, with 11.10 Taler higher than in the in the inside-risk treatment (10.86 Taler), though the difference in ranks was not significant (MWU-test: *p* = 0.65).

Finding 4: *Selectors are not willing to pay higher prices to acquire the right to make a selection decision if it was opaque who made the selection decision. Thus, we find no support for hypothesis 4.*

Table 3.6: The *Influence of Transparency on As' Reservation Prices*

	Inside-risk & non-transparency treatment					
	Model 11: Linear Regression	Model 12: Linear Regression	Model 13: Tobit regression	Model 14: Tobit regression	Model 13: Panel Tobit regression (AME)	Model 14: Panel Tobit regression (AME)
	Dependent variable: ***reservation price***					
Transparency	0.900 (3.035)	0.373 (2.030)	0.320 (4.0731)	0.374 (3.523)	0.320 (4.0731)	0.373 (3.522)
Belief Friend	-0.099 (0.066)	0.822 (0.066))	0.142* (0.0745)	0.082 (0.066)	0.142* (0.075)	0.0822 (0.066)
Belief No-Friend	-0.116* (0.056)	-0.125** (0.062	-0.143* (0.077)	-0.127** (0.062)	-0.143* (0.077)	-0.127** (0.062)
Constant	5.817 (4.635)	12.967* (7.631)	0.482 (6.536)	12.967* (7.632)	–	–
Controls	no	yes	no	yes	no	yes
# of obs.	66	66	66	66	66	66
Left-censorded obs.	_	_	20	20	20	20
Uncensorded obs.	_	_	46	46	46	46

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01 robust standard errors in parenthesis; control variables for altruism, age, gender (see Table 3.3). To elicit the variable "Beliefs Fried" we asked: how many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C)? To elicit the variable "Beliefs Stranger" we asked: how many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D)? The variable "Transparency" will take 1 if the participant is assigned to the inside-risk treatment.*

Regression Analyses: To test to the robustness of finding 4, we run regressions to assess the impact of *As'* beliefs about *Bs* contingent on who made the selection decision as well as who was selected by *A* on *As'* reservation price in the non-transparency condition. In particular, we calculated two ordinary least square models with robust standard errors (*see* Table 3.6, Model 11 and 12), as well as two Tobit models (*see* Table 3.6, Model 13 and 14) presented in Table 3.5. Model 12 and Model 14 include in addition to the main treatment variables, control variables for altruism, age and gender, as some of the previously discussed model. The coefficient of the dummy variable Transparency (Transparency = 1 if the participant is assigned the inside-risk treatment) is not significant throughout all models and thus corroborates finding 6. Overall, while the transparency effect is positive as predicted, the effects is not statistically significant at any conventional significant level. Model 11-14 furthermore corroborates a significant and negative effect on *As*' beliefs regarding *Bs*' contribution conditional on the selection of *D* established in previous models.

5 Discussion, Implications and Conclusion

This paper theoretically and experimentally examined the question *why and to what extent individuals selecting new group members consider current group members' group composition preferences as well as their behavioral responses in case their preferences are met or not met.*

On average and across all treatments, roughly 60% of all subjects were willing to pay to select a particular candidate in the absence of an own taste or statistical reasons for the inclusion of either candidate. A vast majority of those 60% based their actual selection on others' group composition preferences. However, about 20% of all selectors refrained under all circumstances –even in the presence of strategic incentives– to select a candidate in exchange for a small price and adhered to the norm to not treat people differently on the basis of their social relations. We conjecture that in settings in which the norm to not discriminate is more salient (e.g., discrimination on the basis of gender or ethnicity) this share is higher, though this hypothesis requires further empirical investigations.

With that stated, we emphasize that it is often opaque on which personal characteristic (e.g., gender, ethnicity, or character traits) one's group composition preferences rest upon. Current group members' willingness to cooperate was almost 150% higher in groups comprising friends in comparison to strangers. This increase explains –inter alia– why endogenous team formation processes can lead in an optimal setting to output enhancements (c.f. Herbst, Konrad, & Morath, 2015).

The evaluation of our experimental finding, however, is ethically ambiguous: the established effect that selector account for the group composition preferences of others and that the consideration enhances team productivity underlines the value of employee referral programs (Cappellari & Tatsiramos, 2015; Ioannides & Loury, 2004; Topa, 2011) on the one hand. On the other hand, these spill-over effects may lead in social environments in which preferences originate from prejudice or biased perception of minority members' talent to a significant enhancement of discrimination.

By studying discrimination and group composition preference spill-over effects, we introduced a new type of economic discrimination emerging from social preferences or strategic considerations ("strategic discrimination"). Therefore, the reason for discrimination is likely not only in the nature of an atomized individual as conjectured by taste-based and statistical discrimination models. Contrary, discrimination regularly emerges from group dynamics and autopoiesis (Luhmann, 1986; Mingers, 1994) i.e., organizations, groups and networks are often self-reproducing in the sense that individuals have an incentive to recruit new group members from often homogenous networks of current group members and therefore reproduce power structures. We thus suggest that future (field) studies should study the role of group processes to assess the true extent of discrimination and explain how group dynamics alter the "systems of discrimination" (Lang & Spitzer, 2020).

In our experiment in the absence of any potential strategic incentives, 50% of all selectors were willing to pay (often a considerable amount) to actively chose –in the vast majority of cases– the preferred candidate of the other group member. Furthermore, we found strong evidence that social preferences, in particular altruism, partly determines socially embedded individuals' selection decisions.

Secondly, we theoretically predicted and empirically confirmed that individuals correctly anticipate and account for dynamics between newly included group members and third parties to increase their monetary benefits. Consequently, the consideration of others' group composition preferences is partly explained by strategic incentives arising from the prospects of higher cooperation levels. However, the effect size of the social preference effect on the reservation price is on average significantly higher than the effect size of the strategic incentive effect, though about 40% of the decision makers are exclusively prone to strategic incentives and their decisions are not driven by social preferences.

Managers pursuing to alleviate discrimination spill-over effects should therefore tailor mitigation policies to the origin of discrimination spill-over effects: if discrimination spill-over effects emerge from strategic incentives, they might be attenuated by an increase of current employees' willingness to cooperate in heterogenous teams. Such an increase might be achieved by conditioning subordinates' monetary bonuses and performance assessments on their willingness to cooperate. If spill-over effects emerge from managers' social preferences, it is likely not sufficient to change the behavior of subordinates but policies should instead focus on managers themselves as well as on recruitment processes: by imposing a structured recruitment process with strict selection criteria, outsourcing the recruitment of employees, or applying a quota, the discretionary power of managers accounting for the prejudiced taste of his employees can presumably be limited and thereby the consideration of group composition preferences remedied. Finally, directly addressing prejudices and making an impact on preferences in teams by inclusive policies may reduce the source of discrimination in the first place. Thereby, it likely mitigates spill-over effects. With that stated, we point out that the assessment of managerial practices should not be based exclusively on the results discussed but must be evaluated more holistically.

Lastly, we investigated whether individuals selected their group members preferred candidate in exchange for a higher fee to signal their altruistic preferences towards current group members. Thereby, they attempted to trigger current group members' reciprocal behavior. Current group members punished and rewarded selectors for their respective selection decision, though not significantly, presumably because the statistical tests utilized to study this particular effect were underpowered. Therefore, while we conjecture that a deliberate choice in line with present team members preferences might trigger reciprocity, further research is need to either confirm or deny the hypothesized effect. If future research confirms the conjectured effect, managers should rather rely on exogenous group formation policies to enhance the share of minorities. In addition, we found no empirical evidence that decision makers actually utilize their selection choice as a signal. Having stated this, we are aware that testing anticipation effects in artificial one-shot experiments might have only a limited external validity, because current group members could only act reciprocally at the costs or benefits of an uninvolved third party. The co-determination of payoffs therefore alleviates reciprocal behavior.

Our findings and theoretical model imply that discrimination might not only be more widespread but also more persistent than previously predicted by taste-based or statistical discrimination models (see Lang &Lehmann 2012 for overview of conflicting findings on taste-based discrimination), because the composition of the team changes team members' willingness to cooperate and thus the overall output. Therefore, in contrast to a vast range of discrimination models in which discrimination decreases social welfare (Becker

1957), we find that teams in which all team members share a social identity are regularly are more willing to cooperate. Thereby, we provide an explanation why discrimination may persist in markets. In particular, our mathematical model predicts that the overall welfare gains from living up to the preferences of others increase, the more pronounced social preferences towards the current group members, the stronger group members' preference for a certain group composition, and current group members' potential to contribute to the group success are. In our experiment, we focused on disentangling the major behavioral channels. Thus, we leave it for future research to experimentally vary the strength of others group composition preferences, the degree of strategic incentives and assess the impact of these manipulation on the acknowledgment of group composition preferences.

Another limitation of this paper is that while our behavioral model allows to analyze the social dynamics within the entire group, our partial empirical analysis is however restricted to the selectors' and the present team members' behavior. A total analysis would require assessing the behavior of included candidates who have the ability to vary their contributions. If candidates are reciprocal co-operators (Fischbacher et al., 2001) the preferred candidate will contribute larger shares to the public good, because preferred candidates expect a higher level of cooperation than those candidates who are less preferred by current group members. Moreover, people being hired based on recommendations of their friends try to return this favor by providing more effort (see Topa, 2011 for an overview). Contrary, studies also established the effect that minorities or less able workers that are often discriminated against, are more grateful when being finally and therefore might be often more cooperative than majority members (Montinari et al., 2016). Hence, it remains an empirical exercise to analyze the impact of included group members on team dynamics to be addresses in future studies.

Eventually, we suggest as a future direction for research to address the question whether (social) punishment is able to reduce strategic discrimination. There is ample empirical evidence that the ability to monetarily punish and blame non-cooperators in co-operative environments, such as public good games, significantly enhances pro-social behavior and cooperation (e.g., Bicskei, Lankau, & Bizer, 2014; Carpenter, 2007; Fehr & Gächter, 2000; Fischbacher, 2008; Kamijo, 2016; Nikiforakis, 2008; Nikiforakis & Normann, 2008). Therefore, the threat of punishment may decrease the difference in cooperation levels of present team members conditional on who is included in the group. Thereby, it may alleviate strategic discrimination.

Appendix

A1 Supplementary Regression Results

Table A3.1: *Current Group members Contribution Decisions contingent on the selected candidate*

	Inside-risk & non-transparency treatment		**Inside-risk treatment**
	Model A1: Linear Regression	Model A2: Linear Regression	Model A3: Linear Regression
Dependent variable: ***Present team member's contribution to the public good***			
Friend is selected	48.84*** (4.57)	48.84*** (4.57)	48.59*** (8.29)
Inside-Risk	–5.93 (5.67)	–5.79 (5.67)	
Non-transparency	–0.31 (6.09)	–0.49 (6.08)	
Inside-Risk x Friend is selected	5.10 (7.68)	5.10 (7.68)	
Non-transparency x Friend is included	–2.15 (8.17)	–2.15 (8.17)	
Selector's decision is implemented			–10.30 (14.25)
Selector's decision is implemented x friend is selected			21.41 (16.63)
Female		1.13 (6.66)	
Age		2.99*** (1.12)	
Constant	35.76*** (4.25)	–37.48 (38.67)	(31.96)*** (6.69)
# of obs.	264	264	72
# of clusters	66	66	36
Left-censorded obs.	–	–	
Uncensorded obs.	–	–	

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01. Clustered standard errors in parentheses. Baseline = No-Selection treatment. To elicit the variable "Beliefs Fried" we asked: how many Taler did the average participant B invest if the chosen participant is his/her friend (participant C)? To elicit the variable "Beliefs Stranger" we asked: how many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D)? In the inside-risk treatment we additionally varied whether A's or the computer's choice has been implemented.*

A2 Proofs

Proof of Proposition 1a: in order to prove proposition 1a, we have to show that the utility *A* gains from selecting *C* is larger than the utility *A* gains from selecting *D*. When *A* experiences altruism towards *B* *(*$a_{A\to B}$ > 0*)*, *A* has ceteris paribus an incentive to select *C* if $\sigma_{B\to C} > \sigma_{B\to D}$. Therefore, it is sufficient to show that the following inequality is true to prove proposition 1a:

$$\begin{aligned}&\left(1-\frac{1}{3}\,t_A+\frac{2}{3}\,(t_C+t_B)\right)^{\alpha}+\left(\left(1-\frac{1}{3}\,t_B+\frac{2}{3}\,(t_A+t_C)+\,0.5\,(\sigma_{B\to A}+\,\sigma_{B\to C})\right)(a_{A\to B})\right)\\&>\left(1-\frac{1}{3}\,t_A+\frac{2}{3}\,(t_D+t_B)\right)^{\alpha}+\left(\left(1-\frac{1}{3}\,t_B+\frac{2}{3}\,(t_A+t_D)+\,0.5\,(\sigma_{B\to A}+\,\sigma_{B\to D})\right)(a_{A\to B})\right)\end{aligned} \tag{6}$$

By assumption, $t_A = t_C = t_D = 1$ and t_B is constant contingent on who is included into the group. Thus, we can simplify the inequality above as follows:

$$\big(\,(\sigma_{B\to A}+\,\sigma_{B\to C})\big)(a_{A\to B}) > (\sigma_{B\to A}+\,\sigma_{B\to D})(a_{A\to B}) \Leftrightarrow \sigma_{B\to C} > \sigma_{B\to D}, \tag{7}$$

This statement is true by assumption. Hence, *A* chooses *C* if *B* experiences stronger social preferences towards *C* than towards *D* *(*$\sigma_{B\to C} > \sigma_{B\to D}$*)*.■

Proof of Proposition 1b: We want to prove that A is willing to pay a higher price *p* the larger z defined as $\sigma_{B\to C} - \sigma_{B\to D}$ is, under the assumption that $a_{A\to B} > 0, \lambda_A = 0$,$t_A = t_C = t_D = 1$, and t_B is constant $\sigma_{C\to A} = \sigma_{D\to A} = 0$ and $a_{C\to A} = a_{D\to A} = 0$.

In order to do so, we set the utility function in which A interacts with *C* in exchange for a price equal to utility function resulting from a setting in which *A* interacts with *D*. Thereafter, we show that for the equality to hold *p* must increase in z:

$$\begin{aligned}&\left(1-\tfrac{1}{3}\,t_A+\tfrac{2}{3}\,(t_C+t_B)-p\right)^{\alpha}+\left(\left(1-\tfrac{1}{3}\,t_B+\tfrac{2}{3}\,(t_A+t_C)+\,0.5\,(\sigma_{B\to A}+\,\sigma_{B\to C})\right)(a_{A\to B})\right)=\\&\left(1-\tfrac{1}{3}\,t_A+\tfrac{2}{3}\,(t_D+t_B)\right)^{\alpha}+\left(\left(1-\tfrac{1}{3}\,t_B+\tfrac{2}{3}\,(t_A+t_D)+\,0.5\,(\sigma_{B\to A}+\,\sigma_{B\to D})\right)(a_{A\to B})\right),\end{aligned} \tag{1}$$

Besides, it holds by assumption that

$$1-\tfrac{1}{3}\,t_A+\tfrac{2}{3}\,(t_C+t_B) = 1-\tfrac{1}{3}\,t_A+\tfrac{2}{3}\,(t_D+t_B) \tag{2}$$

and $t_A = t_C = t_D = 1$. Next, we simplify the equation by substituting the following term

$$1-\tfrac{1}{3}\,t_A+\tfrac{2}{3}\,(t_C+t_B) = 1-\tfrac{1}{3}\,t_A+\tfrac{2}{3}\,(t_D+t_B) = \tfrac{4}{3}+\tfrac{1}{3}\,t_B = x \tag{3}$$

Hence, the initial equation can be simplified to

$$(x-p)^{\alpha}+\big(\big(\,0.5\,(\,\sigma_{B\to C})\big)(a_{A\to B})\,\big) = (x)^{\alpha}+\big(\big(0.5\,(\,\sigma_{B\to D})\big)(a_{A\to B})\,\big). \tag{5}$$

It directly follows that

$$(x-p)^{\alpha} = (x)^{\alpha} - 0.5z(a_{A\to B}). \tag{6}$$

Trivially, if z = σ$_{B\to C}$ − σ$_{B\to D}$ rises, i.e., B preference to interact with C instead with D increases, the right-hand side becomes smaller. Hence, p has to rise such that also the left-hand side decreases in order that the equality still holds. ■

Proof of Proposition 2a: Assume that $a_{A\to B}=0,\ \lambda_A=0, t_A=t_C=t_D=1,\ \sigma_{C\to A}=\sigma_{D\to A}=0$ and $a_{C\to A}=a_{D\to A}=0$. *A* selects *C* if $a_{B\to C}\geq a_{B\to D}$, since *A* has an incentive to select *C*, if the selection choice maximizes *B's* contribution. In this case the utility function of *A* reduces (irrespective of $\sigma_{B\to C}$ and $\sigma_{B\to D}$) to

$$U_A=\left(1-\frac{1}{3}\,t_A+\frac{2}{3}\,(t_C+t_B)\right)^{\alpha}+\sigma_{A\to B}\ \text{s.t.}\ 0\leq t_i\leq T=1 \tag{1}$$

Or respectively

$$U_A=\left(1-\frac{1}{3}\,t_A+\frac{2}{3}\,(t_D+t_B)\right)^{\alpha}+\ \sigma_{A\to B}\ \text{s.t.}\ 0\leq t_i\leq T=1 \tag{2}$$

Hence, under the given assumptions that $t_A=t_C=t_D=1$ it holds that the A's utility is in both cases given by, though $t_B(C)\neq t_B(D)$

$$U_A=\left(\frac{4}{3}+\frac{2}{3}\,t_B\right)^{\alpha}+\sigma_{A\to B}. \tag{3}$$

The term is strictly decreasing in t_B. *A* is going to select the candidate that maximizes *B*'s contribution. Hence, *A* selects max {argmax $U_B(C)$, argmax $U_B(D)$}. The optimization problem of *B* if *C* is chosen is given by:

$$\begin{aligned} max\ U_B(t_B)=\left(1-\frac{1}{3}\,t_B+\frac{4}{3}\right)^{\alpha}+\ 0.5\,(\sigma_{B\to A}+\ \sigma_{B\to C}))\ +0.5\,(1-\frac{1}{3}+\frac{2}{3}\,(t_B+1)\ +\ 0.5\,(\sigma_{A\to B}+\ \sigma_{A\to C}))a_{A\to B} \\ +0.5\,(1-\frac{1}{3}+\frac{2}{3}\,(1+t_B)\ +\ 0.5\,(\sigma_{C\to A}+\ \sigma_{C\to B}))(a_{B\to C})\ \text{s.t.}\ 0\leq t_i\leq T=1 \end{aligned} \tag{4}$$

The optimization problem of *B* if *D* is chosen is given by:

$$\begin{aligned} max\ U_B(t_B)=\left(1-\frac{1}{3}\,t_B+\frac{4}{3}\right)^{\alpha}+\ 0.5\,(\sigma_{B\to A}+\ \sigma_{B\to D}))\ +0.5\,(1-\frac{1}{3}+\frac{2}{3}\,(t_B+1)\ +\ 0.5\,(\sigma_{A\to B}+\ \sigma_{A\to D}))(a_{B\to A}) \\ +0.5\,(1-\frac{1}{3}+\frac{2}{3}\,(1+t_B)\ +\ 0.5\,(\sigma_{D\to A}+\ \sigma_{D\to B}))a_{B\to D}\ \text{s.t.}\ 0\leq t_i\leq T=1 \end{aligned} \tag{5}$$

The first order condition of the unconstraint utility function represented by equation 4 is given by

$$\frac{\partial\, U_B(t_B)}{\partial t_B} = -\frac{\alpha}{3}\left(\frac{7}{3} - \frac{1}{3}\, t_B\right)^{\alpha-1} + \frac{1}{3}\ (a_{B\to A} + a_{B\to C}) \ = \ 0 \tag{6}$$

Solving for t_B yields.

$$t_B{}^* = 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to C}\ )^{\frac{1}{1-\alpha}}}\right) \tag{7}$$

The solution to constraint first maximization problem (equation 4) is thus given by

$$t_B{}^* = \begin{cases} 0, & if\ 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to C}\ )^{\frac{1}{1-\alpha}}}\right) < 0 \\ 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to C}\ )^{\frac{1}{1-\alpha}}}\right), & if\ \ 0\ \leq 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to C}\ )^{\frac{1}{1-\alpha}}}\right) < 1 \\ 1, & if\ 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to C}\ )^{\frac{1}{1-\alpha}}}\right) < 0 \end{cases} \tag{8}$$

The solution to constraint second (equation 5) maximization problem can be calculated analogously and is thus given by

$$t_B{}^* = \begin{cases} 0, & if\ 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to D}\ )^{\frac{1}{1-\alpha}}}\right) < 0 \\ 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to D}\ )^{\frac{1}{1-\alpha}}}\right), & if\ \ 0\ \leq 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to D}\ )^{\frac{1}{1-\alpha}}}\right) < 1 \\ 1, & if\ 7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A} + a_{B\to D}\ )^{\frac{1}{1-\alpha}}}\right) < 0 \end{cases} \tag{9}$$

Hence, if ceteris paribus $a_{B\to C} \geq\ a_{B\to D}$, B is willing to contribute the same or a higher amount the public good if *C* and not *D* is part of the team, since $7 - 3\left(\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\to A}+a_{B\to D}\ )^{\frac{1}{1-\alpha}}}\right)$ increases in $a_{B\to C}$. Hence, *A* will select *C* if $a_{B\to C} \geq\ a_{B\to D}$ ■

Proof Proposition 2b: We want to show that given that

$$U_A(C;t_B) = \left(\frac{4}{3} + \frac{2}{3}t_B{}^{*} - p\right)^{\alpha} \;=\; U_A(D;t_B) = \left(\frac{4}{3} + \frac{2}{3}t_B{}^{**}\right)^{\alpha} \qquad (1)$$

p increases with $a_{B\rightarrow C} - a_{B\rightarrow D}$. Solving equation 1 for p yield the following function: $1.5\,p = t_B{}^{*} - t_B{}^{**}$. Hence, p increases $t_B{}^{*} - t_B{}^{**}$. B's profit maximizing value conditional on the selection of C is given by (see proposition 2a, equation 8):

$$t_B{}^{*} = \begin{cases} 0, & if\ 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow C})^{\frac{1}{1-\alpha}}}\right) < 0 \\ 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow C})^{\frac{1}{1-\alpha}}}\right), & if\ 0 \leq 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow C})^{\frac{1}{1-\alpha}}}\right) < 1 \\ 1, & if\ 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow C})^{\frac{1}{1-\alpha}}}\right) < 0 \end{cases} \qquad (2)$$

The profit maximizing valued given D is selected is given by (see proposition 2a, equation 9)

$$t_B{}^{*} = \begin{cases} 0, & if\ 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}\right) < 0 \\ 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}\right), & if\ 0 \leq 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}\right) < 1 \\ 1, & if\ 7 - 3\left(\dfrac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}\right) < 0 \end{cases} \qquad (3)$$

Hence, we must show that

$$t_B{}^{*} - t_B{}^{**} = 7{-}3\left[\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow C})^{\frac{1}{1-\alpha}}}\right] - \left(7{-}3\left[\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}\right]\right) = p \qquad (4)$$

increases in $a_{B\rightarrow C} - a_{B\rightarrow D}$. The inequality can be simplified to

$$\begin{aligned} p = t_B{}^{*} - t_B{}^{**} &= 3\left(\left[\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}\right] - \left[\frac{\alpha^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A} + a_{B\rightarrow C})^{\frac{1}{1-\alpha}}}\right]\right) \\ &= 3\alpha^{\frac{1}{1-\alpha}}\ \frac{(a_{B\rightarrow A}+a_{B\rightarrow C})^{\frac{1}{1-\alpha}} - (a_{B\rightarrow A}+a_{B\rightarrow D})^{\frac{1}{1-\alpha}}}{(a_{B\rightarrow A}+a_{B\rightarrow D})^{\frac{1}{1-\alpha}}\,(a_{B\rightarrow A}+a_{B\rightarrow C})^{\frac{1}{1-\alpha}}} \end{aligned} \qquad (5)$$

It follows that if p increases in $a_{B\rightarrow C}$ and decreases in $a_{B\rightarrow D}$ p increases in $a_{B\rightarrow C} - a_{B\rightarrow D}$, because the terms in which the respective parameters are included are additively separable.

The first derivative of p with respect to $a_{B\rightarrow C}$ is given by

$$\frac{\partial p}{\partial a_{B\rightarrow C}} = \frac{3\alpha^{\frac{1}{1-\alpha}}}{1-\alpha}\ (a_{B\rightarrow A} + a_{B\rightarrow C})^{-\frac{2-\alpha}{1-\alpha}} > 0, \text{ since } 0 \leq \alpha, a_{B\rightarrow A}, a_{B\rightarrow C}. \qquad (6)$$

The first derivative of p with respect to $a_{B\to D}$ is given by

$$\frac{\partial p}{\partial a_{B\to D}} = \frac{-3\alpha^{\frac{1}{1-\alpha}}}{1-\alpha}\,(a_{B\to A} + a_{B\to D})^{-\frac{2-\alpha}{1-\alpha}} > 0, \text{ since } 0 \leq \alpha, a_{B\to A}, a_{B\to D}. \tag{7}$$

Hence, since p increases in $a_{B\to C}$ and decreases in $a_{B\to D}$ and decreases in $a_{B\to D}$ p increases in $a_{B\to C} - a_{B\to D}$. ■

Proof Proposition B3: Assume that while $a_{B\to A}$ is known by all team members, $a_{A\to B}$ is A's private information and only the distribution function F$\sim U(-1,1)$ from which the parameter is drawn is general knowledge. This implies that the described public good game becomes a signaling game (see Spence, 1974), which can be solved by the adoption of a Perfect Bayesian Nash Equilibrium (PBNE). Furthermore, assume that $\sigma_{A\to B} = 0$ and $\sigma_{B\to A} = 0$, as well as $a_{B\to A} = 0$ and note that these assumptions will not change the qualitative implications of our model. The signalling game is a sequential game and has the following structure.

A decides whether to select *C* or *D*, before *B* decides on how much he wants to contribute to the public good. *A* 's strategy specifies for each of his types $\mathrm{a}_{\mathrm{A\to B}}$. Assume without the loss of generality that $\sigma_{\mathrm{B\to C}} > \sigma_{\mathrm{B\to D}}$ and note that selecting either *C* or *D* allows *A* to send a signal regarding her type $a_{A\to B}$. *B* specifies for each of her beliefs $E[a_{A\to B}|\text{selection choice}]$ dependent on *A*'s selection decision her contribution level – or respectively strategy– accordingly. Note that in equilibrium *B*'s beliefs regarding $a_{A\to B}$ equals the actual expected value E[$a_{A\to B}$]. We apply the concept of backward induction to analyze the game. If *C* has been chosen, *B* maximizes his utility with respect to t_B:

$$max\ U_B(C) = \left(\frac{7}{3} - \frac{1}{3}\,t_B\right)^{\alpha} + \frac{1}{2}\,(\,\sigma_{B\to C})) + (\frac{1}{3}\,(4 + 2t_B))\lambda_B\, E[a_{A\to B}\mid C] \tag{1}$$
$$\text{s.t.}\, 0 \leq t_i \leq T = 1$$

The optimization problem conditional on the selection of *D* can be derived analogously. The optimization problem is concave in t_B so that if the first order condition is met, it defines an optimum. Furthermore, since $0 \leq t_i \leq T$ it holds that if $E[a_{A\to B}|\text{selection choice}]$ 0, B will contribute nothing to the public good, since this inequality implies that the utility is strictly increasing in t_B.

The first order condition of the contribution decision contingent that ***C*** is chosen is given by

$$-\frac{1}{3}\alpha\left(\frac{7}{3} - \frac{1}{3}\,t_B\right)^{\alpha-1} + \frac{2}{3}\,\lambda_B\, E[a_{A\to B}\mid C] = 0 \Leftrightarrow t_B = 7 - 3\left(\frac{2\,\lambda_B\, E[a_{A\to B}\mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}} \tag{2}$$

Taking the constraint that $0 \leq t_B \leq 1$ into consideration it holds that

$$t_B{}^* = \begin{cases} 0, & if\ \frac{7}{3} < \left(\frac{2\lambda_B\ E[a_{A\to B} \mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}}\ or\ E[a_{A\to B} \mid C] \leq 0 \\ 7 - 3\left(\frac{2\lambda_B\ E[a_{A\to B} \mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}}, & if\ 2 \leq \left(\frac{2\lambda_B\ E[a_{A\to B} \mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}} \leq \frac{7}{3}\ and\ E[a_{A\to B} \mid C] > 0 \\ 1, & if\ 2 > \left(\frac{2\lambda_B\ E[a_{A\to B} \mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}}\ and\ E[a_{A\to B} \mid C] > 0 \end{cases} \tag{3}$$

The first order condition of the contribution decision contingent on the selection of *D* can be calculated analogously. In summary, *B* is willing to reciprocate given that *A* is identified as an altruistic type under the defined assumptions. It can be easily verified that since $\frac{1}{\alpha-1}$ is negative, t_B increases the more *B* values that *A* is of an altruistic type, i.e., the higher λ_B is. Now the utility function of *A* given that *C* or *D* has been chosen are given by:

$$\begin{aligned} U_A(C) &= \left(\frac{4}{3}+\frac{2}{3}\ t_B{}^*\right)^a + 0.5\ \left(\frac{7}{3}-\frac{1}{3}\ t_B{}^* + 0.5\ \sigma_{B\to C}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right)\ \& \\ = U_A(D) &= \left(\frac{4}{3}+\frac{2}{3}\ t_B{}^{**}\right)^a + 0.5\ \left(\frac{7}{3}-\frac{1}{3}\ t_B{}^* + 0.5\ \sigma_{B\to D}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \end{aligned} \tag{4}$$

Now, we can assess the signaling equilibria of the game. In the PBNE equilibrium the ***Bs'*** belief equals the actual expected value of $E[a_{A\to B}]$. There are only four potential equilibria, since they are only two potential signals available (C and D):

i. All types of *A* pool on selecting *C*
ii. All types of *A* pool on selecting *D*
iii. Envious *A*s up to a threshold value of $\overline{a_{A\to B}}$ select *C*, all others select *D*
iv. Envious *A*s up to a threshold value of $\overline{a_{A\to B}}$ select *D*, all others select *C*.

The first two potential equilibria do not constitute actual equilibria, because in any pooling equilibria that $E[a_{A\to B} \mid selection\ decision] = 0$ and hence $t_B = 0$. Hence, in (i) envious types are better off selecting D and thus deviating from the potential pooling equilibrium. Vice versa, in (ii) altruistic types are better of selecting C.

Trivially, (iii) cannot constitute an equilibrium because envious types always would have an incentive to deviate and select *D*, since they would gain utility from the reciprocal behavior of *D* as well as the smaller utility *B* gains interacting with *D*. Therefore, the equilibrium of the game must be a separating equilibrium in which all envious types select *D* and all altruistic types pool on *C*.

The threshold value which above that all decision makers *A* select *C* is denoted by $\overline{a_{A\to B}}$. Note that $a_{A\to B}$ is uniformly distributed with F$\sim U(-1,1)$. Hence, we know that $E[a_{A\to B}|C]$ = 0.5 ($\overline{a_{A\to B}} + 1$) and $E[a_{A\to B}|D]$ = 0.5 ($\overline{a_{A\to B}} - 1$).

The following equation thus defined the threshold value $\overline{a_{A\to B}}$

$$\begin{aligned} U_A(C) &= \left(\frac{4}{3}+\frac{2}{3}\ t_B{}^*\right)^a + 0.5\left(\frac{7}{3}-\frac{1}{3}\ t_B{}^* + 0.5\ \sigma_{B\to C}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \\ &= U_A(D) = \left(\frac{4}{3}\right)^a + 0.5\left(\frac{7}{3}-\frac{1}{3}\ t_B{}^{**} + 0.5\ \sigma_{B\to D}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \end{aligned} \tag{5}$$

It follows from 0.5 $(\overline{a_{A\to B}} - 1) \leq 0,$ that for any value $a_{A\to B} < 1$ is selected $t_B{}^{**} = 0$ if D is selected. Hence,

$$\begin{aligned} U_A(D) &= \left(\frac{4}{3}+\frac{2}{3}\ t_B{}^{**}\right)^a + 0.5\left(\frac{4}{3}+\ t_B{}^{**} + 0.5\ \sigma_{B\to C}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \\ &= \left(\frac{4}{3}\right)^a + 0.5\left(\frac{7}{3} + 0.5\ \sigma_{B\to D}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \end{aligned} \tag{6}$$

Next, we have to consider two different cases

$$t_B{}^* = 0, \text{ since } 7 - 3\left(\frac{\lambda_B\ E[a_{A\to B} \mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}} \leq 0 \text{ and} \tag{7}$$

and

$$7 - 3\left(\frac{\lambda_B\ E[a_{A\to B} \mid C]}{\alpha}\right)^{\frac{1}{\alpha-1}} > 0 \tag{8}$$

Considering equation 7 the utility function simplifies to

$$U_A(C) = \left(\frac{4}{3}\right)^a + 0.5\left(\frac{7}{3} + 0.5\ \sigma_{B\to C}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \tag{9}$$

And respectively

$$U_A(D) = \left(\frac{4}{3}\right)^a + 0.5\left(\frac{7}{3} + 0.5\ \sigma_{B\to C}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right). \tag{10}$$

Thus, it holds that $U_A(C) \geq U_A(D)$ if $\overline{a_{A\to B}} \geq 0$ and $U_A(D) > U_A(C)$ if $\overline{a_{A\to B}} < 0$.

Considering equation 8 and assume that the argmax of B given C has been chosen is given by

$$t_B{}^* = 7 - 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}}, \text{ because } E[a_{A\to B}|C] = 0.5\ (\overline{a_{A\to B}} + 1). \tag{11}$$

If we assume that $t_B{}^* = 1$ (B is enforced to cooperate, as in one of our treatments (no-risk treatment)), we would derive the same results as for case (i). Assume otherwise, then the following inequality constitutes the threshold value.

$$\begin{aligned} U_A(C) &= \left(\frac{4}{3}+\ 7 - 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)^a \\ &\quad + 0.5\left(\frac{7}{3} - 7 + 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}} + 0.5\ \sigma_{B\to C}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right) \\ &= U_A(D) = \left(\frac{4}{3}\right)^a + 0.5\left(\frac{7}{3} + 0.5\ \sigma_{B\to D}\right)\left((1-\lambda_A)\overline{a_{A\to B}}\right). \end{aligned} \tag{12}$$

Next, we rearrange the function in such a way that all potential benefits from selecting C instead of D of an envious A ($\overline{a_{A \to B}} < 0$) are on the left side and all the potential costs are on the right side:

$$\left(\frac{4}{3} + 7 - 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)^{a} - \left(\frac{4}{3}\right)^{a} + (\overline{a_{A\to B}})\ 0.5\ (-7 + 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}}) = (\overline{a_{A\to B}})(\sigma_{B\to C} - \sigma_{B\to D}) \qquad (13)$$

where

i. $\left(\frac{4}{3} + 7 - 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)^{a} - \left(\frac{4}{3}\right)^{a}$ captures the monetary gain from the selection of C

ii. and $((1 - \lambda_A)\overline{a_{A\to B}})\ 0.5\ (-7 + 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}}+1)}{\alpha}\right)^{\frac{1}{\alpha-1}})$ captures the gain from the joy an envious A experiences due to the a relative reduction of B’s income as a consequence of his or her higher contribution to the public good.

Finally, we show that there exists for every $\sigma_{B\to C} - \ \sigma_{B\to D}$ an $\overline{a_{A\to B}} < 0$ such that requirement above is fulfilled. To see this, let’s take the limit of the above function is given by:

$$\lim_{\overline{a_{A\to B}}\to 0} \left(\frac{4}{3} + 7 - 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}} + 1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)^{a} - \left(\frac{4}{3}\right)^{a} + ((1-\lambda_A)\overline{a_{A\to B}})\ 0.5\left(-7 + 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}} + 1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right) \qquad (14)$$
$$> \lim_{\overline{a_{A\to B}}\to 0}((1-\lambda_A)\overline{a_{A\to B}})(\sigma_{B\to C} - \sigma_{B\to D})$$

The limits are given by

$$\lim_{\overline{a_{A\to B}}\to 0} \left(\frac{4}{3} + 7 - 3\left(\frac{\lambda_B\ (\overline{a_{A\to B}} + 1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)^{a} - \left(\frac{4}{3}\right)^{a} + ((1-\lambda_A)\overline{a_{A\to B}})\ 0.5\left(-7 + 3\left(\frac{\lambda_B(\overline{a_{A\to B}} + 1)}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)$$
$$= \left(\frac{4}{3} + 7 - 3\left(\frac{\lambda_B}{\alpha}\right)^{\frac{1}{\alpha-1}}\right)^{a} - \left(\frac{4}{3}\right)^{a} > \lim_{\overline{a_{A\to B}}\to 0}(\overline{a_{A\to B}})(\sigma_{B\to C} - \sigma_{B\to D}) = 0 \qquad (15)$$

This proves that as long as $\alpha \frac{7}{3}^{\ \alpha-1} > \lambda_B$ there exists for every $\sigma_{B\to C} - \ \sigma_{B\to D}$ a type $\overline{a_{A\to B}} < 0$ who is willing to select C. ■

A3 Experimental Instructions

General Instructions

[all participants]

Welcome to todays' experiment and thank you for your participation

Dear participant,

This experiment takes about 45 minutes. Please read the following instructions carefully. If you have any questions during the experiment, please raise your hand. We will come to your cubical and answer your question.

This experiment consists of 3 parts and a questionnaire covering different topics at the end of the experiment. You will receive the instructions for each particular part before it starts. Subsequently, you will answer questions and make various decisions. During this economic experiment you **will be earn payoffs depending on your decisions** during the experiment. In addition, some of **your decisions have real consequences on other participants** within this experiment.

At the end of the experiment one out of three parts will be randomly selected for payments. The payoffs from the entire experiment will be solely dependent on the decisions you and other participants made in this particular part. During the experiment you will receive Taler.

1 Taler is worth €0.05.

At the end of the experiment all Talers will be converted into Euros. In addition to the payoffs you earned in this experiment, you will receive a show-up fee of 4€ for participating in the experiment independent from your decision. You will be paid your show-up fee and the payoffs you earned in the experiment at the of the experiment in private and in cash. No other team member will know how much you have earned in the experiment. This procedure guarantees that we can assure the anonymity of your decisions as well as other participants cannot identify you.

During the experiment you are not allowed to communicate with each other. A violation of this rule is leads to an exclusion form the experiment. In this case you will receive no payoffs and no show-up fee.

Click on “continue” if you have read an understood the instructions.

No-Risk Treatment

Instructions Project Stage

[all participants]

This part of the experiment consists of two stages; the team formation stage and the project stage in which the final payoffs from this part of the experiments are determined. In the team formation phase 3 out of 4 team members are going to build a team. Participant A and participant B are already part of the team. One further participant (either C or D) will be selected as the third team member in the team formation stage.

You are participant A [B; C or D].

The three participants will later have the opportunity to increase their payoffs. The fourth participant, that is not part of the team, will get no additional payoff from this part of the experiment. Before we are going to explain in detail how the team members are selected, we like to explain you how the payoffs are determined in this part of the experiment.

Determination of Payoffs

Every team member receives 200 Taler.
The fourth participant –who is not part of the team– receives 0 Taler.

Instructions Team Formation Stage

[all participants]

Only the three out of the for participants who are part of the team have the opportunity to increase their final payoffs in the team project stage. Two out of four participants (participant A and participant P) are guaranteed to be part of the team. The participant that is not part of the team receives no payoffs from this part of the experiment.

You are participant A [B; C or D].

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

Participant A can alter the decision of the computer and actively select one of the two participants (participant C or participant D. If participant A choose this option, he or she has to pay a fee between 1 and 45 Taler.

All participants will receive the following information

We will inform all participants about who will be part of the team and whether participant A or the computer have selected the third team member. Only participant A will be informed about how much he or she has to pay in order to have the opportunity to select one team member

Comprehension Test

[all participants]

Before the participants will make their decision, we kindly ask you to answer to comprehension questions.

Is participant A able to influence the decision of the third participant?Can participant B influence his or her own payment as well as the payments of the other participant?

Selection Decision Participant A

[only displayed to participant A]

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

You can alter the decision of the computer and actively select one of the two participants (participant C or participant D. If you choose this option, you have to pay a fee that reduce the final payoffs from this part of the experiment.

In order to do so we ask you to decide 10 times between Option A and Option B.
The computer will randomly select one out of the 10 lines and will implement the decision, you have chosen in the respective line.

Remember that every participant in the team receives 200 Taler.

All participants will receive the following information

We will inform all participants about who will be part of the team and whether participant A or the computer have selected the third team member. Only participant A will be informed about how much he or she has to pay in order to have the opportunity to select one team member.

If you want to read the instructions again you can click ***here.***

Option A	Option B
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **1 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **5 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **10 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **15 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **20 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **25 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **30 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **35 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay 40 **Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to **pay 45 Taler.**	The computer will randomly select participant C or participant D to be the third team member.

No-Selection Treatment

[all participants]

This part of the experiment consists of two stages; the team formation stage and the project stage in which the final payoffs from this part of the experiments are determined. In the team formation phase 3 out of 4 team members are going to build a team. Participant A and participant B are already part of the team. One further participant (either C or D) will be selected as the third team member in the team formation stage.

You are participant A [B; C or D].

The three team members will later have the opportunity to increase their payoffs. The fourth participant, that is not part of the team, will get no additional payoff from this part of the experiment. Before we are going to explain in detail how the team members are selected, we like to explain you how the payoffs are determined in this part of the experiment.

Instructions Project Stage

[all participants]

At the beginning of the project stage each team member receives an endowment of 100 Taler. The way in which this endowment can be used depends on your role.

Participant A as well as the third to be determined team member (participant C or D) are passive. They cannot decide on their own how to use the endowment.

Only participant B can decide freely how to use his/her endowment. His or her task is to decide, how to use the endowment. In particular, he or she decided how many Taler her or she wants to invest into a team project and how many Taler he or she wants to keep for himself/herself. Participant A and the third team member (participant C or D) are enforced to invest their entire endowment of 100 Taler into the team project.

The influences of participant B's decision on his or her payoffs

The payoff of participant B consists of two parts

(1) The amount of money that B keeps for himself/herself
(2) The payoffs, she receives as a return from her investments into the team project.

The amount of money she receives from the Team project is calculated as follows (calculations are valid not only for participant B, but for all team members):

All investments of the three team members will be doubled by the laboratory.
The doubled investments will be shared equally among the team members:

payoffs from the team project = 1/3 x 2 x sum of all investments of all team members

How the payoffs of participant A and the third team member (participant C or D) are determined

Participant A as well as the third team member (either participant C or D) only receive payoffs from the team project, because they are enforced to invest their whole endowment into the team project.

payoffs from the team project = 1/3 x 2 x sum of all investments of all team members

Using the slider below you can try how changes in the investment decision of B changes the payoffs of all team members.

Instructions Team Formation Stage

[all participants]

Only the three out of the for participants who are part of the team have the opportunity to increase their final payoffs in the team project stage. Two out of four participants (participant A and participant P) are guaranteed to be part of the team. The participant that is not part of the team receives no payoffs from this part of the experiment.

You are participant A [B; C or D].

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

Comprehension Test

Before the participants will make their decision, we kindly ask you to answer to comprehension questions.

Is participant A able to influence the decision of the third participant?

Can participant B influence his or her own payment as well as the payments?

Team Project Stage

[only displayed to participant B]

We will inform you about the identity of the selected third team at the end of today's experiment. Hence, we kindly ask you to make two decisions:

- Firstly, you will make an investment decision in case that the chosen team member is your friend.
- Secondly, you will make an investment decision in case that you do not now the third team member.

How much would you like to invest into the common team project in case the chosen team member is your friend?

How much would you like to invest into the common team project in case you do not now the third team member?

If you want, you may use the slider below to try out how your investment decision affect the payoffs of all team members.

Your guess regarding the decisions of other participants

[only displayed to participant A]

Lastly, we ask you to make a guess how much B-participants on average invested into the team project. For every answer that is no more than 3 Talers away from the true value, you receive 50 Taler.

How many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C)?

How many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D)?

Inside-Risk

[all participants]

This part of the experiment consists of two stages; the team formation stage and the project stage in which the final payoffs from this part of the experiments are determined. In the team formation phase 3 out of 4 team members are going to build a team. Participant A and participant B are already part of the team. One further participant (either C or D) will be selected as the third team member in the team formation stage.

You are participant A [B; C or D].

The three team members will later have the opportunity to increase their payoffs. The fourth participant, that is not part of the team, will get no additional payoff from this part of the experiment. Before we are going to explain in detail how the team members are selected, we like to explain you how the payoffs are determined in this part of the experiment.

Instructions Project Stage

[all participants]

At the beginning of the project stage each team member receives an endowment of 100 Taler. The way in which this endowment can be used depends on your role.

Participant A as well as the third to be determined team member (participant C or D) are passive. They cannot decide on their own how to use the endowment.

Only participant B can decide freely how to use his/her endowment. His or her task is to decide, how to use the endowment. In particular, he or she decided how many Taler her or she wants to invest into a team project and how many Taler he or she wants to keep for himself/herself. Participant A and the third team member (participant C or D) are enforced to invest their entire endowment of 100 Taler into the team project.

The influences of participant B's decision on his or her payoffs

The payoff of participant B consists of two parts

(1) The amount of money that B keeps for himself/herself
(2) The payoffs, she receives as a return from her investments into the team project.

The amount of money she receives from the Team project is calculated as follows (calculations are valid not only for participant B, but for all team members):

All investments of the three team members will be doubled by the laboratory.
The doubled investments will be shared equally among the team members:

payoffs from the team project = 1/3 x 2 x sum of all investments of all team members

How the payoffs of participant A and the third team member (participant C or D) are determined

Participant A as well as the third team member (either participant C or D) only receive payoffs from the team project, because they are enforced to invest their whole endowment into the team project.

payoffs from the team project = 1/3 x 2 x sum of all investments of all team members

Using the slider below you can try how changes in the investment decision of B changes the payoffs of all team members.

Instructions Team Formation Stage

[all participants]

Only the three out of the for participants who are part of the team have the opportunity to increase their final payoffs in the team project stage. Two out of four participants (participant A and participant P) are guaranteed to be part of the team. The participant that is not part of the team receives no payoffs from this part of the experiment.

You are participant A [B, C, D]

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

Participant A can alter the decision of the computer and actively select one of the two participants (participant C or participant D. If participant A choose this option, he or she has to pay a fee between 1 and 45 Taler.

All participants will receive the following information

We will inform all participants about who will be part of the team and whether participant A or the computer have selected the third team member. Only participant A will be informed about how much he or she has to pay in order to have the opportunity to select one team member.

Comprehension Test

[all participants]

Before the participants will make their decision, we kindly ask you to answer to comprehension questions?

Is participant A able to influence the decision of the third participant?
Can participant B influence his or her own payment as well as the payments of the other participant?

Remember that player B can affect with his/her decision the outcome of all team members.

Selection Decision

[only displayed to participant A]

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

You can alter the decision of the computer and actively select one of the two participants (participant C or participant D. If you choose this option, you have to pay a fee that reduce the final payoffs from this part of the experiment.

In order to do so we ask you to decide 10 times between Option A and Option B.
The computer will randomly select one out of the 10 lines and will implement the decision, you have chosen in the respective line.

Remember that every participant in the team receives 200 Taler.

All participants will receive the following information

We will inform all participants about who will be part of the team and whether participant A or the computer have selected the third team member. Only participant A will be informed about how much he or she has to pay in order to have the opportunity to select one team member.

If you want to read the instructions again you can click ***here.***

Option A	Option B
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **1 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **5 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **10 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **15 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **20 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **25 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **30 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **35 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay 40 **Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to **pay 45 Taler.**	The computer will randomly select participant C or participant D to be the third team member.

Team Project Stage

[only displayed to participant B]

We will inform you about the identity of the selected third team at the end of today's experiment. Hence, we kindly ask you to make two decisions:

- Firstly, you will make an investment decision in case that the chosen team member is your friend.
- Secondly, you will make an investment decision in case that you do not now the third team member.

Please note that depending on the decision of participant A either he/she or the computer had selected the third team member:

Participant A [the computer] has selected the third team member.

How much would you like to invest into the common team project in case the chosen team member is your friend?

How much would you like to invest into the common team project in case you do not now the third team member?

If you want, you may use the slider below to try out how your investment decision affect the payoffs of all team members.

Your guess regarding the decisions of other participants

[only displayed to participant A]

Lastly, we ask you to make a guess how much B-participants on average invested into the team project. For every answer that is no more than 3 Talers away from the true value, you receive 50 Taler.

How many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C) and participant A has made the selection decision?

How many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D) and participant A has made the selection decision?

How many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C) and the computer has made the selection decision?

How many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D) and the computer has made the selection decision?

Non-Transparency Treatment

[all participants]

This part of the experiment consists of two stages; the team formation stage and the project stage in which the final payoffs from this part of the experiments are determined. In the team formation phase 3 out of 4 team members are going to build a team. Participant A and participant B are already part of the team. One further participant (either C or D) will be selected as the third team member in the team formation stage.

You are participant A [B; C or D].

The three team members will later have the opportunity to increase their payoffs. The fourth participant, that is not part of the team, will get no additional payoff from this part of the experiment. Before we are going to explain in detail how the team members are selected, we like to explain you how the payoffs are determined in this part of the experiment.

Instruction Project Stage

[all participants]

At the beginning of the project stage each team member receives an endowment of 100 Taler. The way in which this endowment can be used depends on your role.

Participant A as well as the third to be determined team member (participant C or D) are passive. They cannot decide on their own how to use the endowment.

Only participant B can decide freely how to use his/her endowment. His or her task is to decide, how to use the endowment. In particular, he or she decided how many Taler her or she wants to invest into a team project and how many Taler he or she wants to keep for himself/herself. Participant A and the third team member (participant C or D) are enforced to invest their entire endowment of 100 Taler into the team project.

The influences of participant B's decision on his or her payoffs

The payoff of participant B consists of two parts

(3) The amount of money that B keeps for himself/herself
(4) The payoffs, she receives as a return from her investments into the team project.

The amount of money she receives from the Team project is calculated as follows (calculations are valid not only for participant B, but for all team members):

All investments of the three team members will be doubled by the laboratory. The doubled investments will be shared equally among the team members:

payoffs from the team project = 1/3 x 2 x sum of all investments of all team members

How the payoffs of participant A and the third team member (participant C or D) are determined

Participant A as well as the third team member (either participant C or D) only receive payoffs from the team project, because they are enforced to invest their whole endowment into the team project.

payoffs from the team project = 1/3 x 2 x sum of all investments of all team members

Using the slider below you can try how changes in the investment decision of B changes the payoffs of all team members.

Instructions Team Formation Stage

[all participants]

Only the three out of the for participants who are part of the team have the opportunity to increase their final payoffs in the team project stage. Two out of four participants (participant A and participant P) are guaranteed to be part of the team. The participant that is not part of the team receives no payoffs from this part of the experiment.

You are participant A [B, C, D].

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

Participant A can alter the decision of the computer and actively select one of the two participants (participant C or participant D. If participant A choose this option, he or she has to pay a fee between 1 and 45 Taler.

All participants will receive the following information

We will inform all participants about who will be part of the team**. We will not inform participants whether participant A or the computer have selected the third team member.** Only participant A will be informed about how much he or she has to pay in order to have the opportunity to select one team member.

Comprehension Test

[all participants]

Before the participants will make their decision, we kindly ask you to answer to comprehension questions.

Is participant A able to influence the decision of the third participant?
Can participant B influence his or her own payment as well as the payments of the other participant?

Selection Decision

[only displayed to participant A]

The third team member and the participant that will not be part of the team will be selected from the two remaining participants in the following way:

- The computer randomly selects either candidate C or candidate D as the third team member. Participant C and participant D have equal chances to become the third team member.
- Team member B and participant C are friends. Team member B and participant C came today together to the lab.
- Team member B and participant D are no friends. Team member B and participant D do not know each other.

You can alter the decision of the computer and actively select one of the two participants (participant C or participant D. If you choose this option, you have to pay a fee that reduce the final payoffs from this part of the experiment.

In order to do so we ask you to decide 10 times between Option A and Option B.
The computer will randomly select one out of the 10 lines and will implement the decision, you have chosen in the respective line.

Remember that every participant in the team receives 200 Taler.

All participants will receive the following information

We will inform all participants about who will be part of the team**. We will not inform participants whether participant A or the computer have selected the third team member.** Only participant A will be informed about how much he or she has to pay in order to have the opportunity to select one team member.

If you want to read the instructions again you can click ***here.***

Remember that player B can affect with his/her decision the outcome of all team members.

Option A	**Option B**
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **1 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **5 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **10 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **15 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **20 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **25 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **30 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay **35 Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to pay 40 **Taler.**	The computer will randomly select participant C or participant D to be the third team member.
You have the opportunity to select one of the two participants (C or D) to become part of the team. In order to do so you have to **pay 45 Taler.**	The computer will randomly select participant C or participant D to be the third team member.

Team Project Stage

[only displayed to participant B]

We will inform you about the identity of the selected third team at the end of today's experiment. Hence, we kindly ask you to make two decisions:

- Firstly, you will make an investment decision in case that the chosen team member is your friend.
- Secondly, you will make an investment decision in case that you do not now the third team member.

Please note that depending on the decision of participant A either he/she or the computer had selected the third team member:

Participant A [the computer] has selected the third team member.

How much would you like to invest into the common team project in case the chosen team member is your friend?

How much would you like to invest into the common team project in case you do not now the third team member?

If you want, you may use the slider below to try out how your investment decision affect the payoffs of all team members.

Your guess regarding the decisions of other participants

[only displayed to participant A]

Lastly, we ask you to make a guess how much B-participants on average invested into the team project. For every answer that is no more than 3 Talers away from the true value, you receive 50 Taler.

How many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C) and participant A has made the selection decision?

How many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D) and participant A has made the selection decision?

How many Taler did the average participant B invest if the chosen participant is his/ her friend (participant C) and the computer has made the selection decision?

How many Taler did the average participant B invest if the chosen participant is unknown to him/her (participant D) and the computer has made the selection decision?

Questionnaire – Demographic Questions

[all participants]

Global Preference Survey

Finally, we ask you to answer some questions regarding yourself.

1. Imagine the following situation: you won 1,000 Euro in a lottery. Considering your current situation, how much would you donate to charity? (Values between 0 and 1000 are allowed)
2. How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10.
3. Imagine the following situation: you are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 Euro in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 Euro, the most expensive one 30 Euro. You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you give?

   Respondents can choose from the following options: The bottle for 5, 10, 15, 20, 25, or 30 Euro)

4. How do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly? Please use a scale from 0 to 10, where 0 means you are "not willing at all to incur costs to punish unfair behavior" and a 10 means you are "very willing to incur costs to punish unfair behavior". You can also use the values in-between to indicate where you fall on the scale.
5. Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means you are "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Social Value Orientation Task

We ask you to make the following 6 hypothetical decision that will not be implemented.

You receive	85	□	□	□	□	□	□	□	□	□	85
Other receives	50	□	□	□	□	□	□	□	□	□	15
You receive	85	□	□	□	□	□	□	□	□	□	100
Other receives	50	□	□	□	□	□	□	□	□	□	50
You receive	50	□	□	□	□	□	□	□	□	□	85
Other receives	100	□	□	□	□	□	□	□	□	□	85
You receive	50	□	□	□	□	□	□	□	□	□	85
Other receives	100	□	□	□	□	□	□	□	□	□	15
You receive	100	□	□	□	□	□	□	□	□	□	50
Other receives	50	□	□	□	□	□	□	□	□	□	100
You receive	100	□	□	□	□	□	□	□	□	□	85
Other receives	50	□	□	□	□	□	□	□	□	□	85

Risk Task

Please imagine the following situation: You can choose between a sure payment and a lottery. The lottery gives you a 50 percent chance of receiving 300 Euro. With an equally high chance you receive nothing. Now imagine you had to choose between the lottery and a sure payment. We will present to you five different situations. The lottery is the same in all situations. The sure payment is different in every situation.

1. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 160 Euro as a sure payment?

   (a) lottery → go to question 17
   (b) sure payment → go to question 2

2. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 80 Euro as a sure payment?

   (a) lottery → go to question 10
   (b) sure payment → go to question 3

3. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 40 Euro as a sure payment?

   (a) lottery → go to question 4
   (b) sure payment → go to question 7

4. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 60 Euro as a sure payment?

   (a) lottery → go to question 5
   (b) sure payment → go to question 6

5. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 70 Euro as a sure payment?

   (a) lottery
   (b) sure payment

6. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 50 Euro as a sure payment?

   (a) lottery
   (b) sure payment

7. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 20 Euro as a sure payment?

   (a) lottery → go to question 8
   (b) sure payment → go to question 9

8. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 30 Euro as a sure payment?

   (a) lottery
   (b) sure payment

9. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 10 Euro as a sure payment?

   (a) lottery
   (b) sure payment

10. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 120 Euro as a sure payment?

    (a) lottery → go to question 14
    (b) sure payment → go to question 11

11. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 100 Euro as a sure payment?

    (a) lottery → go to question 13
    (b) sure payment → go to question 12

12. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 90 Euro as a sure payment?

    (a) lottery
    (b) sure payment

13. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 110 Euro as a sure payment?

    (a) lottery
    (b) sure payment

14. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 140 Euro as a sure payment?

    (a) lottery → go to question 15
    (b) sure payment → go to question 16

15. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 150 Euro as a sure payment?

    (a) lottery
    (b) sure payment

16. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 130 Euro as a sure payment?

(a) lottery
(b) sure payment

17. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 240 Euro as a sure payment?

    (a) lottery → go to question 25
    (b) sure payment → go to question 18

18. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 200 Euro as a sure payment?

    (a) lottery → go to question 22
    (b) sure payment → go to question 19

19. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 180 Euro as a sure payment?

    (a) lottery → go to question 20
    (b) sure payment → go to question 21

20. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 190 Euro as a sure payment?

    (a) lottery
    (b) sure payment

21. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 170 Euro as a sure payment?

    (a) lottery
    (b) sure payment

22. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 220 Euro as a sure payment?

    (a) lottery → go to question 23
    (b) sure payment → go to question 24

23. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 230 Euro as a sure payment?

    (a) lottery
    (b) sure payment

24. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 210 Euro as a sure payment?

    (a) lottery
    (b) sure payment

25. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 280 Euro as a sure payment?

    (a) lottery → go to question 29
    (b) sure payment → go to question 26

26. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 260 Euro as a sure payment?

    (a) lottery → go to question 27
    (b) sure payment → go to question 28

27. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 270 Euro as a sure payment?

    (a) lottery
    (b) sure payment

28. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 250 Euro as a sure payment?

    (a) lottery
    (b) sure payment

29. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 300 Euro as a sure payment?

    (a) lottery → go to question 31
    (b) sure payment → go to question 30

30. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 290 Euro as a sure payment?

    (a) lottery
    (b) sure payment

31. What would you prefer: a 50 percent chance of winning 300 Euro when at the same time there is 50 percent chance of winning nothing, or would you rather have the amount of 310 Euro as a sure payment?

    (a) lottery
    (b) sure payment

Demographic Questions

How old are you?

What is your gender?

What is the highest degree of education you have earned?

What is/was your subject of study?

Would you like to tell us something about this study?

4

THE EFFECT OF INTER-GROUP CONTACT ON ECONOMIC TYPES OF DISCRIMINATION[*]

Inter-group contact has been found to increase or to decrease discrimination in the empirical literature. These conflicting results might originate from differences in addressed types of discrimination – i.e., whether discriminatory behavior arises from differences in tastes or beliefs – and differences in contact's capacity to alter tastes and beliefs. This article investigates the causal effect of contact on statistical and taste-based discrimination as well as on the associated anticipation effects of the latter. In our experiment, individuals are assigned to teams comprising out-group members or to remain in homogeneous teams, interact in a co-operative task, and subsequently play different games apt to elicit their discriminatory tastes and beliefs about out-groups. Our contact intervention remedied taste-based discrimination by about 45%, while it fostered statistical discrimination. Derived lessons for policy makers concerned with the reduction of discrimination involve features that inclusive policies should strive for by changing preferences or beliefs, and thereby reducing different types of discrimination.

1 Introduction

Can inter-group contact reduce discrimination and its preconditions? And what are the requirements for policies to successfully achieve this goal? These fundamental questions have been studied empirically in the social sciences, but with contradicting results. While seminal meta-analyses ascertain that inter-group contact reduces discrimination against out-groups (Pettigrew et al., 2011; Pettigrew & Tropp, 2006), they consider foremost survey studies that regularly face self-selection problems (Bertrand & Duflo 2016). On the other side, causal field experiments deliver equivocal results by revealing that inter-group contact decrease (Burns, Corno, & La Ferrara, 2015; Finseraas et al., 2016; Scacco and Warren, 2018, MacInnis & Page-Gould, 2015), has no impact on (Broockman & Kalla, 2016; Finseraas et al., 2016), or even increases discrimination against out-groups (Bhavnani et al., 2014; Enos, 2014; MacInnis & Page-Gould, 2015). These conflicting results might stem from differences in the applied inter-group contact interventions – i.e., how naturalistic, how long and how intense inter-group interactions are (MacInnis & Page-Gould, 2015; Paluck et al., 2018) – or from differences in the type of discrimination being studied and the utilized measurement approaches – i.e., whether implicit attitudes, actual behavior, stereotypes or prejudices are assessed (Fiske, 1998). A recent meta-study by Paluck, Green, and Green (2018) found that the impact of contact on discrimination against out-groups vigorously varies across studies depending on the type of discrimination being measured and prejudice being addressed, though it identified a general positive effect.

[*] We thank Dirk Sliwka, Christoph Engel, Oliver Kirchkamp, Michael Kurschilgen, Sven Hoeppner, Frédéric Gruninger and Carina Hausladen for helpful discussions and comments on earlier versions of the paper. We thank the participants of various workshops and conference session for helpful comments. Financial support by the Cologne Graduate School in Management, Economics and Social Science, is gratefully acknowledged.

We contribute to resolve existing contradictions concerning the impact of contact on discrimination in the form of in-group favoritism[86] by systematically studying for the first time to what extent contact mitigates different types of discrimination considered in economics. We study whether contact attenuates statistical (e.g., Arrow, 1973; Phelps 1972; Aigner & Cain 1977; Lundberg & Startz, 1983) and taste-based discrimination (Becker 1957), as well as effects arising from the anticipation of the latter, thereby leading to lower inter-group trust (e.g., Alesina & La Ferrara 2002; Song, 2009). While a growing number of recent papers inquire into how discrimination operates (Altonji & Pierret, 2001; Knowles et al., 2009; List, 2004), there is a shortage of empirical work analyzing operating principles of attenuating policy interventions in general. In particular, the contact hypothesis itself is agnostic about its underlying behavioral channels (Pettigrew, 1998). Therefore, we inquire into the capacity of contact to mitigate distinct types of discrimination and into how inclusive social policies should differ contingent on whether they aim to attenuate taste-based discrimination, effects caused by its anticipation or statistical discrimination.

Taste-based discrimination models (Becker, 1957) assume that individuals derive utility from the act of discrimination and experience negative preferences towards out-groups arising from prejudice. Thus, contact would need to affect individuals' preferences to change behavior. Contrary, statistical discrimination (Arrow, 1973; Phelps, 1972; Aigner & Cain, 1977) is considered a rational response to uncertainty. It will occur if one uses stereotypes or group averages as proxies for unobservable relevant characteristics to fill an information void, and thus treat otherwise identical members of two groups differently. Lastly, a process we denote anticipated taste-based discrimination captures how individuals might discriminate in response to others practicing taste-based discrimination: if one expects being discriminated against by prejudiced out-group members, one will refrain from interacting with them, to trust them, or to treat them kindly. Anticipated taste-based discrimination is a hybrid of statistical and taste-based discrimination leading to lower levels of inter-group trust (e.g., Alesina & La Ferrara 2002, Song 2009), with a focus on social in contrast to purely financial risks (see Ashraf, Bohnet, & Piankov, 2006; and Bohnet & Zeckhauser, 2004 for a discussion why a distinction is necessary). For this reason, we examine it separately from statistical discrimination. To mitigate anticipated taste-based or statistical discrimination, contact must change one's beliefs about out-group members' taste or personal characteristics in question, such as out-group members' productivity.

To assess the impact of contact on the outlined types of discrimination, we implement a lab-in-the-field experiment which offers more experimental control than field experiments and utilizes behavioral outcome instead of survey measures. Our experiment entails a one-shot inter-group interaction intervention which assigns members with distinct political identities (U.S. citizens either supporting the Democratic or the Republican Party) into teams and fosters communication and interactions within these teams. In the first stage of the experiment, participants were assigned to teams of four and had to solve a (cooperative) group task, namely matching artists with their corresponding paintings. To study the effect of contact, and its distinctive features, we exogenously varied across treatments whether teams include only in-group or a mixture of in-group and out-group members, and whether team members had the opportunity to communicate with each other. To measure the extent of taste-based, anticipated taste-based, and statistical discrimination, participants subsequently played an other-other allocation game in which they should allocate an endowment

[86] In this article we focus on in-group favoritism (see e.g., Hewtone et al., 2002 for an overview) – i.e., treating members of one's own social group preferentially. For an overview on out-group favoritism we refer to Batalha, Akrami, and Ekehammar (2007).

between a passive in-group and a passive out-group member (Chen & Li, 2009), a trust game (Berg, Dickhaut, & McCabe, 1995), and a novel real effort task game in which one could bet on the productivity of other in-group or out-group participants. Our design does not only allow identification of the causal effect of inter-group interactions on different types of discrimination but also generates insights on how to achieve the desired social and normative goals of reducing discrimination best.

Notably, comprising an artist guessing and an other-other allocation game, our research design is most closely related to the experiment of Chen and Li (2009), who studied in their seminal paper how categorization enhances in-group favoritism from the social identity theory perspective (Tajfel & Turner, 1979). They established a positive impact of social identity on social preferences towards in-groups and a negative effect towards out-groups. Contrary, we contribute to the literature by studying how these differences in social preferences can be attenuated by inter-group contact. Thereby, we assess the impact of assignments to homogeneous and heterogeneous teams and inter-group communication on distinct types of discrimination and behavioral channels underlying the contact hypothesis (Allport, 1954). While Chen and Li (2009) only assess social preferences, we additionally consider beliefs and anticipation effects as alternative sources of in-group favoritism. In contrast to Chen and Li (2009), who rely on artificial groups, we utilize naturally occurring groups with a higher ecological validity.

Our experimental design addresses methodological issues of previous studies. First, in contrast to research designs that rely on happenstance data, random control trials – such as our experiment – are less prone to positive selection biases because they abstract from the explanation that those who have more contact with out-groups are merely more tolerant (Bertrand & Duflo, 2007; Scacco & Warren 2018).[87] Second, Paluck et al. (2018) infer from a meta-analysis on experiments assessing the effect of contact on prejudice that the extent of discrimination relies upon the measurement method of prejudice and the used sample. Using behavioral instead of survey measures, political instead of inborn social identities defining groups (such as ethnicity or gender), and a sample that is more diverse and representative than standard student samples, we contribute to the question of the generalizability of previous findings.

Our results reveal that inter-group interactions attenuate taste-based discrimination – measured in excessively allocated money to in-group members in an allocation game – by a remarkable 45%. The effect was caused by a combination of contact and actual communication instead of a mere association with out-groups and vanished if the number of in-group and out-group members was not balanced. However, contact had no significant effect on beliefs about out-group members' pro-sociality – and therefore anticipated taste-based discrimination – as well as on the general level of inter-group trust. With respect to statistical discrimination, contact lowered, in contrast to our initial hypothesis, participants' willingness to bet on out-group members' productivity, presumably because contact might be experienced as positive on the personal level and thereby mitigates animus but simultaneously as negative on the factual level and thereby deteriorate the perception of competence. Overall, differences in social preferences better predicted the discrimination patterns in our experiment and the mitigation effect of contact than variations in beliefs.

Our findings have theoretical and practical implications: speaking to theorists, we contribute to the discussion whether the effect of contact on discrimination originates from changes in preferences or in beliefs.

[87] The selection problem in non-experimental field studies arises because victims of prejudice may deliberately avoid contact with out-groups that discriminate against them while more tolerant people may actively seek contact with out-groups. Therefore, survey studies that link higher self-reported levels of intergroup contact in daily life with lower self-reported levels of prejudice (Cehajic et al., 2008; Dixon et al., 2010; Semyonov & Glikman, 2009) cannot identify causal effects as contact is endogenous.

Standard economic theory assumes that changes in the behavior of an individual can be entirely explained by variations in beliefs and constraints, while preferences are considered to be stable and hardly malleable by alike interventions (Stigler & Becker, 1977). Contrary, Chuang and Schechter (2015) reviewed and empirically assessed the mean-level stability of behavioral social preference measures over time and find on average only mild correlations between social preferences elicited at different points in time. Furthermore, Kosse et al. (2020) and Sutter, Yilmaz, and Oberauer (2015) find that children's actual social and, respectively, time preferences are directly malleable by a long-term or, respectively, a short-term framing intervention holding other decision constraints constant. In both experiments, beliefs are not presumed to impact the measured behavior or vary across treatments. In line with these two studies, we show that contact leads to a decline in discriminatory behavior primarily due to relative changes in social preferences towards out-group members and not due to alterations in beliefs regarding the pro-sociality and productivity of out-group members.

Speaking to policy makers, inter-group contact (Allport, 1954, 1958) is proposed as a rationale for school desegregation policies and peacebuilding interventions (e.g., Kelman, 1998; Maoz, 2010). Understanding contact interventions is therefore currently a pressing issue, given the recent migratory waves to developed countries, and the diversity of policies and proposals for the integration of minorities in plural societies. As the "Black Lives Matter" movement demonstrates, tensions between ethnic groups dominated the news in 2020 in the U.S., and a cure for racial discrimination has not yet been found.

We cautiously argue that behavioral channels affected by inter-group interactions may also reduce taste-based discrimination by altering preferences in the field, if they go beyond simply including participants with distinct social identities into heterogeneous teams. They should create an environment in which different social groups of equal size actively communicate instead. Page-Gould, Mendoza-Denton, and Tropp (2008) found that while inter-group contact is apt to reduce anxiety towards out-groups, participants in their experimental inclusion study showed little initiative to initiating inter-group interactions after having participated in the experiment. Thus, it is likely not sufficient to include minority students, or co-workers in a group where discrimination is rampant, but they should instead be included in after-class activities, spend leisure time together and share the same infrastructure. Policies that do not fulfill the requirements and do not inhibit segregation remain behind their possibilities.

Another necessary requirement for contact to work is that interactions must be perceived as positive (Aberson, 2015; Allport, 1954). This being stated, our results reveal that while participants communicated more in heterogeneous treatments and inter-group contact was apt to reduce taste-based discrimination, it enhanced statistical discrimination. This might stem from the fact that participants in heterogeneous teams might perceive the interaction as ineffective: heterogeneous teams were, in fact, significantly less successful in the team task and coordinated to a lesser extent. Therefore, to what extent intergroup interactions are perceived as positive or negative has to be evaluated differently based on whether they are meant to reduce taste-based or statistical discrimination. For the latter type, policies must create successful and productive groups such that participants review their prior beliefs and update them in a way that they perceive their interactions with out-group members as productive. In school, heterogeneous groups must be productive, which requires extra effort and monitoring by teachers. In firms, team incentive structures fostering cooperation might be put into effect to enhance team success.

This paper is organized as follows: in the next section we briefly introduce the different types of economic discrimination and discuss which underlying behavioral channels are expected to drive this reduction, and under which conditions they are more likely to attenuate discrimination. In Section 3 we discuss the design of our experiment. Section 4 presents our empirical findings. Section 5 concludes.

2 Theory and Hypotheses

When inter-group contact and communication are structured so that inter-group interactions are perceived as positive, they are presumed to attenuate, according to the contact hypothesis, discrimination (Allport, 1954). While the necessity and sufficiency of distinct conditions of contact to be perceived as positive and apt to achieve this goal – namely equal status between groups, common goals, a cooperative environment, and the support of authorities (Allport, 1954) – are frequently discussed in literature (see Pettigrew et al., 2011 for an overview), the contact hypothesis is silent about how different economic types of discrimination are affected. In this section we thus outline how contact is apt to reduce taste-based, anticipate taste-based and statistical discrimination by altering either preferences or beliefs.

2.1 Taste-Based Discrimination

Taste-based discrimination models define a family of formal models that assume that individuals derive (dis-)utility from interacting and being associated with members of certain groups. The original taste-based discrimination model by Becker (1957) assumes that "*if an individual has a 'taste for discrimination', he must act as if he were willing to pay something [...] to be associated with some persons instead of others*" (Becker 1957, p. 14).

To the contrary, numerous experimental studies (Ben-Ner et al., 2009; Chen & Li, 2009; Falk & Zehnder, 2013; Fershtman & Gneezy, 2001; Ockenfels & Werner, 2014) rely on the assumption that taste-based discrimination goes along with differences in outcome-dependent social preferences towards in- and out-group members, though these studies do not state the re-definition of taste-based discrimination explicitly. That is, individuals with a taste for discrimination do not only care about being associated with certain persons instead of others but additionally treat these persons kindlier and care more about their well-being. Allocation games – in which participants have to decide how to allocate money either between themselves and in-group or out-group members (Fershtman & Gneezy, 2001), or between a passive in-group and a passive out-group member (Chen & Li, 2009) are not apt to measure whether someone prefers to be associated with certain persons instead of others because associations are not put up for consideration. Instead, these games elicit revealed social preferences (Chen & Li, 2009) in that an individual allocates a larger share of her endowment to the person for whose well-being she cares more.

Social preference-based explanations of taste-based discrimination imply that contact would need to change individuals' (social) preferences to mitigate discrimination. In appendix A2, we introduce a formal taste-based discrimination model based on a non-linear version of the Fehr-Schmidt model (1999) that demonstrates that alterations in social preference can explain discriminator patterns. In the following, we make use of the social preference-based definition of taste-based discrimination. Chen and Li (2009) studied whether communication in homogeneous minimal groups can enhance in-group favoritism. Contrary, we investigate how inter-group contact might decrease the differences in social preferences towards in-groups and out-groups, and therefore mitigate in-group favoritism.

Hypothesis 1: *Contact induces higher relative levels of social preferences towards out-group members and thus reduces taste-based discrimination.*

Inter-group interactions may alter social preferences towards out-group members in distinct ways: the affective social distance (Bogardus, 1927) is usually higher towards an out-group than towards an in-group member (Bastian et al., 2012; Sherif & Sherif, 1969),[88] and individuals behave less pro-socially towards others, the larger their social distance is (Bohnet & Frey, 1999; Brañas-Garza et al., 2010; Engel, 2011; Hoffman et al., 1996; Leider et al., 2009). However, inter-group interactions may decrease the social distance towards out-groups (Bastian et al., 2012; Stephan & Stephan, 1985) because they either decrease the experienced anxiety toward out-group members or enhance the ability to put oneself in the position of an out-group member (Allport 1954), and thereby change social preferences.

Alternatively, when comparing groups, individuals regularly build a favorable bias towards their in-group, which creates a positive self-concept (Taijfel & Turner, 1979) and simultaneously builds up prejudice against out-groups (Pettigrew & Tropp, 2008). Thereby, individuals may pursuit an enhancement of self-esteem stemming from a positive distinctiveness towards out-groups (Abrams & Hogg, 2004). In such a setting, communication may lead to de-categorization, because it enables individuals to find categorical dimensions that cut across the original in-group and out-group distinction (Brewer and Miller, 1984). Alternatively, inter-group interactions lead to a redrawing of categorical boundaries: contact changes members' cognitive representations of group memberships. Instead of assigning members to two groups, they perceive the two groups rather as an inclusive social entity (Dovidio et al., 1993; Gaertner et al., 1996). Either way, individuals gain less self-esteem by differentiating between in-group and out-group members. Therefore, affective prejudices and negative affections experienced when interacting with out-group members as well as the intensity with which discrimination is used as an outlet are potentially mitigated by de- and re-categorization.

2.2 Anticipated Taste-Based Discrimination and Inter-Group Trust

Anticipated taste-based discrimination – the tendency that one discriminates in response to the presumed taste-based discrimination of others – is closely related to inter-group trust. Trust is an evaluation of social risks contingent on individuals' expectations regarding the behavior and trustworthiness of others and can be explained – in contrast to pure anticipated taste-based discrimination – by social preferences. A rather distressing empirical result is that the more diverse a society is in general, the smaller is the level of interpersonal trust (Alesina & La Ferrara 2002). The empirical literature on inter-group trust based on the minimal group paradigm (e.g., Yamagishi, Cook, & Watabe 1998; Güth, Levati, & Ploner 2008) and naturally occurring groups across nationalities, races or ethnicities (e.g., Bouckaert & Dhaene, 2004; Buchan et al., 2008; Fershtman & Gneezy, 2001; Goette et al., 2006; Greig & Bohnet, 2008) is, however, ambivalent. In a field experiment, Falk and Zehnder (2013) studied the behavior of participants from 12 different residential districts in Zurich in a trust game (Berg, Dickhaut, and McCabe 1995, Glaeser et al. 2000) and found significant evidence for in-group favoritism. Others, deriving results from correlational studies (Stolle et al.,

[88] Social distance captures the affective distance between two people, i.e., how much sympathy one feels for another person. In social distance studies the focus lies on one's emotional reactions and affections toward other persons and groups. According to Sherif and Sherif (1969) "social distance is a dimension of interaction between members of different groups ranging from intimacy to complete separation (no contact). It is defined by norms governing the situation in which interaction with members of the out-groups is permissible.

2008) or survey experiments (Koopmans & Veit, 2014), argue that contact mitigates the negative relationship between diversity and trust.

***Hypothesis 2:** Individuals are more likely to trust out-group members after experiencing contact with out-group members.*

The degree of interpersonal trust – measured by utilizing the trust or investment game as a workhorse – is contingent on two dimensions: beliefs about potential outcomes and their evaluation in terms of the associated utility level. To assess inter-personal trust both dimensions have to be evaluated together, because neither social preferences (Brülhart & Usunier, 2012) nor the anticipation of social risks or, respectively, anticipated taste-based discrimination (Espín et al., 2016) can alone explain why and when individuals trust.

First, trust depends on how an individual evaluates potential outcomes of social interactions by assessing its psychological (i.e., to what extent social preferences are satisfied) and monetary benefits. The more pronounced one's positive (e.g., altruism) and the less pronounced one's negative social preferences (e.g., envy) are, the more likely one trusts, because one's willingness to trust does not only arise from a prospect of personal benefit but, in addition, also from one's growing interest in enhancing the welfare of the trustee. That is, one is more willing to accept outcomes that are more favorable for the trustee and less favorable for oneself.

In the previous section, we have discussed how contact brings about shifts in social preferences towards out-group members that also apply here. A potential increase in the willingness to trust an out-group member induced by contact may partly be explained by a relative increase in social preferences towards out-group members. Notably, the presence of social preferences explains why individuals regularly make upfront investment in the absence of any perceived advantages, because these investments help to satisfy their prosocial or income equality concerns (Ashraf et al., 2006; Berg et al., 1995; Dunning et al., 2014; Fetchenhauer & Dunning, 2009; Ortmann et al., 2000).[89]

Second, whether one trusts depends on probabilities assigned to each potential statistical event. If individuals anticipate that others are not willing to interact and cooperate as a result of taste-based discrimination (a potential effect discussed in the previous section), they may themselves refuse to cooperate and interact with discriminatory people to reduce their social risks. Consequently, the anticipation of taste-based discrimination reinforces the decline of inter-group trust. We denote this phenomenon as anticipated taste-based discrimination. The willingness to accept social risks or financial risks regularly differs within a person, although the sign of the effect is ambiguous (Bohnet & Zeckhauser, 2004; Fairley et al., 2016; Fetchenhauer & Dunning, 2009). Therefore, we evaluate inter-group trust and thereby anticipated taste-based discrimination independently from the statistical discrimination. Contrary, statistical discrimination addresses how individuals deal with monetary risks emerging from hiring or interacting with less productive persons, rather than social risks arising from the betrayal of trust.

[89] Studies find that people with a higher social value orientation (Kanagaretnam et al., 2009) or stronger outcome-based preferences (Cox, 2004) are more likely to show trust. The detected correlation in within-subject designs between the subject's decision in the role of the trustor and in the role of the trustee also substantiates this hypothesis, since second movers' behavior allows to draw conclusions about non-strategic social preferences (Altmann et al., 2008; Chaudhuri & Gangadharan, 2007; Glaeser et al., 2000; Kovacs & Willinger, 2013).

2.3 Statistical Discrimination

Statistical discrimination models (Aigner & Cain, 1977; Arrow, 1973; Phelps, 1972) assume that individuals have imperfect knowledge about others' attributes in question, such as their productivity (for a review see Fang & Moro, 2011). In the absence of complete information, they rely on their beliefs, which can be true or false, about this attribute based, to fill the information void. While the assessment based on the features of the assumed distribution makes the process statistical, the assignment to different (arbitrary) classes makes it discriminatory. To attenuate discrimination, contact thus needs to change the distribution of individuals' posterior beliefs. In this section, we outline how stereotypes, uninformative priors, and sampling process impact posteriors. Thereby, we analyze how contact changes the belief formation process in itself. We follow the suggestions of Lang and Lehmann (2012) and base our reasoning on endogenously evolved beliefs resting upon initially biased priors and data sampling processes which potentially trigger reinforcement effects (Arrow, 1973).

Consider a selfish and risk-averse decision maker who profits from her interaction partner's productivity. If the decision maker expects the interaction with an out-group member to be riskier than the interaction with an in-group member, her willingness to interact and thus her expected utility from an interaction with an out-group member is lower (respectively, her certainty equivalent is higher). If due to the possibility of belief-updating – emerging from inter-group interactions – one perceives interactions with an out-group member as less risky or more comparable to interactions with an in-group member, the willingness to pay to interact with an out-group member will increase and statistical discrimination will be attenuated.

Hypothesis 3: *Contact reduces statistical discrimination and thereby the willingness to pay to interact with an out-group member.*

Contact may impact beliefs and thereby how risky an interaction is perceived. In the following, we describe how distributional features affecting risks are influenced by inter-group interactions. Precise criteria, such as first- and second-order stochastic dominance, are hardly intuitively accessible. Thus, our analysis rests upon the assumption that decision makers prefer risky choices with higher expected values (Phelps, 1972) and lower variances (Aigner and Cain, 1977).[90] A belief formation process can lead to discriminatory behavior, either because an individual has differing (Arrow, 1973) or differently precise beliefs (Aigner and Cain, 1977) about the characteristic of members from different groups.

When assessing information about others prior to an interaction, individuals cannot gather all relevant information due to time, monetary, or cognitive constraints. Therefore, they have prior[91] beliefs about interaction partners' characteristics deduced from group averages and stereotypes[92] – cognitive schemas used to process information (Hilton & Hippel, 1996) – and update their beliefs whenever they gain new information about the interaction partners or their associated group. Thereby, stereotypes utilized as a heuristic rule in

[90] This concept is often used in management science to evaluate risky decisions in risk (often measured in terms of standard deviations) and return evaluation (Markowitz, 1952). Though there are exceptions in which a risk-averse decision maker prefers lotteries with a higher variance and a lower expected value compared to alternative lottery, Bell (1995) has shown that for a wide range of utility functions it holds that given the expected value of a variable is greater and the associated variance of both random variables is equal, a risk-averse utility maker prefers the variable with the greater expected value. Moreover, Ingersoll (1987) formally proves that for any concave, continuous and differentiable utility function it holds that if the expected values of two random variables are the same and the variance of one random variable is smaller than the other and both variables are elliptically distributed, the variable with the lower variance first-order stochastic dominates the other.

[91] Prior beliefs describe one's beliefs about a variable in question before any evidence is taken into account.

[92] Stereotypes are defined as probabilistic generalizations of group attributes (McCauley et al., 1980).

the search for information (Oakes & Turner, 1990) limit time and cognitive capacities needed for encoding information because they facilitate rapid, initial identification of congruent information (Fiske, 1998). Nevertheless, over-generalization and biases can lead to discrimination based on misjudgments. These misjudgments are mitigated by contact in three ways.

First, the higher the amount and accuracy of the data about interaction partners, the less posterior beliefs depend on prior beliefs and the more they depend on available data about (individual) group members. If one has only little contact with out-groups, one will have little possibilities to collect data about out-group members' characteristics as well as information about the group as a whole. Hence, the posterior beliefs about out-groups are associated with a higher variance and are thus perceived as riskier. Policies designed to foster inter-group communication open up opportunities to collect data. Thereby, information about out-groups becomes more precise, information about individuals becomes more important and stereotypes become less generalizable. This leads to a shift from a categorical to an individualistic assessment of out-groups (Fiske & Neuberg, 1989).

Second, contact impacts how priors and stereotypes evolve. Individuals are presumed to categorize people based on two criteria (Turner, 1985): accessibility[93] and fit (Oakes, Turner, & Haslam 1991; Hornsey 2008). While the former is associated with the categorization process' search costs, the latter is optimized if the within-group differences are minimized while between-group differences are maximized. This leads to stereotypical thinking (Taylor, 1981). Inter-group interactions might reveal that factors previously not considered are more predictive. The original categories are perceived as less informative than initially thought (de-categorization). Alternatively, by enhancing the perceived similarity between groups, contact reduces one's desire to draw a clear line between groups. Eventually, the in-group and the out-group are considered two sub-groups of a broader entity (re-categorization).

Third, individuals are likely more eager to seek information that confirms existing stereotypes, to avoid information that contradicts prior beliefs (Johnston et al., 1994; Ruscher & Fiske, 1990), and to detect stereotypical behavior faster if they support pre-existing stereotypes (Payne, 2001). Such confirmation biases or self-serving beliefs (Akerlof, 1973) leverage the importance of initial priors because informative priors lead to a biased data generating process. Decision makers with strong priors likely update their beliefs to a smaller extent when interacting with out-group members.

3 Experimental Design

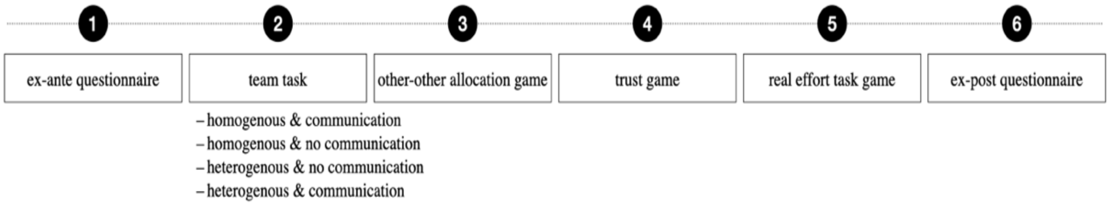
Our oTree-based (Chen, Schonger, and Wickens 2016) lab-in-the-field experiment was conducted in late 2019 and early 2020 at the online labor market Amazon's Mechanical Turk (MTurk) with a total of 440 subjects. It was approved by the ethics committee of the Max Planck Society in 2019. MTurk[94] is a widely used crowdsourcing platform which offers a similar amount of experimental control to lab experiments. We

[93] The application of the accessibility criterion explains why categorization is often associated with salient tags such as physical appearance and easily distinguishable social characteristics such as race, religion or gender (Taylor, 1981) and why priming social identities triggers stereotypical thinking (Hornsey, 2008).

[94] Ample studies covered concerns related to the internal as well as external validity of MTurk studies, and data quality issues such as measurement errors caused by inattentiveness of MTurkers (Bergvall-Kåreborn and Howcroft 2014; Chandler, Mueller, and Paolacci 2014; Paolacci and Chandler 2014; Hauser and Schwarz 2016). Although, the level of inattentiveness and associated measurement errors are a smaller issue in online experiments compared to lab experiments (Hauser and Schwarz 2016), we introduced control questions with respect to the provided instructions as well as response time checks to see whether subjects paid attention to the instructions. The economic decisions made in behavioral economics games on MTurk are comparable to those of standard lab populations (Raihani et al., 2013).

chose to implement our experiment at MTurk because obtained samples are more representative of the country's population than standard student samples (Berinsky, Huber, and Lenz 2012; Shapiro, Chandler, and Mueller 2013; Roulin 2015; Buhrmester, Kwang, and Gosling 2011; Paolacci, Chandler, and Stern 2010).[95] Moreover, university or college students and other highly educated subjects who often experience contact with people from different backgrounds have fewer prejudices. In contrast, our sample allows for the study of the effects of contact on a population not restricted to adult university students under the age of 25, the traditional age range in samples on which the majority of contact hypothesis studies rest upon (Paluck 2018; Pettigrew and Tropp 2006).

Figure 4.1: Experimental Design


1
ex-ante questionnaire
2
team task
– homogenous & communication
– homogenous & no communication
– heterogenous & no communication
– heterogenous & communication
3
other-other allocation game
4
trust game
5
real effort task game
6
ex-post questionnaire

Our experiment included six stages (see Figure 4.1). In the first stage, we assigned subjects to two distinct political groups. To implement this, subjects had to previously state their political affiliation (whether they supported the Democratic or Republican Party, or neither), their social distance towards in- and out-groups, and their level of interest in politics.[96] Only subjects who supported one of the two parties and had at least an average interest in politics (and thus likely a sufficiently pronounced group identity) were allowed to participate. Thereafter, we exogenously introduced inter-group contact intervention by randomly assigning some subjects to heterogeneous teams (consisting of supporters of both parties) and other ones to homogeneous teams (comprising only supporters of the same party).

Naturally occurring groups were, in our experiment, preferable to artificial groups, because they allow us to abstract from the concern of overly high levels of discrimination detected in artificial (minimal) groups (Lane, 2016) – which are likely caused by demand effects (Zizzo, 2010) and high levels of social uncertainty.[97] Moreover, social identities needed to be sufficiently pronounced such that inter-group contact did not induce a new social identity. In fact, our data provided some evidence that after interactions with out-group members, the political identity still accounted for one's self-perception.

In the second stage, all teams had to solve a cooperative team task. While we consider contact to be a combination of the assignment to a heterogeneous team – leading to a temporary association and interaction with out-groups – in combination with inter-group communication, we exogenously altered, beyond the composition of the teams, whether team members were able to chat with each other while solving the team

[95] While in many settings, student samples show similar behavior in comparison to representative samples, this is not the case for experiments dealing with discrimination. In student samples, researchers are less likely to establish a baseline discrimination effect (Henry, 2008), plausibly as students are more prone to social desirability and sensitive to experimenter demand effects.

[96] The survey included four items. Those subjects who did not qualify received a payment of $0.05 cents. All others were guaranteed to receive a fixed payment of $0.50 and a bonus of approximately $3.00 on average. We only recruited those subjects who were located in the United States, have a HIT-approval rate beyond 75% and have not completed more than 5,000 HITs, preferred either the Democrats or the Republicans and answered the question "How much interested in U.S. politics are you? Please state your answer on a scale from 0 to 6 where "0" means "not interested at all" and "6" means "very interested"" at least with a 3.

[97] A plausible reason why the creation of minimal groups induces higher levels of discrimination is that subjects who gain utility from an identification with an artificial in-group may have to signal that they are a legitimate member of the artificial group by practicing in-group favoritism.

task to disentangle the communication from a mere association effect. In the third, fourth, and fifth stage, we measured subjects' tendency for taste-based discrimination, anticipated taste-based discrimination, and statistical discrimination with an other-other allocation game (Chen & Li, 2009), a trust game (Berg et al., 1995), and a new real effort task game in which subjects had the opportunity to bet on the productivity of a matched in-group or out-group partner.

We applied a random stranger matching protocol to guarantee the independence of observations. Subjects were always informed that they would not interact with the same counterparties again in the experiment. To account for order effects, we randomized the order in which subjects played the three games. Lastly, we generated control variables with an ex-post questionnaire covering demographic questions, comprising two items on the social distance towards in- and out-groups and survey items on general tendency for (i) altruism, (ii) positive reciprocity, as well as (iii) negative reciprocity adopted from Falk et al. (2018).[98]

Team task: In the team task stage, subjects interacted in teams of four. Their task was to guess which painter (Paul Klee or Wassily Kandinsky) painted a series of 4 paintings (Chen & Li, 2009). To exclude the possibility of fraud, the paintings had been blurred such that a reverse google image search did not reveal the correct answers. Subjects had four minutes to complete the task. Before the actual team task, they had two minutes to examine two paintings by Klee and Kandinsky. Our payment scheme was meant to foster cooperative behavior as every team member received 10 points for every correct answer provided by each team member.

In the team task stage, we varied whether subjects interacted exclusively with in-group members (*homogeneous condition*) or whether subjects were assigned to a team comprising two supporters of the Democrats and two of the Republicans (*heterogeneous condition*). To assess the robustness of our results, we varied additionally whether in the *heterogeneous* groups the relation of Democrats to Republicans was *2:2, 3:1* or *1:3*. In our main treatments, participants were able to communicate using a chat window (*communication condition*). To disentangle whether the mere association with out-groups (being in a heterogeneous team) or more in-depth effects such as generalized reciprocity or belief updating explain a potential treatment effect, we varied whether subjects were able to communicate using a chat window (*no communication* vs. *communication condition*). In the communication treatments, subjects were informed that they were neither allowed to identify themselves nor to offend or insult other participants. Violation of these rules would lead to the exclusion from the experiment and all payments.

The contact hypothesis presumes that contact more efficiently reduces discrimination towards out-groups if it is structured within a cooperative and egalitarian framework. Such a framework comprises four conditions, namely: (i) those in contact have equal status in the particular context, (ii) they share common goals, (iii) they are in a cooperative environment, and (iv) the contact takes place under some form of authority (Pettigrew, 1998, Pettigrew & Tropp 2006). Our experiment comes very close to fulfilling these four conditions, except for, perhaps, the last one, as in most studies using the minimal group paradigm (Pettigrew & Tropp, 2006).[99]

[98] See the instructions in appendix A1 for the precise wording.

[99] Notably, Pettigrew and co-authors (2011) found that those four conditions merely facilitate but are not essential to reduce prejudice. We moreover ascertain that participants experience inter-group interactions positively, since negative contact likely exaggerates inter-group tensions (Barlow et al., 2012; Graf et al., 2014).

Measuring taste-based discrimination: In the other-other allocation game (Chen & Li, 2009), each subject must distribute 200 points between a passive in-group and a passive out-group subject, with more points allocated to an in-group revealing discrimination against out-groups. Every subject determines the payoffs of two subjects, while her payoffs are determined by two other subjects. Any observed discrimination (providing more points to an in-group than to an out-group) must be, in a one-shot game, taste-based (Lane, 2016). We provide a formal derivation of the relationship between relative social preferences towards out-groups and the share of the endowment towards out-groups in appendix A2.

The other-other allocation game is designed to test our first hypothesis: "inter-group interactions induce higher relative levels of social preferences towards out-group members." To elicit the impact of contact on social preferences, we assess whether participants who interacted in heterogeneous teams allocated larger shares to out-group members when being put into contact with out-groups.[100] To test whether the mere assignment to a heterogeneous group, and thus the association with out-groups, is sufficient to trigger an effect, we also test whether there is a difference between the heterogeneous and the homogeneous treatment without communication.

We hypothesized that differences in social preferences may stem from differences in the social distance towards out-group members, and that inter-group contact increases the social distance towards out-group members. To inquire into this behavioral channel, the social distance before and after the contact intervention was elicited. This allows for the assessment of a potential negative, correlational relationship between social distance and taste-based discrimination and a negative relationship between contact and social distance towards out-group members.

Measuring inter-group trust and anticipated taste-based discrimination: In the trust game, the variable of interest is the amount transferred to the trustee as well as the subject's beliefs on the trustworthiness of the trustee who can be *either* an in-group or an out-group subject. One subject per group was endowed with an amount of 100 points, anonymously paired and assigned to either the role of trustor or trustee as well as informed about the group identity of their respective counterpart. At stage one, the trustor might send nothing or a portion x of the endowment to the trustee. The trustor then kept 100 - x, and the remaining amount was tripled, such that 3x is received by the trustee. In the second stage, the trustee either passed nothing or any portion y of the money received back to the trustor. Trustors could transfer any amount between 10 and 100 points in steps of ten to a trustee who could be either an in-group or an out-group and were subsequently asked about their beliefs on how much in-groups and out-groups will reciprocate trust. The returned amount captures the trustee's trustworthiness, for whom we apply the strategy method (Selten, 1965). Hence, being in the role of the trustee, subjects had to answer how much they would like to return to the trustor dependent on ten possible levels. Every subject in our experiment had to make decisions for two different scenarios that differed in the subject's role. The computer randomly implemented one of the two scenarios. Thereafter, we elicited beliefs how many percentage points in-group and out-group members on average return to trustors in an incentivized manner.

[100] Assessing how decision makers allocate endowments between passive participants allows for abstraction from strategic social preferences and from interaction costs associated with being assigned to an out-group member, since the interaction itself is not avoidable. We also argue that the alternative explanation that changes in the interaction partners' beliefs or norms lead to changes in discriminatory behavior is implausible, because in our experiment the allocators' decision affects participants with whom they have not interacted before.

The trust game is designed to test our second hypothesis: "*individuals are more likely to trust out-group members in the presence of inter-group interactions*." If inter-group interactions enhance inter-group trust, trustors should send higher shares of their endowment to trustees. We hypothesized that an increase in the willingness to trust caused by inter-group interactions might partly be explained by a relative increase in social preferences towards out-group members. Hence, we assess a correlation between social preferences and the willingness to trust out-group members as well as the impact of contact on trustees' behavior. We elicited how much participants returned as trustees and include this social preference measure in some of the regressions assessing trustors' investments. To differentiate whether a potential enhancement of inter-group trust might alternatively be caused by an anticipation effect, we statistically assessed the impact of contact on beliefs about the behavior of trustees and the impact of beliefs on inter-group trust. We test whether there is a positive correlation between inter-group interactions and beliefs and the correlation between beliefs and the amount sent by trustors.

Measuring statistical discrimination: To measure statistical discrimination while abstracting from taste-based discrimination, subjects played a two-staged real effort task game. In the actual real effort task, subjects worked on an encryption task similar to the task presented by Erkal, Gangadharan, and Nikiforakis (2011). Subjects were shown a series of character combinations, each consisting of three letters, on a computer screen. They were asked to encode the presented combination of characters by replacing each of the three letters with a respective three-digit number which can be learned from a coding table presented on the screen and which changed after each new series of characters.[101]

The individual payoff from the real effort task game depended on the number of correct answers provided by the individual and potentially by a partner assigned to the particular subject, though the assignment was non-mutual. That is, person A's payoff might have been co-determined by person B, although B's payoff was not co-determined by A. For every code the subject or her partner encrypted correctly, the subject received 10 points. We elicited the willingness to pay for this opportunity to have one's payoffs co-determined by either an in-group or an out-group member (Becker, Degroot, & Marschak, 1964) and exogenously varied the political orientation of the assigned interaction partner**.** Subjects knew the group identity of the partner they could bet on.

In particular, we asked subjects to decide between option A and option B in 10 different scenarios. When a subject selected option A, she preferred to have her payoffs co-determined by another participant in exchange for a fee of the amount x. When option B is chosen, the participant neither had to pay a fee nor were her payoffs co-determined by another participant. The amount x varied between 1 point and 50 points in steps of 5 points. Subjects were only allowed to select option A in a single scenario if they had not selected option B in a scenario with a lower price x to acquire the selection option. This secured the absence of multiple switching points. If participants at least once selected option A, they had to state whether they prefer that their payoff is co-determined by an in-group or by an out-group member. Finally, the computer randomly chose one of the 10 scenarios and implemented the decision for option A or B that had been made. It was necessary that participants were familiar with the effort task before they faced any partner selection

[101] In the original paper by Erkal, Gangadharan, and Nikiforakis (2011) subjects had to replace letters with one and two-digit numbers. A potential drawback of this procedure is that the encoding of different letters might vary in difficulty. Thus, we use three-digit numbers where all three digits are different from another. The task is easy to understand and requires no previous knowledge. There is little scope for guessing.

decision on this basis. Consequently, we introduced the partner selection stage only after participants got familiar with the task. For solving the test task correctly, subjects received 50 point that could later be used to pay the price to acquire the partner selection option.

The real effort task game is designed to test our third hypothesis: "*inter-group interactions reduce statistical discrimination. Hence, the willingness to pay to interact with an out-group member decreases after an inter-group interaction."* In our experiment, participants betted on the productivity of in-group and out-group members in a real effort task. This allows to elicit participants' willingness to pay to have their earning co-determined by an out-group member. To further analyze the impact of beliefs on the decision makers' willingness to bet on the productivity of an in-group or an out-group member, we assess the role of contact on beliefs and the role of beliefs on selection decision. To assess the general role of priors, we empirically assess whether participants with stronger prejudice (who had a large relative social distance towards out-group members in comparison to in-group members) and thus potentially more biased priors discriminated more. Finally, we also study on a correlational level whether decision makers with strong priors updated their beliefs to a larger extent when interacting with out-group members by analyzing whether subjects who had a large relative social distance towards out-group members in comparison to in-group members did to a smaller extent consider beliefs when deciding how much to bet.

4 Results

4.1 Sample Description

In total, 440 subjects took part in our experiment, with a small majority of supporter of the Democratic party (62%), henceforth Democrats. Average age was about 37 years, and a roughly equal number of females (52%) and males participated. The average age of Democrats was 36.1 years, while Republicans were on average 38.1 years old. The share of females among Democrats (56%) is larger than among Republicans (46%). Hence, we control for age and gender in our regressions. A discussion about differences in further demographics between Democrats and Republicans is provided in the appendix A3.

We elicited, in the final questionnaire, items (*see* Table A4.1 in appendix A3) that are known to predict pro- and anti-social behavior introduced in the global preference study conducted by Falk et al. (2018). Democrats did not significantly differ from Republicans in three out of four items (on altruism and positive reciprocity).[102] However, Democrats are slightly more prone to negative reciprocity, although this effect is only weakly significant according to a Mann-Whitney-U-test (p=0.06). With that stated, negative reciprocity is not considered a determinant of any of our applied games according to the behavioral economic literature.

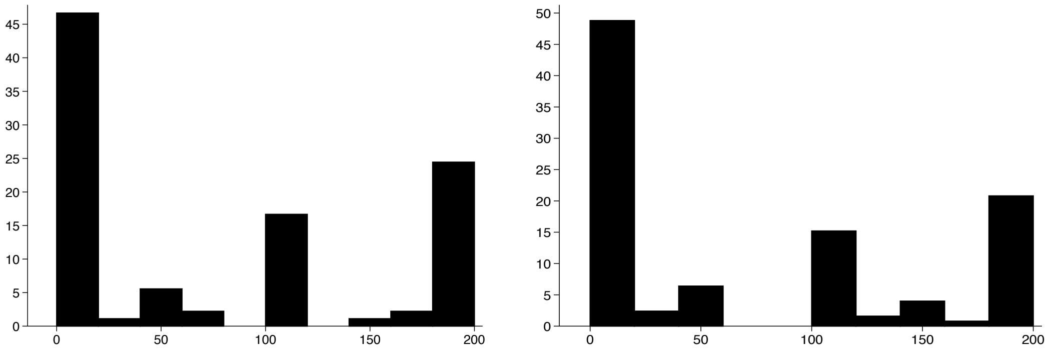
[102] Previous studies have found a correlationals between a high social value orientation and subjects' willingness to vote for liberal instead of conservative parties (Lange et al. 2012). Zettler, Hilbig, and Haubrich (2011) established a correlational effect between self-reported altruism and the willingness to vote for liberal parties. We thus elicited the following four items from the global preference study by Falk et al. (2018): **Altruism 1**: Imagine the following situation: You won 1,000 dollars in a lottery. your current situation, how much would you donate to charity? **Altruism 2:** How do you assess your willingness to share with others without expecting anything in return when it comes to charity? (scale from 0 – 10). **Positive Reciprocity:** Imagine the following situation: You are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 dollars in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 dollars, the most expensive one 30 dollars. You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you want to give? **Negative Reciprocity:** How do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly? (scale from 0 – 10)

Eventually, we elicited a set of proxy variables measuring the identification with in-group as well as out-group members (see Table A4.2 in the appendix A3): the median political interest of both Democrats and Republicans is 5.[103] In addition, measuring the social distance between a participant and a supporter of the Democratic or Republican party allows us to draw conclusions about how much subjects associated with in-group members and disassociated with out-group members.[104] There is no significant difference in the social distances towards in-group and members between Democrats and Republicans. However, Democrats do significantly identify less with out-group members (MWU-test: p=0.01), though the difference is less than half a point on a 7-item Likert scale, and thus relatively small. We find no evidence that Democrats and Republicans identify themselves with in-group members to a different extent (MWU-test: p=0.33).

4.2 Taste-Based Discrimination

Descriptive Statistics – The analysis of the capacity of contact to mitigate taste-based discrimination requires the presence of discriminatory tendencies in the absence of contact. This requirement is satisfied, in our sample, since the majority of subjects discriminated in the other-other allocation game against out-groups by allocating more points to in-group members when teams were homogeneous. Figure 4.2 reveals that around 55% of all participants discriminated when they were assigned to homogeneous groups and able to communicate, with about 20% of all subjects allocating 100 points more to an in-group than to an out-group member and a remarkable 25% of all subjects discriminating to the maximum possible extent of 200 points.[105] Overall, participants distributed on average around 70 points more to in-group members. The average allocation to out-group members is less than 50% of the allocation to in-group members.

Figure 4.2: *Histograms of points excessively allocated to in-group members in the homogenous treatment in the presence (left) or the absence (right) of communication in the other-other allocation game*



Out-group members' payoffs are lower than in a similar other-other allocation game in the paper by Chen and Li (2009). In Chen and Li's (2009) experiment – which rests upon an artificial group assignment procedure instead of political identities – out-group members received around 35% less than in-group members.

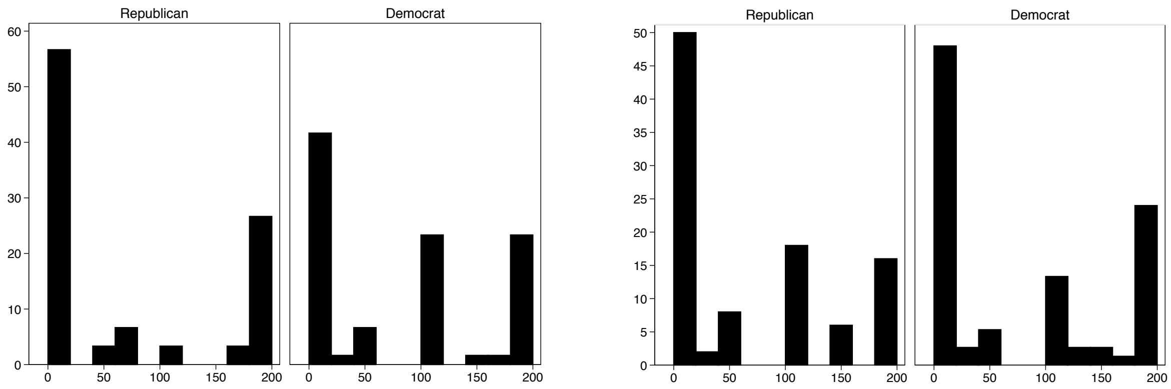
[103] We measured the political interest of subjects using a 7-point Likert Scale. Only those subjects who had a value above 2 were allowed to participate. It is reasonable that the higher the political interest, the more pronounced is the effect of the political affiliation one subject's personal identity and the identification with a certain political group. Indeed, the level of political interest is correlated with the social distance of the in-group member. The Democrats' average political interest (5.0) is only marginally higher than the Republicans average political interest (4.73), though the political interest significantly differs between the two groups (MWU-test: p = 0.01). However, there is no significant correlation between the major dependent variables of our overall analyses and the political interest of the subject using both Person's r as well as Spearman's rank correlation coefficient.

[104] We measured the social distance between a subject and a hypothetical in-group member on a 7-point scale using graphical illustrations of the perceived social distance between the respective subject and either an in-group or an out-group member. On our scale the number 1 indicated the largest social distance, and the number 7 indicated the smallest social distance.

[105] Only two subjects, one Republican and one Democrat, allocated more points to an out-group than to an in-group. These two observations are not considered in the analysis of taste-based discriminatory behavior.

Discrimination based on taste against supporters of the opposing political party is clearly more pervasive and stronger than discrimination against other individuals belonging to an artificial group, reflecting, perhaps, the strong polarization in politics observed today in the United States. Figure 4.3 entails two histograms of points excessively allocated to in-group members comprising observations from Democrats or Republicans in homogeneous teams in the presence and the absence of communication. Democrats discriminated to the same extent as Republicans when subjects could communicate (MWU test: *p*=0.36) as well as when they could not (*id.*, *p*=0.69).[106]

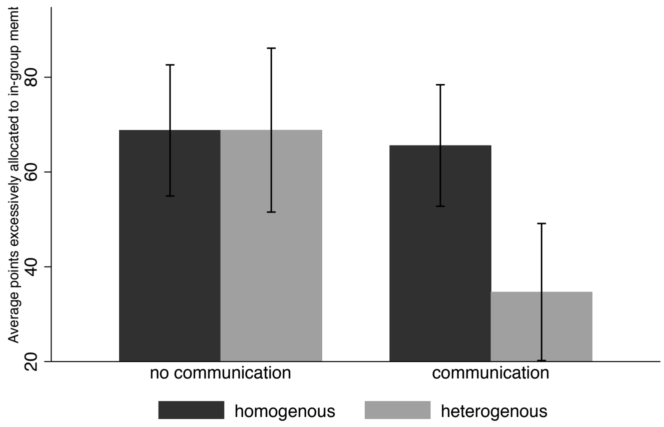
Figure 4.3: *Histograms of points excessively allocated to in-group members in the homogenous treatment in the in the other-other allocation game (without communication: left; with communication: right)*



Assessing the main treatment effect of contact we find that subjects discriminate in the absence of any reason for either anticipated taste-based or statistical discrimination about 45% less when being assigned a heterogeneous group and having the opportunity to communicate (MWU test: *p*<0.001), as depicted in Figure 4.4. In addition, Figure 4.4 reveals and a corresponding MWU test (*p=0.73*) corroborates that the assignment to heterogeneous teams in the absence of communication did not have any effect on taste-based discrimination. We conclude that actual communication triggered the treatment effect and altered social preferences.

Result 1: *Contact reduced taste-based discrimination by 45%. We find support for hypothesis 1.*

Figure 4.4: *Average points excessively allocated to in-group members by treatments (95%-confidence intervals included)*



[106] The baseline results are qualitatively and quantitatively in line with the findings of Chen & Li (2009) who utilized an other-other allocation game to study in-group favoritism in a setting with – in contrast to our setting – artificially generated groups.

Regression Analyses – Next, we assess the treatment effect of contact (defined as communication in combination with the association with out-group members) on taste-based discrimination. Table 4.1 comprises linear regressions with robust standard errors that depict the effect of contact on taste-based discrimination accounting for control variables. Model (1) includes only those subjects who could communicate. Model (2) includes those who could not. Model (3) distinguishes the effect of contact in the absence and in the presence of communication. In all of our models, we control for age and gender. All three models comprise age and gender control variables.

Table 4.1: *Other-Other Allocation Game*

	Model 1 (no com.)	**Model 2** (com.)	**Model 3** (combined)	**Model 4** (com.)	**Model 5** (com.)	**Model 6** (com.)
	Dependent variable: ***Excess money given to in-group members***					
Heterogenous Treatment	1.65 (11.38)	-37.11*** (11.76)	1.75 (11.38)	-26.07** (12.16)	87.85** (36.71)	
Communication			6.07 (11.38)			
Communication X Heterogenous Treatment			-39.62** (16.17)			
Number of Out-group Members = 1						-17.11 (12.98)
Number of Out-group Members = 2						-25.34** (11.87)
Number of Out-group Members = 3						0.057 (24.50)
Difference in Social Distances				13.39*** (3.62)	13.46 *** (3.81)	15.17 *** (3.04)
Words in Conversation					0.02 (0.29)	
Correct Answers in Team without Subject					-19.99* (11.29)	
Words in Conversation X Heterogenous Treatment					0.22 (0.42)	
Correct Answers in Group without Subject X Heterogenous Treatment					25.65* (14.75)	
Age	0.72 (0.49)	0.39 (0.57)	0.60 (0.37)	0.36 (0.51)	0.67 (0.49)	0.41 (0.40)
Female	-3.12 (11.49)	-14.70 (11.70)	-8.37 (8.20)	-23.31** (11.39)	-20.48* (11.62)	-19.04** (9.67)
Constant	42.81 (19.7)	66.43*** (22.22)	50.64*** (15.97)	24.01 (25.93)	47.56 (36.63)	14.15 (20.61)
Number of obs.	202	167	369	167	156	229
R^2	0.01	0.08	0.11	0.16	0.19	0.16

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01; ordinary least square regressions with robust standard errors in parentheses. The variable Communication = 1 if participant is able to communicate with other members in her groups. The variable Heterogenous Treatment = 1 if participant is assigned a team comprising two Democrats and two Republicans. The variable Difference in Social Distances measures the difference between the social distances towards an in-group and an out-group member. Words in Conversation measures the total number of words exchanged in the team. Correct Answers in the team measures the number of correct answers provided by team members other than the observed participant. Participants who have a smaller social distance towards out-group than towards in-group members are excluded.*

We conclude from Model (2) that in the presence of communication being in a heterogeneous team significantly decreased the average amount of points excessively distributed to in-group members by about 37 points. Thus, contact reduced taste-based discrimination and regression results provide additional support for hypothesis 1. Model (1) reveals that the mere assignment to a heterogeneous team had no significant impact on taste-based discrimination in the absence of the opportunity to communicate. Hence, we confirm that communication in heterogeneous teams was causal for the treatment effect on taste-based discrimination and alteration in revealed social preferences. Model (3) confirms the results of Model (1) and Model (2) as well as our non-parametric treatment tests: the effect of being assigned a heterogeneous team was insignificant in the combined model (t-test: *p*=0.878). Contrary, being assigned to the heterogeneous chat treatment reduced the amount excessively distributed to in-group members by about 38 points (t-test: *p*=0.015).

Discussion – Having established a causal attenuation effect of contact on taste-based discrimination, we turn to additional investigation based on correlational, not causal evidence, and survey, instead of behavioral measures. Hence, derived implications must be interpreted more cautiously but they still deliver information useful for understanding the mechanisms behind the observed causal effects. We begin by assessing the impact of the differences in affective social distance on taste-based discrimination. In our theory section, we discussed that alterations in social distance potentially explain why participants may experience stronger social preferences towards out-group members after participating in inter-group interactions, and thus practice taste-based discrimination to a lesser extent. Notably, the difference in the social distance towards in-group and out-group members can be considered a proxy variable for the extent to which one's political identity is considered to be a distinction criterion.

In Model (4) – which relies on the same subsample as Model (2) – we include in addition to the variables considered in Model (2) the variable "difference in social distance."[107] Model (4) reveals that the higher the initial social distance towards in-group and out-group members, the higher is the detected level of in-group favoritism. On average, participants allocated about 15 points more to in-group members for every increase of 1 point on the scale measuring the difference in social distance.[108] Therefore, affections and the perceived social distance between participants are strongly correlated with taste-based discrimination in general, as proposed. The higher the observed social distance, the higher is the extent of taste-based discrimination.

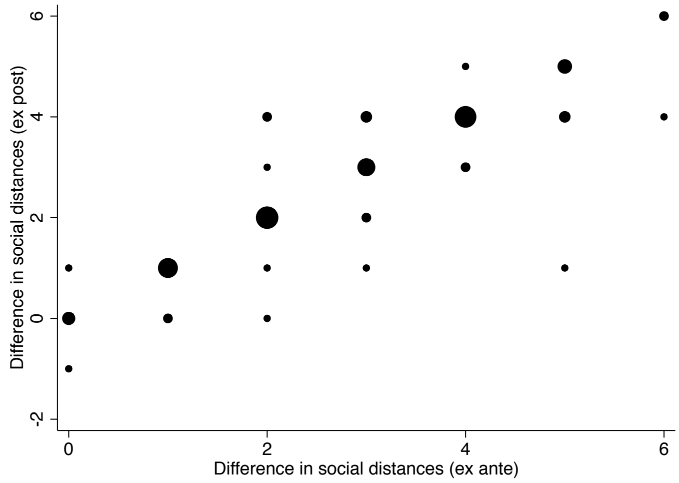
However, other than presumed, changes in the difference between the affective social distance seem not to trigger alterations in revealed social preferences and thereby a reduction of taste-based discrimination. In fact, the difference in social distance elicited before and after the experiment had not changed for the majority of subjects (57 out of 80) who were assigned the heterogeneous teams and were allowed to communicate (Wilcoxon signed-rank test: *p*=0.14). Thus, we find no causal effect of contact on the difference in social distances as well as on social distances per se. Figure 4.5 entails a scatter plot which graphically illustrates that contact has only little, and a non-systematic impact on the difference in social distances. Overall, our correlational result indicates the divergence between proposed underlying behavioral channel, actual discriminatory behavior and mitigation measures (Fiske 1998). This is in line with the findings of

[107] We calculate the difference in social distance towards an arbitrary in-group member and an arbitrary out-group member between the beginning and the end of the experiment. This difference estimator can take values between -6 and +6. It is a proxy for how much more a subject identified ex-ante with her in-group compared to the out-group. The difference is higher the more the subject associates with her own in-group and dis-associates with her out-group. We exclude from the analysis nine outliers whose difference in social distance value was below 0, since we cannot ensure that subjects may accidentally select the opposite social identity. In addition, the behavioral patterns of these 9 subjects seemed to be atypical.

[108] Furthermore, MWU tests reveal that in neither the homogeneous nor the heterogeneous treatments and independent from the possibility of communication, the difference in social distance measured ex-ante as well as ex-post the experiment did not differ at any convenient significance level.

Sacco and Warren (2018), who in an experimental study measuring attitudes as well as behavior find that contact leads to alterations in behavior without changing potential underlying attitudes. A speculative presumption about this divergence – to be addressed by future studies – is that contact changes how relevant affections in comparison to other decision criteria for allocation decisions are, and thus relatively fast changes behavior. To the contrary, contact does not change affections in the short but presumably only in the long run (see Scacco and Warren 2018 for a comprehensive discussion).

Figure 4.5: *Scatter Plot of Difference in Social Distances before and after the Contact Intervention (circle size indicates frequency of obs.)*



Next, we assess the impact of team composition and communication patterns on taste-based discrimination on a correlational level. Model (5) includes, in addition to Model (4), a proxy of the amount of communication (*number of words in conversation*) and the number of correctly solved tasks by either the team or by the particular subject, as well as interaction terms with the heterogeneous team treatment dummy (see Figures 4.6 and 4.7). In heterogeneous teams, participants communicated more with each other (on average 10 words more), though the effect is not significant (MWU test: p = 0.26).

Figure 4.6: Quantity of Communications by Treatment; *(95%-confidence intervals of means included)*

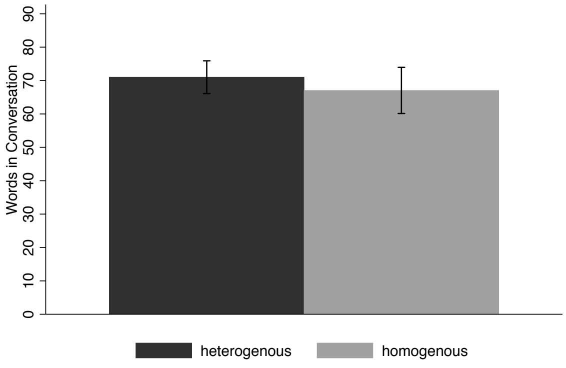
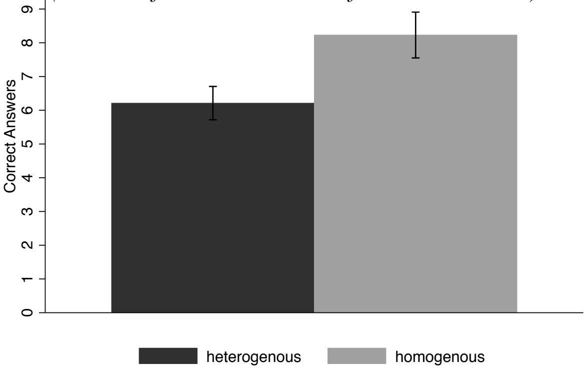


Figure 4.7: Number of Correct Answers in the Presence of Communication by Treatment *(95%-confidence intervals of means included)*



To the contrary, the number of correct answers was with 8 points in homogeneous groups larger than with about 5.5 points in heterogeneous groups (MWU test: p <0.001). Model (5) reveals that the mere amount of communication had no significant effect on taste-based discrimination. However, the number of correctly solved tasks significantly increased taste-based discrimination in heterogeneous in comparison to homogeneous teams, as the interaction effect in Model (5) reveals. We conclude that even though there exists a negative impact of contact on the group success which may lower the capacity of our contact intervention to attenuate taste-based discrimination, the overall mitigation effect on taste-based discrimination was still very pronounced. When considering the effect of contact in itself, holding constant the effectiveness of communication, contact might be apt to reduce discrimination to even larger extents. Potentially, inclusive social policies which include features apt to enhance team outcomes (e.g., including optimal sorting strategies) may even increase the effect size of the established mitigation effect of contact on taste-based discrimination.

Lastly, we assessed whether the positive impact of inclusive policies on taste-based discrimination is also present in groups of unequal size, i.e., one group has the majority status, the other the minority status. Being in a minority or a majority group can be considered as one status dimension. In Allport's (1954) seminal paper, equal group status was considered one of the key requirements for contact in order to be apt to reduce inter-group conflicts. Previous experiments have shown that low-status members being part of a minority tend to give less to high-status members belonging to a majority, while high-status members who feel entitled to the endowment may give in return less to low-status members (e.g., Liebe and Tutic 2010).

In order to test the necessity of an equal group size, we exogenously varied in additional control treatments the number of in-group or respectively out-group members within one team. Thereafter, we regressed the number of out-group members on the excess points given to the in-group members.[109] We find that the causal positive effect on being in a group with two out-group members is roughly twice as large as the positive effect of being in a group with either one or three out-group members (*see* Model 6). The effects between the homogeneous and the heterogeneous treatments with either one-group member or three-group members are not statistically significant.

4.3 Inter-Group Trust and Anticipated Taste-Based Discrimination

Descriptive Statistics – In the following, we analyze the impact of contact on inter-group trust and anticipated taste-based discrimination. In a pilot study applying a within-subject design, none of our participants varied their level of trust contingent on whether being matched with an in-group or an out-group member. To reduce a potential, mitigating experimenter demand effect on the baseline level of discrimination, we applied a between-subject design in our trust game. That is, we elicit a single decision per subject in the role of the trustor while being matched with *either* an in-group or an out-group member.

In the trust game, trustors were less willing to trust out-group than in-group members as depicted in Figure 4.8, both in the absence of communication and in its presence, albeit to a smaller extent in the latter case. In the absence of communication, when being assigned to homogeneous teams, trustors sent with about 37 points about 16 points less to out-group trustees (MWU test: p=0.003). However, the effect was not significant in the presence of communication (MWU test: p=0.28). An ex-post power analysis revealed that for a given α=0.05 the two-sided MWU test was likely not underpowered (1-β = 0.99). Consequently, the detected null effect is likely not the result of a type-II error.

In the homogeneous treatment with communication, Democrats were not significantly more willing to trust Republicans than vice versa (MWU test: p=0.96). Assessing the main treatment effect of interest, we

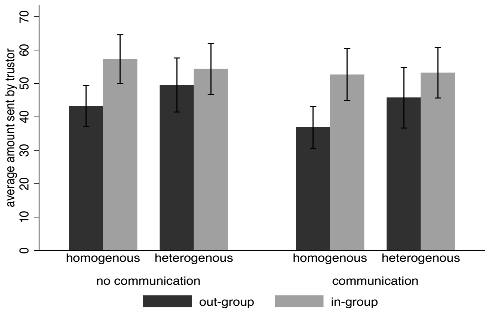
[109] We restricted our regression to those treatments in which subjects were able to chat

find that being assigned to a heterogeneous group, in itself, did not induce higher levels of trust on in-groups (MWU test: p=0.57) or out-groups (MWU test: p=0.31) in the presence of communication. The difference in the trust level between in-group and out-group members was, with 4.3 points, almost identical.[110] In the absence of communication, the level of trust towards in-groups (MWU test: p=0.52) or out-groups (MWU test: p=0.36) did not differ between homogeneous and heterogeneous treatments. In conclusion, there was only limited scope for a significant decrease in discrimination, and the observed increase of inter-group trust, in the presence of communication, was not significant.

Result 2: *Contact did not reduce inter-group trust arising from anticipated taste-based discrimination. We find no causal evidence for hypothesis 2.*

The presence of taste-based discrimination and the simultaneous insignificance of a difference in the level of trust towards in-group and out-group members in the homogeneous communication condition is in line with the findings of Gneezy, List and Price (2012). They find in a series of field experiments that if the object of discrimination is considered to be controllable (such as political identities), animus-based discrimination is stronger than any form of discrimination that relies on beliefs regarding others' characteristics or behavior.[111]

Figure 4.8: Trustors' average transferred points by communication, out-group identity and team heterogeneity, *(95%-confidence intervals of means included)*



Regressions – Table 4.2 presents the results of linear regressions with robust standard errors suited to test hypotheses 2 (contact mitigates differences in trust towards in-group and out-group members). Model (2) is restricted to observations in which participants were able to communicate. Contrary, Model (1) is based on those subjects who were not able to communicate. Model (2) reveals no significant discriminatory behavior of subjects, and contact did not significantly reduce discrimination. Hence, we find no support in Model (2) for hypothesis 2. An ex-post power analysis revealed that the power of the underlying t-test is sufficiently high (1- β = 0.99). Consequently, the detected null result is likely triggered by an insufficiently low sample size. On the other side, Model (1) reveals that out-group members are trusted less but provides no evidence that contact enhances inter-group trust. Eventually, Model (3), which combines observations in the absence and presence of communication, indicates that participants significantly discriminate against out-group

[110] An ex-post power analysis of the applied t-tests revealed that the tests were not only insignificant but also had sufficient power for an alpha of a = 0.05 (1- β > 0.99).

[111] The findings are theoretically justified by the attribution theory. In contrast, if discrimination is based on incontrollable characteristics, such as ethnicity, Fershtman and Gneezy (2001) have found in a study investigating discriminatory behavior by Ashkenazic Jews against Eastern Jews by utilizing a trust game that not animus, but mistaken ethnic stereotypes solely predict discriminatory patters. These findings are corroborated by Gneezy and List (2012). However, a study by Falk and Zehnder (2013) has shown that also for controllable characteristics such as the area within the city of Zurich in which a subject lived, both animus-based in-group favoritisms as well as anticipation effects explained discrimination.

members. Again, neither being assigned a heterogeneous team, nor communication or its interaction significantly remedied the lower levels of trust towards out-group members.

Discussion – We assess to what extent beliefs – and thus anticipated taste-based discrimination – or social preferences better explain patterns of inter-group trust. In Model (4), which is restricted to treatments in which communication is present, we consider whether a change in the social preferences of the subjects and the change in the actual expectations about the behavior of others affected trust. This analysis is based on correlational data. It reveals that in the presence of communication, social preferences and beliefs about the behavior of others positively affected trust in out-group members, though not in in-group members. Beliefs regarding the amount returned by trustees significantly increased trust. Hence, we conclude that the psychological processes underlying trust discussed in the theory section indeed were capable of describing the behavior in the trust game.

To get a more fine-grained picture we also compare whether trustees returned on average more to out-group members given that they were part of a heterogeneous group and were allowed to chat. The average amount returned to the trustor across all 10 potential trust levels was with roughly 41% of the transferred amount only 3% higher in heterogeneous groups. If out-group trustors were willing to send the entire endowment to the trustee the amount returned was, with 136 points, 14 points higher than in homogeneous groups, though the difference in ranks is not statistically significant (MWU test: $p = 0.81$). We conclude that while social preferences significantly predict trust, we find no significant difference between elicited pro-social preferences towards either in-group or out-group members between homogeneous and heterogeneous treatments.

Moreover, we presumed a statistical difference in beliefs regarding the behavior of trustees that potentially mediates the reduction of anticipated taste-based discrimination in the heterogeneous chat treatment. The assumed average amount of the sum returned to the trustor was, with roughly 40%, only about 4% higher than the actual amount sent by trustors. MWU tests indicated that the beliefs regarding the behavior of in-group trustees or respectively out-group trustees do not significantly differ between the heterogeneous and the homogeneous communication treatment (in-group: $p = 0.14$; out-group*:* $p = 0.52$*)*. We conclude that beliefs regarding the transfer payment of trustees significantly predict trust in the trust game. However, the difference in beliefs between homogeneous and heterogeneous treatments regarding the transfer amount of neither in-group nor out-group trustees is not significant.

Lastly, for the purpose of the comparability of results between games, we test in Model (5) whether the number of out-group members in the team has an effect on trustors' willingness to trust. Model (5) only includes subjects who were able to communicate. It reveals that neither the number of out-group members had an impact on the trust in in-group members (see baseline level of number of out-group members dummy) nor the number of out-group members in combination with communication (see interaction terms' coefficients) had a significant effect on trustor's willingness to trust.

***Table 4.2:** Trustors' Behavior in The Trust game*

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
	(no com.)	(com.)	(combined)	(com)	(com.)
Dependent variable: ***Amount sent by trustor***					
In-group Counterpart	13.48***	3.98	14.19***	3.48	3.18
	(4.91)	(5.61)	(4.83)	(5.39)	(5.45)
Heterogenous Treatment	-5.38	-3.65	-5.49	-4.05	
	(4.42)	(6.30)	(64.64)	(6.04)	
Number of Out-group					-5.27
Members = 1					(6.71)
Number of Out-group					-3.88
Members = 2					(6.09)
Number of Out-group					20.87
Members = 3					(13.31)
Number of Out-group Members					0.85
= 1 X In-group Counterpart					(9.89)
Number of Out-group Members					2.69
= 2 X In-group Counterpart					(7.96)
Number of Out-group Members					-5.98
= 3 X In-group Counterpart					(16.33)
In-group Counterpart X	1.53	3.35	1.49	1.75	
Heterogenous Treatment	(7.00)	(8.27)	(7.00)	(7.95)	
Communication			6.57		
			(5.42)		
Communication X			10.16		
In-group Counterpart			(7.46)		
Heterogenous Treatment X			2.44		
Communication X In-group			(10.74)		
Counterpart					
Trustor's beliefs			0.25**	0.25**	0.22**
Average amount returned			(0.10)	(0.10)	(0.09)
as a trustee			21.29*	21.29*	2126***
			(11.5)	(11.5)	(79.98)
Age	0.15	0.32	0.21*	0.12	0.16
	(0.15)	(0.21)	(0.13)	(0.19)	(0.15)
Female	-7.54*	-2.15	-5.07*	-1.24	-7.74**
	(3.91)	(4.27)	(2.88)	(4.14)	(3.66)
Constant	42.07***	29.60***	37.79***	28.55***	29.63***
	(7.35)	(7.74)	(6.19)	(7.79)	(6.88)
# of obs.	202	167	369	167	229
R2	0.24	0.03	0.17	0.12	0.14

***Notes:** *=p≤0.10; **=p≤0.05; ***=p≤0.01; ordinary least square regressions with robust standard errors in parentheses. The variable Communication = 1 if participant is able to communicate with other members in her groups. The variable Heterogenous Treatment = 1 if participant is assigned a team comprising two Democrats and two Republicans. The variable inter-group counterpart = 1 if a Democrats is matched with a Democrat or a Republican is matched with a Republican. Participants who have a smaller social distance towards out-group than towards in-group members are excluded.*

4.4 Statistical Discrimination

Descriptive Statistics – To reduce a potential, mitigating experimenter demand effect on the baseline level of discrimination, we applied a between-subject design in our real effort decryption task. That is, we elicit a single bet decision per subject while being matched with *either* an in-group or an out-group member. Subjects were willing to pay a higher maximum price for having their earnings codetermined by another participant if being matched with an in-group instead of an out-group member in the real effort decryption task (*see* Figure 4.9), though neither in the presence of communication (MWU test: p = 0.11) nor in its absence the effect was statistically significant (MWU test: p = 0.94). In contrast to the allocation game, potential differences cannot, by design, be explained by taste, because the interaction partner on which the subject's payoff potentially depended does not profit from the co-determination.

Figure 4.9: *Reservation Price by Treatments (95%-confidence intervals of means included)*

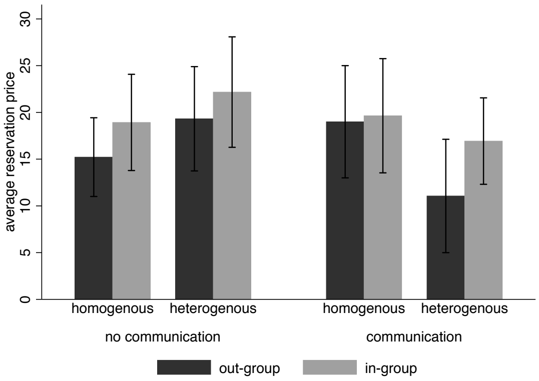
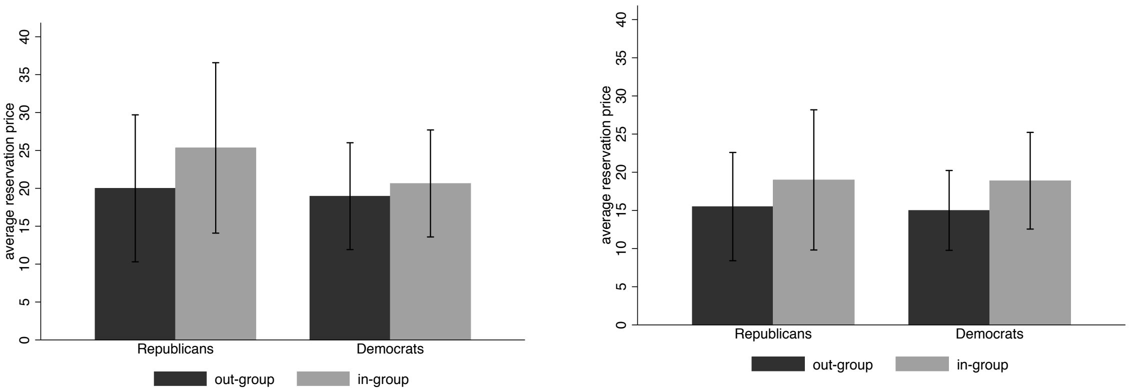


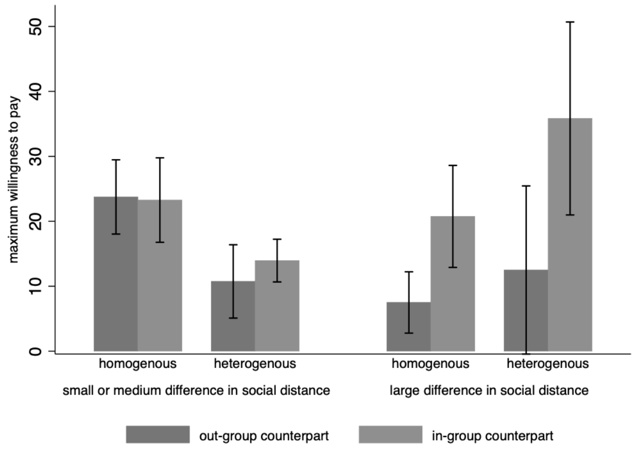
Figure 4.10 and our data reveal no significant difference between the reservation price of Democrats and Republicans being matched with either in-groups (MWU test: p=0.41) or out-groups (id. p=0.49) in the homogeneous groups in the presence and in the absence of communication (MWU test for in-group subsample: p=0.63; MWU test for out-group subsample: p=0.47). Hence, we pool the data in the following analyses.

Figure 4.10: *Reservation prices by political identities (left side: in the presence of communication; right side: in its absence, 95%-confidence intervals of means included)*



Assessing the main treatment effect, Figure 4.9 reveals how the assignment to a heterogenous team had only a minor, and not significant, effect on average reservation price paid for ingroups (MWU-test: $p=0.41$) and outgroup members (id. p=0.49) in the presence, as well as in the absence of communication (MWU-test for in-group subsample: p=0.63; MWU-test for out-group subsample: p=0.47).

Figure 4.11: *Reservation price by difference in social distances in the presence of discrimination (95%-confidence intervals of means included)*



A careful analysis, however, reveals that the effect of contact on the willingness to pay contingent on being matched with an in-group or an out-group member is dissimilar in different subsamples. In particular, we evaluated the behavior on subsamples that we generated on a splitting rule conditional on the variable "*difference in social distance"*. Such a sample split is motivated by the assumption that those who use political identity as a social distinction criterion have stronger priors and thus are less prone to belief updating. Figure 4.11 depicts that there is support for statistical discrimination in the absence of contact for those people whose difference in social distance was above 4. In fact, the subjects in this subsample were willing to pay 35.8 points to have their earnings be co-determined by an in-group in comparison to 12.5 points to an out-group member in the chat conditions (MWU test: *p* = 0.065) when given the opportunity to communicate.

In contrast, subjects whose difference in social distance measure was below 5 revealed no tendency for statistical discrimination in the homogeneous communication treatment (MWU test: *p* = 0.22). These subjects were willing to pay a comparable amount to let their payoffs be co-determined by an in-group member in the homogeneous setting. However, there was a negative effect of being in the heterogeneous communication treatment on the willingness to pay, independent of whether being matched with an in-group or an out-group member (MWU: *p* = 0.004), albeit to a slightly larger extent in the latter case. The overall decline is statistically significant when matched with an out-group member (MWU: *p* = 0.015) but not when matched with an in-group member (MWU: *p* = 0.12)[112]. In conclusion, contact triggered a decline in the willingness to bet on the productivity of an out-group member, though contrary to the initial hypothesis 3 for the subsample of participants with a moderate difference in social distance value.

Result 3: *Contact, in contrast to the initial hypothesis, enhanced statistical discrimination for a significant share of participants. We find no support for hypothesis 3.*

[112] While the MWU test depicts no differences in ranks, a corresponding t-test reveals a significant difference in group averages (*p*= 0.035, two-tailed).

Table 4.3: *Reservation Prices Stated in the Real Effort Task Game*

	Model 1 (without com.)	**Model 2** (with com.)	**Model 3** (combined)	**Model 4** (with com.)	**Model 5** (with com.)	**Model 6** (with com.)
			Dependent variable: ***reservation price***			
In-group Counterpart	3.83	2.26	3.74	-1.63	15.62***	2.24
	(3.41)	(4.14)	(3.43)	(5.37)	(5.58)	(4.13)
Heterogenous Treatment	3.50	-8.14*	3.90	-13.37***	8.00	
	(3.72)	(4.27)	(3.75)	(4.89)	(8.63)	
Communication			4.16			
			(3.61)			
In-group X Heterogenous Treatment	-3.03	4.02	-3.09	5.10	4.32	
	(5.53)	(5.81)	(5.53)	(7.03)	(13.1)	
In-group X Communication			-0.99			
			(5.39)			
Heterogenous Treatment X Communication			-12.1**			
			(5.64)			
Heterogenous Treatment X In-group X Communication			6.28			
			(7.96)			
Number of Out-group Members = 1						0.37
						(4.51)
Number of Out-group Members = 2						-8.15*
						(4.28)
Number of Out-group Members = 3						-18.04***
						(4.72)
Number of Out-group Members = 1 X In-group Counterpart						.36
						(6.92)
Number of Out-group Members = 2 X In-group Counterpart						3.75
						(5.89)
Number of Out-group Members = 3 X In-group Counterpart						6.51
						(8.03)
Age	-0.05	0.17	0.33	0.16	0.25	0.21**
	(0.0)	(0.13)	(0.08)	(0.13)	(0.31)	(0.10)
Female	2.08	-3.85	-0.66	-0.79	-10.7*	-2.54
	(2.73)	(2.94)	(2.01)	(3.43)	(5.64)	(2.48)
Constant	16.09	15.22***	14.29***	18.96***	2.27	13.14***
	(5.02)	(5.21)	(4.00)	(5.38)	(11.80)	(4.45)
Number of obs.	202	167	369	123	44	229
R^2	0.02	0.06	0.02	0.11	0.28	0.08

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01, OLS regressions with robust standard errors in parentheses. The variable Communication = 1 if participant is able to communicate with other members in her groups. The variable Heterogenous Treatment = 1 if participant is assigned a team comprising two Democrats and two Republicans. The variable inter-group counterpart = 1 if a Democrats is matched with a Democrat or a Republican is matched with a Republican. Participants who have a smaller social distance towards out-group than towards in-group members are excluded.*

Regression Analysis – The dependent variable in the regressions reported in Table 4.3 is the maximum willingness to pay in order to receive a piece rate for every real effort task correctly solved by the interaction partner (henceforth: reservation price). We excluded from all models those 9 participants who revealed that the difference in social distance is below 0. Model (1) includes only those subjects who could talk with others. In contrast, Model (2) includes only those who could not. In the former model we do not find a significant difference in the willingness to pay contingent on (i) with whom subjects are grouped (heterogeneous or homogeneous teams) as well as (ii) whether the interaction partner is an in-group or an out-group member. In contrast, in the latter model we find a weakly significant effect of being in the heterogeneous treatment on both in-group as well as out-group members: subjects were on average willing to pay about 8.14 points less in the heterogeneous treatment. However, there was neither a significant effect of the identity of the interaction partner on the reservation price nor did the assignment to a heterogeneous group significantly reduce reservation prices.

Model (3) combines data from Model (1) and Model (2) and corroborates these results by indicating that while all other coefficients are not significant, the interaction effect between communication and being assigned to a heterogeneous team is significant.

Discussion – In a more fine-grained analysis, we inquire whether the willingness to pay differs among participants who experience a comparatively strong identification with the in-group compared to the out-group. Overall, we find evidence for a heterogeneous treatment effect in the descriptive results outlined previously. Model (4) includes the same independent variables as Model (2) but restricts the number to those subjects in the chat treatments, whose elicited difference in social distance has a value of 4 or lower. In Model (5) we include those subjects in the chat treatments whose difference in social distance is above 4. Model (4) reveals a negative effect on being in a heterogeneous group on both the willingness to pay for the opportunity that the own payoffs are co-determined by an in-group and by an out-group member.

In contrast, Model (5) indicates that for those people who feel comparatively much more affection towards in-group members, we find that independent of which treatment those subjects are assigned to a strong tendency to statistically discriminate against out-group members. Being in a heterogeneous treatment had no significant effect on the willingness to pay independent of being matched with an in-group or an out-group member.

In contrast, Model (5) indicates that for those people who feel comparatively much more affection towards in-group members, we find that independent of which treatment those subjects are assigned to a strong tendency to statistically discriminate against out-group members. Being in a heterogenous treatment had no significant effect on the willingness to pay independent on being matched with an in-group or an out-group member.

***Figure 4.12:** Beliefs on the output of the matched counterpart in the real effort task game (95%-confidence intervals of means included)*

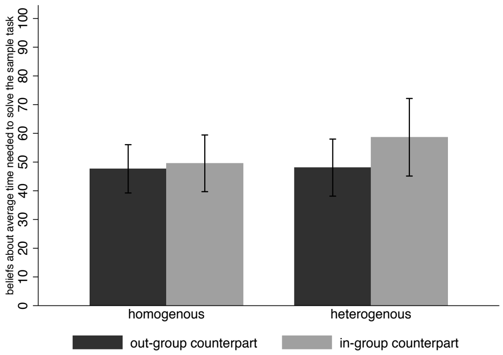
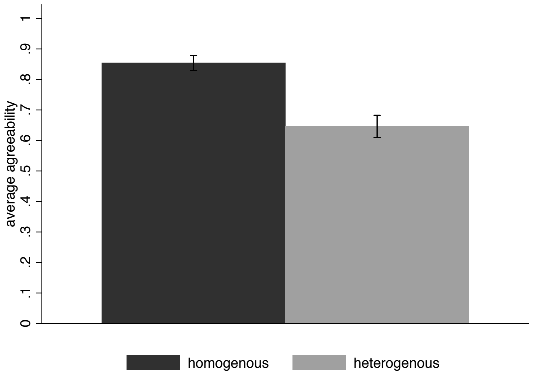


***Figure 4.13:** Agreeability in the team communication task by treatment (95%-confidence intervals of means included)*



The concept of belief-based discrimination is based on the idea that differences in the perception of competence and productivity of in-group and out-group members triggers discriminatory behavior. In our experiment, we elicited first order-beliefs regarding how much time in-group or out-group members needed to solve the task. As shown in Figure 4.12, neither the social identity of the interaction partner nor being matched with an out-group member had a significant effect on the guessed average completion time. The average completion time of one encryption task is, with about 16.44 seconds, more than 50% smaller than the guessed average task to solve the test and in addition does not have a significant effect on stated reservation prices (*see* regressions presented in Table 4.4). Likely, our belief measure was an imprecise proxy of beliefs regarding the competence of in-group as well as out-group members.

However, subjects form beliefs about the general competence of in-group and out-group members based on all available signals about how productive an interaction partner is. In the experiment, another signal is captured by the number of tasks correctly solved by all team members excluding the observed subject. Observing the productivity of homogeneous groups allowed subjects to draw conclusions about either the average in-group or the average out-group member. Although the participants did not receive any feedback on how many tasks the group solved correctly, they likely could infer from the messages sent in the chat how good the group as a whole performed in the group task.[113] A comparison between the mean of the average number of correctly solved tasks by other subjects revealed that heterogeneous groups performed significantly worse than homogeneous groups: in particular, heterogeneous groups answered about 2.5 questions less (*see* Figure 4.10).

We investigate whether cooperation within one group, independent of the final payoff, is higher in homogeneous treatment. The correlation between the number of correctly solved tasks by one individual in comparison to the average answered questions solved by the team members ($r = 0.39$) is higher than the correlation in the heterogeneous treatment ($r = 0.29$). Therefore, subjects in the homogeneous treatments coordinated their answers to a larger extent. To further explore this relationship, we elicited what share of group members selected the most popular answer (either Klee or Kandinsky) in the group task for each question. Subsequently, we calculated the average of this variable regarding the 4 questions asked in the group task and called it variable "agreeability measure". The agreeability measure was on a 1%-level significantly higher in homogeneous than in heterogeneous groups (*see* Figure 4.13), indicating that the willingness to listen to each other and follow the advice of other group members was likely higher in heterogeneous treatments.

Therefore, a negative perception of contact on a factual dimension may explain the unexpected negative effect of being in a heterogeneous treatment on the observed reservation price (*see* Table 4.4, Model 4). In Models (7), (8), and (9) in Table 4.4 we include in comparison to Models (2), (4), and (5) three additional variables: beliefs about the average time an out-group or respectively an in-group member needed to complete the task, the average numbers of correct guesses in the group task excluding the respective subjects and, lastly, the number of correct guesses by the respective subject in the group task. Model (7) comprises all subjects allowed to chat. In Model (8), we restricted this sample further to only those subjects whose elicited difference in social distance was below 5. In Model (9), we restricted the sample to those subjects whose elicited differences in social distance is above 4. In Models (7), (8) and (9) we find no significant effect of the subject's first order beliefs as well as the number of average correct answers on the willingness to pay for the opportunity to have her payoff co-determined by another subject.

[113] Some participants acknowledged in the chat that they have no idea about art. Others communicated that they have some experience with art or provided a comprehensive reasoning why they thought that a painting was drawn by a particular artist.

Table 4.4: *Impact of Communication Features on Reservation Prices*

	Model 7 (with com.)	**Model 8** (with com.)	**Model 9** (with com.)
Dependent variable: ***maximum willingness to pay***			
In-group Counterpart	1.91	-1.1	13.79**
	(4.18)	(5.5)	(5.71)
Heterogenous Treatment	-9.63**	-15.63***	9.04
	(4.85)	(5.1)	(10.69)
In-group Counterpart	3.82	3.93	6.43
X Heterogenous Treatment	(5.90)	(7.21)	(13.46)
Age	0.13	0.13	0.28
	(0.13)	(0.15)	(0.36)
Female	-4.31	-0.55	-12.78*
	(2.99)	(3.61)	(6.35)
FOB about completion time	0.01	-0.003	-0.01
	(0.02)	(0.03)	(0.03)
Number of average correct answers	0.92	.80	4.45
given by other group members	(1.7)	(1.82)	(3.13)
Number of correct answers given by	-2.12	-2.29	-3.73
the respective subject	(1.41)	(1.60)	(2.05)
Agreeability	-6.24	-8.001	-8.05
	(8.90)	(8.72)	(26.5)
Constant	24.1**	30.32***	7.56
	(9.93)	(10.91)	(21.63)
Number of obs.	167	123	44
R²	0.081	0.1246	0.3384

Notes: **=p≤0.10; **=p≤0.05; ***=p≤0.01; OLS regressions with robust standard errors in parentheses. The variable Heterogenous Treatment = 1 if participant is assigned a team comprising two Democrats and two Republicans. The variable inter-group counterpart = 1 if a Democrats is matched with a Democrat or a Republican is matched with a Republican. Participants who have a smaller social distance towards out-group than towards in-group members are excluded.*

5 Conclusion

In this paper we studied which type of economic discrimination is to what extent affected by inter-group contact, as well as the required features of contact to achieve the goal of reducing discrimination. In doing so, we assess to what extent inter-group contact affects preferences and to what extent it affects beliefs.

The results indicate, first, that the causal attenuation effect of contact on taste-based discrimination is significant and pronounced. Contact's impact on taste-based discrimination was stronger than on statistical or anticipated taste-based discrimination and inter-group trust. Overall, affective prejudice and social preferences more precisely predict the reduction of discrimination due to inter-group contact than cognitive components such as beliefs (compare Dovidio et al., 2003; Gaertner et al., 1996).[114] This result stands in

[114] Dovido et al. (1996) concluded in their meta-study that taste-related components significantly reduce racism, but the evidence for cognitive stereotypes are equivocal. In addition, Pettigrew et al. (2000) conclude that presumably affective factors – like empathy (Reich & Purbhoo; 1975) or anxiety (Stephan & Stephan, 1985) – play a critical role explaining the attenuation effect of contact (see also Pettigrew, 1998). In contrast, Fershtman and Gneezy (2001) investigated the topic of discrimination in the segmented Israeli Jewish society utilizing incentivized lab experiments: they conclude that ethnic stereotypes, but not a taste for discrimination, determines discrimination against men of Eastern origin, because statistically significant patterns of discrimination are only observable in their trust game and not in their second dictator game experiment. However, they do neither directly control for taste, nor for beliefs in the analysis of their experiment. Hence, whether irregularities in the sampling process between the two experiments or other factors like the availability of potentially justifiable excuses or differences in context may have alternatively altered the taste for discrimination or the social costs of discrimination in the dictator game is unclear.

contrast to the more traditional economic assumption that changes in behavior can be entirely explained by variations in individuals' beliefs as well as constraints, while preferences are hardly malleable by policy interventions (Stigler & Becker, 1977).

The treatment effect on taste-based discrimination is triggered by actual communication and not triggered by merely being associated with out-groups. Hence, we suggest to practitioners implementing inclusive policies – which may be apt to decrease factual ethnical school segregation or integrate immigrants into the society – to encourage people with distinct social identities to kindly communicate with each other, instead of letting them only attend the same schools or live in the same areas. In fact, Mele (2020) finds that inter-group contact as an inclusive social school policy often fails if schools do not encourage inter-group communication, because in the absence of frequent communication, homophily (the desire to interact with alike people) regularly leads to the creation of segregated, homogeneous groups. Their emergence confounds inclusion efforts. Carell, Sacredote, and West (2013) find in a field experiment about optimal sorting in the United States Air Force Academy that in the absence of communication exogenously imposed inter-group contact might even lead to more segregation.

We concluded cautiously from correlational evidence that the success of communication reduces the likelihood of taste-based discrimination, while the mere quantity of contact had no significant effect. Thus, when designing policies, one should focus on policies that secure a positive perception of the group output, including tailored incentive schemes – which encourage team members to provide optimal effort – as well as the acknowledgement of outstanding team performances – to enhance the visibility of success. Ultimately, a good fit between team members' skill sets and the team task enhances the output and thereby offers answers as to whether inclusive policies are apt to reduce taste-based discrimination. The positive effect of contact is more pronounced if both groups have the same share in the interacting group.

Second, contact had no significant effect on the level of anticipated taste-based discrimination or inter-group trust with respect to opposing political groups. While both social preferences and beliefs about the behavior predict interpersonal trust in general, we find neither a significant difference in beliefs, nor in social preferences towards in-group and out-groups between homogeneous and heterogeneous treatments.

Third, being in a heterogeneous group surprisingly reduces participants' reservation price to let their earnings be co-determined by another participant irrespective of being matched with an in-group or an out-group member – in particular for those subjects who had a small or medium difference in social distance. In contrast, if subjects had a large difference in social distance, we find no statistical support for an effect of inter-group inclusive policies on anticipated taste-based discrimination.

A potential explanation for the latter effect is that participants that strongly identify with in-groups and disassociate with out-groups have stronger, and potentially biased prior beliefs (likely based on prejudice). Consequently, their posterior beliefs are less prone to belief updating. Therefore, additional requirements not considered in our experiment have to be met to mitigate taste-based discrimination if the group identity is considered a pronounced distinguishing feature of a person's personality. Changes in the extensive margin, i.e., increase in the interaction period (MacInnis & Page-Gould, 2015), as well as in the intensive margin, i.e., changes in the intensity of communication, e.g., by allowing for face-to-face interaction, might have decreased the social distance between subjects and out-group members and consequently amplified the mitigation effect of contact on discrimination, though the empirical results concerning this communication feature are inconclusive (see Hasler & Amichai-Hamburger, 2013 for a literature review).

The former effect can partly be explained by a decrease of output quality detected in heterogeneous in comparison to homogeneous teams. Thus, we contribute to the research on the influence of negative contact on discrimination. In particular, we cautiously conclude from our findings that the quality of contact must be assessed on different dimensions: previous research on the impact of negative contact on inter-group behavior considered negative contact as being belittled, intimidated, or insulted by an out-group member (Aberson, 2015). Although we can abstract from obvious forms of negative interaction experiences, we conclude that while contact can be experienced as positive on the personal level and thus reduce taste-based discrimination, it can be perceived as negative on the factual level and hence negatively impact the perception of competence.

To conclude, our finding that contact foremost attenuates taste-based discrimination, while it might even enhance statistical discrimination under some unfortunate circumstance, is apt to explain some of the contradictions in the experimental literature on the contact hypothesis (see Paluck et al., 2018 for a meta-analysis). Thereby, we contribute to solve contradictions in existing experimental studies inquiring into the contact hypothesis. For instance, Scacco and Warren (2018) find that contact leads to increased inter-group generosity. These are indications for a mitigation of taste-based discrimination (Becker, 1957) caused by inter-group interactions. Finseraas et al. (2016) find that contact is apt to reduce discrimination on the basis of sex in hiring decisions, but presumably not by reducing statistical discrimination (Arrow, 1973; Phelps, 1972). Based on happenstance data, Zhang (2017) argues that there is no correlation between stereotypes towards out-groups and inter-group contact between coaches and players in the NBA. In line, we find first indications that if contact is perceived as negative on a factual level it may even enhance statistical discrimination.

We conclude from combining insights from the literature and our experimental results that policies often reduce or increase discrimination because they target and change different elements. When they provide the opportunity for inter-group communication between groups of an equal status, they may reduce people's animus against out-groups. Thereby, they avoid in-group favoritism arising from affective prejudice. At the same time, an inclusive social policy can be inapt to reduce people's prejudices if the communication is not perceived as successful (Finseraas et al., 2016), such as when communication barriers hinder cooperation (Condra & Linardi, 2019). Thus, it might not prevent individuals from disfavoring out-groups when discrimination arises from cognitive-based stereotypes. While discrimination based on taste and social preferences can well be reduced by inclusive social policies, discrimination based on beliefs requires more elaborate inclusive policies or other types of policies to be effectively reduced.

Appendix

A1 Instructions

Enclosed you find the instructions of our experiment. Note that the headlines provided are dedicated to guide you through the different stages of the experiment. Participants received instructions with headlines which don't contain loaded language.

Survey Before the Actual Experiment

Welcome to this Survey

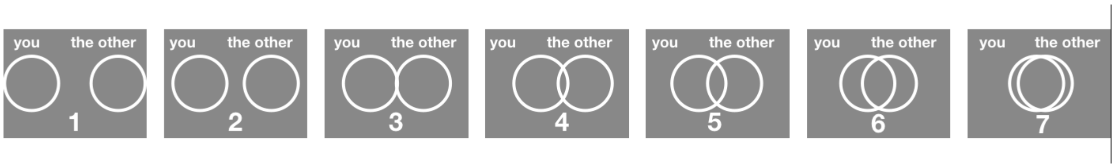
This survey takes **about 1-2 minutes** and is anonymous. The purpose of this survey is to figure out whether you qualify for our economic experiment. Participating in the experiment, you are able to earn more money and gain a significant monetary bonus depending on the decision you and others make during the experiment.

If you qualify for the experiment you will be directly forwarded to the experiment. You will receive a fixed payment of $0.50 for participating in the experiment and on average a significant bonus of $3.00. If you do not qualify, you will receive $0.05 for taking this survey.

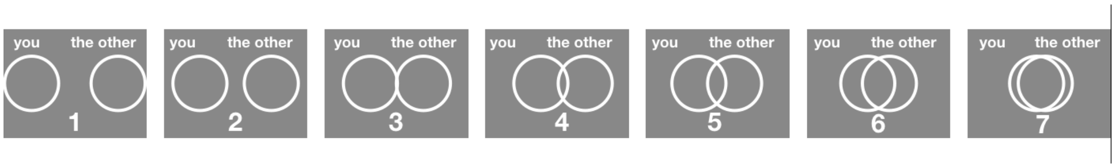
Which party do you most likely support in the next elections?

How much interested in U.S. politics are you? Please state your answer on a scale from 0 to 6 where "0" means "not interested at all" and "6" means "very interested".

Please pick the circle that best describes your relationship towards a voter of the Democratic party.



Please pick the circle that best describes your relationship towards a voter of the Republican party?



[If the participant did not qualify for the experiment, he or she received the following message]

Thank you for participating in this survey.
Unfortunately, you do not fulfill the requirements for participating in our economic experiment.
Please copy and paste the following code into the MTurk HIT and click on submit.

[If the participant qualified for the experiment, he or she received the following message]

Thank you for participating in this survey.
Congratulations, you fulfill the requirements for participating in our economic experiment. The experiment will take about 10 minutes.

In order to continue with the experiment please click on "next".

General Instructions

Dear participant,

you are taking part in an experiment of the University of Cologne and the Max Planck Institute in Munich. Experiments such as today's help us to collect reliable data about human decision making that is needed for scientific publications.

This experiment is anonymous. The data is collected in a way that we cannot link individual responses to participants' identities. Moreover, participants will receive no information about the identity of other participants.

Information Concerning the Course of the Experiment and the Bonus Payments

Please read the following instructions carefully. A clear understanding of the instructions will help you make better decisions. All statements made in these instructions are true. In particular, **all actions will be implemented exactly in the way they are described.**

You will receive a **fixed amount of $0.50** for completing this experiment. You will be **paid additional money depending on your stated decisions.** We will explain you in detail how you can earn a significant bonus. **On average** participants earn in total about **$3.00** and the experiment takes about 10 minutes.

In addition, **your decisions have real consequences on other participants**. The experiment consists of **4 parts**. Each part will be introduced on a screen with the header "instructions". These instructions will explain in detail what the respective part of the experiment is about.
You will receive payoffs from part 1 and from one of the other three parts: the computer will randomly decide whether you receive in addition to the payoffs from part 1 payoffs from part 2, 3, or 4.

Click on "next" if you have read the instructions and consent to take part in this experiment.
In the first part of the experiment 4 people will participate. You may have to wait a moment until 3 other participants are also ready to take part in the experiment.

Painter Guessing Game

You and three other participants will form a group.
The group consists of [# dependent on the treatment] supporters of the Republican party and [# dependent on the treatment] supporters of the Democratic party.

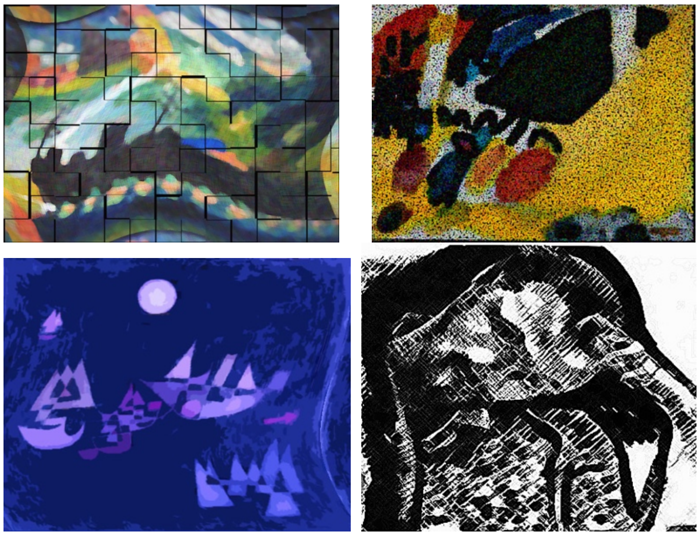
We will show you two paintings from **Paul Klee** and two paintings from **Wassily Kandinsky.** We will tell you which of these painters painted each painting.

[Example paintings; Kandinksy right, Klee left]

Below you will see 4 other paintings. The paintings have been modified. Your task is to select which painter (either Paul Klee or Wassily Kandinsky) painted which original painting. You will have 4 minutes to complete this task.

You and the other members of the group will see the same paintings.

[4 paintings, similar to the ones below]



Task

Your earnings will be calculated as follows: for each painting that you or another member of the group assign to the correct painter, you receive 10 points. The more members of the group answer correctly, the more you earn.

You will only have the opportunity to receive money from the entire experiment if you make at least one decision (select either "Kandinsky" or "Klee") in the first part of the experiment.

Other-other Allocation Game

In this part of the experiment you receive **200 points**.

Your task is to divide 200 points between two other participants.

One of the participants is a supporter of the **Democratic Party**, the other participants is a supporter of the **Republican Party**. You have **not** interacted with one of the two participants before. The choice you make determines the payoffs of the other two participants.

How many of the 200 points do you wish to give to the supporter of the Republicans?
How many of the 200 points do you wish to give to the supporter of the Democrats?

Trust Game

In this part of the experiment, you will interact with another participant. The other interaction partner supports the **Democratic Party** **[alternatively Republican Party]**. You have not interacted with this participant before.

Either you or the other participant will be assigned the role A. The other one will then be assigned the role B.

Participant A receives **100 points**. He or she will have the possibility to send some of these points to participant B. Participant A keeps all points that he/she does not send to participant B. The points sent to participant B will be multiplied by 3.

Participant B can send back points to participant A. Participant B keeps all the points he or she does not send back. Participant A receives in addition to the points he or she initially kept, the points he/she receives from participant B. The points sent back by participant B to participant A will not be multiplied by three.

Example:

If A sends **20 points** to B, B will receive **60 points**.
If A sends **90 points** to B, B will receive **270 points**.

If B receives **60 points**, B can decide to send back to A any amount between 0 and 60.
If B receives **270 points**, B can decide to send back to A any amount between 0 and 270.

B's payoff is equal to the number of points he receives minus the points he sends back to A.
A's payoff is equal to 100 minus the number of points sent to B plus the number of points sent back from B.

Summary If A sends an amount x to B and B sends back an amount y to A, then the earnings of each participants are calculated as following:

A earns $100 = x + y$
B earns $3x = y$

You must make a decision for two different possible scenarios.

Scenario 1: you are assigned the role A and you interact with a participant in the role of B that is a supporter of the **Democrats** **[Republicans]**.

Scenario 2: you are assigned the role B and you interact with a participant in the role of A that is a supporter of the **Democrats** **[Republicans]**.

The computer will then randomly select one of the two scenarios. Your corresponding decision in that situation will then be implemented and will determine your earnings in this part.

Before you can start with the task, you have to answer the following comprehension question correctly:

If participant A sends **[random number]** points and
the participant B returns **[random number]** points,

how much points would participant A receive?

Your Decision Part 1

If you are assigned the role **A**, you will receive **100 points**. You will have the possibility to send some of these points to participant B, who supports the Democrats **[Republicans].** You keep all points that you do not send to participant B. The points sent to participant A will be multiplied by three.

Participant B can then choose to send (some of the) points back to participant A.

Your Decision Part 2

If you are assigned the role B, how much points are you willing to send back to your matched partner who supports the Democrats:

Your beliefs regarding the Decision of the Average Participant B

Please answer the following question: On average how much does a supporter of the Republicans [Democrats] in role B return to a supporter of the Republicans **[Democrats]** in role A? Please state your answer in percentages of the points player A received from B."

Real Effort Task

In this part of the experiment you must correctly solve as many tasks as you can.

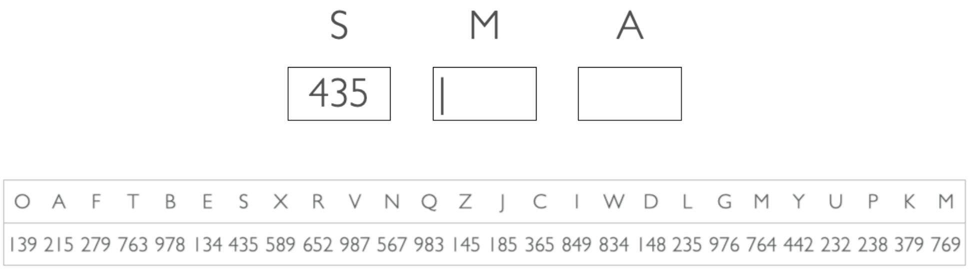
In particular, you have to decrypt codes. You will see three letters. You must assign the correct 3-digit number to each of these letters. The correct 3-digit numbers are located in the table below the letters.

Example: You receive the letters …

After solving the first task, the computer will generate three new letters for you to decrypt. Please note that for each new letter, the numbers and the order of the numbers in the table will be randomly altered. You will receive one training task. Try to solve the task as fast as you can.

For solving this single task, you earn 50 points.

Test Task



S	M	A
435		

O	A	F	T	B	E	S	X	R	V	N	Q	Z	J	C	I	W	D	L	G	M	Y	U	P	K	M
139	215	279	763	978	134	435	589	652	987	567	983	145	185	365	849	834	148	235	976	764	442	232	238	379	769

Payoffs

In this part of the experiment, you must solve as many tasks as you can in 120 seconds. For each task that you solve correctly, you earn 5 points.

Moreover, you can choose to have your earnings co-determined by how many tasks another randomly selected participant correctly solves in 120 seconds.

For each code that this other participant correctly solves, you earn another 5 points. You have never interacted with this assigned participant in the previous parts of the experiment before. If you want to make use of this option, you will have to pay a price.

We ask you to choose between option A and option B for 10 different possible prices.

The computer will randomly select one of these 10 possible prices, and the choice you made contingent on that particular price, will be implemented.

Option A	Option B
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **5 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **10 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **15 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **20 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **25 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **30 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **35 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **40 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **45 points.**	You receive 5 points for every task **you** solve correctly.
You receive 5 points for every task you or a supporter of the Democrats has solved correctly. For the opportunity that your payoff depends on the answers of another participant you pay a price of **50 points.**	You receive 5 points for every task **you** solve correctly

Ex-Post Experimental Questionnaire

Finally, we ask you to answer questions concerning your assessment of four hypothetical situations as well as a few demographic questions. After you have answered all questions, you will receive your validation code.

1. Imagine the following situation: You won 1,000 dollars in a lottery. Considering your current situation, how much would you donate to charity? (Values between 0 and 1000 are allowed)"

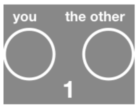
2. How do you assess your willingness to share with others without expecting anything in return when it comes to charity? Please use a scale from 0 to 10, where 0 means you are "completely unwilling to share" and a 10 means you are "very willing to share". You can also use the values in-between to indicate where you fall on the scale.

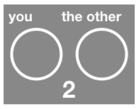
3. Imagine the following situation: You are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 dollars in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 dollars, the most expensive one 30 dollars. You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you give?"

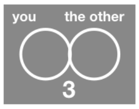
4. How do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly?

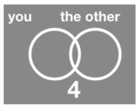
   'Please use a scale from 0 to 10, where 0 means you are "not willing at all to incur costs to punish unfair behavior" and a 10 means you are "very willing to incur costs to punish unfair behavior". You can also use the values in-between to indicate where you fall on the scale'

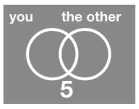
5. Please indicate which of the pairs of circles describes your relationship towards a supporter of the Democratic Party best?

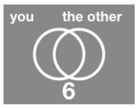


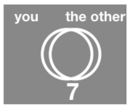




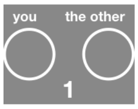


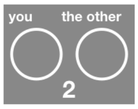


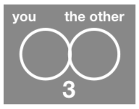


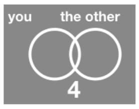


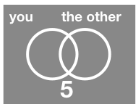
6. Please indicate which of the pairs of circles describes your relationship towards a supporter of the Republican Party best? Which pair of the circles describes it best?

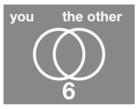


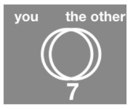












7. How old are you?

8. What is your highest education degree (in case you are still studying, pick the highest one you have already achieved?

9. Please indicate the filed that best described your subject of study Have you ever participated in a similar study

10. Is there something you want to tell about this study?

A2 Taste-Based Discrimination Model

In this appendix, we introduce a taste-based discrimination model that is based on a non-linear version of the Fehr-Schmidt model (1999). While linear models produce corner solutions, non-linear models allow for intermediate values of discriminating behavior which are regularly observed in the field. The aim of the model is to show that differences in social preferences towards in-groups and out-group members explain discrimination patterns detected in the other-other allocation game (Chen & Li, 2009). Formally, consider a set of n players indexed by *i.* The utility function of player *i* is given by

$$U(x_i, x_j) = x_i - \frac{1}{n-1}\sum\nolimits_{j\neq i}^{n} I_{in} \cdot \alpha_{in} \cdot \left[\max(0;\ x_j - x_i)\right]^2 + I_{out} \cdot \alpha_{out} \cdot \left[\max(0;\ x_j - x_i)\right]^2$$

$$+ I_{in} \cdot \beta_{in} \left[\max(0;\ x_i - x_j)\right]^2 + I_{out} \cdot \beta_{out} \left[\max(0;\ x_i - x_j)\right]^2$$

where

x_i = monetary payoff of an individual *i*

x_{jy} = monetary payoff of an individual *j*

α_{in} = disadvantageous inequity towards an in-group member

α_{out} = disadvantageous inequity towards out-group member (with $\alpha_{in} \leq \alpha_{out}$)

β_{in} = advantageous inequity towards an in-group member

β_{out} = advantageous inequity towards an out-group member (with $1 \geq \beta_{in} > \beta_{out}$)

I_{in} = indicator functions, taking the value 1 if counter party is from the in-group

I_{out}= indicator functions, taking the value 1 if counter party is from the out-group.

The assumptions that $\alpha_{in} \leq \alpha_{out}\ and\ \beta_{out} \geq \beta_{in}$ reflect that agents have stronger positive social preferences towards in-groups and stronger negative social preferences towards out-groups. We assume that $0 \leq \alpha_{in,out}$ and $0 \leq \beta_{in,out} < 1$. The latter assumption imply that one does not have higher regards for another person than for oneself.

Our model comprises that, *ceteris paribus*, *i* gains greater disutility from advantageous or disadvantageous inequity, the larger the parameters α and β are. Therefore, if people have a taste for discrimination that manifests itself in differences in social preferences, they suffer less from advantageous ($\alpha_{in} < \alpha_{out}$) and more from disadvantageous inequity ($\beta_{in} > \beta_{out}$) towards in-group members. Their taste is stronger, the larger the difference between the two parameters is. In the following, we prove that if $\beta_{in} > \beta_{out}$ or respectively $\alpha_{in} < \alpha_{out}$, allocators allocate more to in-group than to out-group members in the utilized other-other allocation game irrespective of x_i.

Application of the Taste-based Discrimination Model to an Other-other Allocation Game

Next, we demonstrate that an other-other allocation game –in which an allocator has to divide, without the loss of generality, a standardized amount of T=1 between an in-group and an out-group member– is well suited to detect patterns of taste-based discrimination emerging from differences in social preferences towards in-group and out-group members. The utility maximization problem of i is, in the other-other allocation game in which t_{in} is the distribution to an in-group and $t_{out} = 1 - t_{in}$ to an out-group member, defined as follows:

$$\max U(t_{in}) = x_i - \frac{1}{2} \cdot (\alpha_{in} \cdot [\min(0;\ t_{in} - x_i)]^2 + \alpha_{out} \cdot [\min(0;\ 1 - t_{in} - x_i)]^2$$

$$+ \beta_{in}\,[\min(0;\ x_i - t_{in})]^2 + \beta_{out}\left[\min\left(0;\ x_i - (\,1 - t_{in})\right)\right]^2)$$

$$s.t\ 0 \leq\ t_{in} \leq 1 = T$$

Trivially, if $\alpha_{in}, \alpha_{out}, \beta_{in}, \beta_{out} = 0$ the distributor is indifferent between all possible allocation in the other-other allocation game.

Assume otherwise, that $\alpha_{in} + \alpha_{out} > 0$ and $\beta_{in} + \beta_{out} > 0$. The utility function is not differentiable at all points where $t_{in} - x_i = 0$ and where $1 - t_{in} - x_i = 0$. Thus, in order to find the maximizing argument of the utility function, we have discussed the solution of two different cases and argue which of the three solutions maximizes the overall utility function. The distributor either experiences disutility from disadvantageous inequity or disutility from advantageous inequity towards both the in-group and the out-group member. We furthermore show that if $\alpha_{in}, \alpha_{out}, \beta_{in}, \beta_{out} > 0$ the third possible condition – i experiences disutility from disadvantageous inequity towards the in-group and advantageous inequity towards the outgroup member– does not have to be considered.

Proposition 1: *If the distributor experiences in the optimum disadvantageous inequality towards the in-group and the out-group member and 0 <* $\alpha_{in} < \alpha_{out}$*, the amount allocated to the in-group member* (t_{in}) *increases in* α_{out} *and decreases in* α_{in}.

First, we consider all cases in which the utility maximizing value of $t_{in}{}^*$ is defined such that $t_{in}{}^* - x_i > 0$ and $1 - t_{in}{}^* - x_i > 0$, i.e., the distributor experiences in the optimum disadvantageous inequality towards the in-group and the out-group member. This implies that because $0 \leq t_{in} \leq 1 = T$, the amount $x_i <$ allocated to i is such that $x_i < 0.5$. Under the given assumptions, the utility function reduces to:

$$\max U(t) = x_i - \; 0.5 \; \cdot [\,\alpha_{in} \cdot (t_{in} - x_i)^2 + \; \alpha_{out} \cdot (1 - t_{in} - x_i)^2]$$
$$s.t\; 0 \leq t_i \leq 1 = T \qquad (1)$$

The first order condition of the unconstraint maximization problem with respect to the amount allocated to the in-group member t_{in} is given by

$$0 = \; -\alpha_{in} \cdot (t_{in} - x_i) + \alpha_{out} \cdot (1 - t_{in} - x_i)]\,. \qquad (2)$$

The above defined utility function is strictly concave. Hence, it holds that solving the first order condition for $t_{in}{}^*$ yields the optimal amount allocated to the in-group member

$$t_{in}{}^* = \frac{a_{out}}{a_{in} + a_{out}} + \frac{(a_{in} - a_{out})}{a_{in} + a_{out}} x_i \qquad (3)$$

Which is the maximizing argument of the unconstrained problem. It directly follows that if $a_{out} = a_{in}$, the maximizing argument will be t**=0.5. Taking the constraints that

$$t_{in} \geq x_i \text{ and } 1 - t_{in} \geq x_i \qquad (4)$$

into account, it holds that

$$t_{in}{}^* = \frac{a_{out}}{a_{in} + a_{out}} + \frac{(a_{in} - a_{out})}{a_{in} + a_{out}} x_i \text{ if } x_i \leq \frac{a_{out}}{a_{in} + a_{out}} + \frac{(a_{in} - a_{out})}{a_{in} + a_{out}} x_i \; \leq 1 \qquad (5)$$

Notably, this inequality is true for all $x_i \leq 0.5$: solving $x_i \leq \frac{a_{out}}{a_{in} + a_{out}} + \frac{(a_{in} - a_{out})}{a_{in} + a_{out}} x_i$ for x_i yields $x_i \leq 0.5$ and solving $t_{in}{}^* \; = 1 - \frac{a_{out}}{a_{in} + a_{out}} + \frac{(a_{in} - a_{out})}{a_{in} + a_{out}} x_i \; \geq x_i$ for x_i again yields $x_i \leq 0.5$. Hence, in all cases in which $x_i \leq 0.5$, $t_{in}{}^*$ constitutes the maximizing argument of i's utility function.

Furthermore, it holds under the above described assumptions that $t_i \geq 1 - t_i$ because $a_{in} < a_{out}$. That is, distributors allocate more to in-group members if they suffer less from in advantageous inequality towards in-group members.

We now show that $t_{in}{}^*$ increases in a_{out}. Therefore, we derive the solution of $t_{in}{}^*$ with respect to a_{out} and show that the expression is larger than 0. In fact,

$$\frac{\partial t_{in}{}^*}{\partial a_{out}} = \frac{a_{in}\,(1-2x_i)}{a_{in}+a_{out}} > 0, \text{ since } x_i < 0.5 \text{ and therefore } 1 - 2x_i > 0 \tag{6}$$

In the final step we show that the optimal amount $t_{in}{}^*$ allocated to an in-group member decreases in a_{in}. Therefore, we derive $t_{in}{}^*$ with respect to a_{in} and show that the expression is smaller than 0. In fact,

$$\frac{\partial t_{in}{}^*}{\partial a_{in}} = \frac{a_{out}\ (2x_i-1)}{a_{in}+a_{out}} > 0, \text{ since } x_i < 0.5 \text{ and therefore } 2x_i - 1 < 0. \ \blacksquare$$

Proposition 2: *If the distributor experiences advantageous inequality towards both players, i.e.,* $\beta_{in} > \beta_{out}$ *the amount allocated to the in-group member* (t_{in}) *decreases in* β_{out}.

We consider all cases where $t_{in} - x_i < 0$ and $1 - t_{in} - x_i < 0$. Trivially, this implies that if $0 \leq t_{in} \leq 1 = T$, $x_i > 0.5$. Under the given conditions the utility function reduces to:

$$\max U(t) = x_i - 0.5 \cdot \left[\beta_{in}(x_i - t_{in})^2 + \beta_{out}\big(x_i - (1 - t_{in})\big)^2\right] \tag{7}$$

$$s.t\ 0 \leq t_i \leq 1 = T$$

The first order condition with respect to t_{in} is given by

$$0 = \beta_{in}(x_i - t_{in}) - \beta_{out}\,(x_i - 1 + t_{in}). \tag{8}$$

Rearranging the function yields that

$$t_{in}{}^{**} = \frac{(\beta_{in}-\beta_{out})\,x_i}{(\beta_{in}+\beta_{out})} + \frac{\beta_{out}}{(\beta_{in}+\beta_{out})}. \tag{9}$$

Again, if $\beta_{in} = \beta_{out}$, the maximizing argument is $t_{in}{}^{**}$=0.5. Taking the constraints that

$$t_{in} - x_i < 0 \text{ and } 1 - t_{in} - x_i < 0 \tag{10}$$

into consideration it holds that

$$t_{in}{}^{**} = \frac{(\beta_{in}-\beta_{out})\,x_i}{(\beta_{in}+\beta_{out})} + \frac{\beta_{out}}{(\beta_{in}+\beta_{out})} \text{ if } 1\text{-}x_{i,} < \frac{(\beta_{in}-\beta_{out})\,x_i}{(\beta_{in}+\beta_{out})} + \frac{\beta_{out}}{(\beta_{in}+\beta_{out})} < x_{i,} \tag{11}$$

Notably, this inequality is true for all $x_i \geq 0.5$: solving ${t_{in}}^{**} = \frac{(\beta_{in}-\beta_{out})\, x_i}{(\beta_{in}+\beta_{out})} + \frac{\beta_{out}}{(\beta_{in}+\beta_{out})} < x_{i,}$ for x_i leads to $x_i \geq 0.5$ and solving 1-$x_{i,} < \frac{(\beta_{in}-\beta_{out})\, x_i}{(\beta_{in}+\beta_{out})} + \frac{\beta_{out}}{(\beta_{in}+\beta_{out})} = {t_{in}}^{**}$ for x_i leads to $x_i \geq 0.5$. Hence, in all cases in which $x_i \geq 0.5$, ${t_{in}}^{**}$ constitutes the maximizing argument of i's utility function. Recall that in all cases in which $x_i \leq 0.5$, ${t_{in}}^{*}$ constitutes the maximizing argument of i's utility function. Hence, the maximizing argument of the overall utility function is given by

$$t_{in}^{opt} = \begin{cases} \frac{\beta_{out}}{(\beta_{in}+\beta_{out})} + \frac{(\beta_{in}-\beta_{out})}{(\beta_{in}+\beta_{out})} x_i\,, & x_i \geq 0.5 \\ \frac{a_{out}}{a_{in}+\, a_{out}} + \frac{(a_{in}-a_{out})}{a_{in}+\, a_{out}}\, x_i, & x_i \leq 0.5 \end{cases} \qquad (12)$$

Notably, if $x_i \geq 0.5$, the fact that $t_i > 1 - t_i$ follows from $\beta_{in} \leq \beta_{out.}$ In the next step we prove that t_{in} decreases in β_{out}. Therefore, we derive ${t_{in}}^{*}$ with respect to β_{out} and show that the expression is larger than 0. In fact,

$$\frac{\partial {t_{in}}^{**}}{\partial \beta_{out}} = \frac{\beta_{in}\ (1-\, 2x_i)}{(\beta_{in}+\, \beta_{out})^2} > 0, \text{ since } x_i > 0.5 \text{ and therefor } 1 - \ 2x_i < 0 \qquad (13)$$

In the final step we show that increases in β_{in}. Therefore, we derive ${t_{in}}^{*}$ with respect to β_{in} and show that the expression is smaller than 0. In fact,

$$\frac{\partial {t_{in}}^{**}}{\partial \beta_{in}} = \frac{\beta_{out}\ (2x_i-1)}{(\beta_{in}+\, \beta_{out})^2} > 0, \text{ since } x_i > 0.5 \text{ and therefor } \ 2x_i - 1 > 0. \ \blacksquare \qquad (14)$$

A3 Additional Statistical Analyses

Figure A4.1: *Subjects of study by political orientation (Democrats: right; Republicans: left)*

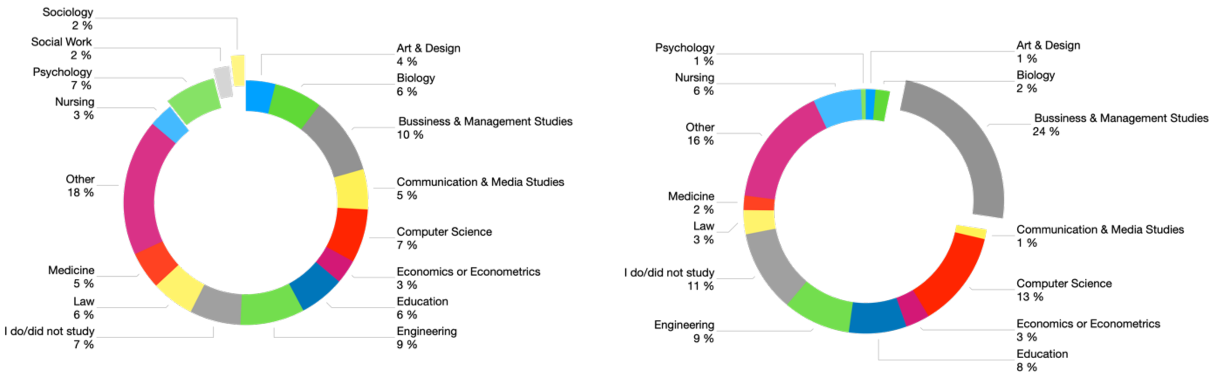


Table A4.1*: Comparison of Social Preferences by Group Identity*

	Average Value	Average Value Democrats	Average Value Republicans	MWU-Test (Prob > |z|)
Altruism*: Imagine the following situation: You won 1,000 dollars in a lottery. your current situation, how much would you donate to charity?*	$ 126.97	$ 120.45	$ 138.61	0.8834
Altruism *How do you assess your willingness to share with others without expecting anything in return when it comes to charity? (scale from 0 – 10)*	6.6	6.7	6.6	0.8232
Positive Reciprocity: *Imagine the following situation: You are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 dollars in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 dollars, the most expensive one 30 dollars. You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you want to give?*	4.0	4.0	4.1	0.5375
Negative Reciprocity: *How do you see yourself: Are you a person who is generally willing to punish unfair behavior even if this is costly? (scale from 0 – 10)*	5.4	5.2	5.7	0.0564

Table A4.2*: Comparison of Different Variables Associated with Subjects' Social Identity*

	Average Values	Average Value Democrats	Average Value Republicans	MWU-Test (Prob > |z|)
Political Interest	4.91	5.0	4.73	0.0099
Social Distance (In-group Member)	5.67	5.71	5.59	0.3173
Social Distance (Out-group Member)	2.66	2.50	2.94	0.0106

R

REFERENCES

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for Truth-Telling. *Econometrica*, *87*(4), 1115–1153.

Aberson, C. L. (2015). Positive intergroup contact , negative intergroup contact, and threat as predictors of cognitive and affective dimensions of prejudice. *Group Processes & Intergroup Relations*, *18*(6), 743–760.

Abrams, D., & Hogg, M. A. (2004). Metatheory: Lessons from social identity research. *Personality and Social Psychology Review*, *8*(2), 98–106.

Aigner, D. J., & Cain, G. G. (1977). Statistical Theories of Discrimination in Labor Markets. *ILR Review*, *30*(2), 175–187.

Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, *121*(4), 543–568.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity. *Quarterly Journal of Economics*, *65*(3), 715–753.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, *19*(1), 9–32.

Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, *85*(2), 207–234.

Allport, G. W. (1954). The Nature of Prejudice. *Reading, MA: Addison-Wesley*.

Allport, G. W. (1958). Personality: normal and abnormal'. *The Sociological Review*, *6*(2), 167–181.

Altmann, S., Dohmen, T., & Wibral, M. (2008). Do the reciprocal trust less? *Economics Letters*, *99*(3), 454–457.

Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, *116*(1), 313–350.

Ambrus, B. A., Mobius, M., & Szeidl, A. (2014). Consumption Risk-Sharing in Social Networks. *American Economic Review*, *104*(1), 149–182.

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the internet: The effect of $1 stakes. *PLoS ONE*, *7*(2), 1–4.

Andreoni, J. (1988). Why free ride?. Strategies and learning in public goods experiments. *Journal of Public Economics*, *37*(3), 291–304.

Andreoni, J. (1990). Impure Altruism and Donations To Public Goods: a Theory of Warm-Glow Giving. *The Economic Journal*, *100*(401), 464–477.

Andreoni, J. (2007). Giving gifts to groups: How altruism depends on the number of recipients. *Journal of Public Economics*, *91*(9), 1731–1749.

Arrow, K. (1973). The theory of discrimination. *Discrimination in Labor Markets*, *3*(10), 3–33.

Arrow, K. J. (1994). Methodological Individualism and Social Knowledge. *American Economic Review*, *84*(2), 1–9.

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208.

Au, W. T., & Kwong, J. Y. Y. (2004). *Measurements and Effects of Social-Value Orientation in Social Dilemmas: A Review.*

Azmat, G. (2019). Gender diversity in teams. *Mimeo*, 1–10.

Bacharach, M., Guerra, G., & Zizzo, D. J. (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision*, *63*(4), 349–388.

Balafoutas, L. (2011). Public beliefs and corruption in a repeated psychological game. *Journal of Economic Behavior and Organization*, *78*(1–2), 51–59.

Balafoutas, L., & Fornwagner, H. (2017). The limits of guilt. *Journal of the Economic Science Association*, *3*(2), 137–148.

Balafoutas, L., Kerschbamer, R., & Sutter, M. (2017). Second-Degree Moral Hazard in a Real-World Credence Goods Market. *The Economic Journal*, *599*, 1–18.

Barberis, N. C. (2013). Thirty Years of Prospect Theory in Economics: A Review and Assessment. *The Journal of Economic Perspectives*, *27*(1), 173–195.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, *11*(2), 122–133.

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R. M., Harwood, J., Rubin, M., & Sibley, C. G. (2012). The Contact Caveat: Negative Contact Predicts Increased Prejudice More Than Positive Contact Predicts Reduced Prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629–1643.

Bastian, B., Lusher, D., & Ata, A. (2012). Contact, evaluation and social distance: Differentiating majority and minority effects. *International Journal of Intercultural Relations*, *36*(1), 100–107.

Batalha, L., Akrami, N., & Ekehammar, B. (2007). Outgroup favoritism: The role of power perception, gender, and conservatism. *Current Research in Social Psychology*, *13*(4), 38–49.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170–176.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin*, *115*(2), 243–267.

Beck, A., Kerschbamer, R., Qiu, J., & Sutter, M. (2013). Shaping beliefs in experimental markets for expert services: Guilt aversion and the impact of promises and money-burning options. *Games and Economic Behavior*, *81*(1), 145–164.

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–232.

Becker, G. S. (1957). *The Economics of Discrimination* (1st ed.). The University of Chicago Press.

Bell, D. E. (1995). Risk , Return , and Utility. *Management Science*, *41*(1), 23–30.

Bellemare, C., Sebald, A., & Strobel, M. (2011). Measruning the willingness to pay to avoid guilt. Estimation using Equilibrium and stated belief models. *Journal of Applied Econometrics*, *26*, 437–453.

Bellemare, C., Sebald, A., & Suetens, S. (2017). A note on testing guilt aversion. *Games and Economic Behavior*, *102*, 233–239.

Ben-Ner, A., McCall, B. P., Stephane, M., & Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. *Journal of Economic Behavior and Organization*, *72*(1), 153–170.

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2012). What do you think would make you happier? What do you think you would choose? *American Economic Review*, *102*(5), 2083–2110.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. In *Games and Economic Behavior* (Vol. 10, Issue 1, pp. 122–142).

Bergmann, B. R., & Darity, W. (1981). Social relations, productivity, and employer discrimination. *Monthly Labor Review*, *104*(4), 47–49.

Bergvall-Kåreborn, B., & Howcroft, D. (2014). Amazon Mechanical Turk and the commodification of labour. *New Technology, Work and Employment*, *29*(3), 213–223.

Bhavnani, R., Donnay, K., Miodownik, D., Mor, M., & Helbing, D. (2014). Group segregation and urban violence. *American Journal of Political Science*, *58*(1), 226–245.

Bicskei, M., Lankau, M., & Bizer, K. (2014). How Peer-Punishment Affects Cooperativeness in Homogeneous and Heterogeneous Groups. *Nimeo*.

Binzel, C., & Fehr, D. (2013). Social distance and trust: Experimental evidence from a slum in Cairo. *Journal of Development Economics*, *103*(1), 99–106.

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338.

Bogardus, E. S. (1927). *Immigration and race attitudes.* Heath.

Bohnet, I., & Frey, B. S. (1999). Social Distance and Other-Regarding Behavior in Dictator Games: Comment. *American Economic Review*, *86*(3), 336–339.

Bohnet, I., & Zeckhauser, R. (2004). Risk and betrayal. *Journal of Economic Behavior & Organization*, *55*(4), 467–484.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(l), 892–896.

Bouckaert, J., & Dhaene, G. (2004). Inter-ethnic trust and reciprocity: Results of an experiment with small businessmen. *European Journal of Political Economy*, *20*(4), 869–886.

Bracht, J., & Regner, T. (2013). Moral emotions and partnership. *Journal of Economic Psychology*, *39*, 313–326.

Brañas-Garza, P., Cobo-Reyes, R., Espinosa, M. P., Jiménez, N., Kovářík, J., & Ponti, G. (2010). Altruism and social integration. *Games and Economic Behavior*, *69*(2), 249–257.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220–224.

Brülhart, M., & Usunier, J. C. (2012). Does the trust game measure trust? *Economics Letters*, *115*(1), 20–23.

Buchan, N. R., Croson, R. T. A., & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior and Organization*, *68*(3–4), 466–476.

Buchan, N. R., Johnson, E. J., & Croson, R. T. A. (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 and Organization*, *60*(3), 373–398.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Burns, J., Corno, L., & La Ferrara, E. (2015). Interaction, Prejudice, and Performance: Evidence from South Africa. *Nimeo*.

Bursztyn, L., Ederer, F., Ferman, B., & Yuchtman, N. (2014). Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions. *Econometrica*, *82*(4), 1273–1301.

Burt, R. S. (1995). Social capital, structural holes and the entrepreneur. *Revue Francaise de Sociologie*, *36*(4), 599.

Burt, R. S., Barnett, W., Baron, J., Bendor, J.-A., Birner, J., Bothner, M., Dobbin, F., Heath, C., Kranton, R., Khurana, R., Pfeffer, J., Podolny, J., Raider, H., Rauch, J., & Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Candelo, N., Eckel, C., & Johnson, C. (2018). Social distance matters in dictator games: Evidence from 11 Mexican villages. *Games*, *9*(4), 1–13.

Cappellari, L., & Tatsiramos, K. (2015). With a little help from my friends? Quality of social networks, job finding and job match quality. *European Economic Review*, *78*(5240), 55–75.

Card, D. (2013). Peer effects of immigrant children on academic performance of native speakers: Introduction. *Economic Journal*, *123*(570), 279–280.

Cardella, E. (2016). Exploiting the Guilt Aversion of Others - Do Agents do it and is it Effective? *Theory and Decision*, *80*(4), 523–560.

Carell, S., Sacredote, B., & West, J. E. (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, *81*(3), 855–882.

Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, *60*(1), 31–51.

Carron, A., & Kevin, S. (1996). The group size-cohesion relationship in minimal groups. *Small Group Research*, *26*(1), 86–105.

Cehajic, S., Brown, R., & Castano, E. (2008). Herzegovina Forgive and Forget ? and Consequences Antecedents of Intergroup Forgiveness in Bosnia and Herzegovina. *Political Psychology*, *29*(3), 351–367.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130.

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. *Neuron*, *70*(3), 560–572.

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2011). Efficiency, Team building, and Identity in a Public-goods Game. *Nimeo*.

Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, *74*(6), 1579–1601.

Charness, G., & Dufwenberg, M. (2011). Participation. *American Economic Review*, *101*(4), 1211–1237.

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

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

Chaudhuri, A., & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, *73*(4), 959–985.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Chen, Y., & Li, S. (2009). Group Identity and Social Preferences. *American Economic Review*, *99*, 431–457.

Chuang, Y., & Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, *117*, 151–170.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), 947–966.

Condra, L. N., & Linardi, S. (2019). Casual contact and ethnic bias: Experimental evidence from Afghanistan. *Journal of Politics*, *81*(3), 1028–1042.

Conlin, M., Lynn, M., & O'Donoghue, T. (2003). The norm of restaurant tipping. *Journal of Economic Behavior and Organization*, *52*(3), 297–321.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281.

Cox, J. C., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, *59*(1), 17–45.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Dannenberg, A., Haita-Falah, C., & Zitzelsberger, S. (2020). Voting on the threat of exclusion in a public goods experiment. *Experimental Economics*, *23*(1), 84–109.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of Law and Economics*, *16*(1), 67–88.

Darley, J. M., & Latané, B. (1968). Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology*, *8*(4), 377–383.

Daskalova, V. (2018). Discrimination, social identity, and coordination: An experiment. *Games and Economic Behavior*, *107*, 238–252.

Dearing, R. L., & Tangney, J. P. (2004). *Shame and guilt*. Guilford Press.

DiDonato, T. E., Ullrich, J., & Krueger, J. I. (2011). Social perception as induction and inference: an integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology*, *100*(1), 66–83.

Diekmann, A. (1985). Volunteer's Dilemma. *Journal of Conflict Resolution*, *29*(4), 605–610.

Dixon, J., Durrheim, K., Tredoux, C., Tropp, L., Clack, B., & Eaton, L. (2010). A paradox of integration? Interracial contact, prejudice reduction, and perceptions of racial discrimination. *Journal of Social Issues*, *66*(2), 401–416.

Dovidio, J. F., Gaertner, S. L., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology*, *4*(1), 1–26.

Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2003). Intergroup Contact: The Past, Present, and the

Future. *Group Processes & Intergroup Relations*, *6*(1), 5–21.

Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, *16*(3), 349–371.

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

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Dunning, D., Anderson, J. E., Schlösser, T., & Ehlebracht, D. (2014). More a Matter of Respect Than Expectation of Reward. *Journal of Personality and Social Psychology*, *107*(1), 122–141.

Eckel, C. C., & Grossman, P. J. (1998). Are Women Less Selfisch Than Men?: Evidence From Dictator Experiments. *The Economic Journal*, *108*, 726–735.

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology*, *29*(23), 371–389.

Ellingsen, T., Johannesson, M., Tjotta, S., & Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, *68*(1), 95–107.

Emons, W. (1997). Credence Goods and Fraudulent Experts Authors. *RAND Journal of Economics*, *28*(1), 107–119.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, *14*(3), 583–610.

Engelmann, D., & Strobel, M. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution. *American Economic Review*, *96*(5), 1918–1923.

Enos, R. D. (2014). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences*, *111*(10), 3699–3704.

Erkal, N., Gangadharan, L., & Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *The American Economic Review*.

Espín, A. M., Exadaktylos, F., & Neyse, L. (2016). Heterogeneous motives in the trust game: A tale of two roles. *Frontiers in Psychology*, *7*(MAY), 1–11.

Fairley, K., Sanfey, A., Vyrastekova, J., & Weitzel, U. (2016). Trust and risk revisited. *Journal of Economic Psychology*, *57*, 74–85.

Falk, A. (2007). Gift exchange in the field. *Econometrica*, *75*(5), 1501–1511.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global Evidence on Economic Preferences. *Quarterly Journal of Economics*, *133*(4), 1645–1692.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, *24*(1), 39–57.

Falk, A., & Zehnder, C. (2013). A city-wide experiment on trust discrimination. *Journal of Public Economics*, *100*, 15–27.

Fang, H., & Moro, A. (2011). Theories of statistical discrimination and affirmative action: A survey. In *Handbook of Social Economics* (Vol. 1, Issue 1 B). Elsevier B.V.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Fehr, E., Naef, M., & Schmidt, K. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. *American Economic Review*, *96*(5), 1912–1997.

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

Fershtman, C, & Gneezy, U. (2001). Trust and Discrimination in a Segmented Society: An Experimental Approach. *The Quarterly Journal of Economics*, *116*(1), 351–377.

Fershtman, Chaim, & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, *116*(1), 351–377.

Fershtman, Chaim, Gneezy, U., & Verboven, F. (2005). Discrimination and Nepotism: The Efficiency of the Anonymity Rule. *The Journal of Legal Studies*, *34*(2), 371–396.

Fetchenhauer, D., & Dunning, D. (2009). Do people trust too much or too little? *Journal of Economic Psychology*, *30*(3), 263–276.

Finseraas, H., Johnsen, Å. A., Kotsadam, A., & Torsvik, G. (2016). Exposure to female colleagues breaks the glass ceiling—Evidence from a combined vignette and field experiment. *European Economic*

Review, *90*, 363–374.

Fischbacher, U. (2008). Learning and Peer Effects Shifting the Blame : On Delegation and Responsibility Shifting the Blame: On Delegation and Responsibility. *Nimeo*, 1–35.

Fischbacher, U., Gachter, S., & Fehr, E. (2001). *Are people conditionally cooperative ? Evidence from a public goods experiment*. *71*, 397–404.

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M., & Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*(4), 517–537.

Fiske, S. T. (1998). Stereotyping, Prejudice, and Discriminiation. *The Handbook of Social Psychology*, *2*(4), 357–411.

Fiske, S. T., & Neuberg, S. L. (1989). Category-Based and Individuating Processes as a Function of Information and Motivation: Evidence from Our Laboratory. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and Prejudice: Changing Conceptions* (pp. 83–103). Springer New York.

Gaertner, S. L., Dovidio, J. F., & Bachman, B. A. (1996). Revisiting the contact hypothesis: The induction of a common ingroup identity. *International Journal of Intercultural Relations*, *20*(3–4), 271–290.

Ghidoni, R., & Ploner, M. (2015). When do the Expectations of Others Matter? An Experiment on Distributional Justice and Guilt Aversion. *CEEL Working Paper 3-14*.

Gillet, J., Schram, A., & Sonnemans, J. (2009). The tragedy of the commons revisited: The importance of group decision-making. *Journal of Public Economics*, *93*(5–6), 785–797.

Glaeser, E. L., Laibson, D., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, *115*(3), 811–846.

Gneezy, U., List, J., & Price, M. (2012). Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments. *NBER Working Paper Series*, 17855.

Goeree, B. J. K., Mcconnell, M. A., & Mitchell, T. (2010). The 1/d Law of Giving. *American Economic Journal: Microeconomics*, *2*(1), 183–203.

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212–216.

Graf, S., Paolini, S., & Rubin, M. (2014). Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries. *European Journal of Social Psychology*, *44*(6), 536–547.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

Granovetter, M. (1992). Problems of explanation in economic sociology. *Networks and Organizations: Structure, Form, and Action*, 25–56.

Greig, F., & Bohnet, I. (2008). Is there reciprocity in a reciprocal-exchange economy? evidence of gendered norms from a slum in Nairobi, Kenya. *Economic Inquiry*, *46*(1), 77–83.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen – Beiträge zum Heinz-Billing-Preis 2003. GWDG-Bericht Nr. 63, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen* (pp. 79–93).

Guerra, G., & John Zizzo, D. (2004). Trust responsiveness and beliefs. *Journal of Economic Behavior and Organization*, *55*(1), 25–30.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2014). On cooperation in open communities. *Journal of Public Economics*, *120*, 220–230.

Güth, W., Levati, M. V., & Ploner, M. (2008). Social identity and trust-An experimental investigation. *Journal of Socio-Economics*, *37*(4), 1293–1308.

Hasler, B. S., & Amichai-Hamburger, Y. (2013). Online Intergroup Contact. In Y. Amichai-Hamburger (Ed.), *The social net: Understanding our online behavior.* (pp. 220–252). Oxford University Press.

Hauge, K. E. (2016). Generosity and guilt: The role of beliefs and moral standards of others. *Journal of Economic Psychology*, *54*, 35–43.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

He, H., & Villeval, M. C. (2017). Are group members less inequality averse than individual decision makers? *Journal of Economic Behavior and Organization*, *138*, 111–124.

Heffetz, O., & Frank, R. H. (2011). Preferences for Status: Evidence and Economic Implications. In J. Benhabib, A. Bisin, & M. O. B. T.-H. of S. E. Jackson (Eds.), *Handbook of Social Economics* (Vol. 1, pp. 69–91). North-Holland.

Henry, P. J. (2008). Student sampling as a theoretical problem. *Psychological Inquiry*, *19*(2), 114–126.

Herbst, L., Konrad, K. A., & Morath, F. (2015). Endogenous group formation in experimental contests. *European Economic Review*, *74*, 163–189.

Hewtone, M., Rubin, M., & Willis, H. (2002). Intergrroup Bias. *Annual Review of Psychology*, *53*, 575–604.

Hilton, J. L., & Hippel, W. Von. (1996). *STEREOTYPES.*

Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social Distance and Other_regarding Behavior in Dictator Games. *The American Economic Review*, *86*(3), 653–660.

Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, *11*(2), 74–91.

Hornsey, M. J. (2008). Social Identity Theory and Self-categorization Theory: A Historical Review. *Social and Personality Psychology Compass*, *2*(1), 204–222.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425.

Huck, S., & Weizsäcker, G. (2002). Do players correctly estimate what others do? Evidence of conservatism in beliefs. *Journal of Economic Behavior and Organization*, *47*(1), 71–85.

Ingersoll, J. E. (1987). *Theory of financial decision making* (Vol. 3). Rowman & Littlefield.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology*, *10*(4), 371–384.

Ioannides, Y. M., & Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, *42*(4), 1056–1093.

Isaac, R. Mark, & Walker, J. M. (1988). Group Size Effects in Public Goods Provision : The Voluntary Contributions. *The Quarterly Journal of Economics*, *103*(1), 179–199.

Isaac, R. Mark, Walker, J. M., & Williams, A. W. (1994). Group size and the voluntary provision of public goods. Experimental evidence utilizing large groups. *Journal of Public Economics*, *54*(1), 1–36.

Isaac, R M, & Walker, J. M. (1988). Group Size Effects in Public Goods Provision: the Voluntary Contributions Mechanism. *Quarterly Journal of Economics*, *103*(February), 179–199.

Jensen, M. K., & Kozlovskaya, M. (2016). Title: A Representation Theorem for Guilt Aversion. *Journal of Economic Behavior & Organization*, *125*, 148–161.

Johnston, L., Hewstone, M., Pendry, L., & Frankish, C. (1994). Cognitive models of stereotype change (4): Motivational and cognitive influences. *European Journal of Social Psychology*, *24*(2), 237–265.

Kahneman, D. (2003). Maps of Bounded Rationality : Psychology for Behavioral Economics. *American Economic Review*, *93*(5), 1449–1475.

Kahneman, D., & Tversky, A. (1979). Prospect Theory : An Analysis of Decision under Risk Linked. *Econometrica*, *47*(2), 263–292.

Kamijo, Y. (2016). Journal of Economic Behavior & Organization Rewards versus punishments in additive , weakest-link , and best-shot contests. *Journal of Economic Behavior and Organization*, *122*, 17–30.

Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology*, *30*(3), 368–380.

Karau, S. J., & Williams, K. D. (1993). Social Loafing: A Meta-Analytic Review and Theoretical Integration. *Interpersonal Relations and Group Processes*, *65*(4), 681–706.

Kawagoe, T., & Narita, Y. (2014). Guilt aversion revisited: An experimental test of a new model. *Journal of Economic Behavior and Organization*, *102*, 1–9.

Kelman, H. C. (1998). Social-psychological contributions to peacemaking and peacebuilding in the Middle East. *Applied Psychology*, *47*(1), 5–28.

Kerschbamer, R., Neururer, D., & Sutter, M. (2016). Insurance coverage of customers induces dishonesty of sellers in markets for credence goods. *Proceedings of the National Academy of Sciences*, *113*(27), 7454–7458.

Khalmetski, K., Ockenfels, A., & Werner, P. (2015). Surprising gifts: Theory and laboratory evidence. *Journal of Economic Theory*, *159*, 163–208.

Knowles, J., Persico, N., & Todd, P. (2009). Reconsidering racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, *109*(1), 203–229.

Koopmans, R., & Veit, S. (2014). Ethnic diversity, trust, and the mediating role of positive and negative interethnic contact: A priming experiment. *Social Science Research*, *47*, 91–107.

Korenok, O., Millner, E. L., & Razzolini, L. (2013). Impure altruism in dictators' giving. *Journal of Public Economics*, *97*(1), 1–8.

Korenok, O., Millner, E. L., & Razzolini, L. (2018). Taking aversion. *Journal of Economic Behavior and Organization*, *150*, 397–403.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., & Falk, A. (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy*, *128*(2), 434–467.

Köszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Kovacs, T., & Willinger, M. (2013). Are trust and reciprocity related within individuals? *B.E. Journal of Theoretical Economics*, *13*(1), 249–270.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*(3), 495–524.

Lane, T. (2016). Discrimination in the laboratory : a meta-analysis of economics experiments. *European Economic Review*, *forthcomin*, 1–28.

Lang, K., & Lehmann, J. Y. K. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature*, *50*(4), 959–1006.

Lang, K., & Spitzer, A. K. (2020). *Race Discrimination: An Economic Perspective*. *34*(2), 68–89.

Latané, B., & Darley, J. M. (1970). *The Unresponsive Bystander: Why Doesn't He Help?, Century Psychology Series*. New York,: Appleton-Century Crofts.

Leider, S., Möbius, M. M., Rosenblat, T., & Do, Q.-A. A. (2009). Directed Altruism and Enforced Reciprocity in Social Networks. *Quarterly Journal of Economics*, *124*(4), 1815–1851.

Levine, D. K. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, *1*(3), 593–622.

Liebe, U., & Tutic, A. (2010). Status groups and altruistic behaviour in dictator games. *Rationality and Society*, *22*(3), 353–380.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics*, *119*(1), 49–89.

List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy*, *115*(3), 482–493.

Luhmann, N. (1986). The autopoiesis of social systems. *Sociocybernetic Paradoxes*, *6*(2), 172–192.

Lundberg, B. S. J., & Startz, R. (1983). Private Discrimination and Social Intervention in Competitive Labor Market Author. *American Economic Review*, *73*(3), 340–347.

MacInnis, C. C., & Page-Gould, E. (2015). How Can Intergroup Interaction Be Bad If Intergroup Contact Is Good? Exploring and Reconciling an Apparent Paradox in the Science of Intergroup Relations. *Perspectives on Psychological Science*, *10*(3), 307–327.

Maoz, I. (2010). Educating for peace through planned encounters between Jews and Arabs in Israel: A reappraisal of effectiveness. In G. S. & E. Cairns (Ed.), *Handbook on peace education* (pp. 303–313). Psychology Press.

Marianne Bertrand, & Duflo, E. (2017). Filed Experiments on Discrimination. *Handbook of Economic Field Experiments*, *1*(May), 309–393.

Markowitz, H. M. (1952). Portfolio Selection. *Portfolio Selection*, *7*(1), 77–91.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112–145.

McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin*, *87*(1), 195–208.

Mele, A. (2020). Does school desegregation promote diverse interactions? An equilibrium model of segregation within schools. *American Economic Journal: Economic Policy*, *12*(2), 228–257.

Menger, C. (1871). Grundsätze der Volkswirtschaftslehre. In *Wirtschaft und Finanzen* (1st ed.). Wilhelm Braumüller.

Merton, R. K. (1972). Insiders and Outsiders : A Chapter in the Sociology of Knowledge. *American Journal of Sociology*, *78*(1), 9–47.

Mingers, J. (1994). *Self-producing systems: Implications and applications of autopoiesis*. Springer Science & Business Media.

Montinari, N., Nicolò, A., & Oexl, R. (2016). The gift of being chosen. *Experimental Economics*, *19*(2), 460–479.

Morell, A. (2017). The Short Arm of Guilt: Guilt Aversion Plays Out More Across a Short Social Distance. *Preprints of the Max Planck Institute for Research on Collective Goods*.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring Social Value Orientation. *Judgment and Decision Making*, *6*(8), 771–781.

Neilson, W., & Ying, S. (2016). Journal of Economic Behavior & Organization From taste-based to statistical discrimination ☆. *Journal of Economic Behavior and Organization*, *129*, 116–128.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1–2), 91–112.

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

Njozela, L., Burns, J., & Langer, A. (2018). The effects of social exclusion and group heterogeneity on the provision of public goods. *Games*, *9*(3), 1–21.

Oakes, Penelone J., & Turner, J. C. (1990). Is limited information processing capacity the cause of social stereotyping? *European Review of Social Psychology*, *1*(1), 111–135.

Oakes, Penelope J, Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, *30*(2), 125–144.

Ockenfels, A., Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, *90*(1), 166–193.

Ockenfels, A., & Werner, P. (2014). Beliefs and ingroup favoritism. *Journal of Economic Behavior and Organization*, *108*, 453–462.

Ondrich, J., Stricker, A., & Yinger, J. (1999). Do Landlords Discriminate? The Incidence and Causes of Racial Discrimination in Rental Housing Markets. *Journal of Housing Economics*, *8*(3), 185–204.

Ortmann, A., Fitzgerald, J., & Boeign, C. (2000). Trust, Reciprocity, and Social History: A Re-examination. *Experimental Economics*, *3*, 81–100.

Page-Gould, E., Mendoza-Denton, R., & Tropp, L. R. (2008). With a Little Help From My Cross-Group Friend: Reducing Anxiety in Intergroup Contexts Through Cross-Group Friendship. *Journal of Personality and Social Psychology*, *95*(5), 1080–1094.

Paluck, E. L., Green, S., & Green, D. P. (2018). The Contact Hypothesis Reevaluated. *Behavioural Public Policy*, 1–30.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188.

Paolacci, G., Chandler, J., & Stern, L. N. (2010). Running experiments on Amazon Mechanical Turk 2 Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181–192.

Pettigrew, T. F. (1998). Intergroup Contact Theory. *Annual Review of Psychology*, *49*(1), 65–85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, *38*, 922–934.

Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations*, *35*(3), 271–280.

Phelps, E. S. (1972). The Statistical theory of Racism and Sexism. *American Economic Review*, *62*(4), 659–661.

Rabin, M. (1993). Incorproating Fariness into Game Theroy and Economics. *American Economic Review*, *83*(4), 1281–1302.

Raihani, N. J., Mace, R., & Lamba, S. (2013). The Effect of $1, $5 and $10 Stakes in an Online Dictator Game. *PLoS ONE*, *8*(8), 3–8.

Reuben, E., Sapienza, P., & Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, *104*(2), 89–91.

Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301.

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics*, *90*(4), 629–649.

Ruscher, J. B., & Fiske, S. T. (1990). Interpersonal competition can cause individuating processes. *Journal of Personality and Social Psychology*, *58*(5), 832.

Scacco, A., & Warren, S. S. (2018). Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria. *American Political Science Review*, *112*(3), 654–677.

Schumacher, H., Kesternich, I., Kosfeld, M., & Winter, J. (2017). One, two, many-insensitivity to group size in games with concentrated benefits and dispersed costs. *Review of Economic Studies*, *84*(3), 1346–1377.

Schumpeter, J. A. (1908). *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie*. Duncker & Humblot.

Selten, R. (1965). *Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes*. Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie.

Semyonov, M., & Glikman, A. (2009). Ethnic residential segregation, social contacts, and anti-minority attitudes in European societies. *European Sociological Review*, *25*(6), 693–708.

Sherif, M., & Sherif, C. W. (1969). 1969 . (1969). *Social psychology* (H. and Row (ed.)).

Song, F. (2009). Intergroup trust and reciprocity in strategic interactions: Effects of group decision-making mechanisms. *Organizational Behavior and Human Decision Processes*, *108*(1), 164–173.

Spence, M. (1974). *Job Market Signaling*. Harvard University Press.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup Anxiety. *Journal of Social Issues*, *41*(3), 157–175.

Stigler, G. J. ., & Becker, G. S. . (1977). De Gustibus Non Est Disputandum. *The American Economic Review*, *67*(2), 76–90.

Stolle, D., Soroka, S., & Johnston, R. (2008). When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions. *Political Studies*, *56*(1), 57–75.

Sülzle, K., & Wambach, A. (2002). Insurance in a market for credence goods. *CESifo Economic Studies*, *677*(9).

Sutter, M., Balafoutas, L., Beck, A., & Kerschbamer, R. (2013). What Drives Taxi Drivers ? A Field Experiment on Fraud in a Market for Credence Goods. *Review of Economic Studies*, *80*(1), 876–891.

Sutter, M., Yilmaz, L., & Oberauer, M. (2015). Delay of gratification and the role of defaults-An experiment with kindergarten children. *Economics Letters*, *137*, 21–24.

Tajfel, H., & Turner, J. (1979). An Integrative Theory of Intergroup Conflict. *The Social Psychology of Intergroup Relations*, 33–47.

Taylor, S. E. (1981). A categorization approach to stereotyping. *Cognitive Processes in Stereotyping and Intergroup Behavior*, *832114*.

Topa, G. (2011). Labor markets and referrals. *Handbook of Social Economics*, *1*(1 B), 1193–1221.

Trautmann, S. T., & van de Kuilen, G. (2015). Belief Elicitation: A Horse Race among Truth Serums. *Economic Journal*, *125*(589), 2116–2135.

Turner, J. (1985). Social Categorization and Self-Concept: A Social Cognitive Theory of Group Behavior. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (Vol. 2, pp. 77–121). JAI Press.

Udehn, L. (2002). The changing face of methodological individualism. *Annual Review of Sociology*, *28*, 479–507.

Vainapel, S., Weisel, O., Zultan, R., & Shalvi, S. (2018). Group moral discount : Diffusing blame when judging group members. *Journal of Behavioral Decision Making*, *17*(1), 1–17.

van Bommel, M., van Prooijen, J. W., Elffers, H., & Van Lange, P. A. M. (2012). Be aware to care: Public self-awareness leads to a reversal of the bystander effect. *Journal of Experimental Social Psychology*, *48*(4), 926–930.

van Lange, P., Bekkers, R., Chirumbolo, A., & Leone, L. (2012). Are Conservatives Less Likely to be Prosocial Than Liberals? From Games to Ideology, Political Preferences and Voting P. *European Journal of Personality*, *26*(3), 461–473.

Visser, M. S., & Roelofs, M. R. (2011). Heterogeneous preferences for altruism: Gender and personality, social status, giving and taking. *Experimental Economics*, *14*(4), 490–506.

Yamagishi, T., Cook, K. S., & Watabe, M. (1998). Uncertainty, trust, and commitment formation in the United States and Japan. *American Journal of Sociology*, *104*(1), 165–194.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, *6*(3), 299–310.

Zettler, I., Hilbig, B. E., & Haubrich, J. (2011). Altruism at the ballots: Predicting political attitudes and behavior. *Journal of Research in Personality*, *45*(1), 130–133.

Zhang, J., & Casari, M. (2012). How groups reach agreement in risky choices: An experiment. *Economic Inquiry*, *50*(2), 502–515.

Zhang, L. (2017). Racial bias and repeated interaction in the NBA. *2017 Annual Meeting of the Academy of Management, AOM 2017*, *2017-Augus*.

Zhao, B., Ondrich, J., & Yinger, J. (2006). Why do real estate brokers continue to discriminate? Evidence from the 2000 Housing Discrimination Study. *Journal of Urban Economics*, *59*(3), 394–419.

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

Eidesstaatliche Erklärung nach §8 Abs. 3 der Promotionsordnung vom 17.02.2015

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, neben den in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im Innoch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

____

Ort, Datum Unterschrift

Lisa Lenz

M.Sc. Economics

Geburtsdatum: 17. Dezember 1991
Geburtsort: Albstadt Ebingen

SCHUL- UND HOCHSCHULBILDUNG

UNIVERSITÄT ZU KÖLN **10 / 2016 – 10/ 2020**	Titel der Dissertation: **“The Impact of Social Embeddedness on Social Preferences, Beliefs and Pro-Social Behavior”** **Akademische Weiterbildung im Rahmen strukturierter Doktorandenprogramme** Abschluss der Cologne Graduate School for Management, Economics & Social Sciences und der International Max Planck Research School
UNIVERSITÄT BONN **10 / 2014 – 09 / 2016**	**Master of Science in Economics („How Guilt and Shame Impact Pro-Social Behavior“)** Abschlussnote: „Sehr Gut“ (excellent, A)
ZEPPELIN UNIVERSITÄT **10 / 2011 – 08 / 2014**	**Bachelor of Arts in Corporate Management & Economics** Abschlussnote: „Sehr Gut“ (excellent, A); Rang 1 von 37
OTTO-HAHN GYMNASIUM **2011**	**Allgemeine Hochschulreife** Abschlussnote: 1,1

AKADEMISCHE BERUFSERFAHRUNG

UNIVERSITÄT ZU KÖLN **2016 – 2020**	**Anstellung am Lehrstuhl für Personalwirtschaftslehre der Universität zu Köln als wissenschaftliche Mitarbeiterin / Hilfskraft**
UNIVERSITÄT BONN **2015**	**Tutorin für das Fach Finanzierung an der an der Rheinischen Friedrich-Wilhelms-Universität Bonn**
ZEPPELIN UNIVERSITÄT **2012 – 2014**	**Studentische Hilfskraft am ZF Friedrichshafen Lehrstuhl für Unternehmensführung und Personalmanagement an der Zeppelin Universität in Friedrichshafen.**

STIPENDIEN

CGS Stipendiatin **2016 – 2019**	**Stipendiatin der Cologne Graduate School for Management, Economics & Social Sciences**
Friedrich-Naumann-Stiftung **2011 –2016**	**Stipendiatin der Friedrich Naumann Stiftung für die Freiheit**