Year : 2017

Social preferences in the lab, in the field, and online

Kistler Deborah

UNIL | Université de Lausanne

FACULTÉ DE DROIT, DES SCIENCES CRIMINELLES, ET D'ADMINISTRATION PUBLIQUE

IDHEAP

Social preferences in the lab, in the field, and online

THÈSE DE DOCTORAT

présentée à la

Faculté de droit, des sciences criminelles, et d'administration publique
de l'Université de Lausanne

pour l'obtention du grade de

Docteur en administration publique

par

Deborah Kistler

Directeur de thèse
Prof. Christian Thöni

Jury

Prof. Andreas Ladner, Président
Prof. Laure Athias, expert interne
Prof. Isabel Günther, expert externe
Prof. Christian Zehnder, expert externe

LAUSANNE
2017

UNIL | Université de Lausanne
IDHEAP
Institut de hautes études
en administration publique

IMPRIMATUR

Le Décanat de la Faculté de droit, des sciences criminelles et d'administration publique, sur proposition d'un jury formé des professeurs Christian Thoeni et Christian Zehnder et des professeures Laure Athias et Isabel Günther, sans se prononcer sur les opinions de la candidate, autorise l'impression de la thèse de Deborah Kistler intitulée :

Social Preferences in the lab, in the field, and online.

Lausanne, le 24 août 2017

Prof. Andreas Ladner
Vice-Doyen de la Faculté de droit,
des sciences criminelles
et d'administration publique

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Christian Thöni. Thank you, Christian, for your immense support during my PhD. You taught me how to conduct good behavioral and experimental research. Your fascination and passion for the topic is contagious.

I am also very grateful to the experts on my thesis committee: Laure Athias, Isabel Günther, and Christian Zehnder. Your valuable comments and suggestions improved my thesis considerably.

I am thankful to my colleagues at the University of Lausanne who supported and encouraged me to continue this journey when I was desperate because of puzzling or inexistent results. A special thanks goes to Jan Schmitz, your enthusiasm and ideas are very inspiring. I enjoyed all our discussions, and I hope that one day we will manage to conduct an experiment together with results that are in line with our predictions.

I would like to thank Monika Hess for initiating the impact evaluation of the Apoyar-Intervention. Thanks to you, I had the possibility to conduct research in Colombia, get to know the Colombian culture and hospitality, discuss a lot about the importance of impact evaluations and randomized control trials, and write the third chapter of my thesis. Moreover, I thank Rafael Lalive. Your guidance and support during this project was very valuable. ¡Muchas gracias!

I am extremely grateful to Herbert. You improved the look of my poster presentations considerably. Thank you for your endless loving support.

Last but not least, I want to thank my family, and friends. Thank you for patiently testing my experiments and encouraging me throughout this journey.

Thank you all!

Contents

General Introduction

Nowadays many scholars agree that the self-interest hypothesis, which postulates that humans care only about their own utility, is not the most accurate assumption to explain decision-making. Countless empirical studies provide evidence that humans have social preferences, meaning that they care about the well-being of other, unrelated individuals. Furthermore, in contrast to the predictions of standard economic theory, concerns about fairness and reciprocity affect behavior in many situations (Fehr & Schmidt, 2003). These social preferences are the central theme of my thesis.

For the functioning of modern societies, social preferences are crucial (see for example Ostrom, 1990; Bowles & Gintis, 2011). They facilitate for instance cooperation in groups of strangers, the provision of public goods or the maintenance of social order, to name just a few. Generally speaking, in many situations, the welfare of the group depends on the individual willingness to behave prosocial. In these situations, individuals have an incentive to free-ride on the others to increase their private benefit. However, if all individuals would act like this, everyone would be worse off. Let's consider the classical example of a group work at the university. If one student free rides and the others behave prosocial and cooperate, the free rider has the highest payoff. He does something else, which increases his utility while the others work on the group project. In the end, however, he still benefits from a good grade thanks to the effort of the others. In contrast, if all group members would free ride the group work would not be done, and hence everyone would be worse off. Charitable giving or volunteering activities are other examples where prosocial behavior is necessary. Thus, for social scientists, it is important to understand the development and determinants of social preferences.

This thesis contains three essays, in which I elaborate on social preferences from distinct angles. Common, besides the topic, is the methodology of experiments, which I applied throughout the thesis. The different types of experiments which we or I conducted rely all on standard methodologies from experimental economics. As the title of the thesis suggests, we conducted one experiment online, one in the laboratory and one in the field. Experiments can be used to either measure preferences or to elaborate factors which causally affect preferences. In chapter one, the main purpose of the experiment was to measure social preferences and their relationship to other socially relevant variables. The results of such correlational studies can reveal patterns which serve as hypothesis for subsequent research. For example, to test whether the observed relationship is causal. Furthermore, based on these correlations one can make predictions. Lastly, such correlations can be evaluated across countries and cultures, to elaborate cross-cultural differences. In contrast, in chapter two and three the aim was to evaluate factors which affect social preferences causally. These designs allow to draw conclusions about the determinants of social preferences. Moreover, insights from such analyses can be very valuable for policy making.

To be more precise, in chapter one, we study the link between moral values and various forms of prosocial behavior. Since decades, cross-cultural psychology examines moral values using data from standardized surveys, assuming that values guide human behavior. However, so far, the claim that moral values influence prosocial behavior has only been demonstrated for activities that respondents self-reported in surveys (Welzel & Deutsch, 2012) but never for directly observed behavior. Moreover, we are interested in a particular set of moral values; namely, emancipative and secular values. Emancipative values reflect the appreciation of equal freedoms and secular values the depreciation of sacred authority. They become more important as the living conditions of individuals improve (Welzel, 2013). To fill this gap, we conducted online behavioral experiments with participants from the sixth wave of the World Values Survey (WVS) in Germany. This allows us to link the respondents' moral values, as measured by the WVS, to the same respondents' prosocial behavior, as observed in the behavioral experiment. In the online behavioral experiment, participants conducted several incentivized decision tasks. Concretely, we elicited measures

for cooperation in a public goods game, the ability to coordinate in a property rights game, and altruism using a donation decision. We test the hypotheses that, (a) emancipative values show a strongly positive association with prosocial behaviors, (b) secular values show a modestly positive association, (c) at the same as they associate strongly negatively with protectionist behavior. The evidence boils down to three findings. Emancipative values relate to higher common pool contributions and larger donations to charitable organizations. Secular values, on the other hand, are linked to more productive and less protective investments. As these results conform to key theories and reach empirical significance in a major postindustrial nation, we conclude that we have significant evidence at hand highlighting the potential of combined survey-experiment methods to establish value-behavior links that are otherwise inexplorable.

In the second chapter, we conducted a laboratory experiment to evaluate the effect of salience on the contributions in a public goods game. According to standard theory in economics, an individual should process all available information before making a decision. This does not mean that all choice attributes receive the same weight in the decision-making process, but they should at least be considered to determine their weight. However, for instance, when I go to Montreal, and I want to buy something, I constantly valuate the items to be cheaper than they actually are. I neglect the taxes because the price tag does not include them. In such cases, the tax is not salient, and salience guides my choice. If and how salience affects choices is the topic of the second chapter. In the first part of the chapter, we introduce a salience factor to the Fehr-Schmidt model (Fehr & Schmidt, 1999) to derive theoretical predictions. All bilateral comparisons that individuals make are weighted by how salient the behavior of other relevant group members is. Based on this extension, we hypothesize that the salience of the contributions of the other players should affect contributions in a standard public goods game systematically. We implement two treatments to test our predictions empirically. In the maximum treatment, the highest contribution is most salient. In contrast, in the minimum treatment, the lowest contribution within a group is most salient. Our results are surprising: We find that people do not adjust their contributions according to our hypotheses. If the lowest contribution in a group is most salient, it does not lead to an immediate decline in the contribution. Vice

versa, if the maximum contribution is salient, it does not result in higher contributions. Based on our results we hypothesize that focusing on the maximum provides an upper bound of the acceptable contribution level, while the minimum serves as lower bound.

Lastly, in chapter three, I elaborate how the social environment of children affects their development in general and in particular the formation of their preferences. Children from low-income families often perform worse with respect to skills compared to children from more advantaged families. Research on early childhood development suggests that providing children with high-quality preschool education potentially closes this skill gap (Heckman, 2006). I elaborate on an intervention which aimed at improving the quality of care in community nurseries in Colombia. These community nurseries are part of a national program, which has been implemented about 40 years ago and targets families from the lowest stratum. Bernal, Fernández, Flórez, Gaviria, et al. (2009) evaluated the program and reported that many caregivers have insufficient knowledge about infant development. Therefore, a local NGO implemented an intervention with four components to improve the quality of care. First, community mothers received a formal vocational training as early childhood teacher. Second, they received support and coaching to integrate what they learned in their daily nursery routines. Third, they learned how to teach parents about issues related to appropriate child care and development. Fourth, the implementing NGO monitored and supported the children in school, once they left the community mother. While the existing research has focused on cognitive and socioemotional skills, I complement the analysis with an elaboration of the effect on individual preference measures. For that, I conducted lab-in-the-field experiments with children who have visited a treated nursery and compare them to children who have been in an untreated nursery. I elicit measures for prosocial behavior, egalitarian preferences, trust, risk and time preferences. Moreover, I do evaluate not only the effect in the short run but also four years after the implementation. The intervention improved children's cognitive, psychosocial, and psychomotor skills. More importantly, the effects persist up to four years after the intervention. We show that children from the treated group have better grades and are more likely to be in the right grade for their age. Interestingly, the intervention also affected social preferences of the children. Children who visited a treated nursery are more

altruistic than children who attended a regular nursery. The evaluation set-up does not allow to fully rule out that unobserved factors drive the effects. Furthermore, we are not able to identify which part of the intervention is responsible for the observed effects. But the pattern of results makes me confident that what I see can be traced to the intervention.

References

Bernal, R., Fernández, C., Flórez, C. E., Gaviria, A., et al. (2009). Evaluación de impacto del programa hogares comunitarios de bienestar del icbf. Serie Documentos Cede, 2009-16.

Bowles, S. & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution.* Princeton, NJ: Princeton University Press.

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

Fehr, E. & Schmidt, K. M. (2003). Theories of fairness and reciprocity-evidence and economic applications. In M. Dewatripont, L. P. Hansen, & S. J. Turnovsky (Eds.), *Advances in economics and econometrics* (pp. 208–257). Volume 1. Econometric Society Monographs, Eighth World Congress.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge University Press.

Welzel, C. (2013). *Freedom rising: Human empowerment and the quest for emancipation.* New York, NY: Cambridge University Press.

Welzel, C. & Deutsch, F. (2012). Emancipative values and non-violent protest: The importance of 'ecological'effects. *British Journal of Political Science*, *42*(02), 465–479.

Chapter 1

Survey Response and Observed Behavior: Emancipative and Secular Values Predict Prosocial Behaviors*

*This chapter is joint work with Christian Thöni and Christian Welzel.

This chapter is published as an article in the *Journal of Cross-Cultural Psychology*. We thank the editor, two anonymous reviewers, Lorenz Goette as well as the conference participants of the IMEBESS 2015 in Toulouse and the seminar participants in Lausanne and Nottingham for helpful comments and suggestions. We thank Philip Büttiker for the programming of the online experiments. We gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft and the Swiss National Science Foundation.

1.1 Introduction

It is widely understood that cooperative behavior among strangers is a key ingredient of thriving societies in which institutions function impartially (Axelrod, 1986; Ostrom, 1990; Coleman, 1990; North, Wallis, & Weingast, 2009; Bowles & Gintis, 2011). The willingness to behave prosocially—to forgo individual gains for the benefit of others—differs considerably across societies and individuals (Gächter, Herrmann, & Thöni, 2010). Understanding the determinants and origins of prosocial behaviors is of central interest to all social sciences.

In this study, we investigate the role of moral values for prosocial behavior. Research in social psychology suggests that values guide prosocial behavior (Kluckhohn, 1951; Eckstein, 1961, 1998; Inkeles, 1969, 1983; Rokeach, 1968, 1973; Axelrod, 1986; Putnam, Leonardi, & Nanetti, 1994; Putnam, 2000; Welzel, 2010). Indeed, there is convincing evidence that "benevolence" and "universalism" values are related to prosocial behavior (Bardi & Schwartz, 2003; Sagiv, Sverdlik, & Schwarz, 2011). According to these findings, people who express concern for the well-being of others are more prosocial (cf. Schwartz, 2004, 2006, 2007).

In a separate strand of scholarship, Inglehart, Norris, and Welzel focus on two particular sets of moral values that grow stronger as the living conditions of the bulk of a population improve: *secular values* and *emancipative values* (R. Inglehart & Norris, 2003; Norris & Inglehart, 2004; R. Inglehart & Welzel, 2005). These values are indeed indicative of a host of life quality indicators, from subjective well-being to life expectancy to education, information access, income equality, physical security, rule of law, democracy, social capital, and ecological sustainability (Welzel, Inglehart, & Klingemann, 2003; R. Inglehart & Welzel, 2005; Welzel & Inglehart, 2010; Welzel, 2013). The emergence of these two sets of values is described as a psychological reaction to the societal transformations that modernization brings about.

Welzel (2013) describes modernization as an empowering process through which the lives of ordinary people improve. Indeed, where and when it happens, modernization makes people's lives safer, longer, and enriches them with more options to pursue a pur-

pose of their choice. Modernization, thus, transforms the nature of life, turning it from a source of threats to suffer into a source of opportunities to thrive. As this happens, societies climb the "utility ladder of freedoms": Tolerating and practicing freedoms becomes increasingly vital to take advantage of the options that a more promising life offers. In recognition of this utility shift, people change their moral values: They begin to see less value in sacred authority and more value in equal freedoms. The first process—depreciation of sacred authority—is reflected in rising secular values and is linked to a growing sense of existential security. The second process—appreciation of equal freedoms—is reflected in rising emancipative values and linked to a growing sense of individual autonomy. The egalitarian component of these values has a strongly antidiscriminatory impetus that favors a deeper internalization of impartiality norms. Welzel (2013) hypothesizes that the antidiscriminatory and humanitarian tendency of emancipative values predisposes their carriers to prosocial behavior vis-à-vis strangers (cf. Welzel, Inglehart, & Deutsch, 2005; Welzel & Delhey, 2015).

As concerns secular values, scholars working in the Inglehart-Welzel framework did not phrase explicit hypotheses about these particular values' behavioral consequences. The evidence, however, that secular values emerge from feelings of existential security suggests various hypotheses. Perhaps, the most intuitively plausible expectation deriving from secular values' link to existential security is that secular oriented people avoid protectionist behavior. Inglehart and Welzel's theory also posits that existential security reduces people's fear of foreigners, which suggests that secular values ease cooperation with strangers.

However, secular values by definition involve lower religiosity. Accordingly, these values' impact on behavior should reflect how religiosity affects behavior. Shariff (2015) provides an overview of studies that see religiosity as a force encouraging prosocial behavior. From this point of view, one would assume that—because of their inverse relation to religiosity—secular values are negatively linked with prosocial behavior. Shariff's study clarifies, however, that the relationship between religiosity and prosocial behavior only holds for self-reported measures of prosocial behavior and vanishes when prosocial behavior is measured by observation. Galen (2012) supports this conclusion. In light of this evidence, the antireligious impetus of secular values should be neutral as concerns these

values' influence on observed prosocial behavior. By contrast, the alleged absence of existential fear among secular oriented people should encourage prosocial behavior. Overall, we thus hypothesize a modestly positive link of secular values with prosocial behavior. Our clearest expectation with respect to secular values, however, is that they discourage protectionist behavior.

So far, the claim that secular and emancipative values influence prosocial behavior has only been demonstrated for activities that respondents *self-report* in surveys (Welzel & Deutsch, 2012) but never for directly *observed* behavior. With this study, we close this gap. We investigate the link between our predictor variables (secular and emancipative values) and outcome variables, measuring three facets of prosocial behavior: (a) cooperation in a social dilemma game, (b) productive versus protectionist investments in a property rights game, and (c) donation to charity. We test the hypotheses that, while (a) emancipative values show a strongly positive association with prosocial behaviors, (b) secular values show a modestly positive association, (c) at the same as they associate strongly negatively with protectionist behavior.

The dominant explanation of prosocial behavior by far is social capital, which is supposed to facilitate interactions between strangers (Coleman, 1990; Putnam et al., 1994; Knack & Keefer, 1997). While definitions of social capital vary, scholars routinely include interpersonal trust and adherence to civic norms. In contrast to the research on values, there exists an extensive literature investigating the link between standard social capital measures and behavior in controlled experiments. Glaeser, Laibson, Scheinkman, and Soutter (2000) investigate the relation between the trust question and behavior in a trust game and do not find a strong relation. In experiments similar to the public goods game reported in this study, Gächter, Herrmann, and Thöni (2004) find only weak evidence for the relation between trust questions and cooperation, while Thöni, Tyran, and Wengström (2012) find a positive link between trust and cooperative behavior in a large and representative sample of the Danish population. Balliet and Van Lange (2013) survey a large number of studies on prisoner's dilemma and public goods games and find a small to moderate positive relation between dispositional trust and cooperation. In this study, we are not interested in the influence of social capital per se. Instead, we include social capital

to see whether the influences of our interest—emancipative and secular values—withstand its control.

To investigate our hypotheses, we invited the respondents of the sixth round of the World Values Survey (WVS) in Germany to participate in an online behavioral experiment. This allows us to link the respondents' moral values, as measured by the WVS, to the same respondents' prosocial behavior, as observed in the behavioral experiment. Germany is a relevant and insightful case for our study. The country includes the biggest national population in Europe and constitutes the fourth-largest economy in the world. It is also a nation with the typical features of postindustrial knowledge societies, arguably the most advanced type of society in human history (R. Inglehart & Welzel, 2005). What we find here in terms of value-behavior links is presumably representative for a wider set of countries at the forefront of modernization.

Moreover, Germany incorporates cleavages of more general interest from a cross-cultural point of view. For one, there is the North-South cleavage between historically Protestant and Catholic regions that characterizes the whole of Europe. No question, secularization, urbanization, and migration have blurred this historic division but culture is an inert force that leaves a lasting imprint. Since the writings of Max Weber, the literature has attributed Protestants a stronger tendency to cooperate with strangers than Catholics.

Equally important, there is the East-West cleavage between ex-communist and market-capitalist societies that is also of a more general cross-cultural relevance, with the East having experienced 40 years of a Soviet-type dictatorship. In light of this division, the literature on "homo Sovieticus" (Kuran, 1997) suggests that "Easterners" tend to trust strangers less and to cooperate to a lesser extent.[1]

Germany embodies these cultural cleavages in a single nation, which allows to study cross-cultural issues while retaining comparability of individuals at a maximum level, as they all belong to the same national culture. Taking advantage of this specific context, we test two subhypotheses: (a) West German respondents behave more prosocially than those from East Germany, and (b) Protestants behave more prosocially than Catholics.

[1]The legendary mutual support networks and neighborhood help in communist systems might suggest otherwise but these phenomena indicate cooperation among people connected through bonds of familiarity. This is not cooperation among strangers.

The remainder of this chapter proceeds as follows: the "Methodological Advantage of Behavioral Experiments" section outlines the added value of combining a behavioral experiment with a nationally representative survey; the "Design and Procedures" section describes the design and procedures; the "Results" section presents the findings; finally, the "Discussion" section concludes.

1.2 Methodological Advantage of Behavioral Experiments

In a seminal article, Brislin (1970) pointed out that one of the major challenges of cross-cultural research is to ensure equivalence across different languages. Besides language, context differs considerably between cultures, which also reduces equivalence (see also Hui & Triandis, 1985). While the task description of our behavioral experiment still relies to some degree on natural language, we argue that the use of behavioral experiments (Camerer, 2003; Camerer & Fehr, 2004) mitigates the equivalence problem substantially. The reason is that the underlying strategic situation is a well-defined game, that is, it consists of a set of formal rules and mathematical functions.[2] These rules and functions are independent of the cultural context and the language of exposition. Hence, such well-defined formal games facilitate the comparison of outcomes across cultures (Gächter et al., 2010; Herrmann, Thöni, & Gächter, 2008).[3]

As a methodological contribution, we develop a novel instrument for an Internet-based behavioral experiment. Key advantages of this tool relative to standard methods in experimental research are (a) high flexibility in terms of the number of participants, their educational background, their local and timely availability; (b) easy adaptation to other participant pools with different languages; and (c) relatively low administrative costs. Furthermore, our experiment does not require a laboratory because respondents can par-

[2] A strategic game is defined by (a) a set of players, (b) a set of strategies for each player, and (c) the players' preferences (Osborne & Rubinstein, 1994, see for example). Although (a) and (b) are under full control of the experimenter, the participants' preferences are not directly observable. The pioneering work by Smith (1976) highlighted the importance of inducing preferences with financial incentives. Smith argues that this can be achieved "[. . .] by using a reward structure to induce prescribed monetary value on actions" (p. 275).

[3] Most of these studies rely on convenient samples from student populations and it is an ongoing debate to what extent the results of such games generalize to entire national populations (Henrich, Heine, & Norenzayan, 2010).

ticipate from any Internet-connected device. For these reasons, our behavioral experiment is ideally suited for cross-cultural research in combination with opinion surveys like the WVS.[4] In the spirit of Hui and Triandis (1985), we see behavioral experiments as a complementary tool for cross-cultural research, which should be used in combination with other methods.

1.3 Design and Procedures

Our main sample consists of respondents who participated in two studies. First, they were interviewed for the WVS in Germany. Their responses in the WVS serve as independent variables of our study. At the end of the WVS interview, we invited all respondents to participate in our online behavioral experiment, which measures three facets of prosocial behavior that form our dependent variables. In this section, we describe our data sources and procedures in detail.

1.3.1 WVS

Since 1981, the European Values Study (EVS) and WVS conduct representative public opinion surveys in scores of countries around the world. The basis is an English-language, fully standardized master questionnaire. The questionnaire covers a host of topics, from social capital to tolerance to trust to happiness to civic engagement. The master questionnaire is translated into national languages with semantic checks through back translation. Questionnaires are administered among randomly selected adult residents, with a minimum sample size of 1,000 people per country. Fieldwork is usually conducted through face-to-face interviews. Average interview length is about 50 min. Thus far, the EVS/WVS has conducted some 300 national surveys in by now 106 societies that represent more than 90% of the world population. Further details on fieldwork, sampling, questionnaires, as

[4]Some of the features of our behavioral experiment come at a cost. The fact that it is Internet-based restricts the potential participant pool to computer-literate individuals with access to an online device. Unlike in the laboratory, we do not have control over the environment in which participants take their decisions. Although this can be an advantage (the lack of direct interaction with the researcher increases cross-subject anonymity and might lower experimenter demand effects or impression management effects), it can also introduce a confound because the researcher does not know whether responses reflect the participant's individual decision, or whether respondents consulted other sources before responding.

well as downloadable data sets are available at www.worldvaluessurvey.org. All data are public domain. Between 2010 and 2014, the WVS finalized its most recent and sixth round of surveys in a total of 60 societies. As part of this project, the sixth German WVS was fielded in fall 2013. Roughly, 2,000 interviews have been realized.

Moral values

Emancipative values and secular values are each measured on the basis of 12 items from the WVS. The 12 items of secular values measure distance from sacred authority in the domains of religion, the nation, the state, and group pressures. The 12 items of emancipative values measure an emphasis on equal freedoms in the domains of personal autonomy, gender equality, the voice of the people, and sexual liberty. These items are combined in a "formative index" procedure, as documented in Welzel (2013, pp. 57-73). These pages also document the cross-cultural validity of these measures as well as their approximately normal distribution.[5] Index scores vary from minimum 0 to maximum 1 on both value indices. The original wording (from the English master questionnaire) and response format of all items used to create the secular and emancipative values are reported in the appendix.

Measures of Social Capital

Trust is measured using the question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" from the WVS. The answer is binary, that is, either "most people can be trusted" or "need to be very

[5]We do not report Cronbach's alpha or any other internal coherence statistics because this would raise the false expectation that our value measures are "reflective scales" that need to be evaluated against the standards of latent variables. We point this out very explicitly, emphasizing that our measures of secular values and emancipative values are "formative indexes". Accordingly, they are not summarized because they reflect interchangeable manifestations of single unifying dimension. Instead, they are summarized because (a) they represent complementary constituents that build on each other in fitting the overarching definition of secularism and emancipation and (b) because, in their very combination, these measures are empirically consequential. To meet these two criteria, interitem correlations should actually be modest at best (which is what they are) because only then do the items incorporate enough complementary variance to make their combination sufficiently distinct from what each single item contributes. As the literature on these values has stressed from the beginning Welzel (2013, pp. 58-73) and more recently in a specifically methodological piece (Welzel & Inglehart, 2016)), this "formative" approach to summarize items is appropriate whenever there are good reasons to assume that items are complementary rather than interchangeable and that what truly matters is indeed the combination of these complementarities. As we will see, our results strongly confirm this approach: When rerunning our regressions with the value measures unpacked into their single components, we generally obtain weaker and less significant results. The supposed effects only surface when the items combine into a single measure of emancipative and secular values each.

careful." For civic norms, we use four items in which respondents indicate whether they find the following behaviors acceptable: (a) "claiming government benefits to which you are not entitled," (b) "avoiding a fare on public transport," (c) "stealing property," and (d) "someone accepting a bribe in the course of their duties." Answers to these items are given on a scale from 1 (never justifiable) to 10 (always justifiable). We average the four items to our measure for civic norms (Cronbach's $\alpha = .74$).

Further Control Variables

In all our regression models, we include age, gender, political interest, and left-right orientation as control variables. Political interest is measured on a 4-point Likert-type scale ranging from very interested to not at all interested. Left-right indicates a person's self-reported political position on a scale from 1 to 10, with lower numbers indicating left and higher numbers right.[6]

1.3.2 Behavioral Experiment

At the end of the WVS interview, all participants were invited to participate in the behavioral experiment. They received a flyer with information about the procedures of the behavioral experiment, the possibility of earning money, and the login information. Each participant received a unique and randomly generated six-digit login code.[7] We use this code to match the answers in the WVS questionnaire to the behavioral data. The behavioral experiment consists of several decision tasks: a public goods game, a property rights game, and a donation decision.[8] The ordering of the tasks is the same for all participants. We administered the behavioral experiment online and designed it such that no

[6]The exact wording of the two political interest and view questions were (a) How interested would you say you are in politics? (b) In political matters, people talk of "the left" and "the right." How would you place your views on this scale, generally speaking?

[7]All interviewers asked participants to indicate their email address. Participants willing to provide us their address received an email with a personalized link to login to the behavioral experiment.

[8]Participants also played a public goods game with punishment, a second property rights game, a risk elicitation task, and a task measuring honesty. To keep the article concise, we restrict our attention to the public goods game, the property rights game, and the donation decision. In short, the analysis of the remaining decision tasks supports the results reported in this article. The link between the moral values and the public goods game with punishment is essentially the same as in public goods game (see "Contributions to the Public Good" section). Honesty seems to be weakly positively correlated with emancipative values. We do not find evidence that moral values have predictive power for risk preferences.

real-time interaction among the participants is necessary. For all decision tasks, involving more than one participant, we matched participants into groups only at the very end of the data collection period. This procedure allows participants to answer the questions at their convenience, that is, they can quit the behavioral experiment anytime and continue at a later point in time.

Following the methodological standards for behavioral experiments, we offer monetary incentives for all decision tasks. Participants' earnings are calculated in an experimental currency unit (ECU; in the experiment we speak of "Talers"). At the beginning of the experiment, participants were informed about the payment procedures. First we communicated the exchange rate from ECU to the local currency, which was 1 ECU to 4 Euros.[9] Second, we explained the random draw. In each decision task, the amount of ECU depends on the participant's decision, and—in case of group decisions—on the decision of the other group member (another randomly selected participant). For each task, we randomly drew a participant to receive the amount he or she earned in the given task. In case the task was a two-player game (public goods game and the property rights game), we randomly drew another participant as the partner. Hence, for these games, we randomly selected a group, for which both participants received the earnings from the decision task. The donation decision became payoff relevant for the participants who had been randomly selected to get paid for one of the other decision tasks. For these participants, we subtracted the percentage according to their donation decision and transferred the money to the charity indicated by the participant. The four participants who have been selected for payment in the public goods game and the property rights game, earned on average € 400 (minimum: € 208, maximum: € 640).[10]

[9]We introduced an experimental currency (ECU) to have a measurement of profits independent of the local currency. This procedure facilitates the adaptation of the experiment to other populations with different currencies because it is sufficient to change the exchange rate to the local currency, while all other payoff parameters remain unchanged.

[10]We calibrated these amounts with the aim to set expected payoffs similar to the financial incentives offered in comparable studies. Participants' expected payoffs depend on the response rate of the behavioral experiment, which introduces an additional element of uncertainty in the incentives. There is a methodological debate about the importance of financial incentives and the effect of different stake sizes in behavioral games. In their survey, Camerer and Hogarth (1999) stress the importance of financial incentives, but they also conclude that there is little evidence suggesting that the size of the incentives importantly influences the results. In total (for all decision tasks) we paid 13 participants of this behavioral experiment an amount of € 263.50 on average.

Public goods game

The public goods game is a standard tool in experimental economics to elicit cooperative behavior (Ledyard et al., 1995). In our case, we randomly group players in pairs of two. Each player is endowed with 100 ECU and has to decide how many ECU to contribute to a common pool, called the *group account*, that is, they simultaneously choose a contribution $c_i \in \{0, 20, \ldots, 100\}$.[11] The contributions of both players are summed up and multiplied by 1.5. The resulting amount is returned to the two players in equal shares. Thus, independent of a player's contribution he or she is entitled to half of the amount in the group account. The payoff function of player i in a group with player j is

$$\pi_i = 100 - c_i + \frac{1.5}{2}(c_i + c_j),$$

Under standard game-theoretic assumptions (selfish preferences and rationality), this game has a unique Nash equilibrium in which both players contribute zero to the public good. Individual rationality commands a contribution of zero because each ECU contributed to the group account yields only 0.75 ECU in return. However, the fact that the other player also receives 0.75 ECU for every ECU contributed means the social return is 1.5 and thus collective rationality calls for full contribution by both players. Much like the famous prisoner's dilemma, this game models a social dilemma.

In the design of the behavioral experiment, we paid great attention to guide participants carefully through the procedures and to give instructions that are intuitive and easy to understand. Interactive graphs illustrate the key features of the game. Participants go through four stages before taking their decision. First, they read the instructions of the public goods game, explained by six bullet points. The bullet points appear consecutively on the screen, and each step is accompanied by a new element in the interactive graph. Participants can go back and forth between the bullet points. Second, participants are shown the payoff consequences of three possible outcomes of the game (both players contribute zero; both players contribute 100; one player contributes 60 and the other

[11] The word "simultaneous" does not mean "at the same time," but should be understood in a game-theoretic sense, namely, that participants choose their actions not knowing the decision of their counterpart.

player contributes 100). Third, participants proceed to an exploration stage, where they can enter any combination of the two contributions and calculate their own and the other group members' hypothetical payoff. Figure 1.1 shows the screen of this exploration stage. Fourth, participants have to answer control questions. In the control questions, they are presented with a randomly selected combination of contributions and have to calculate the resulting payoff. Participants who provide an incorrect answer are asked to try again two times.[12] Finally, participants choose their contribution and proceed to the next game.

Figure 1.1: Screen shot of the public goods game.

Note. Participants can use the + and − buttons to choose the contributions of the two players, which is represented by coins shifted from the personal account to the bowl in the middle. Pressing a button "calculate" (not visible) multiplies the contributions by 1.5 and returns the money to the two players. The figure shows the outcome when the player to the left contributed 60 Talers (ECU) and the player to the right contributed 100 Talers, leading to final profits of 160 Talers and 120 Talers. Participants can click on NEW to reset the situation and calculate another situation. ECU = experimental currency unit.

Property Rights Game

The second decision task aims to capture the strategic situation in an environment with weak external enforcement of property rights. In all modern societies, formal institutions govern protection of property rights. Nonetheless, we observe considerable amounts of private investment in protection of property, suggesting that formal protection is imperfect. Under these circumstances, populations can either coordinate on efficient situations with

[12] Answering correctly was not a requirement to proceed with the experiment, as we did not want to lose respondents not willing or unable to answer the control question. We follow the method of Roux and Thöni (2015) and present the participants with randomly chosen contributions in the control question to avoid a systematic influence on decisions.

high mutual respect for property, or they can coordinate on very inefficient situations with mutual distrust and high private investments in protection.

Like in the public goods game, players are randomly paired into new groups of two. To facilitate understanding, we give the property rights game an agricultural framing, in which participants act as "farmers" who can allocate resources ("working hours") to production, protection, and stealing. In the graphical representation of the game, production refers to planting carrots in fields, and we refer to stealing in a neutral way and label it as "taking." Depending on their decisions and those of their paired counterpart, players end up with more or less carrots. The monetary payoff of the game is equal to the number of carrots they acquire, with each carrot being worth one ECU. Planting carrots is the only productive activity; the other two activities affect only the distribution of wealth. Participants have seven fields to plant carrots and seven units of resources, and they receive a monetary endowment.[13] Players simultaneously allocate their resources to production, protection, and stealing. Similar to the public goods game, participants go through four stages of learning and exploring the game. All parts are accompanied by interactive graphics. Figure 1.2 shows the screen in the exploration stage.

The three activities, production, protection, and stealing have the following payoff implications: each resource for production yields a carrot on each of the seven fields. The marginal return is thus seven ECU if the other player does not steal some of the player's fields. Each resource unit allocated to protection builds a fence around one field, which protects this field against theft. Each unit allocated to stealing results in the annexation of the crop of one field of the other player (the hands in Figure 1.2), up to the number of unprotected fields. Stealing starts from above, protecting starts from below, that is, if a player allocates a unit to stealing, then the first field stolen is the field at the top of the

[13]We implemented three endowment treatments, where the initial endowments varied across players. In one treatment, a player was rich (endowment of 40 ECU) compared with the other player in the group (endowment of 10 ECU). In a second treatment, a player was poor compared with the other player, and in the baseline treatment, both players had the same endowment (10 ECU). Participants were randomly allocated to one of the three treatments. Thus, the effect of the treatment should not influence the analysis with regard to emancipative and secular values. The main effects of the treatment variation on the allocation decision are weak. A one-way ANOVA test indicates the treatment differences are weakly significant for production ($F = 2.24$, $p = .085$), and insignificant for the other two outcomes (protection: $F = 1.41$, $p = .242$, stealing: $F = 0.65$, $p = .585$). In the main article, we do not further discuss these treatment variations. In Tables B3 and B4 in the appendix, we show that none of our results change when we control for differences in the endowment.

screen. The next field stolen is the one to the left in the second row and so on. If a player allocates a unit to protection, then the first field protected is the one at the very bottom of the screen. The next field protected is the one to the right in the second last row.

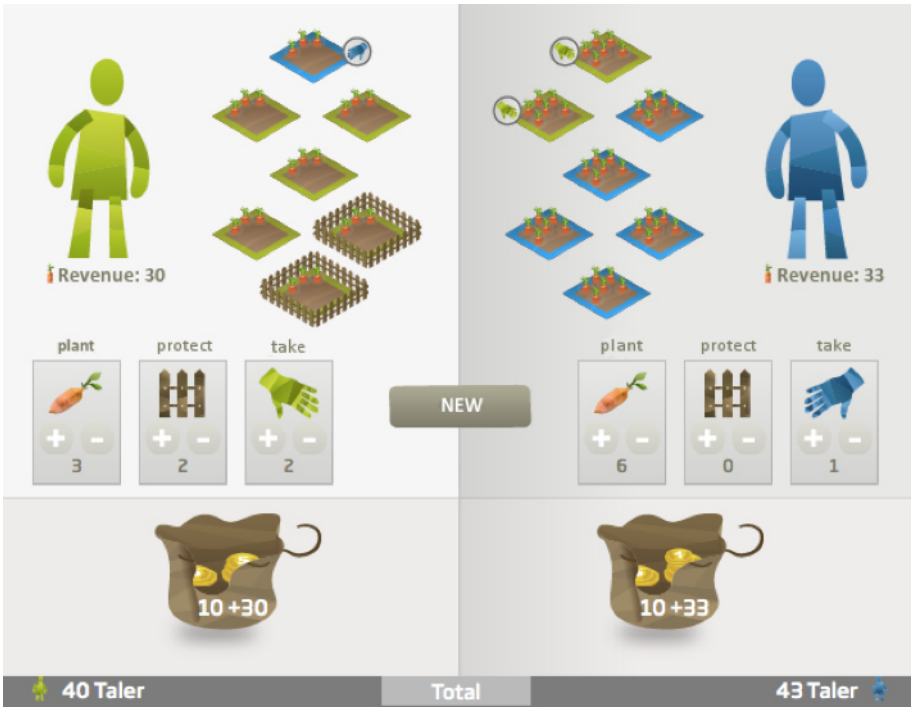

Figure 1.2: Screen shot of the property rights game.

Note. The figure depicts a situation where the player to the left allocates three units to production, two to protection, and two to stealing (take). The player to the right allocates six units to production and one to stealing. Initial endowment is 10 Taler (ECU) for both players. The final payoffs are 40 Talers for player to the left and 43 Talers for the player to the right. ECU = experimental currency unit.

The marginal return of stealing depends on the allocation of resources of the other player. If the other player allocates all her resources to production, then stealing a field yields seven ECU. In case the other participant does not allocate any resources to production, his fields are empty and stealing provides no benefit.

Under standard assumptions, a player's best response is to allocate all her resources to production and stealing if the other player does not protect and allocates less than six units to stealing. If the other player allocates six or more units to stealing, then the best response is to protect three or four fields and use the remaining resources for production. In case the other player protects (and allocates less than six units to stealing), there is a unique best response to allocate all the resources to production. In brief, a player protects if she expects excessive stealing. She allocates all resources to production if she expects

some (minimal) degree of protection, and she is indifferent between stealing and production if the other player does not protect at all. Nash equilibria in this game are all situations in which both players allocate all their resources to production and stealing, and do not steal more than five units. Payoffs in the Nash equilibria range from 14 in the situation where both players steal five units to 49 with no stealing. Coordinating on the Nash equilibrium without stealing and without protection is desirable because it maximizes total surplus. Protection cannot be part of an equilibrium, because player i's best reply to protection by player j is not to steal. But if i does not steal, then it is not the best reply for j to protect.

Our design of the property rights game is similar to Ahn et al. (2016) who study a repeated game in a laboratory environment. Related are also Wilson, Jaworski, Schurter, and Smyth (2012) and Kimbrough, Smith, and Wilson (2010), who show in a series of production and consumption experiments that property rights are respected most when they emerge as informal agreements within civil minded groups.

Donation Decision

The third decision task provides us with a measure for altruism. In the final task of the behavioral experiment, participants are asked to indicate how much of their potential earnings they would be willing to donate to a charity of their choice. Participants' earnings depend on random events and decisions of others, therefore participants do not know their earnings when making this decision. We ask them to indicate the percentage of their potential earnings they are willing to donate. In addition, participants choose a charity from a set of four charities (World Wildlife Fund (WWF), Amnesty International, Red Cross, Doctors Without Borders) or indicate another charitable organization of their choice.

1.4 Results

In total, 252 respondents participated in the behavioral experiment, of whom 55.6% are women. This compares with 50.4% women in the full German WVS sample. Participants are on average 7 years younger in the behavioral experiment sample (42.5, SD =

15.2), compared with the WVS sample (49.5, $SD = 17.7$).[14] We address the issue of self-selection bias in the section "Controlling for Sample Selection Bias." Over the course of the behavioral experiment, dropout leaves us with 179 participants who completed the entire behavioral experiment. Women finish the experiment more often than men do ($p = .056$, χ^2-test). Apart from that, we find no systematic differences between dropouts and finishers.

Table B1 in the appendix provides an overview of the zero-order correlations of all our predictor and outcome variables. Emancipative values correlate positively with secular values ($r = .273$, $p < .001$) and trust ($r = .077$, $p < .001$). Moreover, secular values correlate negatively with civic norms ($r = .527$, $p < .001$). Finally, trust and civic norms correlate only weakly with each other ($r = .041$, $p = .067$). We present our results starting with the public goods game, followed by the property rights game, and finally the donation decision.

1.4.1 Contributions to the Public Good

A total of 252 respondents participated in the public goods game. The mean contribution is 55.1 out of 100, with a standard deviation of 25.4. Figure 1.3 shows the histogram of the contributions. Only four participants (1.6%) contributed zero. Compared to other public goods games, the proportion of minimum contributors is surprisingly low in our data. In an online experiment in Denmark, with a randomly selected sample from the population, Thöni et al. (2012) observe that 15% of the participants contribute zero. However, the average contribution is lower than in the Danish sample, where participants on average contribute 70% of their endowment.

[14] For the emancipative values, the range in the World Values Survey (WVS) sample is [0, 0.875] and the standard deviation is 0.156 compared with a range of [0.13, 0.875] and a standard deviation of 0.132 in the sample of the participants of the behavioral experiment. For the secular values, the range in the WVS sample is [0.0279, 0.943] and the standard deviation is 0.154; in the sample of the behavioral experiment, the range is [0.110, 0.874] and the standard deviation is 0.145.

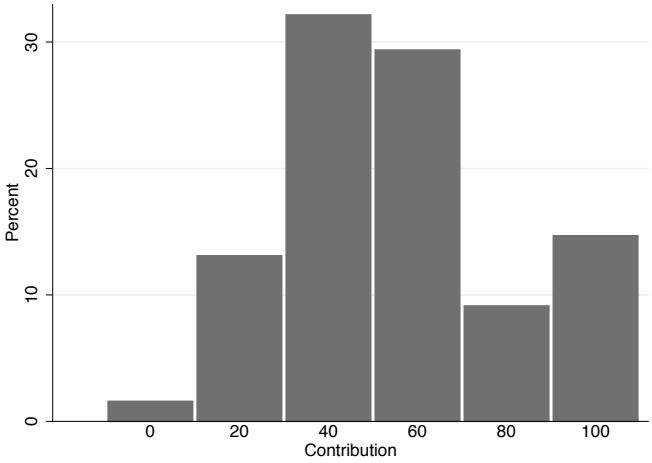

Figure 1.3: Histogram of contributions in the public goods game.

To investigate our hypotheses, we conduct multiple regression analyses. We use ordinary least squares (OLS) estimates with robust standard errors[15] to regress a participant's contribution on her or his values, trust, civic norms, and demographic characteristics. Because of missing responses in some of the predictor variables, we lose some observations ($N = 245$). Model 1 shows the link between emancipative and secular values on one hand and the contribution to the public good on the other hand. For better comparability of the estimates, we normalized all nonbinary predictor variables. Increasing emancipative values by one standard deviation weakly significantly increases the contributions by 3.5 units ($t = 1.84$, $p = .066$). Secular values are not significantly linked with the contribution decision. Due to multicollinearity between the predictor variables, the relationship between emancipative values and the contribution to the public good is underestimated. Indeed, a simple linear regression using emancipative values as the only explanatory variable reveals a significant link with the contribution decision ($\beta = 4.23$, $t = 2.31$, $p = .021$). These results provide evidence for the predictions by Welzel (2013): People with stronger emancipative values contribute more to the public good. Model 2 in Table 1.1 depicts the results for the social capital variables. In line with the results by Thöni et al. (2012), we find that trust is positively related to contributions. Our estimates suggest that people who indicate that "most people can be trusted" contribute roughly seven units more than

[15]In all our regression models, we use Huber-White robust standard errors. The significance levels of our results do not change if we estimate the models without robust standard errors.

participants who state that "one cannot be too careful in dealing with other people" ($t = 2.20$, $p = .029$). However, people indicating stronger civic norms do not seem to contribute more than people indicating weak civic norms. Comparing the overall fit of Models 1 and 2, we find that the values explain more variance than the social capital variables. In Model 3, we estimate the link between values and cooperation controlling for the social capital measures. All coefficients become insignificant, which is presumably due to the positive correlation in the predictor variables. In Model 4, we include further control variables. Aside from the obvious controls for gender and age, we add two control variables capturing general political interest and ideological orientation. We do not find significant effects of these control variables, while trust and emancipative values remain insignificant.[16]

Regarding our subhypotheses, we do not find evidence that participants from West Germany behave more cooperatively than participants from East Germany (Mann-Whitney, $z = 0.179$, $p = .858$). In accordance with our hypothesis, we observe that Protestants contribute significantly more than Catholics (58.2 vs. 47.5, Mann-Whitney, $z = 2.19$, $p = .029$). To conclude, we observe a positive relation between emancipative values and contributions in a simple regression, whereas in multiple regressions the coefficient is at best weakly significant. For secular values, we do not find a significant link with cooperation.

1.4.2 Production and Protection in the Property Rights Game

In total, 188 participants completed the property rights game. Figure 1.4 shows the histograms of the allocation of the seven resource units. On average, participants allocate 4.37 units (62%) to production; 1.94 units (28%) to protection, and 0.7 units (10%) to stealing.[17]

[16] Alternatively, we estimated Model 1 including the Big Five personality measures. In the WVS, the Big Five are measured using Muck, Hell, and Gosling (2007). Previous research (Lu & Argyle, 1991; Ashton, Jackson, Helmes, & Paunonen, 1998) finds positive links to cooperation for extraversion and agreeableness and a negative link for neuroticism. Volk, Thöni, and Ruigrok (2011) find that participants who score high on agreeableness are more likely to have stable cooperation preferences. In our data, we do not find any significant relationship between cooperation and personality traits. Similarly, we included measures for ingroup and outgroup trust in the estimation and find no significant links to contribution. Finally, we check whether our results are driven by a sense of existential security. A set of questions (V170-V191) from the WVS elicits people's feelings of safety. Several of these items load on a first principal component (existential security). We find that our estimates are robust to the inclusion of this variable. See Table B2 in the appendix.

[17] Our results are comparable to Ahn et al. (2016). They report that participants use on average 43% of their resources for production, 29% for protection, and 28% for stealing. The design of the game differed

Table 1.1: OLS Estimations for the Contribution to the Public Good.

	(1) Contribution	(2) Contribution	(3) Contribution	(4) Contribution
Emancipative values	3.498$^{\#}$		2.770	3.007
	[-0.238, 7.234]		[-1.017, 6.557]	[-0.937, 6.951]
Secular values	2.701		3.691	3.268
	[-0.970, 6.371]		[-0.756, 8.138]	[-1.328, 7.863]
Trust		7.039*	4.954	4.516
		[0.725, 13.353]	[-1.442,11.350]	[-2.252, 11.284]
Civic norms		−0.525	1.820	1.364
		[-4.179, 3.129]	[-2.456, 6.096]	[-3.162, 5.890]
Female				−3.126
				[-10.166,3.914]
Age				0.051
				[-0.162,0.265]
Political interest				−0.254
				[-4.004,3.496]
Left-right				−1.383
				[-5.708,2.942]
Constant	52.061**	51.274**	50.016**	49.807**
	[48.641,55.480]	[47.063,55.485]	[45.717,54.316]	[37.985,61.630]
F-test	3.7	2.4	2.8	1.8
Prob $> F$	0.025	0.092	0.025	0.086
R^2	0.029	0.020	0.045	0.051
N	246	245	245	245

Note. OLS estimates. Dependent variable is contribution in the public goods game [0, 20, 40, 60, 80, 100]. Independent variables are standardized except for age, female and trust. Female and trust are dummy variables. Political interest measures how interested a person is in politics. Left-right indicates where a person positions herself with regards to left and right on a scale from 0 to 10, lower numbers indicating left, higher numbers right. Robust standard errors in parentheses. $^{\#}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

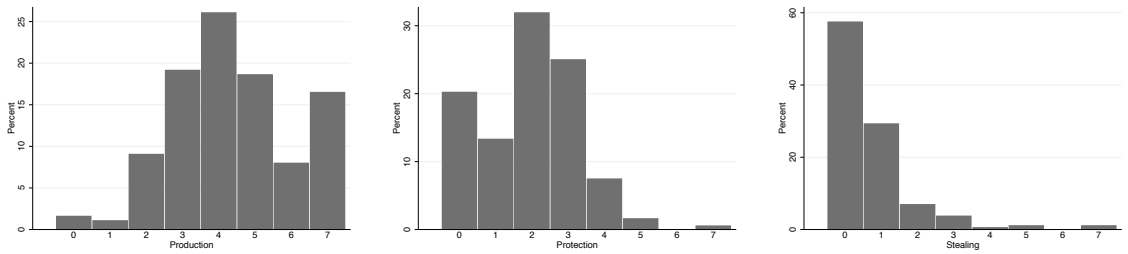

Figure 1.4: Allocation of the resources in the property rights game.

In the following analysis, we focus on production and protection. As participants have to distribute all seven units to the three activities, stealing is merely the residual. Table 1.2 shows regressions with robust standard errors to analyze the relationship between production and our covariates from the WVS. We find a strong relationship between secular values and the production decision, whereas emancipative values do not associate with this decision (Model 1). Participants who score one standard deviation higher on secular values allocate 0.33 units more to production ($t = 2.47$, $p = .014$). Similar to public good contributions, production decisions are positively linked with trust, but not with civic norms (Model 2). A person who trusts allocates roughly 0.6 units more to production ($t = 2.42$, $p = .016$). In Model 3, we include both sets of explanatory variables and find that the link between trust and the production decision weakens, whereas for secular values the link retains its strength. Hence, secular values account for additional variance above and beyond social capital. Effect sizes and significance levels are robust to the inclusion of political and demographic controls (Model 4), and the controls itself are insignificant.[18] Concerning our two subhypotheses, we find that West and East Germans do not differ significantly with respect to production (Mann-Whitney, $z = 1.34$, $p = .179$). However, we observe significant differences across religious groups. Protestants allocate on average 4.5 units to production, versus 3.8 units for Catholics (Mann-Whitney, $z = 1.99$, $p = .046$).

To sum up, secular values show robust predictive power for the production decision, whereas emancipative values seem unrelated. People who trust and attribute less value to sacred authority are more likely to realize efficient outcomes. We take these results as an

on some important dimensions. Most importantly, they implemented a repeated game with 10 periods.

[18] Among the additional sets of covariates, we find that ingroup trust and openness to experience are positively related to the production decision, while all other factors are insignificant (see Table B3 in the appendix).

Table 1.2: OLS Estimates for Production.

	(1) Production	(2) Production	(3) Production	(4) Production
Emancipative values	0.163		0.090	0.087
	[-0.148, 0.474]		[-0.239, 0.419]	[-0.256, 0.430]
Secular values	0.329*		0.356*	0.355*
	[0.066, 0.591]		[0.058, 0.655]	[0.043, 0.668]
Trust		0.591*	0.474#	0.436#
		[0.109, 1.072]	[-0.033, 0.981]	[-0.070, 0.943]
Civic norms		−0.140	0.050	0.024
		[-0.401, 0.120]	[-0.249, 0.350]	[-0.295, 0.342]
Female				−0.041
				[-0.579, 0.498]
Age				0.007
				[-0.009, 0.023]
Political interest				0.028
				[-0.251, 0.306]
Left-right				0.039
				[-0.230, 0.308]
Constant	4.187**	4.069**	4.001**	3.743**
	[3.907, 4.468]	[3.759, 4.378]	[3.685, 4.318]	[2.901, 4.584]
F-test	4.5	3.5	3.6	2.1
Prob $> F$	0.012	0.033	0.007	0.039
R^2	0.047	0.037	0.071	0.076
N	183	182	182	182

Note. OLS estimates. Dependent variable is units allocated to production in the property rights game [0,7]. Independent variables are standardized except for age, female and trust. Female and trust are dummy variables. Political interest measures how interested a person is in politics. Left-right indicates where a person positions herself with regards to left and right on a scale from 0 to 10, lower numbers indicating left, higher numbers right. Robust standard errors in parentheses. # $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

indication that secularization is compatible with desirable social outcomes.

In the next step, we analyze the protection decision. As mentioned above, devoting resources to protection is not part of an equilibrium. However, if excessive stealing is expected, allocating three or four units to protection is a rational response. As shown in Figure 1.4, we observe that 80% of participants allocate at least one unit to protection. The multiple regression analyses in Table 1.3 show the inverse pattern to that of the production behavior. In Model 1, we look at the link between values and the protection decision. Participants, who score high on secular values, use significantly less resources for protection. We find that an increase of secular values by one standard deviation leads to a decrease of protection of roughly 0.35 units ($t = 2.93$, $p = .004$). In Model 2, trust is weakly significantly linked to protection ($t = 1.66$, $p = .099$). In Model 3, we include both sets of variables and the social capital measures become insignificant, whereas secular values retain strength and significance. The same holds if we include the additional controls (Model 4).[19]

While we find that Catholics and Protestants behave differently with respect to production, we do not observe such differences in the protection decision (Mann-Whitney, $z = 1.22$, $p = .224$). Furthermore, we do not find evidence for differences in the behavior between West and East German participants (Mann-Whitney, $z = 0.623$, $p = .533$). To conclude, we observe that behavior in the property rights game is tightly associated with secular values. Participants who score high on those values allocate on average more resources to production and fewer to protection. This confirms our hypothesis about the negative link between secular values and protectionist behavior.

Consequently, the higher the score on secular values, the closer is a participant's behavior to the socially optimal behavior. Interestingly, while we observe that secular values relate positively to production and negatively to protection, they are not significantly linked to stealing. Behavior in this game might be strongly driven by the beliefs participants hold about their partner. People who emphasize secular values find little appeal in sacrosanct authority, which is usually idolized as a force of order in a presumably chaotic

[19]The link between the secular values and the protection decision is robust to the inclusion of the additional control variables discussed in the appendix, while none of the controls is significant (see Table B4).

Table 1.3: OLS Estimates for Protection.

	(1) Protection	(2) Protection	(3) Protection	(4) Protection
Emancipative values	−0.013		0.026	0.016
	[-0.273, 0.248]		[-0.246, 0.298]	[-0.260, 0.293]
Secular values	−0.348**		−0.355**	−0.348*
	[-0.581, -0.114]		[-0.614, -0.096]	[-0.615, -0.081]
Trust		−0.325#	−0.263	−0.260
		[-0.713, 0.062]	[-0.670, 0.143]	[-0.678, 0.158]
Civic norms		0.174	−0.006	0.017
		[-0.037, 0.384]	[-0.239, 0.227]	[-0.224, 0.259]
Female				−0.036
				[-0.469, 0.398]
Age				−0.007
				[-0.020, 0.006]
Political interest				0.062
				[-0.148, 0.273]
Left-right				−0.020
				[-0.229, 0.190]
Constant	2.034**	2.105**	2.135**	2.419**
	[1.810, 2.258]	[1.857, 2.352]	[1.877, 2.393]	[1.727, 3.112]
F-test	5.3	2.8	3.6	1.9
Prob $> F$	0.006	0.066	0.008	0.060
R^2	0.059	0.029	0.072	0.078
N	183	182	182	182

Note. OLS estimates. Dependent variable is units allocated to protection [0,7]. Independent variables are standardized except for age, female and trust. Female and trust are dummy variables. Political interest measures how interested a person is in politics. Left-right indicates where a person positions herself with regards to left and right on a scale from 0 to 10, lower numbers indicating left, higher numbers right. Robust standard errors in parentheses. $^{\#}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

and dangerous world. Hence, adherents of secular values might generally be more relaxed and less fearful, and expect less stealing from others. On the contrary, strong secular values do not seem to be related to a participant's own "immoral" behavior.

1.4.3 Donations

We observe behavior in the donation decision for 179 participants. On average, participants are willing to donate 14.4% of their potential income to a charity ($SD = 20.9$). A bit more than half of the participants (54%) are willing to donate. Among the participants who donate, the mean donation is 26.6% ($SD = 21.9$). In a design very similar to ours (with a student sample), Schulz, Thiemann, and Thöni (2015) find about 44% willing to donate and 16.3% of the earnings donated on average.

Table 1.4 depicts the results of the OLS regressions for the link between values and donations. In Model 1, we regress the amount donated (as percentage) on emancipative and secular values. We find that emancipative values are strongly related to the donation decision. An increase in emancipative values by one standard deviation is associated with an increase in donations by 5.7 percentage points ($t = 3.05$, $p = .003$). In Model 2, we regress the amount donated on the social capital measures. Both predictors account for some variance of the dependent variable. Participants who trust donate on average 7.1 percentage points more ($t = 2.31$, $p = .022$). An increase of one standard deviation in civic norms leads to an increase in the amount donated of about 5 percentage points ($t = 5.00$, $p < .001$). In Model 3, we include both sets of predictors and find that emancipative values and civic norms retain significance, whereas trust becomes insignificant. Compared with Model 2, the adjusted R-squared increases. Model 4 shows that the link between emancipative values and the donation decision is robust to the inclusion of our control variables ($t = 2.17$, $p = .031$). Unlike in the previous estimates, we observe a significant link between age and donation, with older participants being more generous ($t = 2.73$, $p = .007$). Furthermore, stronger interest in politics seems to be positively related to donations ($t = 1.87$, $p = .063$).[20]

[20]Table B5 in the appendix reports the estimates for extended controls. The coefficient for emancipative values is robust in all specifications, while none of the other predictor variables (Big Five, in/outgroup trust, perceived security) reaches significance.

Table 1.4: OLS Estimates for Donation.

	(1) Donation	(2) Donation	(3) Donation	(4) Donation
Emancipative values	5.723**		4.864*	4.119*
	[2.013, 9.432]		[0.784, 8.943]	[0.373, 7.865]
Secular values	−2.135		0.971	0.996
	[-5.632, 1.362]		[-3.007, 4.948]	[-2.780, 4.771]
Trust		7.159*	4.366	2.323
		[1.046, 13.271]	[-2.241,10.974]	[-4.684, 9.330]
Civic norms		5.246**	6.255**	4.634**
		[3.173, 7.319]	[3.337, 9.174]	[1.662, 7.605]
Female				−3.195
				[-9.844, 3.453]
Age				0.315**
				[0.087, 0.544]
Political interest				2.657#
				[-0.146, 5.460]
Left-right				−2.952
				[-6.580, 0.677]
Constant	12.089**	11.398**	10.025**	−0.509
	[9.384, 14.794]	[6.997, 15.799]	[6.128, 13.923]	[-10.260, 9.241]
F-test	4.7	15.8	8.7	5.4
Prob $> F$	0.011	0.000	0.000	0.000
R^2	0.049	0.081	0.118	0.209
N	174	173	173	173

Note. OLS estimates. Dependent variable is the amount donated to a charity of choice in percentage of the earnings [0, 100]. Independent variables are standardized except for age, female and trust. Female and trust are dummy variables. Political interest measures how interested a person is in politics. Left-right indicates where a person positions herself with regards to left and right on a scale from 0 to 10, lower numbers indicating left, higher numbers right. Robust standard errors in parentheses. # $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

We do not observe a significant difference in the donation behavior between East and West German participants (Mann-Whitney, $z = 0.98$, $p = .327$). Nor do we find a significant difference between Protestants and Catholics (Mann-Whitney, $z = 0.96$, $p = .337$).

Overall, the effect size of emancipative values on donation is in the magnitude of an increase between 4.4 and 5.7 percentage points per standard deviation. If we estimate the link between the donation decision and values only for the sample of people who donate, we observe an increase of the effect size to roughly 7 percentage points ($t = 3.07$, $p = .003$).

Summing up, both emancipative values and adherence to civic norms associate positively with the amount donated. If one follows Welzel (2010) and defines "civicness" as people's willingness to cooperate with strangers and to share with unfamiliar others, then

our results provide evidence for the civic implications of certain sets of moral values. Specifically, we see from the donation decisions that people with strong emancipative values share more with strangers compared with what people with weak emancipative values do.

1.4.4 Controlling for Sample Selection Bias

At the beginning of the "Design and Procedures" section, we mentioned that women and young people are overrepresented in our experimental sample compared with the WVS sample. To control for selection bias, we follow Heckman (1979). The first step of the Heckman correction is to estimate a probit model for the individual probability to belong to the selected sample (selection equation). Second, we use the estimated coefficient of the selection equation to predict the inverse Mill's ratio. The final step is to estimate the original regression using the inverse Mill's ratio as a regressor.

It is crucial to this method to identify the selection equation. In practice, this means that the selection equation has to include at least one variable, which matters for selection but is not related to the outcome variable (exclusion restriction). Ideally, this variable does not come from the respondent's answers or characteristics. Fortunately, we have a variable, which strongly predicts participation but is, most likely, unrelated to the outcome variables. The variable measures an interviewer's success in motivating the respondents of the WVS in participating in the behavioral experiment.

The roughly 2,000 WVS interviews were conducted by 134 interviewers. For each interviewer, we calculate an individual success rate as the total number of email addresses collected for the behavioral experiment, divided by the total number of interviews conducted. Interviewers differed a lot in their success rate (mean = 0.24, $SD = 0.22$, minimum = 0, maximum = 1). This variable is highly related to selection into the sample of the behavioral experiment, but unlikely to be related to the dependent variables of interest. Participants who indicated their email address are more likely to participate in our behavioral experiment because we were able to send them a direct link to access our website. However, the success rate of the interviewer is not likely to be related, for example, to contributions in the public goods game.[21] We reestimated all our models from Tables 1.1

[21] We do not observe any statistically significant correlation between our dependent variables and the

to 1.4 using the Heckman procedure. All our main links remain unchanged, which suggests that the results are not driven by selection bias.[22]

1.5 Discussion

We study the individual-level linkage between secular and emancipative values, on one hand, and prosocial behavior, on the other hand, as observed in a nationally representative sample. As concerns secular and emancipative values, we have shown four relations, most of which proved robust against controls for sampling bias, demographic characteristics, public interest, ideological orientations, as well as personality traits and social capital. To begin with, emancipative values seem to relate to (a) larger contributions to the public good and (b) relate strongly to more generous donations to charities. Both relations involve decisions about sharing some part of one's endowment with strangers. But while mutual cooperation produces a personal gain, donating is clearly a sacrifice. The donation decision is, hence, more indicative of altruism, and this is where the link with the emancipative values is strongest. Apparently, the behavioral impulse of these values is more powerful when it comes to decisions about true material sacrifice than in situations that still involve opportunities for material gain. Hence, we confirm Welzel (2010) hypothesis that emancipative values correspond with universal altruism—at least for the German context.

Secular values, for their part, show no behavioral impulse in situations in which social sharing is the issue. They show such an impact, however, in the property rights game, in which the success of a strategy depends on what strategy the others have chosen. Specifically, stronger secular values correspond with (c) more productive investments and at the same time (d) less protective investments. Although these decisions do not involve a joint project, they nevertheless depend on beliefs about the other player's strategy. From this point of view, secular values involve lower fears that strangers behave in an asocial manner, for which reason these values contribute to investment decisions conducive to the social optimum. Despite their significance, the behavioral impulses of secular values proved to

individual success rate of the interviewer. Hence, it seems that the individual success is a valid exclusion restriction.

[22] For the results of the selection models, see the working paper version (Kistler, Thöni, & Welzel, 2015).

be considerably weaker than those of emancipative values, especially in the case of altruism. Our results also support the notion that the positive relation between religiosity and prosociality found in self-reported data does not carry over to behavioral data (Shariff, 2015).

Moreover, our analysis shows that one standard measure of social capital associates with observed behavior while another one does not, or only very weakly so. Specifically, interpersonal trust is linked with larger contributions in the public goods game and bigger investments into production in the property rights game. By contrast, adherence to civic norms is only associated with the donation decision. The lack of relevance of adherence to civic norms confirms Welzel (2010) criticism that this measure often does not keep to the promise of its expected prosocial effects, presumably because it rather captures social desirability than truly internalized civic orientations.

Historically rooted cleavages that characterize the cultural map of Europe, namely, the East-West cleavage and the Protestant-Catholic cleavage did not prove overly important in the German context. The only effect we found with respect to these historic cleavages is that Protestants contribute more to the public good and allocate more units to production in the property rights game than Catholics.

To the best of our knowledge, this is the first study to establish a direct individual-level link between secular/emancipative values and behavior, with respect to observed—not self-reported—behavior. This was possible through a methodological innovation, combining a national representative survey in Germany with a behavioral experiment on three facets of prosocial behavior. The fact that we reached empirically significant and theoretically meaningful results shows that this innovation works and can be extended to other societies with a sufficient degree of computer literacy.

As Germany represents Europe's biggest nation and the fourth-largest economy in the world, the results are relevant in and by themselves. But Germany is also in many respects a typical postindustrial knowledge society, for which reason comparable results might well be obtainable in other such societies, from the United States to Australia to Japan, and even beyond.

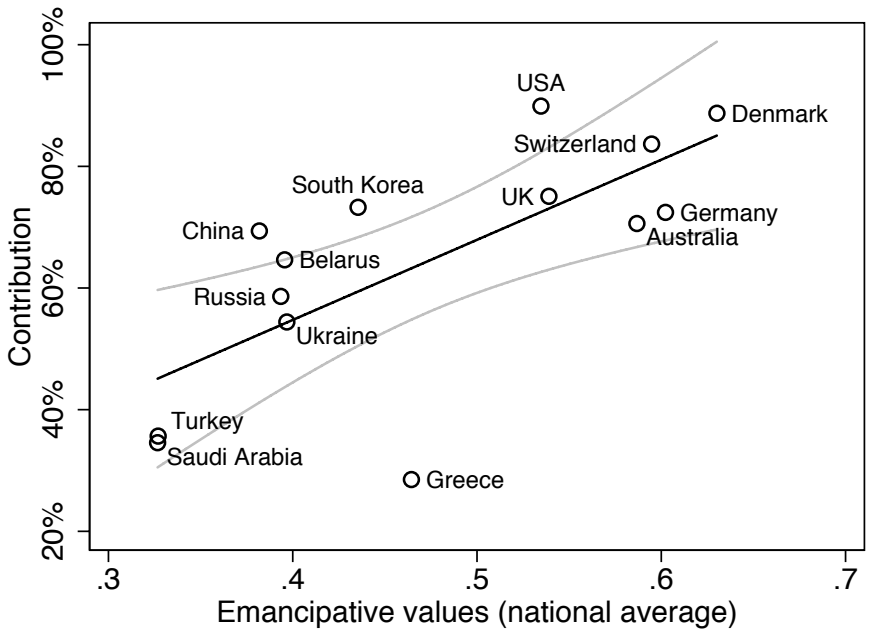

Figure 1.5: Emancipative values and experimentally measured cooperative behavior at the macro level.

Source. Emancipative values: World Values Survey and own calculations. Contributions: Herrmann et al. (2008), Figure 2A.

Preliminary evidence that our approach and findings, especially concerning the link between values and cooperation, might be valid in a wider cross-cultural context than just Germany is provided in Figure 1.5. It plots data from laboratory experiments conducted with student subject pools in 14 countries around the globe (Herrmann et al., 2008), in relation to the average emancipative values in the respective national population, as measured by the WVS. Although this plot only covers 14 different country locations, the relationship is highly significant (standardized linear regression: $\beta = .71$, $t = 3.65$, $p = .003$) and explains a substantial part of the variance in cooperation ($R^2 = .51$).[23] Much like in the behavioral experiment we propose in this article, cooperation shown in Figure 1.5 was measured in a controlled and incentivized environment, following a standardized procedure in all countries. Differences in the observed individual-level behavior can be explained by country-level differences in the cultural environment of the participants. Presumably, the differences in emancipative values as measured by the WVS capture important determinants for cooperative outcomes in social dilemma situations. Against this backdrop, our

[23]The relation between contributions and emancipative values is surprisingly robust to the inclusion of control variables such as trust, the human development index, or a measure for democracy.

German pilot study is truly essential as it tells us that this macro-level link does not seem to be a mere artifact of aggregation but has a valid micro-foundation at the individual level—at least at one of the places covered.[24]

This suggestive evidence calls even more loudly for a replication of our study at other places. To conclude, we are convinced that the combination of behavioral experiments with representative surveys provides a novel and powerful tool to establish mind-set-behavior links in a more valid way than is possible by either one method alone. We see our study as a promising first step to extend this research design to more countries covered in Round 7 of the WVS. This would allow to examine whether and how the individual-level linkages between moral values and cooperative behavior vary by cultural background and country-level characteristics.

References

Ahn, T. K., Balafoutas, L., Batsaikhan, M., Campos-Ortiz, F., Putterman, L., & Sutter, M. (2016). Securing property rights: A dilemma experiment in Austria, Mexico, Mongolia, South Korea and the United States. *Journal of Public Economics*, *143*, 115–124.

Ashton, M. C., Jackson, D. N., Helmes, E., & Paunonen, S. V. (1998). Joint factor analysis of the personality research form and the jackson personality inventory: Comparisons with the big five. *Journal of Research in Personality*, *32*(2), 243–250.

Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, *80*(4), 1095–1111.

Balliet, D. & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112.

Bardi, A. & Schwartz, S. H. (2003). Values and behavior: Strength and structure of relations. *Personality and Social Psychology Bulletin*, *29*(10), 1207–1220.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216.

[24]For a discussion on how behavioral experiments allow to investigate the influence of the "macrosocial" environment on individual behavior, see also Gächter and Thöni (2011).

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Camerer, C. F. & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In J. Henrich, R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, H. Gintis, & R. McElreath (Eds.), *Foundations of human sociality: Experimental and ethnographic evidence from 15 small-scale societies* (97). Oxford, UK: Oxford University Press.

Camerer, C. F. & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1-3), 7–42.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.

Eckstein, H. (1961). *A theory of stable democracy*. Princeton, NJ: Princeton University Press.

Eckstein, H. (1998). Congruence theory explained. In R. Ahl, R. Bova, P. G. Roeder, & H. Eckstein (Eds.), *Can democracy take root in post-soviet russia?* (pp. 3–34). Lanham, MD: Rowman & Littlefield Publishers.

Gächter, S., Herrmann, B., & Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: Survey and experimental evidence. *Journal of Economic Behavior & Organization*, *55*(4), 505–531.

Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B*, *365*, 2651–2661.

Gächter, S. & Thöni, C. (2011). Micromotives, microstructure and macrobehavior: The case of voluntary cooperation. *Journal of Mathematical Sociology*, *35*(1-3), 26–65.

Galen, L. W. (2012). Does religious belief promote prosociality? A critical examination. *Psychological Bulletin*, *138*(5), 876–906.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, *115*(3), 811–846.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–135.

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

Hui, C. H. & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology*, *16*(2), 131–152.

Inglehart, R. & Norris, P. (2003). *Rising tide: Gender equality and cultural change around the world*. New York, NY: Cambridge University Press.

Inglehart, R. & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. New York, NY: Cambridge University Press.

Inkeles, A. (1969). Participant citizenship in six developing countries. *American Political Science Review*, *63*(4), 1120–1141.

Inkeles, A. (1983). *Exploring individual modernity*. New York, NY: Columbia University Press.

Kimbrough, E. O., Smith, V. L., & Wilson, B. J. (2010). Exchange, theft, and the social formation of property. *Journal of Economic Behavior & Organization*, *74*(3), 206–229.

Kistler, D., Thöni, C., & Welzel, C. (2015). Survey response and observed behavior: Emancipative and secular values predict cooperation, protection of property, and pro-social behavior. *World Values Research Paper*, (8/3).

Kluckhohn, C. (1951). Values and value-orientations in the theory of action: An exploration in definition and classification. In P. T & E. Shils (Eds.), *Towards a general theory of action* (pp. 388–433). Cambridge, MA: Harvard University Press.

Knack, S. & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics*, *112*(4), 1251–1288.

Kuran, T. (1997). *Private truths, public lies: The social consequences of preference falsification*. Cambridge, MA: Harvard University Press.

Ledyard, J. et al. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Lu, L. & Argyle, M. (1991). Happiness and cooperation. *Personality and Individual Differences*, *12*(10), 1019–1030.

Muck, P. M., Hell, B., & Gosling, S. D. (2007). Construct validation of a short five-factor model instrument: A self-peer study on the german adaptation of the ten-item personality inventory (tipi-g). *European Journal of Psychological Assessment*, *23*(3), 166–175.

Norris, P. & Inglehart, R. (2004). *Sacred and secular: Religion and politics worldwide*. New York, NY: Cambridge University Press.

North, D., Wallis, J. J., & Weingast, B. R. (2009). Violence and social orders. *Cambridge: Cambridge UP*.

Osborne, M. J. & Rubinstein, A. (1994). *A course in game theory*. Cambridge, MA: MIT Press.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of american community*. New York, NY: Simon and Schuster.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making democracy work: Civic traditions in modern italy*. Princeton, NJ: Princeton University Press.

Rokeach, M. (1968). Beliefs, attitudes and values: A theory of organization and change.

Rokeach, M. (1973). *The nature of human values*. New York: Free press.

Roux, C. & Thöni, C. (2015). Do control questions influence behavior in experiments? *Experimental Economics*, *18*(2), 185–194.

Sagiv, L., Sverdlik, N., & Schwarz, N. (2011). To compete or to cooperate? values' impact on perception and action in social dilemma games. *European Journal of Social Psychology*, *41*(1), 64–77.

Schulz, J. F., Thiemann, P., & Thöni, C. (2015). Defaults in charitable giving. CeDEx Discussion Paper No. 2015-06.

Schwartz, S. H. (2004). Mapping and interpreting cultural differences around the world. *International Studies in Sociology and Social Anthropology*, 43–73.

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology*, *5*(2), 137–182.

Schwartz, S. H. (2007). Value orientations: Measurement, antecedents and consequences across nations. *Measuring Attitudes Cross-nationally: Lessons from the European Social Survey*, 161–193.

Shariff, A. F. (2015). Does religion increase moral behavior? *Current Opinion in Psychology*, *6*, 108–113.

Smith, V. L. (1976). Induced value theory. *American Economic Review*, *66*(2), 274–279.

Thöni, C., Tyran, J.-R., & Wengström, E. (2012). Microfoundations of social capital. *Journal of Public Economics*, *96*(7-8), 635–643.

Volk, S., Thöni, C., & Ruigrok, W. (2011). Personality, personal values and cooperation preferences in public goods games: A longitudinal study. *Personality and Individual Differences*, *50*(6), 810–815.

Welzel, C. (2010). How selfish are self-expression values? A civicness test. *Journal of Cross-Cultural Psychology*, *41*(2), 152–174.

Welzel, C. & Delhey, J. (2015). Generalizing trust: The benign force of emancipation. *Journal of Cross-Cultural Psychology*, *46*(7), 875–896.

Welzel, C. & Inglehart, R. (2010). Agency, values, and well-being: A human development model. *Social Indicators Research*, *97*(1), 43–63.

Welzel, C. & Inglehart, R. F. (2016). Misconceptions of measurement equivalence: Time for a paradigm shift. *Comparative Political Studies*, *49*(8), 1068–1094.

Welzel, C., Inglehart, R., & Deutsch, F. (2005). Social capital, voluntary associations and collective action: Which aspects of social capital have the greatest 'civic' payoff? *Journal of Civil Society*, *1*(2), 121–146.

Welzel, C., Inglehart, R., & Klingemann, H.-D. (2003). The theory of human development: A cross-cultural analysis. *European Journal of Political Research*, *42*, 341–379.

Wilson, B. J., Jaworski, T., Schurter, K. E., & Smyth, A. (2012). The ecological and civil mainsprings of property: An experimental economic history of whalers' rules of capture. *Journal of Law, Economics, & Organization*, *28*(4), 617–656.

Appendix

1.A Items and Response Format

Emancipative Values

1. Autonomy Subindex:

    - Here is a list of qualities that children can be encouraged to learn at home. Which, if any, do you consider to be especially important?
        - Independence (mentioned/not mentioned)
        - Imagination (mentioned/not mentioned)
        - Obedience (mentioned/not mentioned)

2. Equality Subindex:

    - Do you agree, disagree or neither agree nor disagree with the following statements? When jobs are scarce, men should have more right to a job than women. *(agree/disagree/neither agree nor disagree)*
    - For each of the following statements I read out, can you tell me how strongly you agree or disagree with each. Do you strongly agree, agree, disagree, or strongly disagree?
        - A university education is more important for a boy than for a girl. *(strongly agree/agree/disagree/strongly disagree)*
        - On the whole, men make better political leaders than women do. *(strongly agree/agree/disagree/strongly disagree)*

3. Choice Subindex:

    - Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between using this card (10-point scale):
        - Homosexuality *(1 [never justifiable] to 10 [always justifiable])*
        - Abortion *(1 [never justifiable] to 10 [always justifiable])*

- Divorce *(1 [never justifiable] to 10 [always justifiable])*

4. Voice Subindex:

   - People sometimes talk about what the aims of this country should be for the next ten years. On this card are listed some of the goals which different people would give top priority. Would you please say which one of these you, yourself, consider the most important? (. . .) And second most important?
     - Giving people more say in important government decisions
       (most important/second most important/not important)
     - Protecting freedom of speech
       (most important/second most important/not important)
     - Seeing that people have more say about how things are done at their jobs and in their communities
       (most important/second most important/not important)

Secular Values

1. Defiance Subindex:

   - People pursue different goals in life. For each of the following goals, can you tell me if you strongly agree, agree, disagree or strongly disagree with it?
     - One of my main goals in life has been to make my parents proud.
       (strongly agree/agree/disagree/strongly disagree)
   - I'm going to read out a list of various changes in our way of life that might take place in the near future. Please tell me for each one, if it were to happen, whether you think it would be a good thing, a bad thing, or don't you mind?
     - Greater respect for authority *(good/don't mind/bad)*
   - How proud are you to be [German]?
     (very proud/quite proud/not very proud/not at all proud/I am not [German])

2. Agnosticism Subindex:

- For each of the following, indicate how important it is in your life.
  - Religion *(very important/rather important/not very important/not at all important)*
- Apart from weddings and funerals, about how often do you attend religious services these days?
  (more than once a week/once a week/once a month/only on special holy days/once a year/less often/never, practically never)
- Independently of whether you attend religious services or not, would you say you are
  (a religious person/not a religious person/an atheist)

3. Relativism Subindex:

   - Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between, using this card.
     - Avoiding a fare on public transport
       (1 [never justifiable] to 10 [always justifiable])
     - Cheating on taxes if you have a chance
       (1 [never justifiable] to 10 [always justifiable])
     - Someone accepting a bribe in the course of their duties
       (1 [never justifiable] to 10 [always justifiable])

4. Skepticism Subindex:

   - I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?
     - The armed forces *(a great deal/quite a lot/not very much/none at all)*
     - The police *(a great deal/quite a lot/not very much/none at all)*
     - The courts *(a great deal/quite a lot/not very much/none at all)*

1.B Additional Tables

Table B1: Summary Table Zero-Order Correlations.

Measure	1	2	3	4	5	6	7	8	9	10	M	SD
1. Contribution	-										55.08	25.4
2. Production	.3243**	-									4.37	1.66
	.000											
	n = 188											
3. Protection	.3219**	−.7304**	-								1.94	1.33
	.000	.000										
	n = 188	n = 188										
4. Donation	.1775*	−.026	.0243	-							14.41	20.88
	.0175	.7297	.7466									
	n = 179	n = 179	n = 179									
5. Emancipative values	.1427*	.127*	−.0677	.2026**	-						0.55	0.16
	.0252	.0866	.3623	.0073								
	n = 246	n = 183	n = 183	n = 174								
6. Secular values	.1281*	.2025**	−.2429**	−.0336	.2732**	-					0.39	0.15
	.0448	.006	.0009	0.66	.000							
	n = 246	n = 183	n = 183	n = 174	n = 2,041							
7. Trust	.1099#	.1571*	−.1063	.1799**	.0765**	−.0297	-				0.42	0.49
	.0817	.0313	.1464	.016	.0005	.179						
	n = 252	n= 188	n = 188	n = 179	n = 2,048	n = 2,043						
8. Civic norms	−.0198	−.0739	.1162	.2277**	−.0073	−.5267**	.0408#	-			9.24	1.18
	.7583	.3215	.1181	.0026	.7427	.000	.0674					
	n = 245	n = 182	n = 182	n = 173	n = 2,013	n = 2,008	n = 2,013					
9. Political interest	.0654	.0649	.0167	.2254**	.1243**	−.0737**	.0924**	.0503*	-		0.59	0.34
	.3072	.3828	.8228	.0028	.000	.0009	.000	.024				
	n = 246	n = 183	n = 183	n = 174	n = 2,047	n = 2,042	n = 2,050	n = 2,012				
10. Political orientation	−.0598	.0126	−.0053	−.123	−.0398#	−.0998**	.0122	.0322	.0953**	-	4.32	2.52
	.3503	.8656	.9436	.1059	.0717	.000	.5814	.1488	.000			
	n = 183	n = 183	n = 183	n = 174	n = 2,048	n = 2,043	n = 2,051	n = 2,013	n = 2,050			

Note. The first row per entry denotes the estimates of the Pearson product-moment correlation coefficient. The second row denotes the significance level. In the third row, we report the number of observations. The mean and the standard deviation are based on the sample of the behavioral experiment for the measures 1 to 4. For the measures 5 to 10, the mean and the standard deviation are based on the complete sample of the World Values Survey. $^{\#}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Table B2: OLS Estimates for Contribution in the Public Goods Game Including Additional Control Variables.

	(1) Contribution	(2) Contribution	(3) Contribution
Emancipative values	2.253	$3.310^{\#}$	$3.424^{\#}$
	[-1.960,6.466]	[-0.519,7.139]	[-0.471,7.319]
Secular values	2.785	2.711	2.579
	[-1.153,6.723]	[-0.943,6.366]	[-1.179,6.337]
Outgroup trust	1.784		
	[-3.262,6.829]		
Ingroup trust	1.672		
	[-2.860,6.203]		
Perceived security		1.912	
		[-1.286,5.110]	
Extraversion			−0.078
			[-3.878,3.722]
Neuroticism			0.506
			[-2.622,3.634]
Openness to experience			0.711
			[-3.002,4.424]
Conscientiousness			−0.260
			[-3.523,3.004]
Agreeableness			−0.836
			[-4.545,2.872]
Constant	52.712^{**}	51.935^{**}	51.900^{**}
	[49.109,56.315]	[48.545,55.324]	[48.387,55.413]
F-test	1.9	3.1	1.3
Prob $> F$	0.106	0.026	0.251
R^2	0.036	0.033	0.032
N	235	246	245

Note. OLS estimates. Dependent variable is contribution in the public goods game [0, 20, 40, 60, 80, 100]. Independent variables are standardized except for perceived security, age, female. Female and trust are dummy variables. Robust standard errors in parentheses. $^{\#}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Table B3: OLS Estimates for Production in the Property Rights Game Including Additional Control Variables.

	(1) Production	(2) Production	(3) Production	(4) Production
Emancipative values	0.083	0.168	0.094	0.167
	[-0.276,0.441]	[-0.149,0.484]	[-0.211,0.398]	[-0.148,0.481]
Secular values	0.364**	0.329*	0.353*	0.280*
	[0.092,0.635]	[0.066,0.592]	[0.085,0.620]	[0.013,0.547]
Outgroup trust	0.012			
	[-0.357,0.381]			
Ingroup trust	0.313*			
	[0.001,0.625]			
Perceived security		−0.031		
		[-0.265,0.203]		
Extraversion			−0.196	
			[-0.471,0.079]	
Neuroticism			−0.144	
			[-0.375,0.088]	
Openness to experience			0.273*	
			[0.032,0.514]	
Conscientiousness			0.082	
			[-0.166,0.330]	
Agreeableness			−0.001	
			[-0.271,0.270]	
T_rich				−0.364
				[-0.922,0.194]
T_poor				0.344
				[-0.274,0.963]
Constant	4.240**	4.188**	4.191**	4.227**
	[3.945,4.536]	[3.908,4.468]	[3.921,4.462]	[3.860,4.593]
F-test	4.2	3.1	3.5	3.6
Prob $> F$	0.003	0.028	0.002	0.008
R^2	0.076	0.048	0.091	0.069
N	175	183	182	183

Note. OLS estimates. Dependent variable is units allocated to production [0,7]. Independent variables are standardized except for perceived security, age, female and trust. Female and trust are dummy variables. T_rich and T_poor are dummy variables for the endowment treatments. T_rich is the treatment in which the respondent is paired with a person who has a lower endowment in the game. T_poor is the treatment in which the respondent is grouped with a person with a higher endowment in the game. Baseline is the treatment where both respondents have the same endowment. Robust standard errors in parentheses. $^{\#}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Table B4: OLS Estimates for Protection in the Property Rights Game Including Additional Control Variables.

	(1) Protection	(2) Protection	(3) Protection	(4) Protection
Emancipative values	0.079	−0.013	−0.005	−0.009
	[-0.221,0.379]	[-0.282,0.255]	[-0.264,0.253]	[-0.267,0.250]
Secular values	−0.391**	−0.348**	−0.314*	−0.327**
	[-0.642,-0.141]	[-0.582,-0.114]	[-0.553,-0.075]	[-0.561,-0.092]
Outgroup trust	−0.092			
	[-0.379,0.194]			
Ingroup trust	−0.115			
	[-0.368,0.138]			
Perceived security		0.006		
		[-0.196,0.207]		
Extraversion			0.133	
			[-0.078,0.344]	
Neuroticism			−0.095	
			[-0.281,0.091]	
Openness to experience			−0.085	
			[-0.273,0.103]	
Conscientiousness			−0.057	
			[-0.248,0.133]	
Agreeableness			0.053	
			[-0.160,0.265]	
T_rich				0.133
				[-0.296,0.563]
T_poor				−0.262
				[-0.759,0.236]
Constant	1.988**	2.034**	2.042**	2.044**
	[1.750,2.225]	[1.810,2.258]	[1.815,2.268]	[1.748,2.341]
F-test	3.4	3.6	2.4	3.2
Prob $> F$	0.011	0.015	0.022	0.014
R^2	0.077	0.059	0.075	0.070
N	175	183	182	183

Note. OLS estimates. Dependent variable is units allocated to protection [0,7]. Independent variables are standardized except for perceived security, age, female and trust. Female and trust are dummy variables. T_rich and T_poor are dummy variables for the endowment treatments. T_rich is the treatment in which the respondent is paired with a person who has a lower endowment in the game. T_poor is the treatment in which the respondent is grouped with a person with a higher endowment in the game. Baseline is the treatment where both respondents have the same endowment. Robust standard errors in parentheses. # $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Table B5: OLS Estimates for the Donation Decision Including Additional Control Variables.

	(1) Donation	(2) Donation	(3) Donation
Emancipative values	4.579*	5.416**	5.247**
	[0.394,8.764]	[1.574,9.257]	[1.555,8.939]
Secular values	−1.594	−2.141	−1.513
	[-5.247,2.059]	[-5.678,1.396]	[-4.965,1.938]
Outgroup trust	1.237		
	[-3.243,5.718]		
Ingroup trust	2.066		
	[-1.921,6.054]		
Perceived security		1.890	
		[-1.173,4.954]	
Extraversion			−0.609
			[-3.906,2.688]
Neuroticism			−2.104
			[-5.200,0.993]
Openness to experience			1.242
			[-1.708,4.192]
Conscientiousness			2.063
			[-1.238,5.364]
Agreeableness			0.556
			[-2.777,3.890]
Constant	12.654**	12.061**	12.154**
	[9.703,15.605]	[9.353,14.769]	[9.359,14.949]
F-test	2.7	3.9	2.1
Prob $> F$	0.032	0.009	0.048
R^2	0.057	0.055	0.076
N	167	174	173

Note. OLS estimates. Dependent variable is percentage donated [0,100]. Independent variables are standardized except for perceived security, age, female. Female and trust are dummy variables. Robust standard errors in parentheses. # $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Chapter 2

Salience in Public Goods Games*

*This chapter is joint work with Nanxu Su and Christian Thöni.
For helpful comments and discussion we thank Arno Riedl, Urs Fischbacher, Christian Zehnder, Isabel Günther, Laure Athias, the participants of the 7th Thurgau Experimental Economics Meeting, the participants o the ESA 2016 in Bergen, and the participants of Experimental Brownbag Research Seminar at the ETH Zurich. We gratefully acknowledge financial support from the Swiss National Science Foundation.

2.1 Introduction

Standard economic theory usually assumes that agents use all available information when making a decision irrespective of the mode of presentation. However, when we think about decisions in our daily life, most of us would agree that we do not include all available choice attributes when we make a decision. We are often guided by the mode of presentation because our attention is limited. Meaning that usually some characteristics are more salient than others and consequentially we include them in the decision-making process while we neglect others.

Taylor and Thompson (1982) define salience as: "[..] the phenomenon that when one's attention is differentially directed to one portion of the environment rather than to others, the information contained in that portion will receive disproportionate weighing in subsequent judgments."

In economics, Schelling (1960) was among the first to use salience to explain coordination behavior. He asked his students to indicate where they would meet with a friend in New York. A significant majority answered at the Grand Central Terminal. His argument was that the Grand Central Terminal is a more salient choice among all possible meeting points in New York City, because of its tradition to meet at this place. By choosing this salient option, people can solve the coordination problem efficiently. Akerlof (1991) applied the concept of salience to explain why people procrastinate. He postulated that future consequences of today's decisions are not salient and therefore not taken into account when making a decision.

The reasons for why certain choice attributes are salient while others are not, can be endogenous or exogenous. Furthermore, in some cases salience can lead people to attach a disproportional weight to a decision attribute while in other cases salience helps to filter out the relevant attributes of a choice environment. An endogenous factor, for instance, are our political attitudes. Knobloch-Westerwick and Meng (2009) show that participants of an experiment spent more reading time on articles that are in line with their political attitudes. An example for an exogenous factor are policies to promote healthy food choices. These policies aim to design labels to draw the consumer's attention on the health consequences

of her choice. Van Herpen and Van Trijp (2011) show that Traffic-light labels receive more attention than the standard nutrition table and guide consumers to chose healthier options. Presumably, such Traffic-light labels make the health consequences of a food choice more salient and many consumers change their decision because of they take the consequences into account.

It is only recently that economists started to include salience in formal models. A first strand of the literature views salience as arising endogenously within the set of available options (Bordalo, Gennaioli, & Shleifer, 2012). They assume that a characteristic of a good is salient if it stands out compared to the respective feature of a reference good. Meaning that the attribute that is furthest from the reference level is the most salient and will get more weight in the decision process.[1]

The second strand of the literature views salience as an exogenous characteristic of the choice environment. DellaVigna (2009) proposes a simple model to describe why people neglect for example shipping costs or taxes in the valuation of goods. In his model, each object contains a visible and an opaque component. Due to limited attention, the consumer processes the information of the opaque component only partially. Applying this model to consumer taxes, for instance, means that the labeling of the taxes affects the valuation of the respective product. In the extreme case where the price does not include the tax, the model predicts that consumers neglect the tax completely because it is not salient. The contrary is that the price includes the tax. In this case, the model predicts that the consumer processes the tax in the valuation of the object fully. Empirical data supports this claim. In a field experiment with a grocery store Chetty, Looney, and Kroft (2009) observe that when the price includes the tax, demand decreases compared to the case where the tax is separate.

These studies provide evidence that salience can play a role in decision making. Therefore, one of the questions which arises is if salience plays a role beyond consumer and lottery choices. For example, whether behavior in social dilemmas is affected by salience.

[1] Mormann and Frydman (2016) test this model in a laboratory experiment. Besides the lottery choices, they record eye-tracking data from all participants. The authors find that the outcomes are consistent with the theory. However, when they elaborate the eye-tracking data, they observe that participants do not look at the options, which are predicted by the model to be salient. Instead, they discover that the longer the subjects look at any option, the more likely it is that the subject will choose this option.

There are compelling reasons to expect that salience plays a role for contributions in public goods games because many people are conditional cooperators (Fischbacher, Gächter, & Fehr, 2001). For conditional cooperators, the contributions of the other group members are decisive of their own behavior. In this paper, we are interested to see if exogenously increasing the salience of, for instance, the highest contribution in a group affects behavior in the public goods game. To the best of our knowledge, there is no research which evaluates salience in social dilemmas such as the public goods game. Hence, we extend the literature on salience by evaluating formally and empirically if and how salience affects decisions in a social dilemma.

Closest to our design, and a prerequisite for our manipulation to work, is the research on information provision in public goods games. Subjects have to be sensitive to the information provided to them. Otherwise, salience cannot play a role. Engel, Kube, and Kurschilgen (2014) find evidence that providing participants with different pre-play information affects their behavior. In their experiment, subjects of a public goods game receive selected information about the behavior in previous experiments. In one treatment, they showed them high contribution levels, and in the other treatment, participants got information about low contribution levels. They observe that depending on the disclosed information two distinct cooperation norms are established.

Hoffmann, Lauer, and Rockenbach (2013) present participants in a repeated public goods game with exaggerated feedback, i.e., the actual average is multiplied by a factor bigger than one.[2] Surprisingly, subjects do not change their contribution in response to exaggerated feedback. Hartig, Irlenbusch, and Kölle (2015) evaluate how subjects react to different forms of feedback. They compare the provision of the group mean to the provision of the individual behavior of all group members. They find that subjects contribute on average more when they get full information about the behavior of the other group members. Furthermore, they report that the minimum contribution in a group is decisive for the decision of conditional cooperators.

A similar design to ours is by Samek and Sheremeta (2014). They provide participants of a public goods game with additional information about the highest or lowest contributors

[2]Participants received the information that the feedback might deviate from the actual average.

in their group. In the feedback stage participants see, besides the contribution level, a photo, and the name of the other group members. Compared to the standard public goods game, only the treatment which reveals the identity of the lowest contributor in a group affects contributions. Participants in this treatment contribute significantly more compared to the standard public goods game. The mechanism behind the results is unclear; the salience of a certain contribution level could affect the behavior or the identification of contributors.

Summing up, this strand of the literature shows that how information is displayed can have effects on the contribution of subjects in public goods games. However, in most of these papers, the information provided to participants differed depending on the treatments. In contrast, we aim to only shift the focus of attention, while holding the content of the information constant.

In a first step, we extend the Fehr-Schmidt model (Fehr & Schmidt, 1999) to derive theoretical predictions of how salience may affect contributions. The model is set up in the spirit of DellaVigna (2009). In a second step, we conduct a series of three laboratory experiments to test whether we observe systematic differences in behavior due to an exogenous manipulation of the salience of the contribution levels in a standard public goods game.

This chapter proceeds as follows. In section 2.2 we introduce our model. In section 2.3, 2.4, and 2.5 we describe our three experiments and their results. In section 2.6 we discuss the results of the pooled dataset and section 2.7 concludes.

2.2 Theory

In a typical public goods game, each of $n > 2$ subjects in a group receives an endowment of y tokens. Each group member can choose to keep the endowment to herself or to invest parts or all of it into a common pool (the public good). All tokens invested in the public good will be multiplied by a factor a $(1/n < a < 1)$ and equally shared among all group members. The payoff function of a subject i is given by:

$$x_i = y - g_i + a \sum_{j=1}^{n} g_j \quad i, j = 1, 2, 3, ..., n, \tag{2.1}$$

where g_i is the investment of subject i to the public good (with $g_i \in [0, y]$), and a denotes the marginal per capita return of the contribution to the public good.

Under standard assumptions, the only Nash equilibrium in this game is that all players contribute zero tokens to the public good. However, numerous empirical studies provide evidence that people often deviate from the standard prediction and contribute positive amounts. Fehr and Schmidt (1999) propose a utility function, which can account for positive contributions. The function does not only take the player's own payoff into account but also the distance to the payoff of the other players. It represents the observation that individuals often not only care about their payoff but also about the welfare of others. The assumption is that players like equality and increasing the difference between one's own payoff and the payoff of other group members generates disutility. The utility function takes the following form:

$$U_i(x) = x_i - \frac{\alpha_i}{n-1} \sum_{j \neq i} \max\left[x_j - x_i, 0\right] - \frac{\beta_i}{n-1} \sum_{j \neq i} \max\left[x_i - x_j, 0\right] \tag{2.2}$$

where α_i is the aversion factor towards disadvantageous inequality, while β_i stands for the aversion towards advantageous inequality. It is assumed that $\alpha_i > \beta_i$ and $0 < \beta_i < 1$. In the case of $n = 4$ and $a = 0.4$, which is the group size and the MPCR in the three experiments later on, the standard version of the model predicts that the best reply of subject i is to match the lowest contribution within her group ($g_i = R(g_j) = \min\{g_j\}$).[3]

We introduce individual salience factors θ_{ij} in the Fehr-Schmidt model to derive how salience potentially affects behavior. We posit the following utility function:

$$\begin{aligned} U_i(x) = x_i - \alpha_i \sum_{j \neq i} \theta_{ij} \max\left[x_j - x_i, 0\right] - \beta_i \sum_{j \neq i} \theta_{ij} \max\left[x_i - x_j, 0\right], \\ \text{with} \sum_{j \neq i} \theta_{ij} = 1 \text{ and } \theta_{ij} \geq 0 \end{aligned} \tag{2.3}$$

[3]For more details on the predictions of the model see Fehr and Schmidt (1999).

All the payoff-comparisons are now additionally weighted by θ_{ij}. The size of θ_{ij} defines how salient the comparison with j is. Thereby, we assume that overall the attention stays constant, but the proportion of attention that each bilateral comparison gets can be different. If a certain θ_{ij} is equal to 0, it means that this comparison does not receive any attention and is hence not considered in the calculation of the utility. On the other hand, if all θ_{ij} are equal ($\theta_{ij} = \frac{1}{n-1} \quad \forall j \neq i$), function (2.3) is equivalent to the original Fehr-Schmidt utility function.

To derive the best response functions, take player i facing three other players with contributions g_l, g_m, and g_h, such that $g_h \geq g_m \geq g_l$. Then the best response of player i depends not only on her α_i and β_i but also on how salient the contributions of the other players are.

$$g_i = R(g_j) = \begin{cases} 0 & \text{if } a + \beta_i < 1 \\ g_l & \text{else if } \theta_{il} > \hat{\theta}_{il} \\ g_m & \text{else if } \theta_{ih} < \hat{\theta}_{ih} \\ g_h & \text{else if } \theta_{ih} > \hat{\theta}_{ih} \end{cases} \qquad \text{with } \hat{\theta}_{il} = \frac{a + \beta_i - 1}{\alpha_i + \beta_i} \text{ and } \hat{\theta}_{ih} = \frac{1 - a + \alpha_i}{\alpha_i + \beta_i} \tag{2.4}$$

Figure 2.1 depicts the best responses graphically. θ_{ij} denotes the salience level of player i towards the contribution g_j, with $0 \leq \theta_{ij}$. $\hat{\theta}_l$ and $\hat{\theta}_h$ respectively stand for the critical salience levels of the lowest and the highest contributions from the other group members.

If the lowest contribution is sufficiently salient $\hat{\theta}_l \leq \theta_l$, the best response of player i is to match the lowest contribution $g_i = g_l$. In other words, if the lowest contribution is more salient relative to the other contributions, it will receive more weight in the utility function, and player i is best off if he matches the lowest contribution of his group. The more a person dislikes disadvantageous inequality, the lower the threshold for $\hat{\theta}_l$. In contrast, the more a person disrelishes advantageous inequality, the lower the threshold for $\hat{\theta}_l$. Likewise, if the middle contribution becomes more salient relative to the other contributions, the player will match this contribution. Lastly, if $\hat{\theta}_h \leq \theta_h$ the participant will put the most attention on the highest contribution in her group. She will try to minimize the inequality

with respect to the highest contributor and match his contribution.[4]

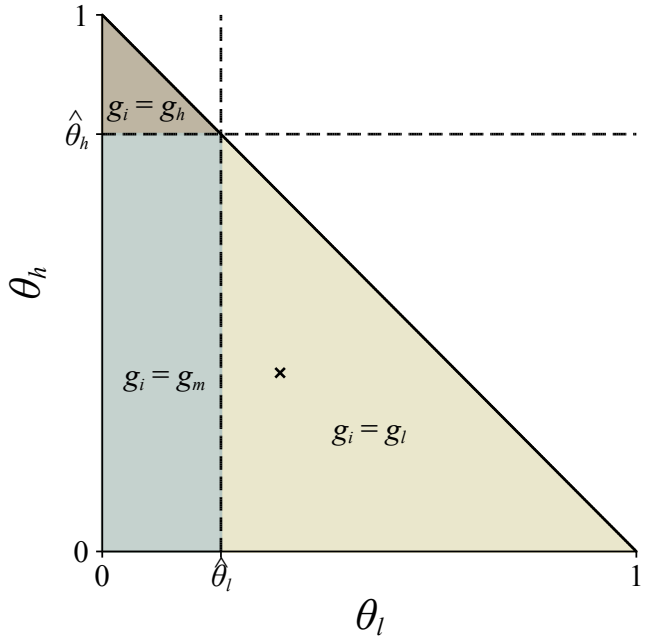

Figure 2.1: Optimal contributions of players under different salience levels.

In this model, we assume that the salience weight is determined exogenously. Obviously, this is the simplest explanation for the determination of the salience weights. Salience can not only be affected by exogenous factors of the choice environment but also through endogenous channels such as the preferences of an individual, her past choices, or strategic behavior. Given that there is no research yet on how salience affects behavior in a public goods game, we believe that it is justified to first start with the question whether salience affects behavior at all. If salience is relevant, our results can then be used to further think about the determinants of the salience weights in the context of social dilemmas.

To conclude, with our extension of the Fehr-Schmidt model, we can show that, depending on the salience of the other contributions within a group, the best response of player i changes. Consequentially, we set up a series of laboratory experiments to test these hypotheses empirically.

[4] A participant will not contribute more than the maximum contribution of the others, because the condition for that would be $a > \alpha_i + 1$ while $\alpha_i \geq \beta_i \geq 0$ and $a < 1$.

2.3 Experiment 1 - The Recall Experiment

2.3.1 Design and Procedures

The first experiment consisted of two sequences. In each sequence, participants played ten consecutive periods of a standard public goods game. At the beginning of the first sequence, we assigned participants randomly into groups of four. The group composition remained the same during the entire experiment. Each group member received an endowment of 20 tokens at the beginning of every period. The marginal per capita return (a) was 0.4. A period consisted of the individual contribution decision followed by the feedback about the individual earnings and the contributions of the other group members.[5]

We based our salience manipulation on results from research in cognitive psychology. The literature shows that attention is often biased towards what is currently in mind (for a review see Olivers, Peters, Houtkamp, and Roelfsema (2011)). Furthermore, according to that literature, recalling an item affects the perception about that item. For example, Dutch students reported using their bike more often after recalling instances of bicycle use.[6] In a similar vein, Janiszewski, Kuo, and Tavassoli (2013) show that paying repeated attention to a product increased the probability of choosing this product in a subsequent choice. Hence, to manipulate the salience of the different contribution levels within a group, we introduced a question at the end of every period. The question implied to recall one of the previously displayed contributions of another group member. The goal was to draw the attention of the participant on a specific contribution level. This central feature of our experimental design is worth emphasizing. Participants memorize a contribution level such that it occupies parts of their working memory and shifts their attention to that level. We provided participants with two stimuli, (i) the minimum and (ii) the maximum stimulus. Under the minimum stimulus, as the name indicates, participants were asked to recall the lowest contribution in their group. On the contrary, under the maximum stimulus, participants had to remember and reproduce the highest contribution of the other group

[5] Participants received the information about the exact contributions of the other group members; they did not get the mean contribution of their group. However, it was easy to calculate the mean from the available information. The screenshots of this experiment are in appendix 2.C.

[6] According to the research, the effect is only present when the recalling task is easy, i.e., when the items to recall come to mind easily (Schwarz, 2004).

members. Subjects earned a point if they recalled the respective contribution correctly.[7] Hence, in both treatments, we did not manipulate the information structure but only the salience of a certain piece of information.

Within a group, all participants received the same stimulus. We assigned the groups randomly to one of the two stimuli at the beginning of the experiment, and they stayed within the assigned stimulus for the first sequence. In the second sequence, all groups switched stimulus. Groups who started under the minimum stimulus switched to the maximum stimulus in the second phase and vice versa. In the instructions, we explained that participants will need to memorize a specific information and that they will receive more information about this task directly on the screen once the experiment started. A translated version of the instructions is in appendix 2.C.

2.3.2 Sample Size Calculation

To know how many independent observations we need to draw statistically meaningful conclusions, we calculate the sample size using the power analysis by Cohen (1988). We want to compare the means under the two stimuli, hence, to calculate the sample size, we need an estimate of the effect size, the standard deviation and to specify the significance level and power. Since our model does not give precise predictions about the size of the treatment effect, we rely on values reported in the related literature. The three papers which are most closely related to our design (Hoffmann et al., 2013; Engel et al., 2014; Samek & Sheremeta, 2014) report treatment effects in the range of 25 and 30 percentage points. Only Engel et al. (2014) report the standard deviation. For the significance level and power, we use the common levels of 0.05 and 0.8 respectively. Based on a treatment effect of 5.3 token and a standard deviation of 5 tokens, the formula indicates that our sample size should be 28 (14 in each treatment cell) to be able to compare the mean under the maximum stimulus to the mean under the minimum stimulus.

[7] We incentivized this question to make sure that participants had an incentive to answer the question correctly and hence took it seriously.

2.3.3 Results

We conducted the experiment at the LABEX of the University of Lausanne in Switzerland. 108 individuals participated, and we used the software zTree to run the experiments (Fischbacher, 2007). For the analysis, the unit of independent observation is the group. Hence, in this experiment, we observe the behavior of 27 independent groups.

In Figure 2.2 we display the behavior of the participants in the first sequence graphically. The thin lines depict the average contributions, which participants had to recall under the two stimuli. We observe that the mean with the maximum stimulus is 12.76 (SD: 6.84) compared to an average of 4.75 (SD: 6.55) with the minimum stimulus. This difference is statistically significant (Wilcoxon rank-sum test, two-sided, N=27, p <0.01) and a prerequisite for our treatment manipulation to work. Because if participants had to recall numbers which are not substantially different, there would be no reason to expect a difference between the minimum and the maximum stimuli.

To our surprise, the two stimuli do not translate at all into significant differences in the contribution behavior. Participants under both stimuli contribute on average similarly (see the thick lines in Figure 2.2). The mean contribution with the maximum stimulus is 8.76 tokens (SD: 7.37) compared to an average of 8.41 tokens (SD: 7.76) with the minimum stimulus (Wilcoxon rank-sum test, two-sided, N=27, p=0.662).[8] Moreover, the variance within the groups over time remains rather constant and is statistically not different between the two stimuli (Wilcoxon rank-sum test, two-sided, N=27, p=0.771).

[8]We elaborated additionally if participants who behave as conditional cooperators behave differently. Conditional cooperators are participants which condition their contribution on the contribution of the other group members. To classify them, participants had to fill out a contribution table at the very beginning of the experiment. In this table, they had to indicate how much they contribute given the average contribution of the other group members. We followed the procedures of Fischbacher et al. (2001), which allow classifying participants into freeriders, conditional cooperators, hump shaped and others. The results are the same if we restrict the analysis at the subgroup of conditional cooperators. Furthermore, the analysis of the second sequence (of all participants) reveals very similar results as in the first sequence. The average contribution, which subjects recalled under the maximum stimulus is 8.71 (SD: 7.86) compared to 4.52 (SD: 6.27) under the minimum stimulus. The difference is not as pronounced as in the first sequence. Moreover, it is only weakly statistically different (Wilcoxon rank-sum test, two-sided, N=27, p=0.093). Again, we do not observe differences in the subsequent contribution behavior (see Figure B1 in the appendix). The average contribution with the minimum stimulus is 6.93 (SD: 7.30), which is even higher than the average of 5.80 (SD: 7.58) with the maximum stimulus. The difference is statistically not significant (Wilcoxon rank-sum test, two-sided, N=27, p=0.599).

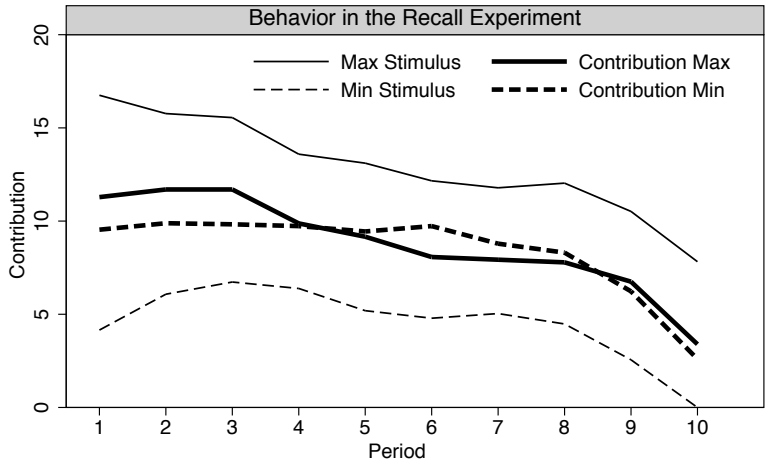

Figure 2.2: Behavior in the Recall Experiment, first sequence.

One possibility for not finding differences could be that participants did not recall the correct contribution and therefore had different numbers on top of their head. However, in 93% of the cases, participants recalled the contribution correctly.[9] Hence, it is fair to assume that people focused on the correct number.

We further elaborate whether the stimuli might affect groups differently depending on the homogeneity of the contributions within the group. It could be that if the contributions in a group are very similar the stimulus leads to a different effect compared to the situation where the behavior within a group is very different. However, this seems not to be the case. We do not find a difference between the stimuli in neither homogeneous nor heterogeneous groups.

So far we discussed the between-subject effects. It might be that, even though we do not observe differences between subjects, subjects themselves react to the stimuli. Our design allows to evaluate whether the behavior of the subjects is different between the first and the second sequence. Based on a sign-rank test we do not find a significant within-subject difference (Signed-rank test, N=16, p=.918).

Thus, the results of this experiment lead us to conclude that either the contributions of subjects in the public goods game are more stable than we thought or that our salience stimuli were too weak. We suspected the second option and decided to manipulate salience more radically in a second experiment.

[9] 50 % of the participants answered the recall question correctly in all ten periods. 33% replied to the question correctly in 9 periods. The remainder of the subjects answered between 6 to 8 times correctly.

2.4 Experiment 2 - The Limited Information Experiment

2.4.1 Design and Procedures

The basis of experiment two is the same standard public goods game as in the first experiment. However, we manipulated salience more drastically. After the contribution decision, we provided participants only with very limited feedback about the behavior of the other group members. Under the maximum stimulus, we only revealed the highest contribution in the group. To be specific, we stated: "The highest contribution in your group was X." Under the minimum stimulus, we only revealed the lowest contribution in the group. In both stimulus-treatments, participants did not receive information about their individual earnings until the very end of the experiment. Hence, they could not infer any information about the behavior of the group except for either the minimum or the maximum contribution. Compared to the previous experiment, participants only received very limited information. We argue that this is a very strong form of salience because it is the only available information. Hence, the focus of attention must lie by construction on the minimum or the maximum contribution respectively.

We also elicited the beliefs of a subset of participants. These participants had to indicate their belief about the average contribution level of the other group members at the same time as they decided about their contribution. We incentivized the belief elicitation with one additional token per correctly estimated average contribution. Again, we implemented an AB/BA design; some groups started with ten periods under the maximum stimulus and then played ten periods under the minimum stimulus and vice versa.[10]

2.4.2 Results

In this experiment, a total of 80 subjects participated. Thus, we have 20 independent observations for the two sequences of the experiment. It is important to note, that these are different subjects than in experiment one.

Figure 2.3 provides an overview of the results. The thin lines depict the stimuli that participants received. With the maximum stimulus, in most of the groups and during the

[10]The instructions are in appendix 2.C.

first seven periods, the contribution displayed was the maximum contribution (20 tokens). Whereas when the stimulus was the minimum, the average of the displayed contribution was 4.65 token with a standard deviation of 6.71 tokens.

To evaluate whether our manipulation has a chance to affect behavior, we first look at the beliefs of the participants. We observe that at the beginning of the experiment beliefs under the two stimuli are on average similar. Given that our manipulation is at the end of each period, the beliefs in the first period should and are not different.[11] However, over the course of the experiment, the beliefs under the two stimuli start to differ. All periods taken together, the beliefs are statistically significantly different (Wilcoxon rank-sum test, two-sided, N=11, p=0.045).[12] Furthermore, we observe that the change in the beliefs between the periods is significantly bigger under the minimum stimulus, where participants on average reduce their belief by 1.15 tokens, compared to a decrease of 0.43 tokens under the maximum stimulus (Wilcoxon rank-sum test, two-sided, N=11, p=0.100). These results suggest that the stimuli affect the beliefs.

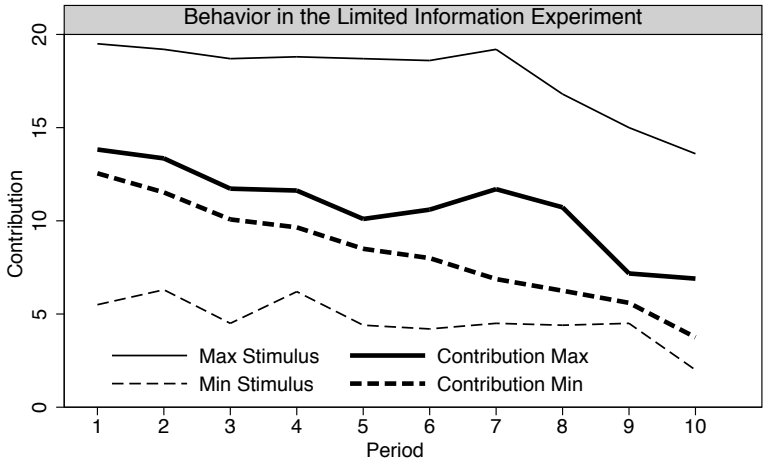

Figure 2.3: Displayed contributions and behavior in the Limited Information Experiment.

However, beyond the manipulation of the beliefs, we do not observe a statistically significant difference in the contribution decisions of the participants (see Figure 2.3).

[11] Analog to the setup of the Recall Experiment, participants do not know that two separate treatments exist. We instruct them that they will get some feedback after each period and that they will get more detailed information once the experiment started. Thus, participants learn at the end of the first period that they receive a minimum or maximum stimulus.

[12] From the fifth period and onward the differences between the two stimuli are even more pronounced (Wilcoxon rank-sum test, two-sided, N=11, p=0.028).

The mean contribution with the maximum stimulus, depicted by the thick line, is 10.77 (SD: 8.8), with the minimum stimulus, the mean is 8.28 (SD: 7.76), represented by the dashed thick line. The difference is not significant (Wilcoxon rank-sum test, two-sided, N=20, p=0.112).[13] Contrary to the findings in the recall experiment, we observe that the variance within the groups is different under the two stimuli. The standard deviation under the maximum stimulus amounts to 6.6 tokens under the minimum stimulus it amounts to 4.6 tokens (both across periods). This difference is significant (Wilcoxon rank-sum test, two-sided, N=20, p=0.016). Under both stimuli, the standard deviation is constant across the periods.

Again, we do not observe significant within-subject differences (Signed-rank test, N=20, p=0.767).

To sum up, in this experiment, we observe that providing participants with very limited feedback about the group behavior affects their beliefs and the variance within the group, but to our surprise, it hardly affects the average contributions.

2.5 Experiment 3 - The Belief Experiment

2.5.1 Design and Procedures

Our third attempt to manipulate salience builds on the successful manipulation of the beliefs in the previous experiment. The basic set-up was the same as in the first two experiments - the standard version of the public goods game. All participants had to indicate their beliefs at the same time as they indicated their contribution. We implemented the third salience stimulation in the following way; under the maximum stimulus, we asked participants to indicate their beliefs about the maximum contribution in their

[13]The behavior in the second sequence of the experiment, where the groups switched treatments, reveals a similar picture. We find that under the maximum stimulus the average displayed contribution is 15.85 (SD: 5.32) and under the minimum stimulus, it is 8.28 (SD: 9.35). While in the first sequence of this experiment, the beliefs were affected by the stimuli, we do not observe different beliefs in the second sequence. With the maximum stimulus the average belief is 8.77 token (SD: 8.09), and with the minimum stimulus, it is 8.73 token (SD: 6.73), (Wilcoxon rank-sum test, two-sided, N=11, p=0.715). Given that the beliefs are already not different from each other it would be very surprising if the behavior under the two stimuli would differ. In fact, we even observe a reversed trend; the average contribution under the maximum stimulus is lower than the average contribution under the minimum stimulus (Maximum: M=8.51, SD=8.71 and Minimum: M=9.75, SD=8.34). However, this disparity is statistically not significantly different (Wilcoxon rank-sum test, two-sided, N=20, p=0.305).

group. Consequentially, under the minimum stimulus, they had to indicate their belief about the lowest contribution in their group. We incentivized the belief elicitation with an additional point per correct estimation. After their contribution, participants received the full information about the behavior of the other group members (analog to experiment one).[14]

2.5.2 Results

We conducted the experiment with 108 subjects, which leaves us with 27 independent observations. Again, the subjects which participated in this experiment did neither participate in the recall experiment nor in the limited information experiment.

In Figure 2.4 we depict the average beliefs and contributions under the two stimuli. Concerning beliefs, we observe that with the maximum stimulus subjects estimated that the highest contribution is on average 8.56 tokens (SD: 7.04). This number seems rather low, but note that it is across all periods. In the first period, subjects estimated that the maximum contribution is 15.32 tokens (SD: 4.94), compared to 6.05 token on average (SD: 6.97) for the last period. Under the minimum stimulus, the overall estimation was on average 5.42 (SD: 6.34). At the beginning of the experiment, participants estimated that the lowest contribution in their group is 7.38 (SD: 6.66) on average, whereas in the last period the average of the estimations dropped to 1.90 (SD: 4.25).[15] We take these results as an indication that participants under the maximum stimulus do not expect their group members to contribute fully. In contrast, under the minimum stimulus, subjects do not expect their group members to freeride completely. Actually, under the minimum stimulus subjects have rather high beliefs about the lowest contribution of their group members at the beginning of the experiment.

[14]The instructions for this experiment are in appendix 2.C.

[15]The difference in the estimation between the stimuli is statistically significantly different (Wilcoxon rank-sum test, two-sided, N=27, p=0.047).

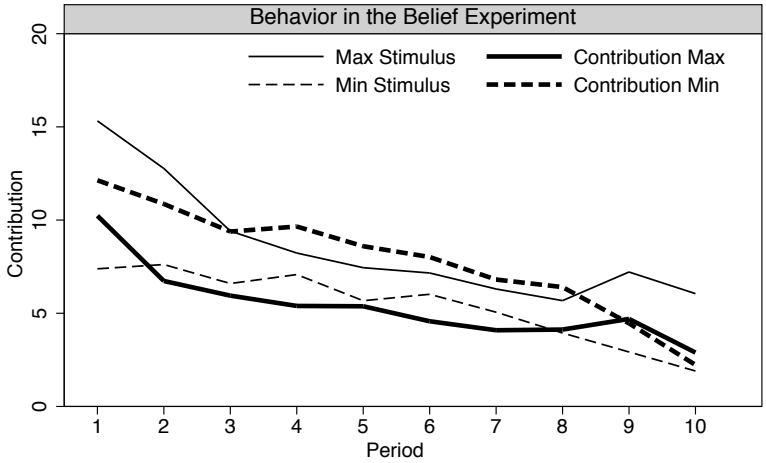

Figure 2.4: Beliefs and behavior in Belief Experiment.

Since we observe that the beliefs are significantly different under the two stimuli, we think that these differences lead participants to adjust their behavior accordingly. As depicted in Figure 2.4 by the thick lines, participants do change their behavior in response to our stimuli. However, the behavioral change is not in the direction that we expect it rather goes into the opposite direction. With the maximum stimulus, participants contribute on average 5.40 (SD: 6.76) under the minimum stimulus the average contribution over all periods is 7.86 (SD: 7.72). This difference is weakly statistically significant (Wilcoxon rank-sum test, two-sided, N=27, p=0.058).[16]

Anew, we evaluate whether the two stimuli affected the within group variance. We find that the standard deviation under the maximum stimulus is 4.3 tokens compared to 5.3 tokens under the minimum stimulus (Wilcoxon rank-sum test, two-sided, N=27, p=0.081). In contrast to the Limited Information Experiment, the difference is the opposite. Under the minimum stimulus, the standard deviation is bigger than under the maximum stimulus. Under both stimuli, the standard deviation remains constant across the periods.

Furthermore, we do not observe that the stimuli lead to a within-subject effect (Signed rank test, N=27, p=.838).

[16]We also analyzed the behavior in the second sequence of this experiment. Probably again, due to the experience in the previous phase, the beliefs under the two stimuli are only weakly significantly different (Wilcoxon rank-sum test, two-sided, N=27, p <0.073). As depicted in Figure B3 in the appendix we do not observe the reversed trend from the first phase. On average with maximum stimulus subjects contribute 6.66 tokens (SD: 7.72). With the minimum stimulus, the average is with 5.33 (SD: 6.78) slightly lower. The difference is not significant (Wilcoxon rank-sum test, two-sided, N=27, p=0.308).

In conclusion, we notice that letting participants think about what they belief will be either the highest or the lowest contribution in their group affects their behavior weakly. This leads us to hypothesize that the two stimuli might generate a lower or upper bound for the contribution. Meaning that if a participant receives the minimum stimulus, she tries to contribute just a little more. In contrast, under the maximum stimulus, a participant might perceives it as legitimate to contribute just a little less. Thus overall there is no difference or as we observed in this experiment a difference which is against our predictions. To elaborate this hypothesis in more detail, we pooled the data of the three experiments and describe the results in the next section.

2.6 Pooled Data

According to our power calculation we are short of the number of independent observations in all three experiments. Hence, to increase the statistical power and to further elaborate the relationship between our stimuli and the contribution decision, we pool the data of the three experiments and estimate OLS regressions with robust standard errors clustered at the group level. In total, we have 74 independent observations.

In Table 2.1 we display the results of the OLS estimations. In column (1) we regress the contribution decision on the period, the stimulus dummy (*Maximum*= 1, *Minimum*= 0), the sequence (*Second Sequence*), the dummy for the Limited Information experiment (*LimInfo*) and the dummy for the Belief Experiment (*BelExp*). The recall experiment serves as baseline.

We confirm the finding of the separate analyses; overall the stimuli do not lead to significant differences in the contributions. Only the period has predictive power for the contribution decision; the coefficient indicates a clear negative time trend. Furthermore, although insignificant, the dummy for the stimulus is negative. This tendency is contrary to what we expect based on the predictions of our extension of the Fehr-Schmidt model.

To further understand this negative tendency, we add control variables for the individual contribution in the previous period (*Contribution (t-1)*), the so-called salience point, and the interaction between the salience point and the stimulus (*Salience Point*Max*) in column

(2). With salience point, we mean the contribution on which an individual was forced to focus on due to the stimulus. For experiment one, the stimulus is the contribution which an individual had to recall. In experiment two, the stimulus is the information about the highest or lowest contribution in the group. For the third experiment, the stimulus is the belief about the highest or lowest contribution of the other group members.

Once we control for these additional variables, we observe that the coefficient of the stimulus dummy is still negative and becomes significant. Participants under the maximum stimulus contribute on average 2.27 tokens ($p<0.001$) less to the public good than participants under the minimum stimulus. This negative effect becomes stronger in the second sequence of the experiment. Not surprisingly, the contribution in the previous period is highly predictive for the contribution in the current period. Indicating that a positive relationship between the lagged contribution and the current contribution exists. We observe the same link between the contribution and the salience point. An increase of one token in the salient contribution or belief, resp. leads to an increase of the contribution by 0.24 tokens ($p <0.001$). Additionally, the interaction term indicates that the effect is slightly stronger under the maximum stimulus, where an increase by one point leads to an increase in the contribution by 0.30 tokens.

Summing up, we find that once we control for the previous contribution and the salience point, participants contribute significantly less under the maximum stimulus than under the minimum stimulus. This result might explain why we do not observe an effect in the first place. Both stimuli lead subjects to adjust their contributions towards the middle. Let us illustrate this point with a numerical example. If a subject, for instance, sees a contribution of 15 tokens as the salience point under the maximum stimulus, our estimation predicts that she will contribute roughly 11 tokens. In contrast, if a subject under the minimum stimulus gets a salience point of 5 tokens, the regression model predicts a contribution of roughly 10 tokens.[17] Thus, under both stimuli, participants will contribute similarly, but adjust their contributions differently. Under the maximum stimulus, our model estimates that participants decrease their contribution in respect to the salience

[17]We calculated the predictions for the second period and with a contribution of 10.8 tokens in the previous round.

Table 2.1: Contribution in the public goods game.

	(1)	(2)
Period	-0.804^{***}	-0.390^{***}
	(0.060)	(0.050)
Maximum (D)	-0.107	-2.275^{***}
	(0.580)	(0.370)
Second Sequence (D)	-1.018^{**}	0.306
	(0.476)	(0.197)
LimInfo (D)	1.720	0.024
	(1.472)	(0.439)
BelExp (D)	-1.360	-0.254
	(1.242)	(0.317)
Contribution (t-1)		0.506^{***}
		(0.026)
Salience Point		0.248^{***}
		(0.023)
Salience Point * Max		0.075^{**}
		(0.033)
Constant	12.621^{***}	4.283^{***}
	(1.088)	(0.581)
F-test	53.8	182.2
Prob $> F$	0.000	0.000
R^2 adjusted	0.116	0.528
N	5480 (74)	4636 (74)

Note. OLS estimates. The dependent variable is the contribution to the public good [0,20]. The table depicts the results for the pooled dataset. *Maximum* is the dummy variable indicating treatment status. *Second sequence* controls for order effects. *LimInfo* is a dummy for the limited information experiment. Analogously *BelExp* is the dummy variable for the belief experiment. *Contribution* indicates the contribution of the subject in the previous period, and *Salience point* controls for the contribution which a subject had to focus on due to the salience manipulation. Robust standard errors, clustered on the group level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

point. In contrast, under the minimum stimulus participants increase their contribution. Hence, salience seems to play a role in the public goods games, but the effect is subtle, and the direction is contrary to our predictions in section 2.2.

2.7 Conclusions

Researchers in cognitive psychology are concerned with the topic of salience for many years. In economics, on the other hand, this topic is relatively new, and research on it is scarce. The literature started to show how salience affects decisions under risk and consumer choices. To the best of our knowledge, there is no research on how salience affects behavior in social dilemmas. Inspired by the work of DellaVigna (2009), where he models how salience influences consumer choice, we extended the Fehr-Schmidt model to include salience. Based on this extension we can derive clear predictions for how salience affects behavior in a standard public goods game. If the lowest contribution in a group is most salient compared to the other contributions, this contribution level will get more weight in the decision-making process. Hence, contributions will be biased towards the minimum in the group. The predictions for the case where the highest contribution in a group is most salient are vice versa.

Consequentially, we conducted a series of laboratory experiments to verify our theoretical predictions. In all three experiments, we manipulated the salience of the contributions of the other group members exogenously; in one treatment (stimulus) the highest contribution was made salient, whereas, under the other stimulus the lowest contribution was made salient. In the Recall Experiment, we manipulated salience through an additional recall question. We did not find that salience affects contributions as predicted. Consequentially, we conducted a second experiment with a much stronger salience manipulation - the Limited Information Experiment. In this experiment, participants received as feedback only the information about either the highest or the lowest contribution in their group. To our big surprise, the results stayed the same; i.e., the salience stimuli did not affect contributions in the public goods game. In the last experiment, the Belief Experiment, we asked participants to indicate their belief about either the highest or the lowest contribution of

the other group members. Again, the experiment did not confirm our hypotheses. On the contrary, we find weak evidence that participants under the maximum stimulus contributed on average less than participants under the minimum stimulus. Indicating that with this salience manipulation we were able to alter behavior, but not in the direction of our predictions.

A more detailed analysis of the pooled dataset of all three experiments revealed that the two stimuli might created an upper or lower bound for the contribution decision. We hypothesize that subjects under the maximum stimulus perceived the maximum contribution as some kind of upper bound and thought that it is fair to contribute slightly less than the maximum. In contrast, participants under the minimum stimulus perceived the minimum as the minimum acceptable contribution and thus contributed slightly more than the minimum. However, to confirm this interpretation, we need further experiments. A simple test would be to conduct the same experiments but to not label the salience stimuli as minimum or maximum. This would allow to disentangle the effect generated by the labeling from the effect of the raw number. Additionally, more research is needed to understand how participants perceived the salience manipulation. This would also help to unravel the effects observed in Samek and Sheremeta (2014). Based on our results, we suppose that the loss of anonymity drives their results and not the labeling of the contributions.

Another interpretation of our results is that the stimuli were not powerful enough. Especially in our set up with groups of four, it is easy for a participant to remember the contributions of all other group members. Possibly, in a larger group where participants do not keep all contributions in mind easily, making them focus on a specific contribution would be more effective.

Nevertheless, from a policy perspective, our results provide initial evidence that focusing on a role model, might not be as effective as intuitively thought so far. It might be that individuals perceive a role model as the upper bound, which then serves as justification for a cooperation level which is considerably lower. As we stated before, this explanation is only a hypothesis, and its confirmation depends on further research.

To sum up, our series of experiments is an important starting point to elaborate how salience affects behavior in social dilemmas. We add to the literature by showing that

salience affects the beliefs and that the effect on the behavior is subtle and contrary to our predictions.

References

Akerlof, G. A. (1991). Procrastination and obedience. *American Economic Review*, *81*(2), 1–19.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics*, *127*(3), 1243–1285.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, *99*(4), 1145–1177.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–372.

Engel, C., Kube, S., & Kurschilgen, M. J. (2014). Can we manage first impressions in cooperation problems? an experiment. (2011/5). Working Paper. MPI Collective Goods Preprint.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404.

Hartig, B., Irlenbusch, B., & Kölle, F. (2015). Conditioning on what? heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics*, *55*, 48–64.

Hoffmann, M., Lauer, T., & Rockenbach, B. (2013). The royal lie. *Journal of Economic Behavior & Organization*, *93*, 305–313.

Janiszewski, C., Kuo, A., & Tavassoli, N. T. (2013). The influence of selective attention and inattention to products on subsequent choice. *Journal of Consumer Research*, *39*(6), 1258–1274.

Mormann, M. M. & Frydman, C. (2016). The role of salience and attention in choice under risk: An experimental investigation. Working Paper, December 2016.

Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334.

Samek, A. S. & Sheremeta, R. M. (2014). Recognizing contributors: An experiment on public goods. *Experimental Economics*, *17*(4), 673–690.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348.

Taylor, S. E. & Thompson, S. C. (1982). Stalking the elusive "vividness" effect. *Psychological Review*, *89*(2), 155–181.

Van Herpen, E. & Van Trijp, H. C. (2011). Front-of-pack nutrition labels. their effect on attention and choices when consumers have varying goals and time constraints. *Appetite*, *57*(1), 148–160.

Appendix

2.A Proposition and Proof

Propositions

Suppose that a player i is facing a contribution vector $(g_1, g_2, ..., g_{n-1})$ of $n-1$ group members, in which $0 \leq g_1 \leq g_2 \leq ... \leq g_{n-1}$

Proposition 1 A player with $a + \beta_i < 1$ has a dominant strategy to choose $g_i = 0$.

Proposition 2 Use k $(0 \leq k < n)$ to represent the amount of players which display $a + \beta_i < 1$. If $k \geq \frac{\sum_{j=k+1}^{n-1} \theta_j^i - (1-a) \sum_{j=1}^{n-1} \theta_j^i}{\theta_1^i}$, there will be a unique equilibrium where all $g_i = 0$, $i \in \{1, 2, \ldots, n\}$. In the inequality above, $j \neq i$ and θ_1^i refers to the salience assigned to the zero contribution of freeriders $g_1 = g_2 = ... = g_k = 0$. We assume that in an anonymous environment contributions of the same amount will get the same attention.

Proposition 3 If $k < \frac{\beta_i \sum_{j=k+1}^{n-1} \theta_j^i - (1-a) \sum_{j=1}^{n-1} \theta_j^i}{\alpha_i \theta_1^i} (j \neq i)$, there is an equilibrium in which players with $a + \beta_i > 1$ make positive contributions.

The contribution of player i will be:

$$g_i = R(g_j) = \begin{cases} 0 & \text{if } a+\beta_i < 1 \\ [0, g_1] & \text{if } a+\beta_i = 1 \\ g_1 & \text{if } a - \frac{\theta_1^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\sum_{j=2}^{n-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\beta_i < 1 \text{ and } a+\beta_i > 1 \\ [g_1, g_2] & \text{if } a - \frac{\theta_1^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\sum_{j=2}^{n-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\beta_i = 1 \\ g_2 & \text{if } a - \frac{\theta_1^i+\theta_2^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\sum_{j=3}^{n-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\beta_i < 1 \text{and } a - \frac{\theta_1^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\sum_{j=2}^{n-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\beta_i > 1 \\ \dots & \dots \\ g_z & \text{if } a - \frac{\sum_{j=1}^{z}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\sum_{j=z+1}^{n-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\beta_i < 1 \text{and } a - \frac{\sum_{j=1}^{z-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\sum_{j=z}^{n-1}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\beta_i > 1 \\ \dots & \dots \\ g_{n-1} & \text{if } a - \frac{\sum_{j=1}^{n-2}\theta_j^i}{\sum_{j=1}^{n-1}\theta_j^i}\alpha_i + \frac{\theta_{n-1}^i}{\sum_{j=1}^{n-1}\theta_{n-1}^i}\beta_i > 1 \end{cases} \tag{2.5}$$

where $j \neq i$ and $z \neq i$. θ_j^i denotes how salient the contribution of player j is for player i. We define $0 \leq \theta_j^i \leq 1$.

Proposition 1 indicates that, after contributing one token to the public pool, a player i earns a tokens from the public good in monetary terms, and at the same time, he also can get at most β_i subjective benefit from reducing inequality. This is the same as the Fehr-Schmidt model. Hence, if the player displays $a + \beta_i < 1$, his dominant strategy will be to contribute zero.

Proposition 2 indicates that if there are only a few players with $a + \beta_i > 1$, they will never choose to contribute a positive amount, as long as there are too many freeriders. Therefore, there will be only one equilibrium in which all players choose $g_j = 0$, $\forall j \in \{1, 2, \dots, n\}$. Nevertheless, if high contributions from the other group members become more salient compared to those from the freeriders, the tolerance with respect to the

amount of freeriders increases.

Proposition 3 shows that if there are enough players with $a + \beta_i > 1$, it is possible to sustain an equilibrium with positive contributions. If the positive contributions are salient enough relative to the saliency of the behavior of the freeriders, the utility of conditional cooperators does not decrease that much by the inequality created by the behavior of the freeriders.

Considering all the contributions from the other group members, the respective decisions made by player i are listed in proposition three. By making high contributions more salient, a player with sufficient α and β levels will respond with a higher contribution. However, it is impossible for a participant to make a contribution larger than the maximum contribution of others, because the condition for that would be $a > \alpha_i + 1$, while $\alpha_i \geq \beta_i \geq 0$ and $a < 1$.

Proof

Re-label and fix all players, except for player i, such that $0 \leq g_1 \leq g_2 \leq \cdots \leq g_{n-1}$. The utility function of player i can be described as follows:

$$\begin{aligned} U_i(g_i, g_{-i}) = \quad & y - g_i + a \sum g_j - \alpha_i \sum_{j \neq i} \frac{\theta_j^i}{\sum\limits_{j \neq i} \theta_j^i} \max\{g_i - g_j, 0\} \\ & -\beta_i \sum_{j \neq i} \frac{\theta_j^i}{\sum\limits_{j \neq i} \theta_j^i} \max\{g_j - g_i, 0\} \end{aligned} \tag{2.6}$$

which is a continuous function.

If all players $j, (j \neq i)$ contribute zero, it is optimal for player i to choose $g_j = 0$ as well, since $U_i(0, 0, \ldots, 0) < U_i(g_i, 0, \ldots, 0) = y + (a - 1)g_i - \alpha_i g_i$, where $a - 1 < 0$ and $\alpha_i \geq 0$.

Now we calculate the partial derivative of the utility function of player i with respect to his contribution:

$$\frac{\partial U_i}{\partial g_i} = \begin{cases} -1 + a + \beta_i & g_i \in [0, g_1) \\ -1 + a - \frac{\theta_1^i}{\sum_{j=1}^{n-1} \theta_j^i}\alpha_i + \frac{\sum_{j=2}^{n-1} \theta_j^i}{\sum_{j=1}^{n-1} \theta_j^i}\beta_i & g_i \in (g_1, g_2) \\ \dots & \dots \\ -1 + a - \frac{\sum_{j=1}^{n-2} \theta_j^i}{\sum_{j=1}^{n-1} \theta_j^i}\alpha_i + \frac{\theta_{n-1}^i}{\sum_{j=1}^{n-1} \theta_{n-1}^i}\beta_i & g_i \in (g_{n-2}, g_{n-1}) \\ -1 + a - \alpha_i & g_i \in (g_{n-1}, y) \end{cases} \tag{2.7}$$

where $j \neq i$.

We can see from above that: $\frac{\partial U_i}{\partial g_i}$ is decreasing from $g_i = 0$ to y. At the same time, player i will never choose a contribution $g_i \in (g_{n-1}, y]$. If player i has a dominant strategy to contribute 0 token, $\frac{\partial U_i}{\partial g_i}$ has to be smaller than 0, which means $a + \beta_i < 1$.

Consider that there are k players with $a + \beta_j < 1$, who will certainly choose to keep all of their endowments for themselves. Player i with $a + \beta_i > 1$ would not choose a positive contribution if:

$$\frac{\partial U_i}{\partial g_i} = -1 + a - \frac{k\theta_1^i}{\sum_{j=1}^{n-1} \theta_j^i}\alpha_i + \frac{\sum_{j=k+1}^{n-1} \theta_j^i}{\sum_{j=1}^{n-1} \theta_j^i}\beta_i \leq 0$$

$$\Leftarrow -1 + a + \frac{\sum_{j=k+1}^{n-1} \theta_j^i - k\theta_1^i}{\sum_{j=1}^{n-1} \theta_j^i}\alpha_i \leq 0$$

$$\Leftrightarrow \frac{\sum_{j=k+1}^{n-1} \theta_j^i - k\theta_1^i}{\sum_{j=1}^{n-1} \theta_j^i}\alpha_i \leq 1 - a$$

$$\Leftarrow \frac{\sum_{j=k+1}^{n-1} \theta_j^i - k\theta_1^i}{\sum_{j=1}^{n-1} \theta_j^i} \leq 1 - a$$

$$k \geq \frac{\sum_{j=k+1}^{n-1} \theta_j^i - (1-a) \sum_{j=1}^{n-1} \theta_j^i}{\theta_1^i}$$

In proposition three, if it is possible for a player i to make a positive contribution, the derivative of his utility should be non-negative in a small interval around $g_i = 0$, which means if:

$$-1 + a - \frac{k\theta_1^i}{\sum_{j=1}^{n-1} \theta_j^i} \alpha_i + \frac{\sum_{j=k+1}^{n-1} \theta_j^i}{\sum_{j=1}^{n-1} \theta_j^i} \beta_i \geq 0$$

$$\Leftrightarrow k \leq \frac{\beta_i \sum_{j=k+1}^{n-1} \theta_j^i - (1-a) \sum_{j=1}^{n-1} \theta_j^i}{\alpha_i \theta_1^i}$$

where $j \neq i$, is met, there will be an equilibrium with positive contributions. We can also elicit proposition three from our derivative function 2.7.

2.B Additional Tables

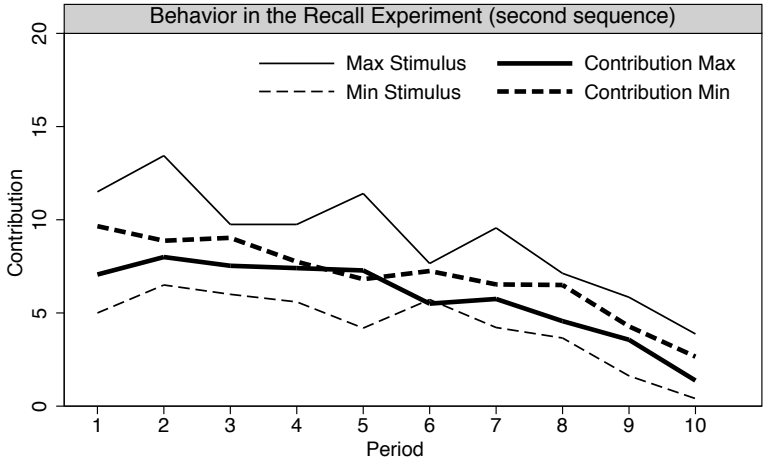

Figure B1: Behavior in the Recall Experiment (second sequence).

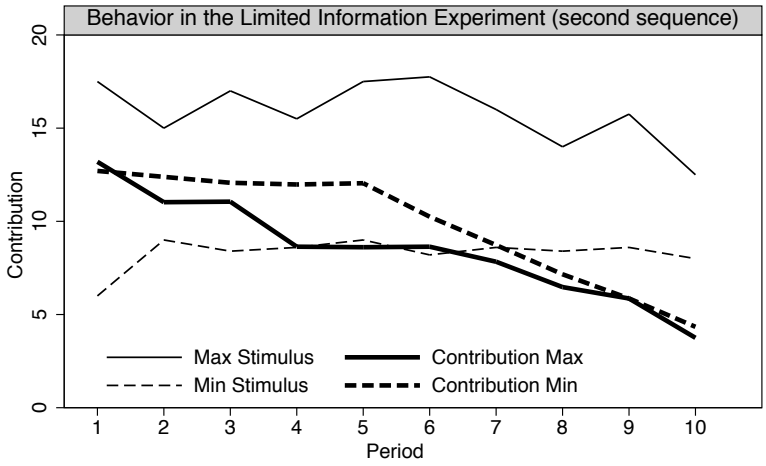

Figure B2: Displayed contributions and behavior in the Limited Information Experiment (second sequence).

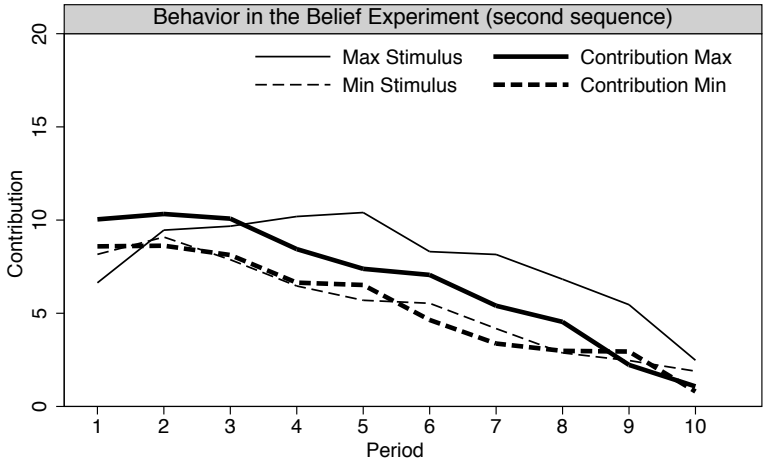

Figure B3: Beliefs and behavior in the Belief Experiment (second sequence).

2.C Translated Instructions

Note. Text in normal font indicates that it was part of the instructions of all three experiments. *Cursive text indicates that the corresponding part was only in some instructions, depending on the experiment.*

Introduction

You are now taking part in an economic experiment financed by various foundations for research. If you read the following instructions carefully, you can depending on your decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. It is prohibited to communicate with the other participants during the experiment. Should you have any questions, please ask us. If you violate this rule, you will be dismissed from the experiment and forfeit all payments.

During the experiment, we will not speak in terms of Swiss Francs, but in Points. During the experiment, your entire earnings will be calculated in Points. At the end of the experiment, the total amount of Points you have earned will be converted to Swiss Francs at the following rate:

$$1 \text{ Point} = 0.08 \text{ CHF}$$

At the end of the experiment, your entire earnings from the experiment plus the show-up fee will be paid to you in cash.

Detailed Information About the Experiment

The experiment is divided into 10 separate periods. In each period all participants are divided into groups of four. You will, therefore, be in a group with 3 other participants. The composition of the groups will stay the same for all ten periods. In the following pages, we describe the experiment in detail.

At the beginning of each period, each participant receives 20 tokens. We call this his or

her endowment. Your task is to decide how to use your endowment. You have to decide how many of the 20 tokens you want to contribute to a project and how many of them to keep for yourself. The consequences of your decision are explained in detail below. At the beginning of each period the following input-screen will appear:

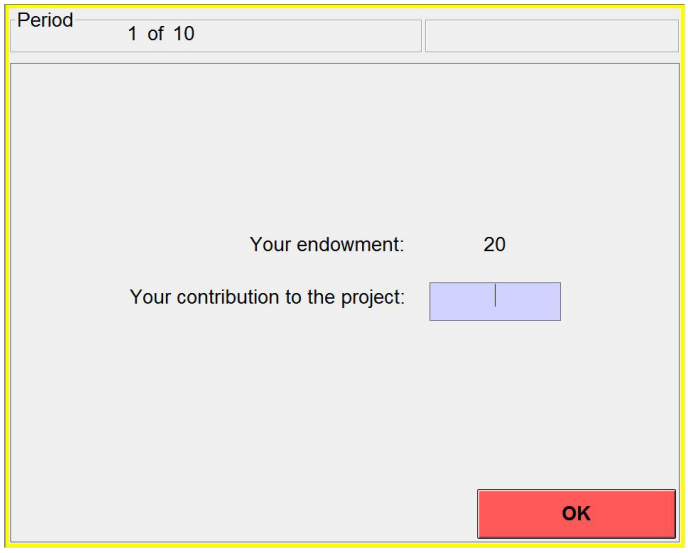

Figure C4: Decision screen in the first and second experiment.

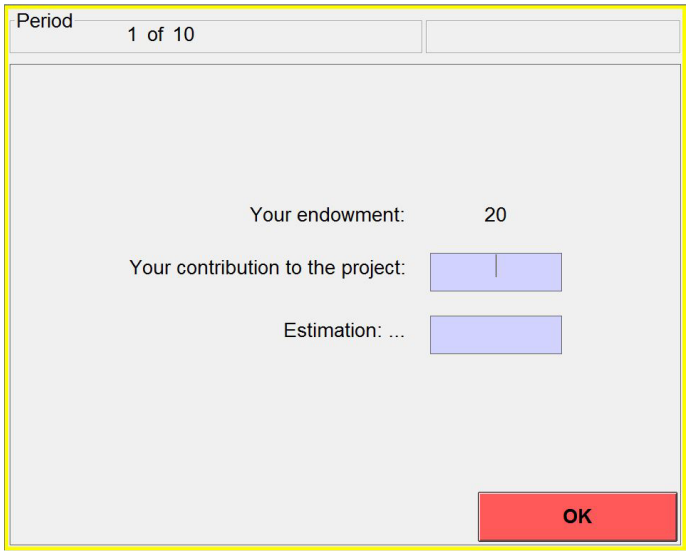

Figure C5: Decision screen in the third experiment.

The period number appears in the top left corner of the screen. In the top right corner, you can see how many more seconds remain for you to decide on your contribution. You will have 90 seconds in the first two periods and 60 seconds in the remaining periods. Your decision must be made within the time limit.

Your endowment in each period is 20 tokens. You have to decide how many tokens you want to contribute to the project by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. As soon as you have decided how many tokens to contribute to the project, you have also decided how many tokens you keep for yourself: This is (20 - your contribution) tokens. After entering your contribution, you must click the O.K. button. Once you have done this, your decision can no longer be revised.

Belief Experiment: *Below the contribution input field there is another input field called "Estimation". You will have to enter an estimation about the behavior of a group member. You will get the detailed information directly on the screen.*

Recall Experiment and Belief Experiment: *After all members of your group have made their decision the following screen will show you the total amount of tokens contributed by all four group members to the project (including your contribution). This screen also shows you how many Points you have earned in this period.*

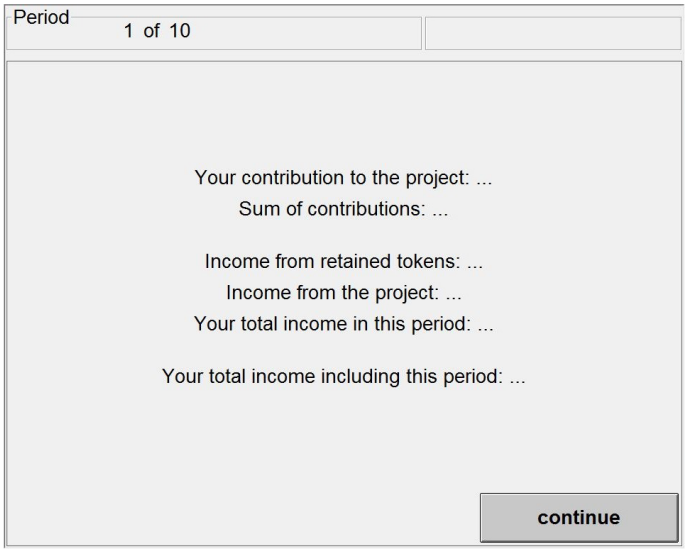

Figure C6: Feedback screen in the first and third experiment.

Limited Information Experiment: *After all members of your group have made their decision you will get some information about the contributions of the other group members.*

Your income consists of two parts:

(1) The tokens which you have kept for yourself ("Income from retained tokens") whereby

1 token = 1 Point.

(2) The "income from the project". This income is calculated as follows:

Your income from the project = 0.4 times the total contributions to the project.

Your income in Points of a period is, therefore:

(20 - your contribution to the project) + 0.4*(total contributions to the project)

The income of each group member from the project is calculated in the same way, i.e., each group member receives the same income from the project. Assume, for example, that the sum of the contributions of all group members is 60 tokens. In this case, each member of the group receives an income from the project of: 0.4*60 = 24 Points. If the total contribution to the project is 9 tokens, then you and all other members of the group receive an income of 0.4*9 = 3.6 Points from the project.

For each token, which you keep for yourself you earn an income of 1 Point. Suppose you contributed this token to the project instead, then the total contribution to the project would rise by one token. Your income from the project would rise by 0.4*1=0.4 Points. However, the income of the other group members would also rise by 0.4 Points each, so that the total income of the group from the project would rise by 1.6 Points. Your contribution to the project therefore also raises the income of the other group members. On the other hand, you earn an income for each token contributed by the other members to the project. For each token contributed by any member you earn 0.4*1=0.4 Points.

In the first two periods you have 45 seconds and in the remaining periods 30 seconds to view this income screen. If you are finished before the time is up, please click the "continue"-button.

Recall Experiment and Belief Experiment: *Next, the information screen appears, which reveals the contributions of the other group members.*

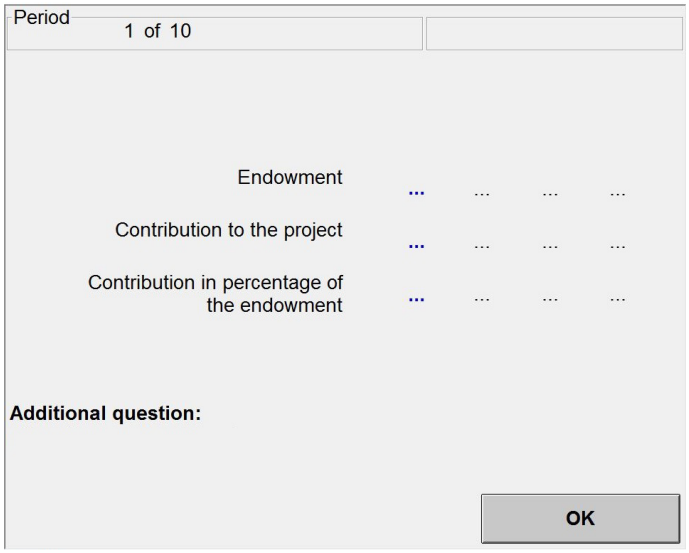

Figure C7: Information screen in the first experiment with the additional question.

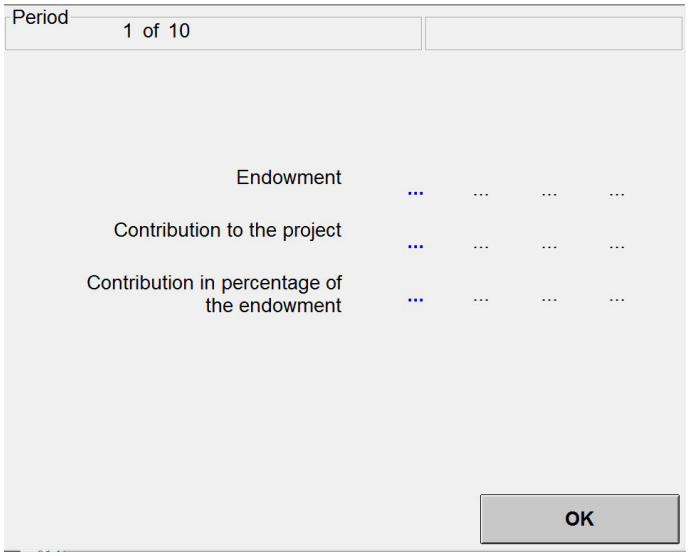

Figure C8: Information screen in the third experiment.

This screen shows how much each of the other group members contributed to the project. Your contribution is displayed in blue in the first column, while the contributions of the other group members are shown in the remaining three columns. Please note that the order in which contributions are displayed is changed randomly in each period. The contribution in the second column, for example, in general, stems always from a different group member. The same holds for the contributions in the other columns. Besides the absolute contributions, the contributions as a percentage of the endowment are also displayed.

Limited Information Experiment: *The profit will be calculated each period, but it will*

not be displayed to you. The profits of all periods will be added up, and you will get informed about your final profit at the end of the last period.

Recall Experiment:

ADDITIONAL QUESTION

In each period you can earn an additional Point by responding correctly to the additional question. You will need to memorize some information about the behavior of the other group members. You will receive the detailed information directly on the screen.
In the following screen, you are asked to enter your response. You will receive an additional Point if you answer the question correctly.

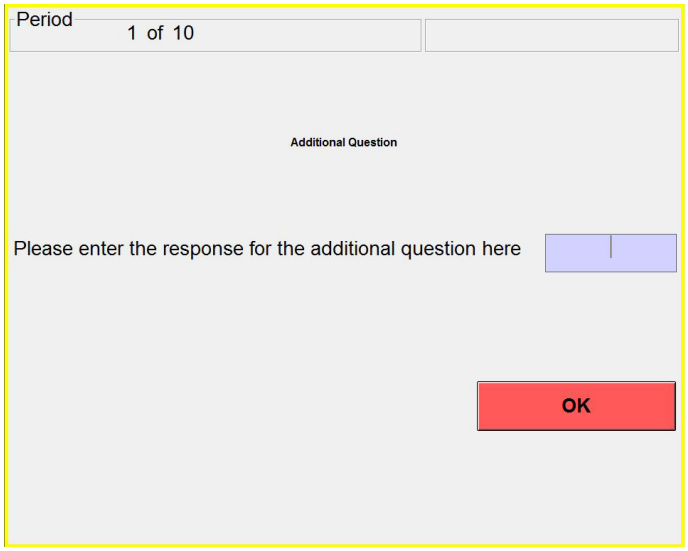

Figure C9: Screen with the additional question in experiment one

Do you have any questions?

2.D Examples of the zTree Screens

Recall Experiment

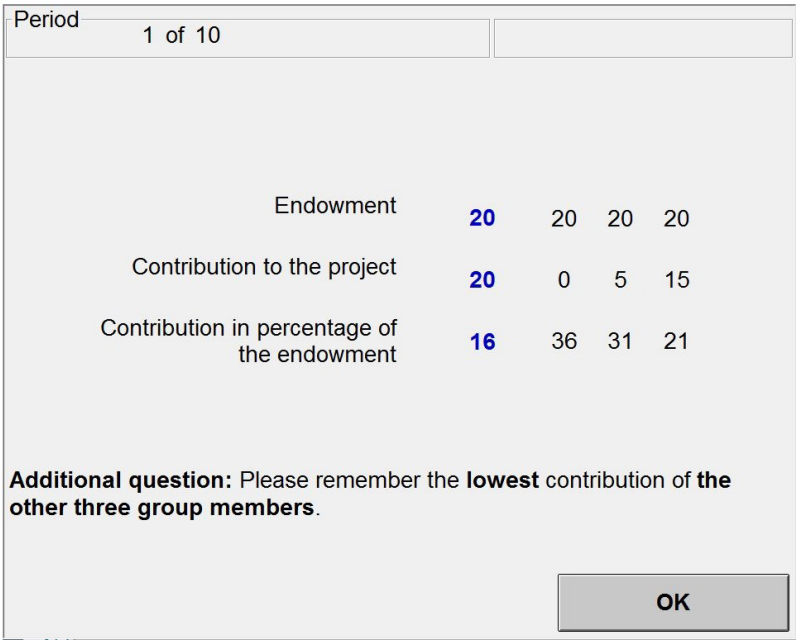

Figure D10: Example of the screen with the additional question [MINIMUM STIMULI]

Limited Information Experiment

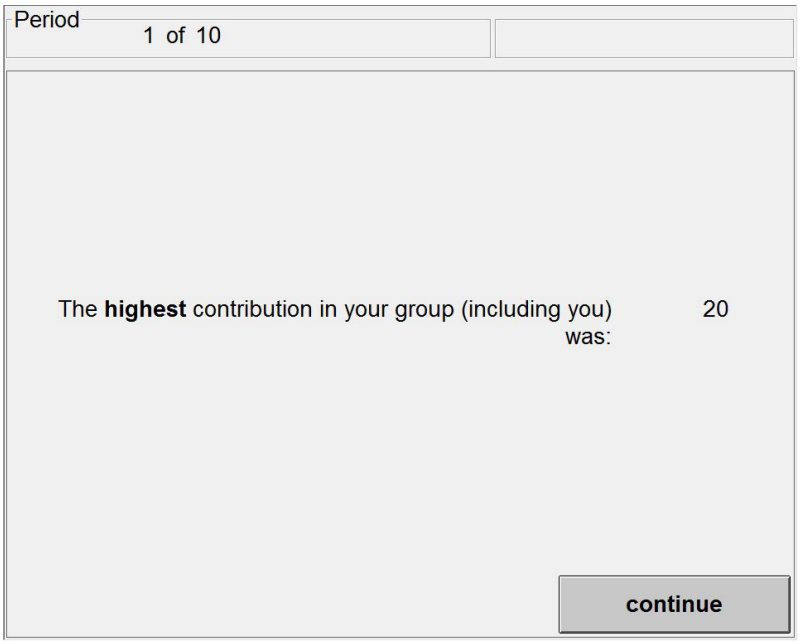

Figure D11: Example of the feedback after the contribution stage [MAXIMUM STIMULI]

Belief Experiment

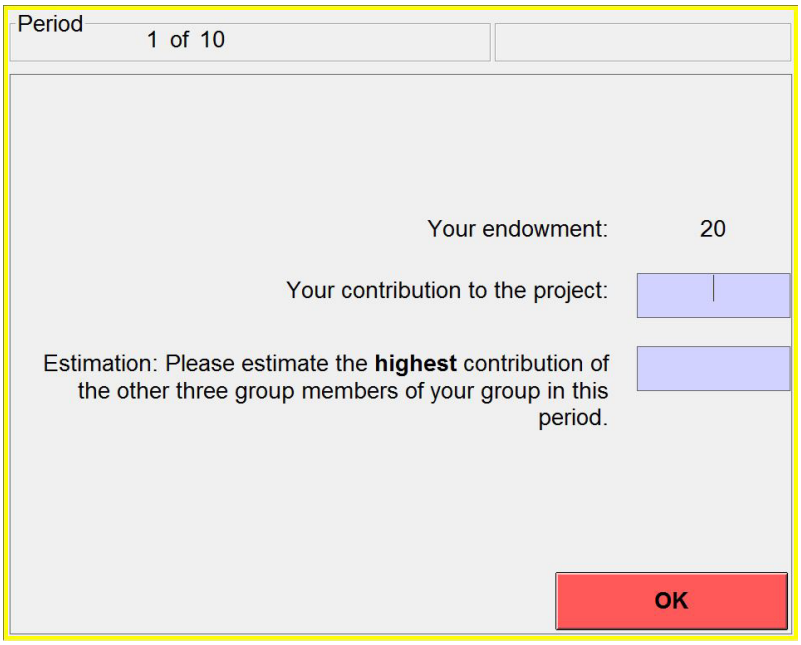

Figure D12: Screen with the additional belief estimation [MAXIMUM STIMULI]

Chapter 3

Positive Effects of High-Quality, Community-Based Child Care. Evidence from a Lab-in-the-Field Experiment in Colombia*

*For helpful comments and discussion, I thank Rafael Lalive, Christian Thöni, Raquel Bernal, Juan Camilo Cardenas, Laure Athias, Isabel Günther, Christian Zehnder, Philippe Ruh, Monika Hess, Laura Metzger, Bruno Martorano, Herbert Elsener, and Vivamos Mejor. Special thanks for the tremendous local support goes to the whole Apoyar Team. I gratefully acknowledge financial support from the Swiss Agency for Development and Cooperation.

3.1 Introduction

In most developing countries, children face risks, which prevent them from developing their full potential (Walker et al., 2011). Many children lack adequate cognitive and non-cognitive stimulation, do not have access to basic health and hygiene, are malnourished, their families have fewer resources, and some even experience violence (Walker et al., 2011). These risks are held responsible for the skill gap which we observe later in life between children from disadvantaged families and children from more advantaged families (Duncan & Brooks-Gunn, 2000). The development of a child in its first five years of life is decisive for her future and depends highly on the quality of its environment and the relationship to parents or other caregivers (Carneiro & Heckman, 2003). Children who grow up with more responsive caregivers, which interact with them frequently and in an affectionate way and who are surrounded by a stimulating environment, have better prerequisites to develop their full cognitive, social and emotional potential (Shonkoff & Phillips, 2000). Therefore, all children, and especially children facing poverty and hardship, are likely to benefit greatly from high-quality preschool education. One of the most prominent examples is probably the Perry Preschool Program; a randomized intervention, which demonstrated that preschool education for disadvantaged African-Americans had long lasting effects. Compared to non-treated individuals, individuals in the program benefited along many dimensions, such as lower crime rates, higher salaries, and a higher probability of graduation (Schweinhart et al., 2005).

Advocates of preschool education as Heckman for example, argue that investment in early childhood development is so efficient because of self-productivity and dynamic complementarity. Self-productivity means that skills acquired in one period persist into future periods of life. Dynamic complementarity refers to the fact that skills produced in one stage rise the productivity of investments in skills in later periods of life (see Heckman (2006) for a short and comprehensive overview).

Since 1990, preschool education and its benefits have gained increasing political attention in many Latin American countries, and governments implemented such programs (Araujo, López Bóo, & Puyana, 2013). The long-term impact of early child care programs

depends highly on the quality of care provided in the respective institution (Barnett, 1995). Bernal et al. (2009) note about a program in Colombia that the quality of care is low and that the knowledge of the caregivers about infant development is not sufficient.

Researchers recognized that the quality of care is crucial, nevertheless, the literature on how to effectively improve it is scarce. Moreover, most of the existing studies focus on the immediate effects, and the question whether the effects persist in the long run remains open often (Behrman, Fernald, & Engle, 2013). Three of the rare studies evaluating the long-term effect of improving the quality of preschool education in the context of developing countries are Raine, Mellingen, Liu, Venables, and Mednick (2003), Yoshikawa et al. (2015), and Özler et al. (2016). Raine et al. (2003) observe that children who benefited from an enriched environment at the age of 3 to 5 display less antisocial behavior at puberty age and are less prone to develop schizotypal personalities. Yoshikawa et al. (2015) evaluate a quality enhancing intervention in Chile. Yoshikawa et al. (2015) do not observe that the intervention improved the schooling outcomes of the children. The authors conclude that more intensive curricular approaches are needed to ensure that children benefit in the long run. Eventually, Özler et al. (2016) elaborate an intervention in Malawi. They report that improving teacher quality in combination with parental education had positive effects on the children in the short run: their language skills improved and they exhibited more prosocial behavior.[1] However, 36 months after the treatment, the effect is insignificant.

Engle et al. (2011) propose that the quality of preschool education in low and middle-income countries can be improved by appropriate design, adapting the curriculum, including practice for parents, training for childcare workers, monitoring and assessment, improved governance, and supervision. These recommendations have been taken into account by Apoyar, a local Colombian non-governmental organization (NGO) when designing an intervention to improve the quality of care offered in *hogares comunitarios*, a popular preschool program in Colombia. It is a community-based program, where so-called community mothers provide day care services to children in their proper home. The intervention to improve the quality lasted for one year and consisted of four components. First, com-

[1] Prosocial behavior in this study was measured using the Strengths and Difficulties Questionnaire, which was filled out by the primary caregiver of each child.

munity mothers received a formal vocational training as early childhood teacher. Second, they received support and coaching to integrate what they learned in their daily nursery routines. Third, they learned how to teach parents about issues related to appropriate child care and development. Fourth, the implementing NGO monitored and supported the children in school, once they left the community mother. Hence, this intervention targeted community mothers, children, and their parents.

The main aim of this paper is to elaborate if this quality improvement of preschool education has measurable effects on a broader set of outcomes of the children. The paper adds to the relatively small body of literature on the effects of enhancing the quality of preschool education in developing countries. Moreover, in the literature on the effectiveness of early childhood interventions, researchers traditionally evaluated the effect on weight, height, cognitive skills, socio-emotional development, and proxies for success in life such as crime rates or income. I am convinced that it is important to expand the analysis to individual preferences because the literature shows that these preferences are good predictors of different economic and health outcomes such as labor market participation, saving decisions or health choices (Heckman & Rubinstein, 2001; Dohmen et al., 2011; Becker, Deckers, Dohmen, Falk, & Kosse, 2012; Sutter, Kocher, Glätzle-Rützler, & Trautmann, 2013; Falk et al., 2017). Furthermore, research on the formation of these preferences shows that they form in early childhood (Fehr, Bernhard, & Rockenbach, 2008; Angerer, Glätzle-Rützler, Lergetporer, & Sutter, 2015). Thus, it is important to evaluate how the social environment such as preschool education affects the formation of these preferences. Moreover, assessing the improvement along a number of child development outcomes allows drawing a more complete picture of the potential benefits of high-quality preschool education.

To assess how the program affected the children, I use two different datasets. First, I analyse a dataset which contains monitoring data elicited by the implementing NGO. Second, I run lab-in-the-field experiments four years after the implementation of the intervention. I elicit the preferences of the children and compare children who benefited from the intervention with children who left for school just before the intervention started.

I find that enhancing the quality of the existing preschool program leads to large skill gains in the short-run which can be observed four years after the intervention. Immediately

after the treatment, children who visit a treated community mother display improved cognitive, psychosocial, and psychomotor skills compared to children from the control group. Four years after the intervention, children who visited a quality improved hogar perform better in school, are more likely to be in the right grade, and exhibit higher levels of social preferences compared to children from a standard hogar. The evaluation design set-up does not allow to fully rule out that unobservable factors drive the effects. However, I conduct different robustness checks, which provide further support for the observed results. Moreover, I cannot identify which component or combination of components of the intervention drives the results. Nevertheless, I am confident that my results add to the literature on how high-quality preschool education affects the cognitive abilities of children and the formation of their preferences.

The chapter continues as follows: Section 3.2 provides a detailed explanation of the intervention. In section 3.3 I outline the conceptual framework. In section 3.4 I present the study design and discuss the strategy for the data analysis. Section 3.5 contains the results and section 3.6 concludes.

3.2 Intervention

The main aim of the intervention under consideration was to improve the quality of care in public community nurseries (called hogares comunitarios or simply HC) in Colombia. The Colombian state agency ICBF (Instituto Colombiano de Bienestar Familiar) introduced these HC in the mid 1970ies. An HC is run by a so-called community mother (MC) and targets children in preschool age (from birth up to entering school). HCs target families from the lowest socioeconomic stratum. The task of the community mother is to attend children during the day, cook them lunch and prepare the children for their schooling career. She is allowed to take care of up to 14 children in her proper home. The state subsidizes the HCs by paying the salary of the MCs and the food for the children. Parents have to pay a minimal fee (around 5 USD per month) to bring their children to an HC. Today this program is the most widespread preschool program in the country reaching roughly a million children by 69'000 MCs. Most often children who attend an HC are from

very poor families.[2]

Bernal et al. (2009) reports that the community mothers, have often no more than primary education and no particular training in providing childcare services.[3] In response to that, the Fundación Apoyar[4] implemented an intervention to improve the quality of care with the ultimate goal of improving the situation of vulnerable children in Colombia.[5] Apoyar implemented the complete program in collaboration with the local state agency ICBF.

They implemented the intervention *professional and comprehensive community child care in Las Ferias* in the city of La Dorada, department of Caldas. More specifically, the project was rolled out in Las Ferias, a densely populated neighborhood with about 25'000 inhabitants and a high incidence of poverty and many internally displaced families. Children in this neighborhood grow up in home environments with only limited learning opportunities. Moreover, children are often left unattended. Hogares have therefore the potential to improve the living conditions of these children considerably.

The main objective of the intervention by Apoyar was to increase the quality of care offered in these HCs. The program offered an overarching bundle of education and support for the community mothers, the parents, and their children. It included four main components. The core of the intervention consisted of a vocational education training for the community mothers. Two private educational institutions[6] in collaboration with the ICBF provided the formal training, which lasted for one year. During that year, MCs learned how to provide age adapted childcare services so that they prepare the children adequately for school. Community mothers who finished the training received an official degree in early childhood teaching.

The second component of the intervention was assistance of the community mothers in

[2]The families of the children of this study are all from the lowest stratum according to the national classification SISBEN

[3]Baseline data collected by Apoyar shows that more than half of the children had uncompleted vaccination cards, and one-third lacked access to health care and civil registration. Furthermore, many children suffered from malnutrition.

[4]Apoyar is an NGO dedicated to support the development of children and their families through community projects. Apoyar works mainly in two regions in Colombia.

[5]The Swiss foundation Vivamos Mejor funded the intervention.

[6]Instituto de Conocimiento para el Trabajo y el Desarrollo Humano (ICT) and Instituto Unidades Tecnicas de Boyacá (UTB)

their household.[7] During the treatment year, Apoyar supported the MCs to implement the national pedagogical model in her daily nursery routine. Trained pedagogues of Apoyar visited the HC regularly and provided coordinated coaching to reinforce the strategies and concepts.

Third, Apoyar supported the MCs with the organization of parental workshops. The ICBF obliges the community mothers to hold monthly parental workshops, where they discuss different topics related to care, education and development of children. Apoyar invited the community mothers regularly to discuss the topics themselves and to discuss how to transmit them to the parents efficiently. Furthermore, Apoyar provided the community mothers with well-elaborated materials to organize and hold the parental workshops. For each topic, they designed a clear schedule with hands-on teaching activities and discussions of the issues with the parents. Issues considered in these workshops include lactation, illnesses of children, nutrition of children, importance of early childcare education, relationship between parents and their children, children's rights, sexual education and education about the rights in case of sexual abuse, childbearing guidelines, diminution of aggression within a family, quality time with the family, and so on.

The fourth component of the intervention was the supervision of the children who left for school. Apoyar supervised the school attendance of these children and provided them with extra classes if needed through a mentor who was an intern at Apoyar.

Due to limited implementation capacities from the side of Apoyar, the program was implemented step-wise. In 2011, 20 community mothers started with the program. In 2012, Apoyar extended the program to a further set of 20 community mothers, and in 2013 the last set of 20 community mothers received the treatment. Selection into cohort was not random, but it did not follow a clear pattern either. The local ICBF agency, which approved the program, created the order of implementation in agreement with Apoyar. It is important to note that the community mothers were not able to opt-in, and none opted out of the program. Moreover, according to Apoyar children did not change the HC because of the treatment. Still, assignment to cohort was not random, and hence there

[7] This component has been shown to be effective to improve child outcomes in developed countries (Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010).

might be imbalances at baseline. I will elaborate these potential imbalances at baseline in detail in subsection 3.4.3.

3.3 Conceptual Framework

In this section, I outline the channels through which the intervention potentially affected the outcomes of the children.

The formal education in combination with the coaching, which ensured that the community mothers implemented what they have learned in their daily nursery routines, can be beneficial for the children for three reasons. First, the infants receive the required cognitive and non-cognitive stimuli, and importantly these stimuli correspond to their age. Thus, they are in a more appropriate learning environment, which fosters their development. Second, the MC will be able to recognize malnutrition or other health issues and act upon them. This will improve the nutritional and basic health condition of the children, which in turn can affect their cognitive development (cf. Glewwe, Jacoby, & King, 2001; Walker et al., 2007; Grantham-McGregor et al., 2007; Martorell et al., 2010). Lastly, harsh parenting styles are still common in many Latin American countries (cf. Fontes, 2002; Ramírez, 2006). The formal education provides the community mothers with knowledge about child care, which should translate to a more benign parenting style. Treating children more affectionate can have a positive effect on their psychosocial development. The literature on the role of parenting styles on child development notes that harsh parenting styles relate to children behaving more anti social (Roopnarine, Krishnakumar, Narine, Logie, & Lape, 2014). Moreover, Carlo, Mestre, Samper, Tur, and Armenta (2011) report that parental warmth is predictive for prosocial behavior of children.[8] Hence, each of the three mechanism affects the child separately, but presumably the amplify each other as well.

As part of the coaching component, the pedagogues of Apoyar visited all community mothers regularly. Thus, they were also in direct contact with the children and the children experienced affectionate treatment from them too. This might reinforced the effects

[8]Note, these are correlational studies.

mentioned above.

The third component of the intervention were the parental workshops, which aimed at improving the knowledge of the parents. Through this element, I expect a further benefit for the children. The mechanisms are the same as for the community mothers. Because parents have more knowledge about infant development, they provide children with more appropriate stimuli, they value health and nutrition of their children more, and they employ more affectionate parenting styles. These improvements should lead to ameliorated parent-child interactions. According to Engle et al. (2011), stable parent-child interactions are responsible for a better health status of the child, increase attachment and encourage learning through appropriate stimuli. Analog to the effect via the community mother, these improvements might affect not only the cognitive development of a child but also its preferences.

Lastly, I expect a positive impact on the children through the monitoring component of the intervention. Children with learning difficulties in school received extra classes. Obviously, this should lead to better schooling performances. Moreover, these children received more attention; this can affect their psychosocial development.

Traditionally, evaluations of such programs have focused on the development of cognitive skills. Some studies evaluated the effect on psychosocial development, but to the best of my knowledge, none has assessed the effects on the preferences of children. As outlined above, I expect the intervention to affect children on different developmental dimensions such as their cognitive skills as well as their individual preferences. Therefore, it seems crucial to extend the analysis to social preferences, trust, risk and time preferences.

Lastly, it is important to evaluate if and how the intervention affected the parents.[9]

In the next section I will describe the monitoring data (short run) which I received from Apoyar, and the data I collected additionally, to analyse the effects of the intervention four years after the treatment.

[9]Obviously it would also be interesting to assess how the intervention affected the community mothers. However, this research has been planned ex-post and hence all community mothers in the program are treated by now. Thus, I cannot elaborate in a meaningful way how the knowledge of the community mothers improved.

3.4 Study Design and Data Analysis Strategy

I start with the description of the design and the analysis of the monitoring data (short run). Subsequently, I will explain the design and the analysis of the grades and the preference measures (longer run). Note, I used two different datasets for these two parts of the analysis. Hence, the control and the treated group of the monitoring dataset do not correspond to the control and the treatment group for the analysis of the grades and the preference measures.

3.4.1 Monitoring Data

During the treatment, Apoyar collected data to monitor the developmental status of the children in the project. They used the *escala abreviada de desarrollo* which is a standardized survey[10] widely used in Latin America to measure the developmental status of children from birth to the age of six. The original questionnaire contains four dimensions.[11] UNICEF adopted the measure to the Colombian context. It comprises four slightly different dimensions; literacy-numeracy, fine and gross psychomotor skills, and psychosocial abilities. Apoyar merged the two indexes for psychomotor skills into one. Hence, they evaluate the developmental status of the children in the project with three indexes: cognitive development, psychosocial development, and psychomotor development. All three indexes consist of several items, which are sorted by the ability for age. All items are rated on a binary scale: fulfilled or not fulfilled. For example, the first item on the index for psychosocial development measures whether the child follows with the eyes, movements in front of her face. The examiner then judges whether the child fulfills the item or not. Depending on the age in months, a child has to be able to fulfill a given number of items. For instance, if a child is two months old, it should be able to fulfill the first four of the 31 items of the cognitive index. If a child is ten months old, it should reach eight additional items, and so on.

For each index and child, Apoyar collected the number of items, which the child sat-

[10] UNICEF created and validated this survey based on the review of existing tests.
[11] The original dimensions are literacy-numeracy, physical, social-emotional, and learning.

isfied.[12] The index for literacy-numeracy measures the cognitive development of a child. It starts with whether a child reacts to movements and sounds and ends with whether a child can name all days of the week properly. I listed all items of the three indexes in the appendix. The psychosocial index measures the emotional and social status of the child. Lastly, the psychomotor index assesses the gross and fine motor skills of the children.

During the intervention, Apoyar elicited each index three times. At the very beginning, they collected the so-called baseline measurements, i.e. the current developmental status of the children before the treatment. For the first cohort (i.e. the first 20 treated community mothers) the baseline data is from January 2011. For the second cohort, they gathered the data in January 2012 and for the third cohort at the beginning of 2013. During the treatment, they collected a second measurement of the three indexes. At the end of each treatment period, Apoyar elicited the endline measurement of the three indexes. Hence, for all children who have been in one of these community nurseries, I have data on their cognitive, psychosocial, psychomotor status for three points in time. Furthermore, the monitoring dataset also contains information about the age and the gender of the child and data about the socioeconomic characteristics of all treated community mothers.

Estimation Strategy for the Monitoring Data

As outlined in the previous section, Apoyar collected the monitoring data at the beginning and during the treatment. This means that for the three cohorts the overlap of the measurement is displaced. Hence, it is not possible to estimate the effect with a difference-in-difference design. Therefore, I use the baseline measurements of the second cohort as a control group for the effect of the treatment on the first cohort (sample 1). Analogously, I use the baseline measurements of the third cohort as a control group to estimate the treatment effect on the second cohort (sample 2). I call the control group from both samples the *Monitoring Data (MD) control group* and correspondingly the treatment group the *MD treatment group*. Figure 3.1 depicts the situation graphically.

[12] The data only contains the total number of items, which a child reached, no information on the item level.

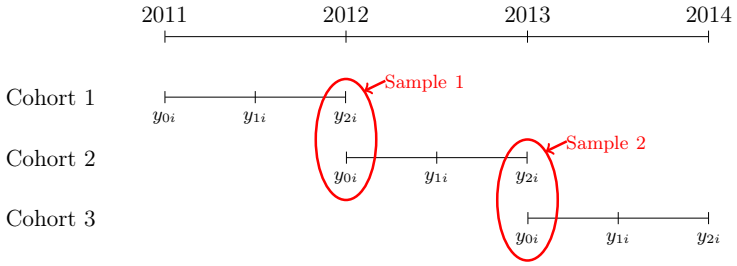

Figure 3.1: Implementation of the treatment by Apoyar.

Note. Apoyar implemented the treatments stepwise. The first cohort was treated in 2011, the second in 2012 and the third in 2013. They collected monitoring data only during the treatment.

Furthermore, I restrict the sample of children for the first sample, to children born in the years 2007 up to 2009, because on average children in the first cohort are nine months older at baseline than children in the second cohort.[13] At the point of comparison (January 2012), children in cohort 1 are roughly two years older on average than children of the second cohort. Since the number of potential items of the developmental indexes increases with age, including all children would distort the measurement. One item less at a younger age leads to a much bigger drop in percentage points than if a child scores one item less at the age of five for example. Likewise, I restrict the second sample, i.e., cohort two versus cohort 3, to children born in the years 2008, 2009, and 2010.

I pool the two samples to estimate ordinary least squares regressions and include all available control variables to increase the precision of the estimates. The regressions take the following form:

$$y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \beta_4 MC_{mi},$$

whereby y_i is the outcome variable of interest of child i. T_i is the dummy variable which indicates the treatment status, i.e., it is set to one to indicate that the child was in the MD treatment group and it is zero if the child belonged to the MD control group. Hence, β_1 captures the treatment effect. X_i is a vector of variables which includes the demographic characteristics of the child and the cohort. Additionally, MC_{mi} is a vector

[13] I depict the characteristics at baseline of the complete sample and the community mothers in Table B1 in the appendix.

of variables including the demographic variables of the respective community mother m. Furthermore, in all estimations, I cluster the standard errors at the level of the community mothers.

The exclusion of the oldest and youngest children allows controlling for the age effect. However, I cannot disentangle the treatment effect from a potential time effect.

To verify the results of the OLS regressions, I will conduct two additional robustness checks. First, I will estimate the effect using inverse probability weighted regression adjustment. Second, I will elaborate the relative degree of selection using the methodology proposed in Altonji, Elder, and Taber (2005) and Oster (2014).

3.4.2 Grades and Preference Measures

Apoyar did not continue to collect systematic information about the children after the treatment. Thus, to evaluate whether the program had persistent effects, I had to collect new data.

The children who started school immediately after the intervention are the most long-term that I can evaluate. Today, these children are in third or fourth grade, meaning that they were in one of the hogares three or four years ago.

The *escala abreviada de desarrollo* which Apoyar uses in the monitoring is eligible for children up to the age of six. Hence, I cannot apply the same measurement instruments to evaluate their development. To assess the effect on the cognitive abilities, I collected the grades of the children. I described in the introduction that individual preferences are important from an individual and societal perspective. I will explain them in more detail in the next paragraph. Let me first elaborate shortly on the methodology of experiments.

We conducted the lab-in-the-field experiments to elicit the individual preference measures and self-esteem with the children in their school. The data collection took place in July and August 2016. Ahead of the data collection, we let all parents of the children sign an informed consent.[14] In each school, we set up a room with three to five separate tables. We assigned each child randomly to one of the five Colombian research assistants who then

[14]The ethics committee of the University of Lausanne approved the study. In the informed consent, parents received a short information about the aim of the study, the treatment of their own and their children's data, and the information about the possibility to quit the study at any given point in time.

conducted the experiment with the child in private.[15] Following the methodological standards in behavioral economics, we incentivized all decisions of the children. We prepared a small kiosk with gifts and informed the children at the very beginning that they would earn stars during the experiment, which they can exchange at the end for little presents. All items in the kiosk had a price in stars.[16] An example of how such a kiosk looked like is in the appendix (Figure C4).

Moreover, to analyze the effect of the program in more detail, I collected data of a randomly selected sub-sample of parents. As outlined in the contextual framework, the intervention potentially also affected the parents via the enhanced bi-monthly parental workshops. Therefore, by collecting information about the parents, it might be possible to shed light on different channels through which the intervention affects children. Thus, we elicited sociodemographic characteristics of the parents, as well as a measure of their parenting style, and the same individual preferences as collected for their children. We conducted the experiments with the parents at their proper home.[17] Likewise to the kiosk for the children, we incentivized their decisions in the experiments with money.

In the next paragraphs, I will explain the elicitation methods for the outcome variables in more detail. I will start with the explanation of the individual preferences measures.

Social preferences

Social preferences, measure the degree to which a person cares about the welfare of others. They are a strong motive for behavior, and in modern societies, the existence of social preferences is of eminent importance. Social preferences facilitate cooperation among unrelated individuals especially in situations with asymmetric information and incomplete contracts (Bowles, 2004). Furthermore, social preferences are a predictor for charitable giving and volunteering activities (de Oliveira, Eckel, & Croson, 2012). So far only limited evidence exists on how the family situation and the social environment affects the formation of social

[15] The protocol of the experiment is in the Appendix.

[16] We ensured that all children would earn in any case enough to buy the most expensive item in the kiosk to avoid that their behavior is biased due to the need of a certain item.

[17] Due to limited capacities, we were not able to elicit the measures for all parents. Therefore, we randomly selected 48 parents from each sample. Although it was hard to coordinate the interviews, all of the selected parents participated.

preferences. Deckers, Falk, Kosse, and Schildberg-Hörisch (2015) demonstrate that children from high-income families in Germany behave more prosocial compared to children from a low-income household. Kosse, Deckers, Schildberg-Hörisch, and Falk (2016) provide evidence that a mentoring program in primary school can close the gap in the level of social preferences of children from low socioeconomic status families compared to children from high socioeconomic backgrounds. Moreover, Cappelen, List, Samek, and Tungodden (2016) are the first to provide causal evidence that a preschool program led children to have more egalitarian fairness views. To the best of my knowledge, no research exists on how preschool education affects social preferences in the context of developing countries.

I elicited social preferences with two measurements. First, with four binary allocation tasks, and second, with a standard dictator game. For the four binary tasks I followed the procedures by Fehr et al. (2008) and Bauer, Chytilová, and Pertold-Gebicka (2014). In all four tasks, subjects had to decide between two allocations that assign resources to oneself and another recipient. For the children, this was another child from the same school. For the parents, the recipient was another parent. We explicitly mentioned that we assigned them a new recipient in each of the four allocation tasks.

In the costless prosocial task, subjects had to choose between the egalitarian choice of (1,1) and the allocation (1,0). The egalitarian choice implies one experimental currency unit (ECU)[18] for the decision maker and one ECU for the recipient. In contrast, the (1,0) allocation implies one ECU for the decision maker and nothing for the recipient. In this task behaving fair came at no cost for the decision maker. The costly prosocial task offered the opportunity to choose between (1,1) and (2,0). In this task, being prosocial comes at a cost, i.e., the decision maker has to give up one ECU to implement a fair allocation. The so-called costless envy task consisted of the following two allocation possibilities; (1,1) vs. (1,2). In the fourth task, the costly envy task, decision makers had to decide between (1,1) and (2,3). Being envious in this task means that the decision maker forgoes one ECU to establish equality.

I randomized the order of the four tasks to control for order effects. Depending on

[18]For the children one ECU corresponded to one star. For the parents, one ECU corresponded to 100 COP.

their choices, individuals can be categorized to be either altruistic, egalitarian, spiteful or selfish. Furthermore, individuals classified as selfish can be sub-classified as being weakly egalitarian, weakly altruistic or weakly spiteful. Table 3.1 describes the classification.

Table 3.1: Definition of types.

	Game			
Type	costly prosocial (1/1) vs. (2/0)	costless prosocial (1/1) vs. (1/0)	costly envy (1/1) vs. (2/3)	costless envy (1/1) vs. (1/2)
Egalitarian	(1/1)	(1/1)	(1/1)	(1/1)
Altruistic	(1/1)	(1/1)	(2/3)	(1/2)
Spiteful	(2/0)	(1/0)	(1/1)	(1/1)
Selfish	(2/0)	∼	(2/3)	∼
Weakly egalitarian	(2/0)	(1/1)	(2/3)	(1/1)
Weakly altruistic	(2/0)	(1/1)	(2/3)	(1/2)
Weakly spiteful	(2/0)	(1/0)	(2/3)	(1/1)

Note. Definition of types according to the choices in the four games. (1/1) stands for the egalitarian option of one ECU for oneself and one ECU for the other person. Whereas for example (1/0) indicates one ECU for oneself and zero ECU for the other person.

The second measure of social preferences is a standard dictator game. Subjects were endowed with 10 ECU and had to decide how to split it between themselves and another new recipient. The amount which they allocate to the other person is a proxy for their altruism.

Trust

Besides social preferences, trust is another important factor for efficient interactions among strangers. Research links trust to various beneficial outcomes for societies. Cardenas and Carpenter (2008) report that developing countries with higher levels of trust exhibit higher growth rates, less poverty, less unemployment, and a more equal division of economic gains.

To measure trust, I followed the protocol proposed by Evans, Athenstaedt, and Krueger (2013). We paired subjects in a group of two, the trustor and the trustee. Analog to the social preferences, we told the children that the other person was a child from the same school and for the parents, it was another parent. The trustor receives one ECU, the trustee did not get any endowment. In the first step, the trustor had to decide to trust and transfer the endowment to the trustee or not to trust and keep the endowment. In the

case of trust, the trustee will receive four ECUs. Then, in a second step, the trustee has to decide whether or not to transfer two ECUs back to the trustor. I applied the strategy method; children and parents first decided in the role of the trustor and then in a second decision they decided in the role as trustee. The proxy for trust is the behavior of the trustor.

Time Preferences

Another relevant measure are time preferences. Research shows that more patient individuals attain higher levels of education and earn higher wages (Golsteyn, Grönqvist, & Lindahl, 2014). Moreover, patient individuals are also more likely to be in a better health condition (Sutter et al., 2013). Research on the formation of time preferences indicates that children who grow up in low-income households display lower levels of patience (Deckers et al., 2015).

I followed the procedures proposed by Angerer, Glätzle-Rützler, et al. (2015) to measure time preferences (patience). In this task, subjects received five ECUs. They had to decide how many ECUs to consume immediately and how many ECUs to invest into the future. Each ECU which they choose to invest was doubled and paid out two weeks after the experiment. To make this future payment credible for the children, the children had the opportunity to select the present immediately after the experiment. However, we handed it to their teacher, who would give it to them only in two weeks time. The amount invested in the future serves as a measure of individual patience.

Risk

Finally, the literature on risk preferences provides evidence that risk seeking individuals are more likely to smoke, to be self-employed, and to invest in the stock market (Dohmen et al., 2011). In contrast to the other preferences, being more risk seeking is not only beneficial. Nevertheless, it is important to evaluate if and how preschool education affects the formation of risk preferences.

We elicited risk preferences using the measure proposed by Gneezy and Potters (1997). Subjects had an endowment of 5 ECU. They had to decide how many of these ECU to

invest into a risky lottery. The lottery paid the triple of the investment with probability 0.5 and nothing with probability 0.5.

For the children, we used a wheel of fortune (see Figure C6 in the appendix). Half of the area was colored and indicated the winning area. If the wheel stopped in this area, they received the triple amount of their investment. We use the amount invested as a proxy for risk preferences. The expected value of investing is higher than the value of not investing. Hence, even a risk-neutral subject should invest the full amount. A risk averse person might invest less, depending on her degree of aversion.

Self-confidence

To estimate self-confidence, all children had to fill out the standardized EDINA questionnaire (Mérida, Serrano, & Tabernero, 2015). It contains items such as "I like my body", "I am an important child", and so on. For all items, children had to answer on a three-point Likert scale (no, somewhat, yes). Besides a measure for general self-esteem, the questionnaire allows evaluating the self-esteem of a child with respect to his body, the family, friends, school, and socio-emotional well-being.

Parenting style

For the parents, we obtained a measure of their parenting style. We employed a short and validated version of the full survey developed by Robinson, Mandleco, Olsen, and Hart (1995). Based on the seminal work by Baumrind (1971) it consists of 18 items and allows to characterize the parenting style into authoritative, authoritarian, and permissive. Authoritative means that parents are demanding and responsive. Meaning that these parents usually have high expectations for their children, but at the same time, they provide their children with the support which they need to succeed. Authoritative parents listen to their children, encourage independence, and employ fair and consistent rules. Furthermore, they allow children to express their own opinions and to discuss them. Authoritarian parenting style relates to a strict, rather controlling parenting style. Authoritarian parents focus on adherence and control. They use punishment with little or no explanation. Lastly, permissive parenting style signifies that parents accept all types of behavior of their child.

Such parents usually employ little or inconsistent rules and provide little structure for the child.

Estimation Strategy for the Effect on the Grades and the Preference Measures

To evaluate whether the treatment had persistent effects I had to create a control group ex-post. In Colombia, the school year starts at the beginning of a year. Apoyar started the treatments also at the beginning of the respective year. We used this incidence to create the so called *Grades and Preferences Measures (GPM) control group*.[19] We identified children who started school in 2012 from cohort 2 and 3. These children have been in a hogar, which had not yet received the treatment. We do the same for children who started school in 2013 from one of the hogares in the third cohort. The *GPM treatment group*, in contrast, consists of children from cohort one and two who started school in either 2012 or 2013. Hence, children in the GPM control and GPM treatment group started school at the same time, and were in comparable conditions, if we accept the assumption that the hogares in the three cohorts did not differ systematically apart from being treated or not. In Figure 3.2, I illustrate the construction of the GPM control and treatment group.

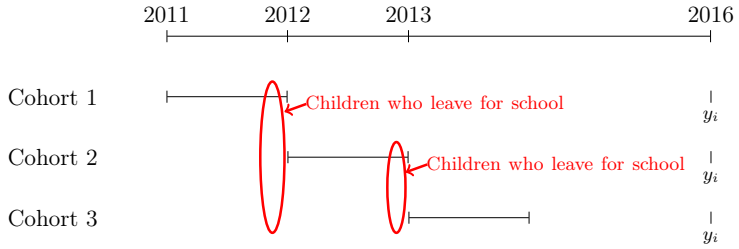

Figure 3.2: Construction of control and treated group.

Again, I estimate the impact of the treatment using regressions of the following form:

$$y_i = \beta_0 + \beta_1 T_i + \beta_3 X_i + \beta_4 MC_{mi} + \beta_5 P_{ji},$$

where y_i is one of the outcome variables explained in the previous section. T_i is the dummy variable for the treatment; consequently, β_1 captures the effect of the treatment.

[19] Note the GPM control group does not contain the same children as the MD control group.

X_i is a vector of variables controlling for the socioeconomic characteristics of the child. Furthermore, I include control variables to adjust the estimate for observable characteristics of the community mother (MC_{mi}) and for the subset where we collected information about the parental characteristics, I include socioeconomic characteristics of the parents as well (P_{ji}). I cluster standard errors at the level of the school and then within a school on community mother to account for within-cluster dependencies.

Anew, I will employ the two robustness checks to provide further support for the results from the OLS regressions.

3.4.3 Imbalances at Baseline

My proposed estimation strategy depends crucially on the assumption that the community mothers in the three cohorts do not differ systematically with respect to unobservable characteristics. Thus, before I turn to the results, I want to render evidence that my estimation strategy is a valid approach with an additional dataset.

The METRIX dataset contains information on the development status of the children attended by a community mother in the city of La Dorada in 2006.[20] In this dataset, I can identify 31 community mothers, which participated in the program of Apoyar. I can identify 13 MCs from the first, ten from the second, and nine from the third cohort. The dataset contains information about the ratios of weight to height, height to age, and weight to age of all children attended by these community mothers.[21] Each ratio is rated on a scale from one to four; one indicating adequate status and four indicating critical status. In Figure 3.3, I depict the ratios of the children attended by one of theses community mothers across the three cohorts.

[20]The METRIX dataset is a national database which stores the information collected by the local ICBF on the development of the children. All community mothers are obliged to gather data about the children in their hogar. They have to submit this data to the local ICBF agency. Unfortunately, the only available data from the ICBF agency of La Dorada is the dataset at hand which is from 2006.

[21]The dataset does not include the height, weight, and age. It only contains the evaluation of the ratio.

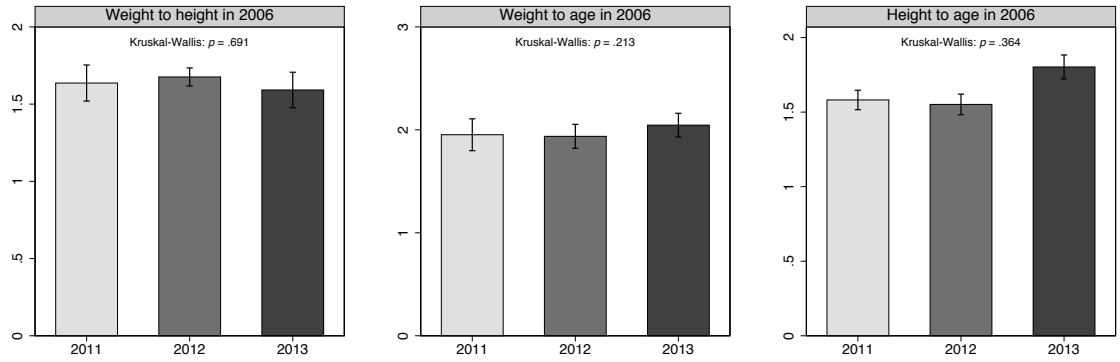

Figure 3.3: Differences in the development status of children across the three treatment cohorts.

I do not observe significant differences in the status of the children in the three cohorts in 2006. I take these results as suggestive evidence that the community mothers in the three cohorts did not differ before the intervention such that it affected the children systematically.

3.5 Results

3.5.1 Development after one Year of Treatment

Descriptive statistics

In Table 3.2, I present the characteristics of the children and the community mothers for the three cohorts from the monitoring dataset. This sample includes the data of 511 children. Across the three samples, we observe that the children display cognitive, psychosocial, and psychomotor levels of between 68% and 78%. Indicating that according to this UNICEF scale children in my sample fall behind. Furthermore, while the cognitive and psychomotor developmental indexes are not significantly different at baseline, the psychosocial development of the children in the third cohort is significantly lower compared to the other two cohorts. Figure 3.4 depicts the average of the three indexes at baseline and at the end of the intervention (endline) separately for the three cohorts.[22]

[22]The cognitive and the psychosocial development at baseline between the three cohorts are not significantly different (based on two tailed t-tests). The same holds true for the psychosocial index between the first and the second cohort. However, the children from the third cohort display a psychosocial development which is significantly lower compared to the other two cohorts. I depict the separate indexes in the

For the community mothers, I observe a significant difference with respect to their age and the years of experience. The community mothers in the first cohort are slightly older than the community mothers in the other two cohorts. This is also reflected in the experience of the community mothers, which refers to the years that a women worked as community mother. In cohort one they worked on average for 20 years as community mother. In cohort two the average years of experience are 11 years, and in cohort three their average experience amounts to 15 years. The income of the community mothers in the first cohort is 377'000 COP ($\sim$ 124 USD). This is lower than the income of the community mothers in the second and the third cohort (cohort 2: 421'000 COP ($\sim$ 139 USD) and cohort 3: 430'000 ($\sim$ 142 USD)). In all three cohorts, the income is less than the legal minimum salary in Colombia which is 690'000 COP ($\sim$ 240 USD). The educational levels of the community mothers are comparable across the cohorts. The variable education captures the level of education of the community mother. 2.5 indicates that on average a community mother finished secondary school (level two) but not tertiary school (level three).

As outlined in the previous section, I will account for all these observable differences in the regression analysis by including them as control variables.

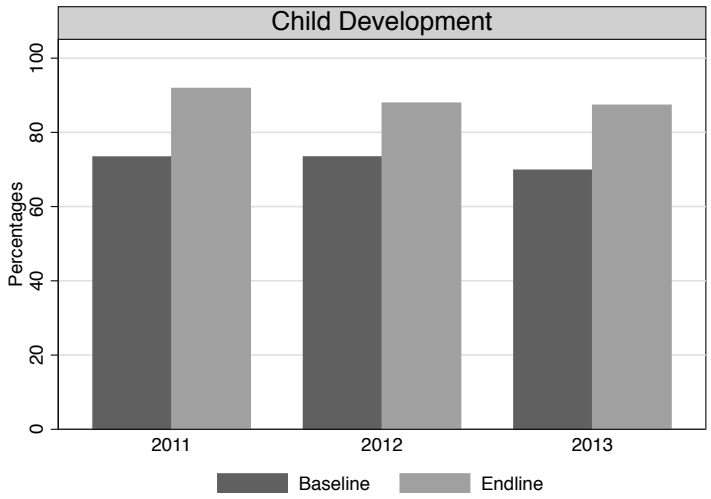

Figure 3.4: Development of the children in the three cohorts (in percentages).

appendix in Figure B1, B2, and B3.

Table 3.2: Descriptive statistics.

	Cohort 1 (2011)		Cohort 2 (2012)		Cohort 3 (2013)	
	Mean	SD	Mean	SD	Mean	SD
Children:						
Age (in months)	32.39	9.21	29.38	9.66	29.56	9.52
Female (D)	0.46	0.50	0.44	0.50	0.55	0.50
Cognitive level (%)	76.23	14.38	74.74	8.52	75.74	6.13
Psychosocial level (%)	75.70	13.11	76.93	7.17	59.83	6.18
Psychomotor level (%)	68.34	15.81	68.64	14.51	74.01	7.38
Observations	136		328		164	
Community Mother:						
Age (in years)	53.32	7.35	46.83	8.92	48.85	8.21
Experience (in years)	19.42	6.28	11.22	5.99	15.54	8.74
Income (in 1'000 COP)	377.24	118.63	421.34	165.82	430.97	159.50
Education	2.57	0.79	2.49	0.75	2.17	0.90
Observations	20		20		20	

Note. Descriptive statistics for the three cohorts. The children from the first cohorts are born between 2007 and 2009. The children from the second cohort are born between 2007 and 2010, and the children from the third cohort are born between 2008 and 2010. Both age variables depict the respective age at the beginning of the intervention. The developmental status of the children are from the baseline measurement. For cohort 1, baseline was elicited in January 2011, for cohort 2 the baseline was elicited in January 2012. For cohort 3 the baseline was elicited in January 2013. The income of the community mothers corresponds to the income at the beginning of the treatment.

Regression results

Table 3.3 depicts the results of the OLS estimations. The first column shows the treatment effect on cognitive development. Children in the MD control group reach on average 70% of the cognitive items. In contrast, treated children score on average 11 percentage points ($p < 0.001$) higher on the cognitive development index. In column (2) I regress the psychosocial development status on the treatment and the control variables. I find that the intervention affected the psychosocial status of the children considerably. Children in the MD control group reach on average 70% of the items. Whereas children from the MD treatment group attain on average 12 percentage points more on the psychosocial index ($p < 0.001$). I observe a similar pattern for the psychomotor development (column (3)). Children from the MD control group accomplish 60% of the items on average, and the intervention accounts for an increase of 13 percentage points ($p < 0.001$).[23]

Interestingly, girls score on average significantly lower on the cognitive and the psy-

[23]The separate results (sample 1 and sample 2) are comparable. The corresponding tables are in the appendix.

chosocial index. Additionally, I observe that age has predictive power for all three indexes.

Table 3.3: OLS estimates for the development of the children.

	Cognitive	Psychosocial	Psychomotor
Treated (D)	10.957***	12.212***	13.387***
	(0.467)	(0.546)	(0.868)
Age (in months)	0.247***	0.209***	0.459***
	(0.075)	(0.075)	(0.090)
Female (D)	−1.857***	−1.551**	−1.508
	(0.663)	(0.689)	(0.928)
Age (in years)	0.134	0.125	0.101
	(0.114)	(0.116)	(0.125)
Experience (in years)	−0.241**	−0.217	−0.197
	(0.117)	(0.134)	(0.146)
Income (in 1'000 COP)	0.003	0.004	0.003
	(0.004)	(0.004)	(0.004)
Education	−0.071	0.124	0.001
	(0.540)	(0.466)	(0.645)
Cohort 2 (D)	−4.206*	−2.064	−4.138*
	(2.138)	(2.148)	(2.278)
Cohort 3 (D)	−3.535	−19.649***	−2.790
	(2.114)	(2.066)	(2.413)
Constant	69.660***	69.594***	60.876***
	(4.289)	(3.679)	(4.609)
F-test	116.8	280.9	62.7
Prob $> F$	0.000	0.000	0.000
R^2 adjusted	0.387	0.692	0.423
N	511	511	511

Note. OLS estimates. The unit of all dependent variables is percentages. Dependent variable in column 1 is cognitive development. In column 2 the dependent variable is psychosocial development and in column 3 it is psychomotor development. Robust standard errors, clustered on community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In a further set of estimations (reported in Table B4 the appendix) I estimated the effect including all interaction terms. I observe that the intervention was less effective for girls with respect to cognitive and psychosocial development. Furthermore, it seems that for the psychosocial and the psychomotor development, the effect of the intervention is smaller the older the child.

To sum up, based on ordinary least squares regressions, I find that the program led to large skill gains for the treated children. There are two potential caveats with these results. First, the implementing NGO was also the NGO which elicited these develop-

mental measures. Therefore, these measures might be biased. With the data at hand, I cannot elaborate this caveat further. However, the second evaluation (four years after the treatment) will not be subject to this bias and hence depending on the conclusion, strengthen the results of this part. Second, it is possible that unobservable characteristics of the community mother or the child are responsible for the differences and not the intervention itself. Hence, in the next section, I will assess the validity of the results using two alternative estimation methods.

Robustness checks

Given that the assignment to treatment was not random, I conduct two robustness checks. First, I estimate the effects of the program using inverse probability weighted regression adjustment (IPWRA) as suggested by Imbens and Wooldridge (2009) and in line with the analysis conducted by Bernal (2015).

The estimation procedure is as follows: first, I estimate for each child the probability to belong to the MD treatment group using a probit regression. For that, I include all observable characteristics of the child and the community mother.

I present the results in Table 3.4. This estimation serves as the basis for the subsequent regression adjustment. They indicate that the income and the experience of the community mother predict treatment participation. In the second step, I estimate the propensity score. Figure 3.5 shows the propensity scores to check the common support condition. For the subsequent estimation, I include all observations which are within the common support. It is not surprising that all observations fall into the common support, given the small number of control variables.

Table 3.4: Probit of participation in treatment.

	Prob. treated
Age (in months)	0.009
	(0.006)
Female (D)	−0.074
	(0.114)
Age (in years)	0.001
	(0.010)
Experience (in years)	0.017
	(0.011)
Income (in 1'000 COP)	−0.001*
	(0.000)
Education	0.285***
	(0.076)
Constant	−0.970**
	(0.483)
χ^2-test	24.2
p	0.000
Pseudo R^2	0.034
N	511

Note. Probit estimates for the probability to belong to the treated group. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

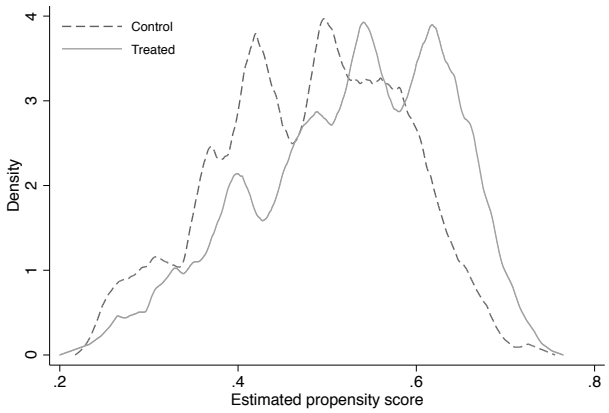

Figure 3.5: Common support.

Finally, to assess the treatment effect, I estimate a linear regression where I weight the observations of the treatment by the inverse probability of the propensity score and the observations of the control group by one minus the inverse probability of the propensity score (Imbens & Wooldridge, 2009). Table 3.5 summarizes the average treatment effects for the three dependent variables estimated with IPWRA. The point estimates of the treatment effect are very similar to the estimates of the OLS regressions and support the results reported in the previous section.

Table 3.5: Treatment effects estimated by IPWRA.

Outcome variable	Control group		Treatment effect	Standard error	N
	Mean	SD			
Cognitive development	76.461	7.163	12.574	0.819***	511
Psychosocial development	69.528	11.038	21.823	0.953***	511
Psychomotor development	73.757	11.284	14.582	0.937***	511

Note. Treatment effects estimated using inverse probability weighted regression adjustment.

Second, I try to estimate the magnitude of the omitted variable bias using the method proposed by Altonji et al. (2005) and by Oster (2014). This methodology bases on the assumption that selection on unobservables is equal to selection on observables. It allows to calculate the relative degree of selection (δ). The authors propose $\delta = 1$ to be an appropriate cutoff. Table 3.6 depicts the estimated δ's.

Table 3.6: Selection on unobservables.

	δ	N
Cognitive	1.24	511
Psychosocial	1.83	511
Psychomotor	1.33	511

Note. Amount of selection on unobservables relative to selection on observables required to attribute the entire high-quality preschool education effect to selection bias. Calculation based on Oster (2014).

I find that all values for δ reach the critical threshold. Hence, the observed effects are likely to hold if we assume that selection on unobservables is equal to selection on observables.

To sum up, the analysis of the monitoring data provides evidence that the program had a considerable positive effect in the short run on all three outcome variables. Thanks to the intervention, the children are able to catch up with the national averages on these three indexes.[24] Furthermore, the results are in line with an evaluation by Bernal (2015). She assessed a similar program from Bogotá in 2007[25] and observes that children's psychosocial, psychomotor and cognitive skills increased by up to 0.3 standard deviations compared to children who visited a standard community mother.

These results are very promising, but the most important question remains, namely whether these differences persist in the longer run, which is the scope of the next section.

3.5.2 Effect four Years after the Treatment

Descriptive results

To assess the persistence of the treatment effect, we collected data from 296 children. Of these children, 173 belong to the GPM control group and 123 to the GPM treatment group. Table 3.7 depicts the summary statistics of the characteristics of the children. Of the group

[24] According to the Tablas de normas para la población colombiana in Ortiz Pinilla (1999) the average for cognitive development in the corresponding age is 85%, for psychomotor it is 87%, and for psychosocial 88%.

[25] The intervention evaluated in Bernal (2015) consisted of providing the community mothers with formal preschool teaching education. The program did not include the other three components (coaching, parental workshops, mentoring) of the intervention by Apoyar.

of treated children, we were not able to find four of them, and the school did not know either if they moved away or if they dropped out.[26]

Table 3.7: Characteristics of the children in the treated and the control group.

	Treated		Control		Difference	
	Mean	SD	Mean	SD	Δ	t
Female	0.41	0.49	0.56	0.50	-0.15***	(-2.64)
Age	9.15	1.07	9.67	1.29	-0.52***	(-3.70)
SISBEN	40.47	14.85	38.56	16.35	1.91	(1.00)
Weight	28.64	6.72	29.96	9.16	-1.32	(-1.35)
Height	132.57	7.02	136.05	8.43	-3.48***	(-3.75)
Mean grade (%)	74.15	13.77	69.89	13.40	4.27***	(2.63)
Right grade for age	.642	.481	.486	.501	-.156***	(-2.695)
Token given away (DG)	2.55	2.04	2.05	2.12	0.51**	(2.06)
Egalitarian	0.04	0.20	0.02	0.15	0.02	(0.86)
Altruistic	0.06	0.23	0.02	0.15	0.03	(1.52)
Spiteful	0.12	0.33	0.14	0.35	-0.02	(-0.56)
Selfish	0.59	0.49	0.70	0.46	-0.11*	(-1.89)
Trust	0.33	0.47	0.23	0.42	0.10*	(1.95)
Trustworthiness	0.70	0.46	0.63	0.48	0.07	(-1.24)
Risk	2.60	1.35	2.50	1.31	0.10	(0.63)
Patience	1.27	1.75	1.08	1.59	0.19	(0.99)
General self-esteem	0.85	0.11	0.85	0.09	-0.00	(-0.15)
Observations	123		173		296	

Note. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

I find that children in the GPM treatment group are slightly younger and shorter. Moreover, the proportion of girls is smaller in the GPM treatment group. The variable SISBEN is a number which each Colombian household gets based on its socioeconomic characteristics. It serves to determine whether a family is eligible for state subsidies. Children in the two groups seem to live in households with a similar socioeconomic status. If we look at the dependent variables, the table provides a first indication of possible differences. To be specific, I observe that children in the GPM treatment group differ significantly (based on a two-tailed t-test) with respect to their grades, the probability of being in the grade which corresponds to their age, altruism, trust, and the probability to be categorized as selfish. I do not observe statistically significant differences for the other

[26] Apoyar collected the data of the children in the GPM control group by collecting the lists with the names of all the children which are currently in school. Unfortunately, this approach does not allow to identify how many children from the GPM control group dropped out of school or moved away.

dependent variables.[27]

Table 3.8 provides the summary statistics for the parents. At first sight, the characteristics of the parents in the two groups are very comparable. Only the probability of being classified as an altruist is significantly higher in the sample of the treated parents.[28]

Table 3.8: Characteristics of the parents in the treated and the control group.

	Treated		Control		Difference	
	Mean	SD	Mean	SD	Δ	t
Age	34.87	7.25	37.43	11.30	-2.55	(-1.30)
Highest degree	1.54	0.62	1.48	0.62	0.06	(0.50)
Monthly income	490.43	266.47	466.17	274.67	24.26	(0.43)
SISBEN	42.30	13.36	39.72	16.59	2.59	(0.79)
Homework help	0.82	0.32	0.82	0.28	0.00	(0.00)
Hours of help	8.96	5.14	8.47	5.71	0.49	(0.43)
Token given away (DG)	4.06	1.98	3.73	2.02	0.33	(0.82)
Egalitarian	0.15	0.36	0.15	0.36	0.00	(0.00)
Altruist	0.25	0.44	0.06	0.24	0.19**	(2.59)
Spiteful	0.06	0.24	0.06	0.24	0.00	(0.00)
Selfish	0.23	0.42	0.29	0.46	-0.06	(-0.69)
Trust	0.56	0.50	0.44	0.50	0.13	(1.22)
Risk	2.40	1.65	2.96	1.71	-0.56	(-1.64)
Patience	3.50	2.06	3.40	2.04	0.10	(0.25)
Observations	48		48		96	

Note. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Regression results

For all outcome variables of interest, I first estimate a model which only contains the dummy variable for the treatment. In the second column, I add the variables to control for the child's characteristics. In the third column, I include the characteristics of the community mother, and in the fourth column, I add the parental characteristics. Note, that I only have the parental characteristics for the subset of children where we collected data on their parents (N=96), this is why the number of observation drops considerably in the last column.

I start the analysis of the long-term effects with the elaboration of the cognitive skills.

[27] As described in subsection 3.4.2, we incentivized all decisions of the children. For the gifts we paid in total 280 USD.

[28] For the experiments with the parents, we paid out in total 250 USD, corresponding to average earnings of 2.60 USD.

The grade refers to the grade obtained in the first quarter of 2016. I transform all the grades to percentages. Table 3.9 provides evidence that children from the GPM treatment group have higher grades on average than children from the GPM control group. In the GPM control group, children reach on average a grade of 69.9 (SD : 13.4) compared to a grade of 74.2 (SD : 13.8) in the GPM treatment group. This raw difference of roughly four percentage points is significant and robust to the inclusion of the child characteristics (column 2). Interestingly, age seems to be negatively related to grades. Furthermore, on average girls achieve roughly four percentage points higher grades than boys. The treatment dummy drops in size and loses its significance once we add the control variables of the respective community mother ($p = 0.152$, column 3). In the last column, the estimate of the treatment remains insignificant ($p = 0.286$), which is no surprise given the low number of observations. Besides the main effect, I observe that socioeconomic status of the family becomes predictive for the grade. Children from lower SES families have on average lower grades.

Next, I analyze whether the children are in the grade which corresponds to their age. I find that in the GPM control group only 48% of the children are in the "right" grade. In the GPM treatment group, 64% of the children are in the grade which corresponds to their age. A probit estimation reveals that the probability of being in the right grade is affected significantly by the treatment. Children in the GPM treatment group are on average 11 percentage points more likely to be in the right grade, controlling for all observable characteristics (see B5). This difference can occur during two points in time. Either children start school too late, or they fall behind during school. I find that presumably, the lag occurs during school because children in both groups start school on average with the correct age of five years. Unfortunately, I do not have the data about grade repetition.

I continue with the analysis of the treatment effect on the individual preference measures.

First, I evaluate the effect of the intervention on the continuous measure for altruism. The dependent variable is the amount given to the recipient in the dictator game. On average, children transferred two stars out of ten to the other child. This is comparable to

Table 3.9: OLS estimates for the grade.

	(1)	(2)	(3)	(4)
Treated (D)	4.233**	3.885*	2.892	4.898
	(1.838)	(2.104)	(2.006)	(4.543)
Age		−4.145***	−4.094***	−1.796
		(1.457)	(1.458)	(2.115)
Female		3.798**	3.509*	4.768
		(1.847)	(1.893)	(3.827)
Height (cm)		0.189	0.185	0.043
		(0.161)	(0.170)	(0.367)
Weight (kg)		−0.004	−0.019	0.111
		(0.101)	(0.103)	(0.279)
Entry into school		0.361	0.439	−1.438
		(1.631)	(1.655)	(3.552)
Current class		0.006	0.007	−0.001
		(0.010)	(0.010)	(0.026)
Sisben		0.035	0.038	−0.265**
		(0.061)	(0.063)	(0.101)
MC: Age			0.160	0.603*
			(0.166)	(0.314)
MC: Experience (y)			−0.215	−0.582
			(0.198)	(0.361)
Parent: Age				−0.038
				(0.228)
Parent: Income				0.009
				(0.005)
Parent: Degree				0.632
				(2.993)
Constant	69.921***	76.871***	73.126***	68.577
	(1.171)	(17.016)	(19.350)	(43.769)
F-test	5.3	5.0	4.5	4.6
Prob $> F$	0.023	0.000	0.000	0.000
R^2 adjusted	0.020	0.105	0.092	0.109
N	286	233	220	72

Note. OLS estimates. Dependent variable is the average grade (measured in percentages). Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

what other researchers observe with infants of the same age in other countries.[29] Table 3.10 summarizes the results for the effect of the intervention on altruism. I find that children in the GPM treatment group allocate on average between 0.5 and one stars more to another anonymous child.[30] The effect is robust to the inclusion of all control variables.

Besides the continuous measure for altruism, I elicited social preferences with the four binary decisions described in the subsection 3.4.2. Using the approach by Fehr et al. (2008), I classify all the children according to their type. I find that in the control group, 2.3% of the children classify as egalitarian, 2.3% as altruists, 14.5% as spiteful, and 70% as selfish. In the group of treated children, 5.7% classify as egalitarian, 4% as altruists, 12.2% as spiteful, and 59.3% as selfish. The selfish ones can be further sub-classified into weakly egalitarian, weakly altruistic, and weakly spiteful. I observe that in the GPM control and the treatment group roughly 13% are sub-classified as weakly altruistic (13.2% in the control group and 12.1% in the treated group), 17.9% are weakly egalitarian in the GPM control group and 13% in the GPM treatment group. 26% in the GPM control group are sub-classified as weakly spiteful compared to 20.3% in the GPM treatment group. In contrast to children from Europe (Austria and Czech Republic) children in both GPM groups are much more likely to be classified as selfish or spiteful.[31] The comparison with children from more developed contexts, confirms the observation of Deckers et al. (2015), that the socioeconomic background is related to selfish behavior. Put differently, the children in my sample are from very vulnerable households, the children in the sample of Fehr et al. (2013) and Bauer et al. (2014) are from developed countries, and presumably, the socioeconomic status of their families is much higher. A further regression analysis provides evidence that children from the treatment group are significantly less likely to be

[29] Harbaugh, Krause, and Liday (2003) observe in a sample of 9-year-old children in the US an average of 1.7 tokens and Benenson, Pascoe, and Radmore (2007) observe an average of 3 tokens in a sample of children from the UK. Furthermore, Kogut (2012) reports an average of 3.24 for second and fourth graders in Israel.

[30] I further evaluated whether the effect differs by gender, age or the characteristics of the parents and community mothers. However, I do not find any significant interactions, which indicates that the treatment effect is rather homogeneous.

[31] Fehr, Glätzle-Rützler, and Sutter (2013) report that in a sample of 8 to 9-year-old children in Austria, 3% of the children behave egalitarian, 34% weakly egalitarian, 17% weakly altruistic, and 42% spiteful. Bauer et al. (2014) find 15.6% of the children to be altruists, 8.9% egalitarian, 6.3% spiteful, and 39.8% selfish. Within the selfish category, they find 10.8% to be weakly altruistic, 11.5% weakly egalitarian and 10% weakly spiteful.

Table 3.10: OLS estimates for altruism.

	(1)	(2)	(3)	(4)
Treated (D)	0.515**	0.602**	0.647**	1.018*
	(0.240)	(0.286)	(0.295)	(0.553)
Age		0.235	0.143	0.329
		(0.164)	(0.172)	(0.301)
Female		0.125	0.110	0.217
		(0.260)	(0.267)	(0.670)
Height (cm)		0.026	0.014	0.025
		(0.033)	(0.034)	(0.053)
Weight (kg)		−0.029	−0.022	−0.002
		(0.022)	(0.022)	(0.051)
Entry into school		−0.504**	−0.402*	−0.611
		(0.222)	(0.220)	(0.562)
Current class		0.001	0.001	0.000
		(0.002)	(0.002)	(0.004)
Sisben		−0.003	−0.002	0.005
		(0.008)	(0.008)	(0.020)
MC: Age			0.017	−0.041
			(0.023)	(0.061)
MC: Experience (y)			−0.003	0.075
			(0.032)	(0.067)
Parent: Age				−0.040
				(0.032)
Parent: Income				0.002
				(0.001)
Parent: Degree				−0.680
				(0.495)
Constant	2.018***	−0.998	−0.125	0.043
	(0.168)	(3.545)	(3.565)	(7.035)
F-test	4.6	2.1	1.5	1.7
Prob $> F$	0.034	0.039	0.151	0.081
R^2 adjusted	0.012	0.017	0.002	−0.006
N	291	235	222	73

Note. OLS estimates. Dependent variable is altruism (tokens given to the other person in the dictator game) measured from [0,10]. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

classified as selfish (see Table B6 in the appendix). For the other three categories (altruist, egalitarian, spiteful) I do not observe significant differences. Hence, even though that the intervention led to a decrease in the probability of being selfish, it is still high in comparison with children from more advantaged backgrounds.

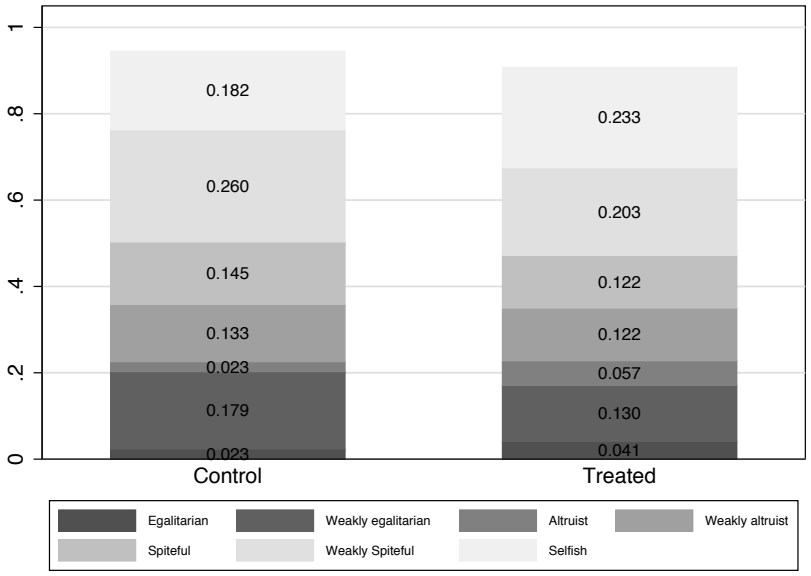

Figure 3.6: Distribution of types in the treated versus the control group.

I turn now to the analysis of the trust behavior. In the GPM control group, 23% of the children decided to trust and send the star to the other child. In contrast, in the GPM treatment group, 33% of the children decided to trust. A probit estimation reveals that the observed difference is just statistically insignificant (p=0.115, complete Table B7 in the appendix). Interestingly, I observe that older children are less likely to transfer the star. Moreover, once I include the parental characteristics, I find that children with a higher socioeconomic background are significantly more likely to trust. The comparison of trust behavior to studies with samples of other children provides further evidence for the hypothesis that individual preferences relate to the socioeconomic background of the family. Evans et al. (2013) find in a sample of children from Austria trust rates of 55%, which is considerably higher compared to what I observe in my sample. The behavior of the second mover in the trust game is not different in the two groups. Roughly 70% of the subjects reciprocate trust and decide to send the two stars back.

In a next step, I elaborate on the effect of the treatment on risk preferences. On average children invested half of their endowment (2.5 stars) into the wheel of fortune. However, the behavior does not differ depending on being exposed to the intervention (for the regressions, see Table B8 in the appendix).

The last individual preference measure under consideration is patience. On average children invested 3.45 stars for later payout. Again, I do not observe that the treatment affected time preferences of the children (see Table B8). The level of patience observed in both GPM groups is much higher compared to the time preferences of children of the same age in Austria (Angerer, Lergetporer, Glätzle-Rützler, & Sutter, 2015). They observe that children aged 7 to 11 invest on average between 1.14 and 2.15 tokens (out of 5).

Eventually, I evaluate whether the treatment affects the self-esteem of the children. As already indicated in the descriptive statistics, children in the two GPM groups do not differ significantly (Table B10 in the appendix). I further assessed whether the treatment might affected only certain self-esteem categories. However, I do not observe differences for any of the subscales.

So far I provide evidence that the treatment had an effect on the children, which is observable four years after being in one of the treated nurseries. The question which component of the intervention or which combination of components is responsible for this difference is still open. The only possibility with the data at hand is to evaluate whether parents of the children in the GPM treatment group differ compared to the parents of children from the GPM control group.

Table 3.11: Parenting style.

	Treated		Control		Difference	
	Mean	SD	Mean	SD	Δ	t
Authoritative PS	4.50	0.47	4.52	0.41	-0.02	(-0.23)
Authoritarian PS	2.91	0.81	3.00	0.91	-0.09	(-0.51)
Permissive PS	2.98	0.82	2.98	0.66	-	-
Observations	48		48		96	

Based on Table 3.11 I do not find that parents in the GPM control and the GPM treatment group apply different parenting styles. Compared to parents from Western

countries, the means for authoritarian and permissive parenting style are considerably higher. This suggests that on one hand parents in the sample employ a more authoritarian but at the same time also more permissive parenting style. In the appendix, I report all regression estimations of the parental characteristics on the treatment. I observe that parents from the GPM treatment group are less risk seeking, but the effect is only weakly significant. For all other variables, the treatment does not account for differences.

Hence, I do not observe that the intervention affected parents directly. Still, it might be that the effect on the parents is more subtle. To test this, I elaborate how parents' and their offspring's traits correlate. Socialization models provide theoretical rationales for these correlations (for example see Doepke and Zilibotti (2014) or Bisin and Verdier (2011)). Not the number of observations for this explorative analysis is rather low. Nevertheless, I hypothesize that well-functioning parent-child interactions are necessary to transmit traits from one generation to another.

In Table 3.12 I depict the results of the estimation of a child's altruism on the altruism of its parents controlling for the child's characteristics. Interestingly, I find that the estimate is positive and significant within the GPM treatment group while negative and non-significant in the GPM control group.

In line with the previous analysis, I evaluate whether trust displays a similar pattern. I find the same systematic differences between the GPM control and the treatment group. For parent-child pairs in the GPM treatment group the relation is significantly positively (column 1), whereas, for parents and their offsprings in the GPM control group, the relation is weakly significantly negative (column 2). Put differently, parents in the GPM treatment group who trust are likely to have children who trust as well; this observation is not the case for parents in the GPM control group.

Table 3.12: OLS estimates for altruism, separated by treatment.

	(1) Treated	(2) Control
Parent: Altruism	0.302**	0.118
	(0.139)	(0.187)
Age	0.067	0.023
	(0.302)	(0.179)
Female	−0.005	0.293
	(0.582)	(0.806)
Height (cm)	0.185***	−0.043
	(0.059)	(0.059)
Weight (kg)	−0.097	0.014
	(0.074)	(0.067)
Constant	−21.105***	6.387
	(6.138)	(6.227)
F-test	4.4	0.8
Prob $> F$	0.003	0.567
R^2 adjusted	0.177	−0.078
N	47	47

Note. OLS estimates. Dependent variable is altruism (tokens given to the other person in the dictator game) measured from [0,10]. Column 1 is the estimation within the treated, column 2 for the control group. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

For the other two preference measures, I do not find any systematic relations between the parents and their offsprings measure. Obviously, these results come from an explorative analysis and are only a first interesting indication that there might be an effect of the intervention on the parents of the treated children. However, without further research, this remains only a hypothesis.

To sum up, based on linear regression models, I show that the intervention affected grades and altruism. Next, I apply the same procedures as in the analysis of the monitoring data to check the robustness of the results.

Table 3.13: Probit estimates for trust, separated by treatment.

	(1) Treated	(2) Control
Parent: Trust	1.043^{**}	-1.099^{*}
	(0.470)	(0.623)
Age	0.036	-0.547^{**}
	(0.227)	(0.239)
Female	0.071	-0.221
	(0.395)	(0.337)
Height (cm)	0.042	0.044
	(0.042)	(0.050)
Weight (kg)	0.001	0.010
	(0.043)	(0.053)
Constant	-7.361	-1.674
	(4.560)	(6.583)
χ^2-test	12.6	14.4
p	0.027	0.013
Pseudo R^2	0.146	0.223
N	47	47

Note. Probit estimates. Dependent variable is trust measured with a dummy variable. Column 1 is the estimation within the treated, column 2 for the control group. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Robustness checks

Again, I start with the estimation of the effect using IPWRA. I display the results of the estimation of the probability to belong to the treatment group in Table 3.14.

Table 3.14: Probit regression for participation in treatment.

	Prob. Treated
Age	−0.097
	(0.082)
Female	−0.413**
	(0.166)
Height (cm)	−0.036**
	(0.016)
Weight (kg)	0.018
	(0.013)
Sisben	0.007
	(0.005)
MC: Age	0.004
	(0.013)
MC: Experience (y)	−0.043***
	(0.017)
Constant	5.442***
	(1.750)
χ^2-test	38.1
p	0.000
Pseudo R^2	0.107
N	264

Note. Probit estimates. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

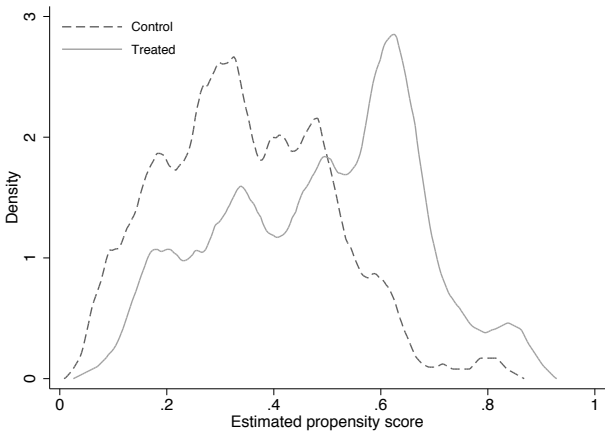

Figure 3.7: Common support.

In Figure 3.7 I depict the common support. I excluded 20 observations because they do not fall into the common support.

Table 3.15 summarizes the effects estimated by IPWRA. The estimation confirms the results observed with the OLS regressions, which indicate that the treatment had a significant and positive effect on the grades of the treated children and their prosocial behavior.

Table 3.15: Treatment effects estimated by IPWRA.

Outcome variable	Control group		Treatment effect	Standard error	N
	Mean	SD			
Grade	70.15	13.30	3.01*	1.68	264
Right grade	.52	0.04	.16**	0.06	264
Altruism	1.93	2.07	.74***	.27	264
Trust	.24	.44	.05	.06	264
Risk	2.41	1.28	.15	.16	264
Patience	1.13	1.59	.10	.21	264
Self-esteem	.86	.09	-.02	0.01	264

Note. Treatment effects estimated using inverse probability weighted regression adjustment.

Second, I evaluate the amount of selection on unobservables relative to selection on observables required to attribute the entire effect to selection bias. Table 3.16 contains the results. Except for trust, the relative degree of selection reaches the critical threshold. These results provide further support that omitted variables do not entirely drive my results.

Table 3.16: Selection on unobservables.

	δ	N
Grade	1.04	297
Right grade	1.07	297
Altruism	1.73	297
Trust	0.36	297

Note. Amount of selection on unobservables relative to selection on observables required to attribute the entire high-quality preschool education effect to selection bias. Calculation based on Oster (2014).

3.5.3 Cost-Benefit Analysis

Lastly, I compare the costs to the benefits estimated in this chapter. According to Bernal (2013), the cost of a standard (non-treated) hogar is 288 USD per child and year. The intervention of Apoyar costs 5464 USD per community mother and year, which amounts to 453 USD per child and year. Note that a community mother, once she has received training, will probably apply her new skills on children outside this evaluation. The 453 USD represent an upper bound on the cost of this investment.

To translate the increase of the cognitive skills to the expected wage increase, we rely on numbers published in Chetty et al. (2011) and Bartik, Gormley, and Adelstein (2012). I expect the rise in cognitive skills to lead to an increase of the wage by 3%. The expected effects on the wage due to an increase in non-cognitive skills is based on the numbers reported in Flossmann, Piatek, and Wichert (2007). It amounts to an increase of the wage by 1%. Based on these conservative estimates, the intervention is likely to increase the children's future wages in adult life by around 4%.

3.6 Conclusions

In this chapter, I elaborated if enhancing the quality of a community-based preschool program in Colombia leads to better outcomes for the treated children. Research on preschool education in Colombia revealed that the knowledge about infant development of many caregivers is insufficient. In response to this issue, a Colombian NGO implemented an intervention, which consisted of four components. The first component was the provision of a formal education as preschool teachers to the caregivers. Second, the NGO supported the caregivers to implement their newly acquired knowledge in their daily nursery routines. Third, the caregivers received assistance for the organization of parental workshops, which aimed at increasing the knowledge of parents about infant care and development. Fourth, the NGO monitored the children in school and supported them if needed.

I find that children in the program benefited on several dimensions. First, in the short run, children who visited a treated community mother display improved cognitive, psychosocial, and psychomotor skills. Second, the effects of the intervention seem to be

persistent. Four years after the treatment, children from treated nurseries have on average higher grades, are more likely to be in the right grade, and behave more prosocial. For the other behavioral measures (trust, risk, and time preferences) I do not observe significant differences. Moreover, I provide suggestive evidence that the treatment affected the parent-child interaction positively. This last result stems from an explorative analysis and hence further research is needed to confirm it. Moreover, I elicited the parenting style via self-reported survey. It might be that this leads to a bias because parents want to appear as good parents. Thus, it would be important to employ more objective methods to evaluate if parenting style was affected by the treatment.

Given that in this study the assignment to the treatment was not random, I cannot make causal claims because I cannot entirely rule out that unobservable characteristics of the community mothers or the children are responsible for the differences. For this reason, I conducted different robustness checks. They provide evidence for the observed results. Another caveat of this study is that I cannot disentangle the effects of the different components of the program. For the further development of this intervention, it would be very interesting to know how the different components contributed to the treatment effect. At last, further research is needed to elaborate to what extent the program had spill-over effects within the family and the school environment. Meaning to elaborate whether the intervention had effects beyond the treated children. It is likely that the positive implications of the intervention extended to the siblings of a child who visited a treated nursery and to their peers in school.

Summing up, the results lead me to believe that enhancing the quality of preschool education, by providing community mothers with formal education, supporting them in the implementation, and including the parents is a promising policy option to improve the school readiness and thus, further success in life of disadvantaged children in Colombia.

References

Angerer, S., Glätzle-Rützler, D., Lergetporer, P., & Sutter, M. (2015). Donations, risk attitudes and time preferences: A study on altruism in primary school children. *Journal of Economic Behavior & Organization*, *115*, 67–74.

Angerer, S., Lergetporer, P., Glätzle-Rützler, D., & Sutter, M. (2015). How to measure time preferences in children: A comparison of two methods. *Journal of the Economic Science Association*, *1*(2), 158–169.

Araujo, M. C., López Bóo, F., & Puyana, J. M. (2013). Overview of early childhood development services in latin america and the caribbean.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, *5*(3), 25–50.

Bauer, M., Chytilová, J., & Pertold-Gebicka, B. (2014). Parental background and other-regarding preferences in children. *Experimental Economics*, *17*(1), 24–46.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology*, *4*(1, Pt.2), 1.

Behrman, J., Fernald, L., & Engle, P. (2013). Preschool programs in developing countries. In P. Glewwe (Ed.), *Education policy in developing countries* (pp. 65–105). Chicago, IL: University of Chicago Press.

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior*, *28*(3), 168–175.

Bernal, R. (2015). The impact of a vocational education program for childcare providers on children's well-being. *Economics of Education Review*, *48*, 165–183.

Bisin, A. & Verdier, T. (2011). The economics of cultural transmission and socialization. In J. Benhabib, M. O. Jackson, & A. Bisin (Eds.), *Handbook of social economics* (pp. 339–416). San Diego, CA: North-Holland.

Bowles, S. (2004). *Microeconomics: Behavior, institutions, and evolution*. Princeton, NJ: Princeton University Press.

Cappelen, A. W., List, J. A., Samek, A., & Tungodden, B. (2016). The effect of early education on social preferences. NBER Working Paper Series, Working Paper 22898.

Cardenas, J. C. & Carpenter, J. (2008). Behavioural development economics: Lessons from field labs in the developing world. *Journal of Development Studies*, *44*(3), 311–338.

Carlo, G., Mestre, M. V., Samper, P., Tur, A., & Armenta, B. E. (2011). The longitudinal relations among dimensions of parenting styles, sympathy, prosocial moral reasoning, and prosocial behaviors. *International Journal of Behavioral Development*, *35*(2), 116–124.

de Oliveira, A., Eckel, C., & Croson, R. T. (2012). The stability of social preferences in a low-income neighborhood. *Southern Economic Journal*, *79*(1), 15–45.

Deckers, T., Falk, A., Kosse, F., & Schildberg-Hörisch, H. (2015). How does socio-economic status shape a child's personality? IZA Discussion Paper, No. 8977.

Doepke, M. & Zilibotti, F. (2014). Parenting with style: Altruism and paternalism in intergenerational preference transmission. NBER Working Paper Series, Working Paper 20214.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Duncan, G. J. & Brooks-Gunn, J. (2000). Family poverty, welfare reform, and child development. *Child Development*, *71*(1), 188–196.

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., ... Ertem, I., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339–1353.

Evans, A. M., Athenstaedt, U., & Krueger, J. I. (2013). The development of trust and altruism during childhood. *Journal of Economic Psychology*, *36*, 82–95.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–1083.

Fehr, E., Glätzle-Rützler, D., & Sutter, M. (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review*, *64*, 369–383.

Fontes, L. A. (2002). Child discipline and physical abuse in immigrant latino families: Reducing violence and misunderstandings. *Journal of Counseling & Development*, *80*(1), 31–40.

Glewwe, P., Jacoby, H. G., & King, E. M. (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics*, *81*(3), 345–368.

Gneezy, U. & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*(2), 631–645.

Golsteyn, B. H., Grönqvist, H., & Lindahl, L. (2014). Adolescent time preferences predict lifetime outcomes. *The Economic Journal*, *124*(580), F739–F761.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., Group, I. C. D. S., et al. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60–70.

Harbaugh, W. T., Krause, K., & Liday, S. J. (2003). Bargaining by children. University of Oregon Economics Working Paper No. 2002-4.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Kogut, T. (2012). Knowing what i should, doing what i want: From selfishness to inequity aversion in young children's sharing behavior. *Journal of Economic Psychology*, *33*(1), 226–236.

Kosse, F., Deckers, T., Schildberg-Hörisch, H., & Falk, A. (2016). The formation of prosociality: Causal evidence on the role of social environment. IZA Discussion Paper, No. 9861.

Martorell, R., Horta, B. L., Adair, L. S., Stein, A. D., Richter, L., Fall, C. H., ... on Health Orientated Research in Transitional Societies Group, C. (2010). Weight gain in the first two years of life is an important predictor of schooling outcomes in pooled analyses from five birth cohorts from low-and middle-income countries. *The Journal of Nutrition*, *140*(2), 348–354.

Mérida, R., Serrano, A., & Tabernero, C. (2015). Diseño y validación de un cuestionario para la evaluación de la autoestima en la infancia. *Revista de Investigación Educativa*, *33*(1), 149–162.

Özler, B., Fernald, L. C., Kariger, P. K., McConnell, C., Neuman, M. J., & Fraga, E. (2016). Combining preschool teacher training with parenting education: A cluster-randomized controlled trial. World Bank Policy Research, Working Paper No. 7817.

Raine, A., Mellingen, K., Liu, J., Venables, P., & Mednick, S. A. (2003). Effects of environmental enrichment at ages 3–5 years on schizotypal personality and antisocial behavior at ages 17 and 23 years. *American Journal of Psychiatry*, *160*(9), 1627–1635.

Ramírez, C. (2006). El impacto del maltrato en los niños y las niñas en colombia. *Revista Infancia, Adolescencia y Familia*, *1*(2), 287–301.

Robinson, C. C., Mandleco, B., Olsen, S. F., & Hart, C. H. (1995). Authoritative, authoritarian, and permissive parenting practices: Development of a new measure. *Psychological Reports*, *77*(3), 819–830.

Roopnarine, J. L., Krishnakumar, A., Narine, L., Logie, C., & Lape, M. E. (2014). Relationships between parenting practices and preschoolers' social skills in african, indo, and mixed-ethnic families in trinidad and tobago: The mediating role of ethnic socialization. *Journal of Cross-Cultural Psychology*, *45*(3), 362–380.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The high/scope perry preschool study through age 40*. Ypsilanti, MI: High/Scope Press.

Shonkoff, J. P. & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, *103*(1), 510–531.

Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., ... Group, I. C. D. S. (2007). Child development: Risk factors for adverse outcomes in developing countries. *The Lancet*, *369*(9556), 145–157.

Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., ... Richter, L. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, *378*(9799), 1325–1338.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M. C., Weiland, C., ... Arbour, C. M. (2015). Experimental impacts of a teacher professional development program in chile on preschool classroom quality and child outcomes. *Developmental Psychology*, *51*(3), 309.

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). Toward the identification of features of effective professional development for early childhood educators. literature review. Office of Planning, Evaluation and Policy Development, US Department of Education.

Appendix

3.A Early Development Instrument

Cognitive Development

- 0 months:

    1. Startles in response to loud noise

- 1 to 3 months:

    2. Indicates with her eyes that she recognized where the noise comes from
    3. Makes two different guttural sounds
    4. Babbles with persons

- 4 to 6 months:

    5. 4 or more different sounds
    6. Laughs loudly when stimulated to do so
    7. Reacts when he is called by his name

- 7 to 9 months:

    8. Pronounces 3 or more syllables
    9. Makes a sound with the bell
    10. One clear word (M)[32]

- 10 to 12 months:

    11. Shakes her head (M)
    12. Calls his mother or his responsible caregiver
    13. Understands simple orders

- 13 to 18 months:

[32](M) indicates that the mother or the responsible caregiver of the child reports whether the child is able to fulfil the item.

14. Recognizes three objects
15. Combines two words
16. Recognizes six objects

- 19 to 24 months:

17. Names five objects
18. Makes phrases with three words (M)
19. More than 20 clear words (M)

- 25 to 36 months:

20. Says her complete name
21. Knows high/low, big/small
22. Uses complete sentences

- 37 to 48 months:

23. Defines five objects by their use
24. Repeats three digits
25. Describes a picture in detail

- 49 to 60 months:

26. Counts the fingers of his hands
27. Distinguishes back/front and under/above
28. Names 4 to 5 colors

- 61 to 72 months:

29. Expresses her opinion
30. Knows left and right
31. Knows the days of the week

Psychosocial Development

- 0 months:

    1. Follows movements in front of his face

- 1 to 3 months:

    2. Recognizes her mother
    3. Smiles when the mother turns the attention him
    4. Tries to turn around when she hears her mother talk

- 4 to 6 months:

    5. Tries to grab the hands of the examiner
    6. Takes a toy if it is to him
    7. Pays attention to a conversation

- 7 to 9 months:

    8. Tries to help holding a cup when drinking (M)
    9. Recognizes herself in the mirror
    10. Imitates applause

- 10 to 12 months:

    11. Brings a toy to the examiner
    12. Asks for an object or a toy
    13. Drinks alone from a cup

- 13 to 18 months:

    14. Identifies one set of articles of clothing correctly
    15. Identifies two parts of the body
    16. Indicates when he needs to go to the bathroom (M)

- 19 to 24 months:

    17. Identifies five parts of the body
    18. Tries to talk about experiences (M)
    19. Controls urination during the day (M)

- 25 to 36 months:

    20. Differentiates between girl and boy
    21. Knows the name of her mother and father
    22. Washes his face and hands on his own

- 37 to 48 months:

    23. Can take some cloths off on her own (M)
    24. Shares toys with other children (M)
    25. Has special friends

- 49 to 60 months:

    26. Can dress and undress on his own completely (M)
    27. Knows her age
    28. Organizes/initiates playing with other children (M)

- 61 to 72 months:

    29. Executes orders to help at home (M)
    30. Knows his exact address
    31. Talks about her family

Psychomotor Development

- 0 months:

    1. Struggles strongly

2. Follows with the eyes vertical and horizontal movements in front of his head

- 1 to 3 months:

3. Lifts her head when lying in prone position
4. Lifts his head and chest when lying in prone position
5. Holds her head
6. Opens/closes his hands and looks at them
7. Holds an object in her hands
8. Puts an object into his mouth

- 4 to 6 months:

9. Controls the movements of her head when sitting
10. Is able to turn around (M)
11. Tries to sit on his own
12. Tries to grab objects
13. Holds objects in both hands simultaneously
14. Moves an object from one hand to the other

- 7 to 9 months:

15. Can stay seated on her own for a few seconds
16. Crawls on his belly
17. Can sit down on her own
18. Plays with various objects at the same time
19. Grabs an object with the hands
20. Grabs a cube with the thumb and trigger finger

- 10 to 12 months:

21. Crawls on hands and knees

22. Stands up with the help of a table or chair
23. Can stand on his own without holding himself for at least 15 seconds
24. Takes objects out of a box
25. Can hold up to three objects in her hands
26. Searches for hidden objects

- 13 to 18 months:

27. Makes a few steps on his own
28. Walks on her own
29. Runs
30. Builds towers with three cubes
31. Turns around pages of a book
32. Anticipates where an object appears after being put in a tube from one side

- 19 to 24 months:

33. Kicks a ball without loosing the equilibrium
34. Throws a ball with the hands
35. Jumps on two feet
36. Puts toys in a box and closes the box
37. Makes round scribbles
38. Builds towers with five cubes

- 25 to 36 months:

39. Can stand on his toes without loosing the equilibrium
40. Stands up without using her hands (M)
41. Walks a few steps backward
42. Threads six or more beads

43. Draws horizontal and vertical lines after being shown how to do so
44. Separates small from big objects

- 37 to 48 months:

45. Walks on the tips of his toes
46. Can stand on one foot
47. Throws and catches a ball
48. Draws a rudimentary human figure
49. Cuts paper with a scissor
50. Draws a cube and circle after being shown how to do so

- 49 to 60 months:

51. Walks in a straight line
52. Jumps three or more steps on one foot
53. Catches and throws the ball back to the examiner
54. Draws a more detailed human figure
55. Groups objects by form and color
56. Draws a ladder after being told to do so

- 61 to 72 months:

57. Jumps down from something with the height of 25 cm
58. Hops and alternates the feet
59. Jumps down from something with the height of 60 cm
60. Groups objects by form, color and size
61. Has to reconstruct a stair with 10 cubes after it has been shown to her
62. Draws a house

3.B Additional Tables

Table B1: Descriptive statistics separated by cohort.

	Cohort 1 (2011)		Cohort 2 (2012)		Cohort 3 (2013)	
	Mean	SD	Mean	SD	Mean	SD
Children:						
Age (in months)	42.80	15.46	33.18	14.68	34.50	15.33
Female (D)	0.44	0.50	0.42	0.49	0.51	0.50
Cognitive level (%)	78.58	11.77	74.82	8.34	75.22	6.15
Psychosocial level (%)	78.16	11.18	76.89	7.12	58.35	7.28
Psychomotor level (%)	72.52	13.80	68.59	13.74	73.46	6.88
Observations	236		235		235	
Community Mother:						
Age (in years)	53.32	7.35	46.83	8.92	48.85	8.21
Experience (in years)	19.42	6.28	11.22	5.99	15.54	8.74
Income (in 1'000 COP)	377.24	118.63	421.34	165.82	430.97	159.50
Education	2.57	0.79	2.49	0.75	2.17	0.90
Observations	20		20		20	

Note. Descriptive statistics for the three cohorts. The children from the first cohorts are born between 2007 and 2009. The children from the second cohort are born between 2007 and 2010, and the children from the third cohort are born between 2008 and 2010. Both age variables depict the respective age at the beginning of the intervention. The developmental status of the children depict the baseline measurement. For cohort 1, baseline was elicited in January 2011, for cohort 2 the baseline was elicited in January 2012. For cohort 3 the baseline was elicited in January 2013. The income of the community mothers corresponds to the income at the beginning of the treatment.

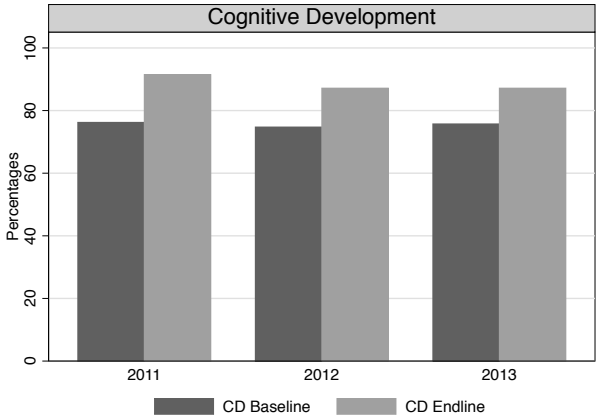

Figure B1: Cognitive development of the children in the three cohorts in percentages.

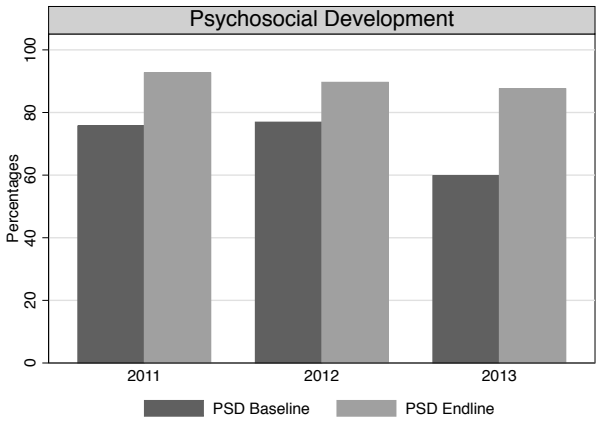

Figure B2: Psychosocial development of the children in the three cohorts in percentages.

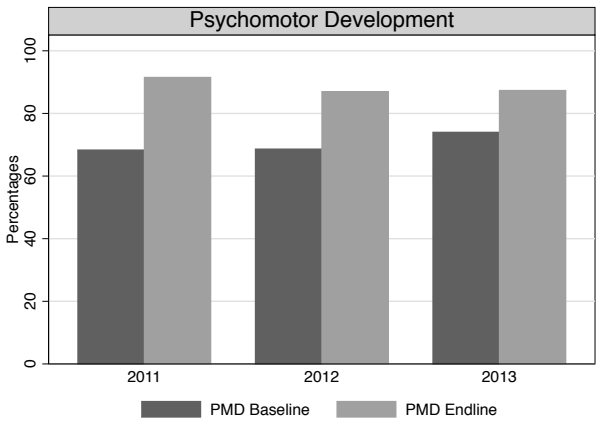

Figure B3: Psychomotor development of the children in the three cohorts in percentages.

Table B2: OLS estimates for sample 1.

	Cognitive	Psychosocial	Psychomotor
Treated (D)	16.222***	15.408***	17.982***
	(2.672)	(2.596)	(2.871)
Age (in months)	0.233	0.181	0.584***
	(0.144)	(0.146)	(0.185)
Female (D)	−2.723**	−2.596**	−2.311
	(1.060)	(1.101)	(1.467)
Age (in years)	0.163	0.261	0.200
	(0.178)	(0.172)	(0.185)
Experience (in years)	−0.355	−0.438*	−0.343
	(0.233)	(0.226)	(0.253)
Income (in 1'000 COP)	0.009	0.008	0.010
	(0.008)	(0.008)	(0.008)
Education	0.094	−0.526	−0.360
	(1.048)	(0.927)	(1.199)
Constant	60.590***	62.874***	41.604***
	(6.186)	(5.444)	(7.277)
F-test	21.4	15.8	19.3
Prob $> F$	0.000	0.000	0.000
R^2 adjusted	0.341	0.320	0.399
N	297	297	297

Note. OLS estimates. The unit of all dependent variables is percentages. Dependent variable in column 1 is cognitive development. In column 2 the dependent variable is psychosocial development and in column 3 it is psychomotor development. Robust standard errors, clustered on community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: OLS estimates for sample 2.

	Cognitive	Psychosocial	Psychomotor
Treated (D)	10.556***	29.963***	12.218***
	(0.758)	(0.704)	(0.876)
Age (in months)	0.255***	0.240***	0.338***
	(0.041)	(0.050)	(0.044)
Female (D)	−0.897	−0.382	−0.752
	(0.545)	(0.572)	(0.809)
Age (in years)	0.100	0.005	0.017
	(0.073)	(0.074)	(0.077)
Experience (in years)	−0.156***	−0.039	−0.076
	(0.053)	(0.090)	(0.087)
Income (in 1'000 COP)	−0.002	0.001	−0.002
	(0.002)	(0.003)	(0.003)
Education	−0.095	0.590	−0.017
	(0.470)	(0.479)	(0.627)
Constant	67.808***	52.005***	65.960***
	(4.363)	(3.252)	(3.978)
F-test	65.4	337.9	55.3
Prob $> F$	0.000	0.000	0.000
R^2 adjusted	0.514	0.892	0.534
N	331	331	331

Note. OLS estimates. The unit of all dependent variables is percentages. Dependent variable in column 1 is cognitive development. In column 2 the dependent variable is psychosocial development and in column 3 it is psychomotor development. Robust standard errors, clustered on community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: OLS estimates for the development of the children including interaction terms.

	Cognitive	Psychosocial	Psychomotor
Treated (D)	12.023*	16.060***	24.130***
	(6.362)	(5.942)	(7.034)
Age (in months)	0.347***	0.371***	0.737***
	(0.052)	(0.038)	(0.087)
Female (D)	−0.683	−0.345	−0.390
	(0.700)	(0.617)	(1.232)
Age (in years)	0.067	0.025	0.020
	(0.079)	(0.075)	(0.105)
Experience (in years)	−0.151**	−0.082	−0.082
	(0.062)	(0.091)	(0.122)
Income (in 1'000 COP)	0.002	0.003	0.002
	(0.004)	(0.004)	(0.004)
Education	−0.529	−0.071	−0.333
	(0.598)	(0.540)	(0.917)
Cohort 2 (D)	−5.120*	−3.268	−5.692**
	(2.595)	(2.648)	(2.782)
Cohort 3 (D)	−4.902*	−21.287***	−4.857
	(2.727)	(2.779)	(2.964)
Age × Treat	−0.203	−0.331**	−0.565***
	(0.129)	(0.136)	(0.154)
Gender × Treat	−2.318*	−2.302*	−1.975
	(1.204)	(1.239)	(1.479)
Age (MC) × Treat	0.134	0.226	0.204
	(0.209)	(0.202)	(0.221)
Exp (MC) × Treat	−0.238	−0.355	−0.347
	(0.277)	(0.279)	(0.294)
Educ (MC) × Treat	0.995	0.299	0.628
	(0.999)	(0.967)	(1.241)
Constant	70.398***	69.294***	57.380***
	(5.112)	(4.133)	(5.945)
F-test	67.9	179.5	61.1
Prob $> F$	0.000	0.000	0.000
R^2 adjusted	0.395	0.704	0.463
N	511	511	511

Note. OLS estimates. The unit of all dependent variables is percentages. Dependent variable in column 1 is cognitive development. In column 2 the dependent variable is psychosocial development and in column 3 it is psychomotor development. Robust standard errors, clustered on community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Probit estimates for being in the right grade.

	(1)	(2)	(3)	(4)
Treated (D)	0.401^{***}	0.436^{***}	0.410^{**}	0.808^{**}
	(0.153)	(0.160)	(0.172)	(0.322)
Age		-0.272^{***}	-0.275^{***}	-0.238
		(0.092)	(0.095)	(0.182)
Female		0.176	0.182	1.035^{***}
		(0.156)	(0.159)	(0.348)
Height (cm)		0.010	0.014	0.003
		(0.017)	(0.017)	(0.036)
Weight (kg)		-0.014	-0.016	-0.055
		(0.013)	(0.013)	(0.035)
Sisben		0.008^{*}	0.010^{**}	-0.004
		(0.005)	(0.005)	(0.011)
MC: Age			-0.008	0.055^{*}
			(0.011)	(0.029)
MC: Experience (y)			0.011	-0.041
			(0.015)	(0.033)
Parent: Age				0.015
				(0.017)
Parent: Income				0.000
				(0.001)
Parent: Degree				-0.295
				(0.264)
Constant	-0.022	1.206	0.911	0.630
	(0.092)	(1.627)	(1.775)	(3.541)
χ^2-test	6.9	36.8	35.3	21.6
p	0.009	0.000	0.000	0.027
Pseudo R^2	0.018	0.086	0.083	0.238
N	291	282	264	83

Note. Probit estimates. Dependent variable is if the child is in the right grade. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Probit estimates for being selfish.

	(1)	(2)	(3)	(4)
Treated (D)	-0.287^{*}	-0.230	-0.210	-1.179^{***}
	(0.157)	(0.189)	(0.203)	(0.410)
Age		0.026	-0.026	-0.258
		(0.105)	(0.102)	(0.283)
Female		0.423^{**}	0.427^{**}	0.651
		(0.168)	(0.177)	(0.463)
Height (cm)		-0.007	-0.013	0.003
		(0.015)	(0.016)	(0.043)
Weight (kg)		0.008	0.008	-0.025
		(0.015)	(0.015)	(0.048)
Entry into school		-0.036	0.034	-0.376
		(0.144)	(0.142)	(0.357)
Current class		-0.002^{**}	-0.002^{**}	-0.003
		(0.001)	(0.001)	(0.003)
Sisben		0.001	0.000	0.014
		(0.006)	(0.006)	(0.014)
MC: Age			-0.005	0.019
			(0.015)	(0.037)
MC: Experience (y)			0.008	0.018
			(0.019)	(0.040)
Parent: Age				-0.001
				(0.022)
Parent: Income				-0.001
				(0.001)
Parent: Degree				0.637^{**}
				(0.287)
Constant	0.536^{***}	1.634	2.736	4.109
	(0.106)	(1.731)	(1.873)	(5.166)
χ^2-test	3.3	16.4	16.3	36.1
p	0.068	0.037	0.092	0.001
Pseudo R^2	0.009	0.043	0.047	0.346
N	291	235	222	73

Note. Probit estimates. Dependent variable is being selfish, measured with four binary decision tasks. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B7: Probit estimates for trust.

	(1)	(2)	(3)	(4)
Treated (D)	0.271	0.297	0.270	0.695
	(0.172)	(0.190)	(0.198)	(0.467)
Age		-0.205^{*}	-0.237^{**}	-0.537^{**}
		(0.115)	(0.120)	(0.217)
Female		0.115	0.083	−0.404
		(0.172)	(0.172)	(0.428)
Height (cm)		0.005	0.002	0.023
		(0.016)	(0.017)	(0.035)
Weight (kg)		−0.007	−0.007	−0.028
		(0.014)	(0.015)	(0.032)
Entry into school		0.097	0.190	0.271
		(0.158)	(0.162)	(0.404)
Current class		0.001	0.001	0.004
		(0.001)	(0.001)	(0.003)
Sisben		−0.001	0.000	0.031^{**}
		(0.006)	(0.006)	(0.014)
MC: Age			0.015	0.017
			(0.015)	(0.040)
MC: Experience (y)			0.000	0.055
			(0.019)	(0.036)
Parent: Age				0.032
				(0.022)
Parent: Income				0.001
				(0.001)
Parent: Degree				0.089
				(0.320)
Constant	-0.717^{***}	−0.071	−0.625	−5.810
	(0.120)	(1.887)	(2.067)	(4.567)
χ^2-test	2.5	8.1	9.4	18.2
p	0.115	0.420	0.493	0.151
Pseudo R^2	0.009	0.027	0.031	0.210
N	291	235	222	73

Note. Probit estimates. Dependent variable is trust measured with a dummy variable. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B8: OLS estimates for risk.

	(1)	(2)	(3)	(4)
Treated (D)	0.121	0.173	0.261	0.344
	(0.171)	(0.189)	(0.199)	(0.413)
Age		0.178*	0.201*	0.048
		(0.104)	(0.103)	(0.148)
Female		0.141	0.205	0.417
		(0.171)	(0.176)	(0.401)
Height (cm)		−0.031**	−0.029*	0.041
		(0.015)	(0.016)	(0.041)
Weight (kg)		0.013	0.011	−0.043
		(0.010)	(0.011)	(0.032)
Entry into school		−0.007	−0.068	0.116
		(0.163)	(0.175)	(0.289)
Current class		−0.000	−0.000	−0.004
		(0.001)	(0.001)	(0.002)
Sisben		−0.012**	−0.014**	−0.015
		(0.005)	(0.006)	(0.012)
MC: Age			0.001	0.003
			(0.013)	(0.038)
MC: Experience (y)			−0.010	0.003
			(0.017)	(0.037)
Parent: Age				−0.001
				(0.019)
Parent: Income				0.000
				(0.001)
Parent: Degree				0.117
				(0.360)
Constant	2.485***	5.015***	5.020***	−1.838
	(0.110)	(1.722)	(1.915)	(5.182)
F-test	0.5	1.6	1.6	0.5
Prob $> F$	0.479	0.143	0.123	0.919
R^2 adjusted	−0.001	0.010	0.017	−0.126
N	291	235	222	73

Note. OLS estimates. Dependent variable is risk (starts invested) [0,5]. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B9: OLS estimates for patience.

	(1)	(2)	(3)	(4)
Treated (D)	0.161	0.303	0.186	0.260
	(0.190)	(0.208)	(0.218)	(0.367)
Age		−0.139	−0.125	−0.194
		(0.168)	(0.169)	(0.205)
Female		−0.130	−0.143	−0.182
		(0.217)	(0.231)	(0.465)
Height (cm)		0.013	0.011	−0.012
		(0.021)	(0.021)	(0.041)
Weight (kg)		−0.013	−0.013	0.043
		(0.017)	(0.017)	(0.039)
Entry into school		0.238	0.254	0.218
		(0.211)	(0.222)	(0.268)
Current class		0.002	0.002	0.003
		(0.001)	(0.001)	(0.003)
Sisben		−0.008	−0.010	0.003
		(0.008)	(0.008)	(0.012)
MC: Age			0.024	0.026
			(0.021)	(0.036)
MC: Experience (y)			−0.047	−0.019
			(0.028)	(0.037)
Parent: Age				0.017
				(0.019)
Parent: Income				0.002**
				(0.001)
Parent: Degree				−0.548
				(0.377)
Constant	1.077***	−0.384	−0.662	−0.407
	(0.102)	(2.201)	(2.414)	(5.711)
F-test	0.7	0.9	0.8	1.3
Prob $> F$	0.398	0.537	0.643	0.245
R^2 adjusted	−0.001	0.002	0.007	0.018
N	291	235	222	73

Note. OLS estimates. Dependent variable is patience (stars invested into future payoff) [0,5]. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B10: OLS estimates for self-esteem.

	(1)	(2)	(3)	(4)
Treated (D)	−0.001	−0.004	−0.012	−0.019
	(0.012)	(0.014)	(0.014)	(0.027)
Age		−0.029***	−0.027***	−0.006
		(0.007)	(0.007)	(0.013)
Female		0.026**	0.022*	0.057
		(0.012)	(0.013)	(0.034)
Height (cm)		0.000	−0.001	−0.001
		(0.001)	(0.001)	(0.002)
Weight (kg)		−0.000	−0.000	−0.000
		(0.001)	(0.001)	(0.003)
Entry into school		0.014	0.010	0.002
		(0.010)	(0.009)	(0.027)
Current class		−0.000	−0.000	−0.000
		(0.000)	(0.000)	(0.000)
Sisben		0.000	0.000	−0.000
		(0.000)	(0.000)	(0.001)
MC: Age			−0.000	0.003
			(0.001)	(0.003)
MC: Experience (y)			−0.001	−0.002
			(0.001)	(0.003)
Parent: Age				−0.002
				(0.002)
Parent: Income				−0.000
				(0.000)
Parent: Degree				−0.011
				(0.025)
Constant	0.851***	1.061***	1.167***	1.025***
	(0.007)	(0.135)	(0.143)	(0.309)
F-test	0.0	6.2	5.4	1.2
Prob $> F$	0.960	0.000	0.000	0.338
R^2 adjusted	−0.004	0.087	0.097	−0.057
N	272	219	206	65

Note. OLS estimates. Dependent variable is self-esteem measured in a questionnaire and averaged across all questions [0,3]. Robust standard errors, clustered on school and community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B11: OLS Regression for parental differences.

	(1) Altruism	(2) Risk	(3) Later	(4) Authoritative	(5) Authoritarian	(6) Permissive
Treated	0.114	−0.802*	0.223	−0.041	0.014	0.015
	(0.369)	(0.411)	(0.342)	(0.093)	(0.180)	(0.159)
Parent: Age	−0.015	−0.033*	−0.000	0.004	0.015	−0.008
	(0.026)	(0.019)	(0.032)	(0.005)	(0.009)	(0.010)
Parent: Income	0.002*	0.000	0.001	0.000	−0.001*	−0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Parent: Degree	−0.137	−0.421	0.054	−0.044	0.221	0.167
	(0.400)	(0.367)	(0.379)	(0.103)	(0.160)	(0.158)
Sisben	0.038***	0.008	−0.011	0.004	0.001	−0.001
	(0.013)	(0.011)	(0.015)	(0.004)	(0.006)	(0.006)
Constant	2.230	4.278***	3.361**	4.217***	2.450***	3.219***
	(1.399)	(1.183)	(1.649)	(0.314)	(0.543)	(0.640)
F-test	5.1	1.1	0.3	0.5	2.8	0.8
Prob > F	0.001	0.370	0.891	0.788	0.028	0.568
R^2 adjusted	0.117	0.032	−0.046	−0.034	0.042	−0.031
N	84	84	84	82	81	82

Note. OLS estimates. In column 1 the dependent variable is altruism [0,10]. In column 2 the dependent variable is the measure for risk preferences [0,5]. In column 3 the dependent variable is patience [0,5]. Column 4 to 6 are the effects on the parenting styles [0,5]. Robust standard errors, clustered on community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B12: Probit regression for parental differences.

	(1) Trust	(2) Egalitarian	(3) Alturist	(4) Spiteful	(5) Selfish
Treated (D)	0.023	−0.233	0.685	0.271	−0.212
	(0.271)	(0.349)	(0.440)	(0.445)	(0.306)
Age	−0.011	0.020	−0.026	0.014	−0.006
	(0.020)	(0.020)	(0.022)	(0.022)	(0.017)
Income	0.001	0.001	0.001	−0.001**	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Degree	−0.227	0.249	−0.073	−0.363	−0.183
	(0.263)	(0.252)	(0.336)	(0.317)	(0.327)
Sisben	0.015*	0.009	0.042***	0.012	−0.006
	(0.009)	(0.012)	(0.016)	(0.013)	(0.011)
Constant	−0.424	−2.811**	−2.801*	−1.554	0.252
	(1.126)	(1.180)	(1.497)	(1.252)	(1.035)
χ^2-test	6.7	2.8	17.4	15.6	2.9
p	0.244	0.735	0.004	0.008	0.723
Pseudo R^2	0.055	0.047	0.232	0.143	0.020
N	84	84	84	84	84

Note. Probit estimates. In column 1 the dependent variable is trust. Column 2 to 5 depict the results of the binary measures for social preferences. In column 2 it is the probability to be classified as egalitarian, in column 3 to be classified as altruist, in column 4 as spiteful, and in column 5 as selfish. Robust standard errors, clustered on community mother, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.C Translated Instructions for the Children

Short introduction

Hi! I'm *name of the research assistant*

What's your name?

Now we're going to play some games. In all games, you can earn stars. For each star that you earn during these games, you can buy one item at the "shop". I will show you the shop now.

Figure C4: Example of a shop.

Explanation for the egalitarian choices

You get some stars, and you can give stars to another child as well.

Here I have two envelopes. One is for you.

Let's write your name on it.

The other envelope is for another child that is of your age and goes to the same school. But both of us do not know exactly who that other child is. So we will not write a name on the second envelope. The envelopes are for the stars that you and the other children are going to get in this game.

Instructions for the (1/1) vs. (1/0) choice

Note. The instructions for the four choices were identical except for the amount of stars. As an example, I depict the instructions for the (1/1) vs. (1/0) choice.

Okay! Here are two sheets of paper. *(Place both sheets of paper in front of the child (right and left).*

In each sheet, you can see two circles with arrows. On both sheets, one arrow is pointing to the bag of the other child, and the other arrow is pointing to you and your bag.

On the first sheet, I place one star in the circle that is closer to the bag of the other child, and I place one star in the circle that is closer to you. *(Place the stars in the circles now.)*

On the second sheet, I place one star in the circle that is closer to you. In the other circle that is pointing to the bag of the other child, I place no star.

You can now choose one of those sheets.

(Point to the respective sheets.) If you choose this one, this arrow is pointing at you. That means you are getting what is placed in this circle, one star.

The other circle is pointing to the bag of the other child. This means that the other child is getting what is placed in this circle, one star.

If you choose this sheet, you get one star and the other child gets no star.

Before you can decide, I have some questions for you.

Questions

(Point to the first sheet.) If you choose this sheet, what will the other child get? And what will you get?

(Point to the second sheet) If you choose the second sheet, what will the other child get? And what will you get?

Do you have any questions?

Decision

Alright, now you can choose a cardboard. Which sheet do you choose? *(Put the stars into the envelopes)*

Dictator game

Now we will again play a game in which you can win stars. As I told you, you can exchange the stars for a present later on. The more stars you win, the more presents you can buy at our "shop".

You will now get some stars, and you can give stars to another child as well.

Here I have two envelopes. The first one is for you. Let's write your name on it.
The other envelope is for another child that is of your age and goes to the same school. But both of us do not know exactly who that other child is. So we will not write a name on the second envelope. The envelope are for the stars that you and the other child are going to get in this game.
(Put both envelopes side by side on the table in front of the child.)
Look! Here are ten stars.
(Put the stars in front of the child between the two envelopes.)
Now you can decide how many stars you put on your envelope and how many you put on the envelope of the other child. The stars on your envelope *(point to the envelope)* are for you. The stars on the other envelope are for the other child. I will make sure that the other child gets something nice for the stars.

Decision

You can decide how you want to divide the stars. You can split the stars, or you can put all stars on one envelope.
Now, please put all ten stars on the envelopes in a way you would like to have it.
(Child puts the stars on the bags.) Okay. How many stars do you get? And how many stars does the other child get? Fine. Let's now put the stars in the envelopes.
(In case the answers are not correct: explain the correct answer and ask for a new suggestion. In case both answers are correct: Remove the envelopes.)

Patience

You will receive five stars *(please place the five stars in front of the child).*
You have to decide how many of these five stars you want to put in the box labeled "NOW *(point at the left box)* and how many stars you want to put in the box labeled "IN TWO WEEKS" *(point at the right box).*
You will receive the stars that you put in the box "NOW" immediately after the game, and you can use these stars for buying presents in our present shop. You can take these presents home today. Each star that you put in the box "IN TWO WEEKS" will be doubled and you will receive the presents that you choose with these stars in 2 weeks only.

Let's consider an example: If you, for instance, want to receive two stars today, what do you have to do? *(Answer of the child: I have to put two stars in the left box)*

And what happens with the other three stars? *(Answer: I have to put these stars in the right box; please let the child demonstrate this)*

How many stars will be added to this box? *(point at the right box; answer of the child: 3; please demonstrate!)*

How many stars are in the box in total? *(Answer of the child: 6)*

When will you receive the presents which you can choose with these six stars? *(Answer of the child: in two weeks)*

And what happens if you put five stars in that box? *(point at the left box; Answer of the child: then I will receive five stars immediately after the game, and I can choose presents with these five stars which I can take home today)*

And what happens if you put all five stars in that box? *(point at the right box; Answer of the child: then these stars will be doubled and I can choose presents with the ten stars which I will receive only in two weeks.)*

Could you please repeat the rules of the game?

Decision

Ok, now you can take your decision, remember the stars which you want to receive today, you have to put them in this box *(point to the NOW box)*. And the stars which you will receive in two weeks you have to put them in the IN TWO WEEKS box *(point to the XXX box)*.

Take as much time as you need for your decision. In the meantime, I will turn around, so I don't disturb you. Just call me when you are done.

Figure C5: Example of how the boxes for the patience decision looked like.

Risk

In the beginning, you will receive five stars. *(Place the five stars in front of the child)*
You have to decide how many of these five stars you want to keep for sure and with how many of these stars you want to play the "wheel of fortune"-game. You have to put the stars you want to keep for sure in this box *(point at the left box)*.
Likewise, you must put the stars with which you want to play the treasure-game in that box *(point at the right box)*.
I will triple the amount of stars that you put into the wheel of fortune box. Then you can turn the wheel of fortune. If it stops in one of the colored areas, for example, *(point at these areas)*, then you will receive all the stars from this box. On the other hand, if it stops at a gray area *(point at this area of the wheel)* then you will lose all the stars from this box.
At the end you will receive the stars that you keep for sure *(point at the left box)* and the stars that you win in the wheel of fortune game *(point at the right box)*.
Let's consider an example: If you, for instance, want to keep one star for sure and play the wheel of fortune game with the other four stars, what do you have to do? *(Answer of the child: I have to put 1 star in the left box and four stars in the right box; please let the child demonstrate this)*

How many stars will be added to this box? *(point at the right box; answer of the child: 8; please demonstrate!)* What happens next? How does the wheel of fortune work? *(Child has to repeat the rules of the game).*

How many stars will you win if the wheel stops in one of the colored areas? *(Answer of the child: 12 stars).*

And how many stars will you receive in total? *(Answer of the child: 13).*

Exactly. You will receive 12 stars from the wheel-of-fortune-game plus one additional star which you kept for sure.

What happens if the wheel stops in a gray area? *(Answer of the child: I lose all the stars of the wheel of fortune game)*

Exactly.

How many stars will you receive in total? *(Answer of the child: 1)*

Exactly. This was only an example. Let's consider another example: Could you please explain the rules of the game if you want to keep four stars for sure and play the wheel-of-fortune-game with 1 star? *(The child has to recapitulate the game with the new example).*

What happens if you, for instance, put all your five stars in this box? *(point at the right box; let the child recapitulate the game)*

What happens if you, for instance, put all your five stars in this box? *(point at the left box; let the child recapitulate the game)*

Decision

Please take your decision now. You have to put the stars which you want to keep for sure in this box *(point at the left box)* and the stars with which you want to play the wheel-of-fortune-game have to be put in that box *(point at the right box).* Take as much time as you need for your decision. In the meantime, I will turn around, so I don't disturb you. Just call me when you are done.

Figure C6: The wheel of fortune.

Trust

In this game, you will be playing with another child that is of your age and goes to the same school. You will now get one star, and you can give this star to the other child, or you can keep it for yourself.

If you keep it, the other child gets nothing. If you give the star to the other child, the other child will get four stars.

Then the other child can decide whether to keep all four stars or whether to keep two stars and give the other two stars back to you.

Let's consider an example. What happens if you keep the star? *(Answer of the child: I will have one star, and the other child will not get anything).*

And what will happen if you decide to transfer the star? *(Answer of the child: The other child will get four stars and can decide then whether to transfer two stars back to me or whether to keep all four stars.)*

Exactly, and how many stars will you have if the other child decides to not transfer any stars back to you? *(Answer of the child: zero)*

Exactly.

And how many stars will you have if the other child decides to transfer two stars back to you? And how many stars will the other child have? *(Answer of the child: two each).*

Decision

Do you want to transfer the star to the other child?

Now, let's consider that another child transferred the star to you. Hence, you have four stars now. Do you want to transfer two stars back?