DISS. ETH NO. 22135

SOCIAL VALUE ORIENTATION: AN ANALYSIS OF MEASUREMENT, FORM, PREDICTIVE POWER, AND MALLEABILITY OF SOCIAL PREFERENCES

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. Sc. ETH Zurich)

presented by
KURT ALEXANDER ACKERMANN
M. Sc. UZH, University of Zurich

born on 18.06.1982

citizen of
Egnach (TG) and Frauenfeld (TG)

accepted on the recommendation of

Prof. Dr. Ryan O. Murphy
Prof. Dr. Andreas Diekmann
Prof. Dr. Urs Fischbacher
Prof. Dr. Robert Böhm

2014

The best way to come to truth being to examine things as really they are, and not to conclude they are, as we fancy of ourselves, or have been taught by others to imagine.

— John Locke, 1690, An Essay Concerning Human Understanding, Book II, Chapter XI

Acknowledgements

The completion of this dissertation would not have been possible without the help and support of several people. First of all, I would like to express my genuine and deepest gratitude to my supervisor, Prof. Dr. Ryan O. Murphy, who helped me accomplish goals I sometimes thought I could not possibly reach.

Ryan, thank you so much for your excellent guidance and mentoring, your patience, and especially for believing in me at times when I had doubts that I would be up to the task. You've always kept on pushing me out of my comfort zone but never beyond the cliff of frustration. I admire your personal and scientific integrity as well as your intellectual honesty, and I feel honored that I have been given the opportunity to learn from such an outstanding teacher and brilliant mind.

Also, I would like to thank the further members of the doctoral committee, namely Prof. Dr. Andreas Diekmann, Prof. Dr. Urs Fischbacher, and Prof. Dr. Robert Böhm for insightful comments and discussions, challenging and inspiring questions, and reliable support.

Additionally, I thank the Swiss National Science Foundation for financial support. Concretely, the research reported in this dissertation has been supported to a large extent by SNF grant 100014_143199/1.

Furthermore, I would like to express my appreciation to several colleagues for valuable conversations, critical comments, constructive suggestions, and warm encouragement. In particular, I want to thank Robert ten Brincke, Jürgen Fleiß, Karsten Donnay, and Joël Berger. Special thanks also to my dear friends Daniel Geeler, Stefan Brandner, Gerry Strasser, and Patrick Bürgi for reminding me from time to time that there is a world outside academia.

Clearly, I also want to thank my family for their unconditional support and affectionate companionship. My parents Kurt W. Ackermann and Ursula Ackermann; my sisters Claudia Zahn and Yvonne Jürgens; my brothers-in-law Stefan Zahn and Alexander Jürgens; and my nephew and nieces Dennis, Sophie, and Lucy Zahn.

Last but not least, I want to thank my beloved partner, Sylvie Pantano, for constant support, patient encouragement, and a myriad of other things. The value of love in life is priceless, and so are you.

Zürich, November 13, 2014 K. A. A.

Abstract

The general topic of this dissertation is an interdisciplinary construct termed Social Value Orientation (SVO), which captures the idea of how much weight a person attaches to the welfare of others in relation to his or her own. The first substantive (i.e. succeeding the introduction and preceding the conclusion) chapter of the following manuscript is concerned with reviewing the literature on SVO with a focus on how the construct has been theorized and measured. For the past forty years, SVO has been assessed almost exclusively –and often thought of, too– as a categorical variable, such that people are simply classified as being prosocial, individualistic, or competitive. We argue that SVO is ultimately a continuous construct that allows for a fine-grained assessment of people's concerns for the well-being of others. We discuss the interplay between SVO measurement and theory from a historical perspective and review the measures that have been used to assess this construct so far. Furthermore, we evaluate these measures according to a set of predefined criteria, including reliability, validity, and output resolution.

The second substantive chapter introduces a novel SVO measure, which is called the SVO Slider Measure, that assesses the construct in congruence with its theoretical conceptualization, namely on a continuous scale. The measure is explicated in detail and the study conducted to validate it is reported. We show that the SVO Slider Measure does not only outperform the previously most commonly used measures in terms of output resolution, but also in terms of reliability, while being at least as valid. We highlight the measure's additional strengths –such as the possibility to check for random responding, for instance– and new opportunities it brings about. For example, now the construct can not only be used as an independent variable with higher resolution –translating into higher statistical power–, but can also be used as a dependent variable. That is, the measure allows for investigating the malleability of SVO and how SVO changes in response to changes in the environment, for example.

The third substantive chapter is concerned with the mathematical modeling of the SVO construct. In both psychology and economics, utility functions have been proposed and used to estimate parameter values that are indicative of how different allocations of resources between the self and another person are perceived and evaluated. Over the last decades, numerous functions of this kind have been suggested and tested, but no attempts have been made to compare them systematically and rigorously. Furthermore, parameters have predominantly been estimated on the aggregate level, i.e. across subjects. However, there is accumulating evidence that people are very heterogeneous in their preferences over own-other resource allocations. Hence, it makes sense to estimate parameters on the individual level. Therefore, we fit prominent models of social preferences to data from more than 200 subjects at the individual level according to three different statistical fitting procedures. This method allows for a comparative evaluation of the models both in terms of goodness-of-fit and psychological interpretability of individual best fitting parameter values. We found that there are substantial differences in the models' abilities to explain choice behavior and that there is variance across the models as to how well best fitting parameter values can be interpreted psychologically. We propose an additional model that yields a good compromise between model complexity, goodness of fit, and psychological interpretability.

The final two substantive chapters of the following manuscript report on studies that used SVO both as an independent and as a dependent variable. Concretely, in the first of these two studies, we investigated the relative impact of preferences and beliefs on contribution levels in a public goods game, and how the contribution levels of interaction partners feed back on individual preferences and beliefs. We show that people have –on average– a tendency to be imperfect conditional cooperators, i.e. they contribute a little less than they believe their interaction partners are going to contribute in a particular period. This observation is also expressed in the result that the best single predictor of a person's contribution level is that person's belief about the contribution levels of others in a particular period. However, SVO explains variance in contribution levels above and beyond of what beliefs can account for throughout a repeated public goods games. Furthermore, we show that people do not only update their beliefs according to observed contribution levels of interaction partners, but do also update their preferences. That is, people who happened to interact with uncooperative others were likely to become less prosocial throughout the repeated play, while people who happened to interact with cooperative others were likely to become more prosocial. In the second study –reported in the final substantive chapter– we show that many people change their concerns for others depending on information provided about the other. Concretely, when people were informed that they are matched with a competitive or individualistic other, they were likely to become less prosocial compared to their baseline SVO. And likewise, if knowingly matched with a prosocial or altruistic other, people tend to become more prosocial towards these others compared to the baseline measurement. These shifts in SVO we refer to as reciprocity, and the SVO Slider Measure can be used for measuring this individual difference score directly and unambiguously.

Zusammenfassung

Das Kernthema der vorliegenden Dissertation ist ein interdisziplinäres Konstrukt names Social Value Orientation (SVO). Dieses Konstrukt erfasst das Gewicht, welches eine Person auf das Wohlergehen anderer Leute im Verhältnis zum eigenen Wohlergehen legt. Das erste substanzielle (d.h. der Einleitung folgende und der Konklusion vorangehende) Kapitel der vorliegenden Arbeit befasst sich mit der Besprechung der Literatur zu SVO mit einem Fokus auf die Thematik, wie das Konstrukt gewöhnlich theoretisiert und gemessen wurde. Im Verlauf der letzten 40 Jahre wurde SVO fast ausschliesslich auf Nominalskalenniveau gemessen und konzeptionell auch oft als kategoriales Konstrukt aufgefasst. Das heisst, Individuen wurden schlicht als einer der drei folgenden Kategorien zugehörig klassifiziert: Kompetitiv, individualistisch (bzw. egoistisch), oder prosozial. Wir argumentieren allerdings, dass SVO letztendlich ein kontinuierliches Konstrukt ist, welches theoretisch eine unendlich feine Gradierung sozialer Präferenzen erlaubt. Wir diskutieren das Zusammenspiel von Theorie und Messung im Bezug auf SVO aus historischer Perspektive und besprechen die Instrumente, welche bis anhin zur Messung des Konstrukts verwendet wurden. Ferner bewerten wir die Instrumente hinsichtlich einer Liste vordefinierter Kriterien, wie etwa Reliabilität, Validität, und Skalenniveau der Messung.

Im zweiten substanziellen Kapitel wird ein neues SVO Messinstrument vorgestellt –das SVO Slider Measure–, welches das Konstrukt in Übereinstimmung mit seiner theoretischen Konzeptionierung erfasst, namentlich als kontinuierliche Variable. Das Instrument wird im Detail erklärt und die Studie, welche zum Zwecke seiner Validierung durchgeführt wurde, wird geschildert. Wir konnten zeigen, dass das SVO Slider Measure den zuvor am häufigsten verwendeten SVO Massen nicht nur bezüglich Skalenniveau überlegen ist, sondern bei gleichwertiger Validität auch eine höhere Reliabilität aufweist. Wir zeigen die weiteren Stärken des Instruments auf –wie etwa die Möglichkeit zur Detektion zufälligen Antwortverhaltens– und weisen darauf hin, welche neuen Gelegenheiten sich durch das Instrument eröffnen. Das Konstrukt kann nun zum Beispiel nicht mehr nur als unabhängige Variable mit höherer Auflösung –und dadurch mehr statistischer Power– verwendet werden, sondern auch als abhängige Variable. Das heisst, das Instrument erlaubt es, die Formbarkeit sozialer Präferenzen zu untersuchen und zu erforschen, wie sich SVO in Abhängigkeit von Kontextvariablen verändert.

Das dritte substanzielle Kapitel befasst sich dann mit der mathematischen Modellierung des SVO Konstrukts. Sowohl in der Psychologie wie auch in der Ökonomie sind verschiedene Nutzenfunktionen vorgeschlagen und verwendet worden um Parameterwerte zu schätzen, welche

indizieren, wie unterschiedliche Individuen die Aufteilung von Ressourcen zwischen sich selbst und einer anderen Person wahrnehmen und bewerten. In den letzten Jahrzehnten hat sich die Anzahl vorgeschlagener Modelle akkumuliert, ohne dass die verschiedenen Nutzenfunktionen systematisch miteinander verglichen worden wären. Des Weiteren wurden Parameterwerte mehrheitlich auf Aggregatsebene geschätzt, d.h. über Individuen hinweg. Es haben sich allerdings Befunde akkumuliert, welche zeigen, dass individuelle Präferenzen bezüglich der Aufteilung von Ressourcen zwischen sich selbst und anderen Menschen überaus heterogen sind. Es macht daher Sinn, Parameterwerte auf der Individualebene zu schätzen. Aus diesen Gründen haben wir passende Parameterwerte von einer beträchtlichen Anzahl prominenter Nutzenfunktionen zu Daten von über 200 Versuchspersonen auf der Individualebene mit drei unterschiedlichen Verfahren geschätzt. Dieses Vorgehen erlaubt es, die verschiedenen Modelle komparativ hinsichtlich Güte der Passung und psychologischer Interpretierbarkeit zu bewerten. Wir haben substanzielle Unterschiede im Ausmass gefunden, in dem die Modelle eine psychologische Interpretation der Parameterwerte zulassen und dahingehend, wie gut die Modelle individuelles Entscheidungsverhalten erklären können. Zudem schlagen wir ein neues Modell vor, welches einen guten Kompromiss zwischen Modellkomplexität, Güte der Passung, und psychologischer Interpretierbarkeit schliesst.

In den letzten beiden substanziellen Kapiteln der vorliegenden Arbeit wird über zwei Studien berichtet, in welchen SVO sowohl als unabhängige, wie auch als abhängige Variable verwendet wurde. In der ersten dieser beiden Studien untersuchten wir den relativen Einfluss von Präferenzen und Erwartungen auf das Beitragsniveau in einem "öffentlichen-Guts-Spiel" auf Individualebene, und wie das Beitragsniveau von Interaktionspartnern wiederum die individuellen Präferenzen und Erwartungen beeinflusst. Wir konnten zeigen, dass Individuen tendenziell imperfekt konditional kooperativ sind, was bedeutet, dass sie etwas weniger als den subjektiv erwarteten durchschnittlichen Beitrag der Interaktionspartner zum öffentlichen Gut transferieren. SVO erklärt jedoch zusätzlich zu subjektiven Erwartungen signifikant Varianz in den individuellen Beiträgen über die gesamte Spieldauer hinweg. Ferner konnten wir zeigen, dass Individuen aufgrund des beobachteten Verhaltens der Interaktionspartner nicht nur ihre Beitragserwartungen aktualisieren, sondern auch ihre sozialen Präferenzen: Individuen, welche zufällig mit eher unkooperativen Versuchsteilnehmern interagierten tendierten dazu, egoistischer zu werden, wohingegen die Interaktion mit kooperativen Versuchsteilnehmern zu einer Tendenz in Richtung Prosozialität führte. In der zweiten Studie – über welche im letzten substanziellen Kapitel berichtet wird – konnten wir zeigen, dass viele Individuen ihre sozialen Präferenzen in Abhängigkeit von Informationen über einen jeweiligen, spezifischen Interaktionspartner verändern. Konkret zeigte sich, dass viele Individuen dazu tendieren, weniger prosozial (d.h. egoistischer) zu werden, wenn sie wissentlich mit einer erwiesenermassen kompetitiven oder egoistischen anderen Person interagierten. Umgekehrt tendieren viele Individuen dazu, im Bezug auf den Interaktionspartner prosozialer zu werden, wenn dieser erwiesenermassen prosozial oder gar altruistisch ist. Diese Veränderungen in den sozialen Präferenzen als Antwort auf offenbarte Information über einen Interaktionspartner definieren wir als Reziprozität, wobei das SVO Slider Measure dazu verwendet werden kann, diese Differenzwerte eindeutig und präzise zu erheben.

Contents

List of Figures

List of Tables

1 Introduction

The high degree of concern for the well-being of unrelated conspecifics is one of the major attributes by which humans differ substantially from other animals. However, people are heterogeneous with respect to the direction and intensity of care for others. Some people, for instance, are willing to give up parts of their own personal resources in order to benefit other people they do not know and will never knowingly interact with. That is, some people are willing to donate money – even blood– to unrelated and anonymous others (e.g. Bennett, 2006; Gillespie & Hillyer, 2002), or invest time in voluntary activities for the benefit of strangers (Smith, 1994). Other people, in contrast, are willing to give up personal resources for the detriment of other humans they do not know and will never knowingly meet, such as investing time to program computer malware (Slade, 2004), for instance. And yet other people are only concerned about their own material well-being, hence not willing to give up anything for the good, or the detriment of anonymous others. The diversity of peoples' concerns for others has been studied extensively by researchers from various disciplines under a variety of technical terms, such as *social motivation*, *welfare tradeoff ratios*, *social preferences*, *other regarding preferences* or *social value orientation* (SVO).

However, the assumption that peoples' behavior may not be driven solely by narrow self-interest, but rather by diverse motives concerning the well-being of others has been accepted surprisingly recently in certain disciplines, such as economics, for instance. For decades, standard economic theory has modeled humans as exclusively selfish beings (*Homo economicus*) whose only concern is the maximization of own material well-being, irrespective of the well-being of others. Furthermore, the phenomena of human cooperation and apparently genuinely altruistic acts have been perceived as puzzling by many scientists across various disciplines (see e.g. Gintis, 2003). There is a large discrepancy between the scientific perspective and the popular perception regarding social preferences and human cooperation. While these phenomena are often perceived as surprising and puzzling in the sciences, they are usually perceived as natural and self-evident in everyday life. In fact, expressions of social preferences in daily life are ubiquitous and examples readily come to mind:

- leaving a tip in a restaurant for rewarding the waiter/waitress,
- helping a person to lift the baby buggy into the bus or tram,
- helping a tourist to find the way to the location he desires to visit,
- making one's seat available to an elderly person in public transportation,
- etc.

These types of daily observable behavior are expressions of prosocial preferences, since all of them require the decision makers to spend own resources (money, time, or effort) for the benefit of an anonymous other person. The expenditures in the above examples are all very small. However, in a world of perfect selfishness, even tiny expenditures of that kind are not expected to occur. In real life, however, at least some people sometimes spend considerable resources to benefit strangers. Expenditures under this rubric involve donations to charity, blood donations, or time and effort for rendering first aid as a casual bystander of a traffic accident, for instance. The latter behavior is not unambiguously an expression of social preferences, though. In several countries, the failure to render assistance to a person in danger constitutes a violation of the law, and a behavior that results solely from the motivation to avoid prosecution is not an expression of social preferences, but of narrow self-interest. It is hard to believe, however, that every person who renders first aid to a stranger does so because of selfish concerns. Furthermore, the failure to render assistance to a person in danger is generally not prosecuted if helping the person in need poses unreasonable demands on the aider, such as risking his or her own life. Hence, in a world of perfect human selfishness, potentially fatal helping behavior is not expected to occur. However, instances of heroic helping behavior have been reported (e.g. Allison & Goethals, 2010; Becker & Eagly, 2004; Shellenbarger, 2012). Examples of heroic acts in this respect are plenty and range from individuals who jumped into a river to save another person from drowning to individuals who saved Jews from the Holocaust in World War II at the risk of losing their own lives. Heroism and altruism are distinct concepts (Franco & Zimbardo, 2011), though, because it clearly takes more than just prosocial preferences to risk one's life to save another one. But the consideration and valuation of an anonymous other person's welfare is a natural prerequisite for any kind of heroic helping behavior. In a world of perfect human selfishness, where no human considers another human's well-being in the first place, heroes do not exist.

In order to investigate a phenomenon scientifically it is not sufficient, however, to just have an intuition about it, or that anecdotes about it are available. The scientific investigation of a phenomenon requires its quantitative measurement, or as William Thomson (Lord Kelvin, see Thomson, 1889, p.73-74) put it:

> [...] When you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.

Hence, for the phenomenon we discuss here, the challenge is to assess in a quantitative way how "nice" a person is, that is, how much weight a person attaches to the welfare of others in relation to the own. One obvious way to address this question would be to just ask people how "nice" they are on a Likert-type scale from, say, 1 to 10. However, with such an approach, one would likely not measure what is intended to be measured. The lack of validity would stem from the fact that people would necessarily answer the question "How nice *do you think* you are?", rather than "How nice are you?". That is, one would not measure "niceness", but would let the addressee of the question subjectively assess his or her "niceness" and then record the result of this introspective subjective estimate. In other words, each addressee of the question would use his or her own subjective scale of measurement to assess "niceness", and comparisons between subjects would consequently be invalid. Another problem with such an approach would be the subjectivity of the term "nice". One person may consider it "nice" to not run around and kill the people one dislikes, while another person with higher standards would consider it "nice" to donate half of his income to charity. Again, comparisons between people would be invalid due to the lack of an objective scale with an unambiguously defined criterion.

In order to get rid of both the problem of subjective scaling and ambiguous criteria, one could let people make decisions rather than having them indicate their opinions. As has been done in studies of helping behavior, for instance, one could make people believe that they are encountering a person in need, and then see whether they help or not (e.g. Bryan & Test, 1967). However, the rigorous quantification of a person's niceness would still not be feasible. First, one could not quantify the degree of a person's niceness. For instance, a comparison between two people who both did help would not be possible because the dependent variable assessed is dichotomous and has the same value for both people. Second, if we compare one person who helped on the one hand to another person who did not help on the other hand, we may make a mistake in assuming that the former person is nicer. The latter person may have been in a hurry or may have misinterpreted the situation. In other words, one's measure of niceness is likely confounded by an indeterminate number of other variables that can not be controlled for. So, ideally we would like to assess the weight people attach to the welfare of others by letting people make real decisions with real consequences in a controlled and unambiguous setting where confounding variables can be ruled out. Furthermore, it is necessary to ensure that the scale used for assessing the variable of interest is well calibrated, such that identical scores obtained by different people on that scale are, in fact, accurately reflecting identical degrees of the concern for others. One way to ensure this is to use a unit of measurement that is the same for all people, i.e. something all people value to the same degree and can thus serve as an objective scale. A scale that fulfills this requirement is money, precisely because it serves as a common measure of value (see e.g. Jevons, 1875, chapter three).

Social preferences encompass narrow self-interest and deviations from it in two directions: altruism and spite. Following biological definitions, altruistic behavior is exclusively costly for the self and exclusively beneficial for others, while spiteful behavior is both exclusively costly for the self and others (for a discussion of this issue, see Bowles, 2004, chapter three). Hence, if a person assigns a positive weight to the welfare of others, we would expect that this person is willing to incur costs in order to benefit some other person. Contrary, if a person assigns a negative weight to the welfare of others, we would expect that this person is willing to incur costs in order to harm some other person. If we employ a common measure of value such as money, then we can assess how large a cost –if any– people are willing to incur in order to benefit or harm another person in an unambiguous and objective way. The question “how nice are you?” then becomes “how much are you willing to pay in order to increase (or decrease) another person’s budget by one monetary unit?". The derivation and utilization of this idea for the purpose of measuring social preferences in a valid, reliable, and efficient way is what concerns us in chapters two and three of this dissertation. Chapter four is then focusing on the mathematical modeling of what is being measured, and chapters five and six provide studies on how the reliable and efficient measurement of social preferences facilitates using the construct both as an independent and as a dependent variable. Concretely, the thesis is structured as follows.

Chapter two is a review of how social preferences have been conceptualized and measured across the social sciences, while a focus is put on a psychological account when discussing theoretical issues. Henceforth, we will use the label *social value orientation* (SVO) when talking about social preferences for reasons that will be highlighted in chapter two. Chapter three introduces a new measure of SVO –the SVO Slider Measure– that is an improvement over existent ones in terms of both output resolution and reliability, which opens up new opportunities for investigating the nature of peoples’ concerns for others. The measure is then explicated in detail and evidence for its strong psychometric properties is provided. In chapter four, it is shown how SVO can be modeled mathematically and different mathematical representations of SVO are compared to each other in a horse-race manner both with respect to how well they can explain empirical data and in terms of how well best fitting parameter values can be interpreted psychologically. Chapter five reports on a study where SVO is used both as an independent and as a dependent variable. More precisely, SVO as assessed by means of the SVO Slider Measure was used for predicting behavior in a public goods game and the relative impact of SVO and beliefs about the behavior of interaction partners on individual behavior was estimated. Furthermore, it was shown that the behavior of interaction partners can alter people’s concern for others, which provides evidence that people do not only update their beliefs, but do also update their preferences in response to the behavior of others. Chapter six then reports on the malleability of SVO and its relation to reciprocity. It is shown how the weight people attach to the welfare of another person changes depending on revealed information about the other person’s preferences. The dissertation then closes in chapter seven with concluding remarks on the topics discussed and investigated.

2 Social Value Orientation: Theoretical and measurement issues in the study of social preferences

Abstract[1]

What motivates people when they make decisions and how those motivations are potentially entangled with concerns for others are central topics for the social, cognitive, and behavioral sciences. According to the postulate of narrow self-interest, decision makers have the goal of maximizing personal payoffs and are wholly indifferent to the consequences for others. The postulate of narrow self-interest –which has been influential in economics, psychology, and sociology– is precise and powerful but is often simply wrong. Its inadequacy is well known and efforts have been made to develop reliable and valid measurement methods to quantify the more nuanced social preferences that people really have. In this paper, we report on the emergence and development of the predominant conceptualization of social preferences in psychology: social value orientation (SVO). Second, we discuss the relationship between measurement and theory development of the SVO construct. We then provide an overview of the literature regarding measurement methods that have been used to assess individual variations in social preferences. We conclude with a comparative evaluation of the various measures and provide suggestions regarding the measures' constructive use in building psychologically realistic theories of people's social preferences.

> No man is an island, entire of itself; every man is a piece of the continent, a part of the main. If a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thy friend's or of thine own were.
>
> –John Donne
> Devotions Upon Emergent Occasions
> Meditation XVII

[1] This chapter is an edited preprint version of the following paper.
Murphy, R.O. and Ackermann, K.A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review. 18*(1), 13-41

2.1 Introduction

People often cooperate with each other –they help others in need, volunteer their efforts, contribute money and goods to causes, and even donate blood –all to benefit unrelated and anonymous others without the expectation of gain or reward. Why people choose to cooperate and act prosocially is a deep question that has fostered active research areas across many disciplines, including biology and all the social and behavioral sciences. A prerequisite for many kinds of cooperative behavior is that decision makers (DMs) consider the wellbeing of others when contemplating their options. That is to say that people may have social preferences and that these preferences promote behavior that is beneficial to others even though it is costly to the actor.

To illustrate the notion of social preferences, consider the following choice (Table 2.1) between two options. In this example the decision maker (DM) is selecting between certain distributions of resources, some amount to herself, and some amount to be allocated to some other randomly determined person. The DM and the other person will remain mutually anonymous during and after the decision is made, and there is nothing the other person can do to affect the DM in any way. Hence this is not a strategic decision (i.e., not within the purview of game theory, as only one DM influences the payoffs for both people) but rather this is a one-shot individual decision under certainty, free of potential repercussions or reprisals. Nonetheless this choice has a social dimension, as the DM will have an effect on another person and the DM is aware of this potential effect. Choices in this austere context can reveal a great deal about a decision makers' social preferences. It is these preferences that affect behavior in situations of interdependence (Kelley & Thibaut, 1978), and choices like this can provide some insight into how much (if at all) a DM cares about her own payoff in conjunction with the payoff for another person.

Table 2.1: A simple binary choice between two allocation options.

Option 1	Option 2
$85 to the DM $85 to another person	$100 to the DM $50 to another person

The "rational" solution to this choice is trivial; a payoff-maximizing DM (Homo economicus) would select Option 2 as it results in a larger individual payoff. That, by choosing Option 2 over Option 1, an extra $15 is gained at a cost of $35 to another person is inconsequential from the normative vantage point – the only pertinent consideration is the DM's individual payoff, irrespective of the payoff to the other. In this instance, the normative account clearly diverges from actual behavior. We find that about 65% of incentivized DMs from a large representative sample, in an anonymous one-shot decision context, choose the prosocial option, a finding that is consistent with other empirical results (Van Lange et al., 1997; Au & Kwong, 2004; Bogaert et al., 2008; Balliet et al., 2009).

Clearly these choice results are incongruent with the postulate of narrow self interest. Social preferences however are often more complex than those assumed under this rubric. The individual difference of how much concern a person has for others has been of interest to a wide range of researchers in different fields. This construct has been studied in parallel under a variety of different names, including: social preferences,[2] other-regarding preferences, social motives, welfare trade-off ratios, altruism, collective interest, and social value orientation (SVO). The rich lexical variety for the same concept is on one hand heartening (as there is widespread interest in the idea of human non-selfishness) but disheartening as well (as different cliques of active researchers operate largely unbeknownst to each other, all the while sharing common intellectual interests). One of our intentions here is to bring together these related but independent lines of research by examining the different ways the elemental construct of social preferences has been measured. The persistent balkanization of research on this topic can in part be attributed to measuring the same thing in different ways; we hope to bridge existing divides by suggesting common measurement methods to establish commensurability. The structure of the paper is as follows. First we briefly discuss the historical developments that resulted in a well-established theoretical framework for considering social preferences and subsequently describe this framework in detail. Then we offer a broad review of different literatures from across the social sciences, discussing different existent measurement methods of social preferences, roughly in chronological order. Strengths and weakness of each of the methods are discussed and the measures are evaluated according to a predefined set of criteria. We conclude with a brief discussion of how a reliable metric of social preferences can inform and support psychologically realistic and descriptively accurate theories of social decision making with an emphasis on the use of high-resolution measures.

2.2 Theoretical background and the emergence of the SVO concept

Early theoretical work on interdependent decision making either had primarily focused on characterizing situations of social encounters in terms of their potential to provoke cooperation or competition (e.g. Deutsch, 1949), or had focused on how peoples' attitudes and emotions shape the relationships they are involved in (e.g. Heider, 1958). It was a natural next step to analyze both the situation and intrapersonal processes when examining the behavior of interdependent decision makers, consistent with Lewin's 1936 suggestion to conceptualize behavior as a simultaneous function of both person and situation. With the emergence of game theory (e.g. von Neumann & Morgenstern, 1944; Luce & Raiffa, 1957), a formal way to describe situations of interdependent decision making, normative predictions of rational behavior in a given situation became possible. Nonetheless, this precision came at the cost of often unrealistically strong assumptions about people's preferences.

Strongly influenced by the concepts and approach of game theory, researchers have built theories of social interactions that take into account the incentive structures that characterize situations of interdependence while also - at least implicitly- assuming that people may vary in how they perceive and evaluate a given incentive structure (e.g. Homans, 1961; Thibaut & Kelley, 1959).

[2] In this paper, we will use the term social preferences or SVO for the sake of consistency when referring to this important transdisciplinary construct.

The explicit assumption that people enter situations of interdependence with individual goals and that different goals may lead to different behavior in the same interdependent situation was stated and studied by Deutsch (1960), who proposed three different motivational orientations: cooperative, individualistic, and competitive. This terminology was later adopted by Messick & McClintock (1968) in their motivational theory of choice behavior in experimental games which was stimulated by a series of studies showing that people do not strictly endeavor to maximize own payoffs when making choices in experimental games, but rather tend to take into account the other player's payoff as well (McClintock & McNeel, 1966a,c,b, 1967; Messick & Thorngate, 1967). In this theory, the three motivations mentioned by Deutsch (1960) were operationally defined as the goals to maximize joint gains (cooperative), maximize own gain (individualistic), and maximize relative gain (competitive). Messick & McClintock (1968) further showed that choice options in formal games may dominate others with respect to one or more of the three stated motivational orientations, and that it is possible to assess a person's primary motivational orientation by observing her choices in a series of what they called *decomposed games* (see also Pruitt, 1967). Basically, any unilateral choice among different allocations of resources for the self and another person is a decomposed game. If two DMs would each make such a choice and each would receive both what he allocated to the self and what the other person allocated to the other, the situation would constitute a proper (i.e. recomposed) game. The purpose of presenting people with decomposed games is that the element of direct interdependence is removed from the situation such that options chosen in these tasks express peoples' social preferences alone rather than their preferences confounded with strategic considerations.[3] Hence, Messick & McClintock's seminal work has both led to a conceptualization of social preferences (McClintock, 1972; Griesinger & Livingston, 1973) that was later termed Social Value Orientation (SVO) and the use of decomposed games with a few discrete options as a method for assessing these preferences.

The general notion that individual differences are both noteworthy and crucial for explaining behavior in situations of interdependence was also adopted in broader theoretical frameworks. For example, in the goal/expectation theory by Pruitt & Kimmel (1977) it is assumed that the choices people make in experimental games depend both on their motives and their beliefs about the behavior of their interaction partner.[4] Highlighting the importance of both of these determinants of behavior, Pruitt & Kimmel (1977, p. 385) recommended that "measures of goals and expectations should be routinely introduced into gaming studies". The assumption that social preferences affect choices in experimental games is also inherent to the theory of interdependence by Kelley & Thibaut (1978) in which people are postulated to vary in their perceptions of a given situation due to individual differences in the goals they attempt to pursue. Concretely, when a person decides which strategy to use when engaged in an interdependent situation represented by a matrix game, she is hypothesized to transform the *given matrix* into a subjective *effective matrix*, which then serves as the basis for her final choice. For instance, if a person had the goal

[3] This only holds, of course, if people are not directly paired with each other when making decisions in a decomposed game. Otherwise the situation is a proper (recomposed) game.

[4] This assumption is also fundamental to comprehensive theories of rational action which model behavior as a function of preferences, beliefs, and constraints (e.g. Gintis, 2007; Hedström, 2005).

to maximize joint payoffs - thereby expressing a cooperative motivation - she would sum up the payoffs for the self and for the other per outcome from the *given matrix*, then internally represent the *effective matrix* containing the computed sums of payoffs as outcomes, and finally choose an option based on this subjective representation of the joint payoffs. Thus, the SVO concept is implicitly embedded in Kelley & Thibaut's theory as the driver of payoff matrix transformation.

Subsequent theoretical work on SVO has focused on issues such as: linking the SVO concept with rules of fairness (McClintock & Van Avermaet, 1982); integrating it into an evolutionary perspective on behavior in situations of interdependence (McClintock, 1988); and embedding it into a broader context of social interactions in general (Van Lange et al., 2007). However, many theoretical advancements regarding SVO have been promoted and achieved on a more basic level, that is refining the concept itself (e.g. Van Lange, 1999) and developing and testing theories of its formation (Van Lange et al., 1997) or its relation with other concepts, such as beliefs, perceptions, or attitudes concerning others (see, for instance, Bogaert et al., 2008; Kelley & Stahelski, 1970; Liebrand et al., 1986; Van Lange & Liebrand, 1991).

Although there are several excellent reviews of SVO and substantial findings associated with it (Au & Kwong, 2004; Bogaert et al., 2008; McClintock & Van Avermaet, 1982), to date there is no unified, overarching *Theory of SVO* that combines the micro with the macro level and provides an extensive and coherent set of general hypotheses. Besides the fact that the investigation of relations between SVO and other variables is still in process and yields interesting discoveries even recently (e.g. with respect to non-verbal behavior, see Shelley et al., 2009, 2010; Shug et al., 2010), another reason why such an ambitious endeavor has not been undertaken so far may be that there is still ambiguity about how to measure this basic construct well.

We contend that the relationship between theory and measurement is bilateral and dynamic. Measurement methods influence how theories develop (or devolve as is the unfortunate case sometimes). As a continuous theoretical construct, SVO is conceptualized as a continuum that reflects the degree to which a DM will choose to sacrifice his or her own resources to benefit another. Furthermore, recent evidence strongly supports the continuous nature of SVO. For instance, Murphy, Ackermann, & Handgraaf (2011) showed that there is rich and reliable variance in people's concerns for others and that categorization destroys valuable information about real and persistent individual differences. Moreover, evidence shows that gradual differences in SVO are accompanied by gradual differences in the search behavior for information concerning outcomes for oneself and another person (Fiedler, Glöckner, Nicklisch, & Dickert, 2013). These findings are incompatible with a categorical conceptualization of SVO. However, the continuous SVO construct has often been diminished and distorted by the stubborn use of categorical measurement methods that yield only nominal data. This low-resolution treatment of evidence has, in our opinion, constrained the way in which SVO has been considered, discussed, and developed. It has also limited the statistical power of studies looking for the interrelations between SVO and other factors, leading to Type II errors that may have undermined the evidence for the importance of non-selfish motivations in human decision making. The intertwined history of theory about SVO and the measurement of SVO provides an interesting example of a back and forth process between measurement methods and theory, and we hope to shed light on this reciprocal process by systematically delineating the development of different measurements of this important construct over time. In addition, we agree with Bogaert et al. (2008, p. 472) that in light of the vast SVO literature, not much effort has been dedicated to a discussion and comparative evaluation of measurement methods for social preferences. The present paper is therefore not only intended to bridge between different scientific disciplines concerned with the investigation of social preferences, but also to contribute to filling a gap in the SVO literature and foster theoretical as well as methodological developments from a broad perspective.

2.3 Social Value Orientation framework

SVO provides a framework for characterizing how a decision maker values joint outcomes (McClintock, 1972; Messick & McClintock, 1968; Griesinger & Livingston, 1973; Liebrand, 1984). A graphical representation of this framework, similar to the one provided by Liebrand (1984, p. 246) is depicted in Figure 2.1 and shows different motivations associated with different joint outcomes.

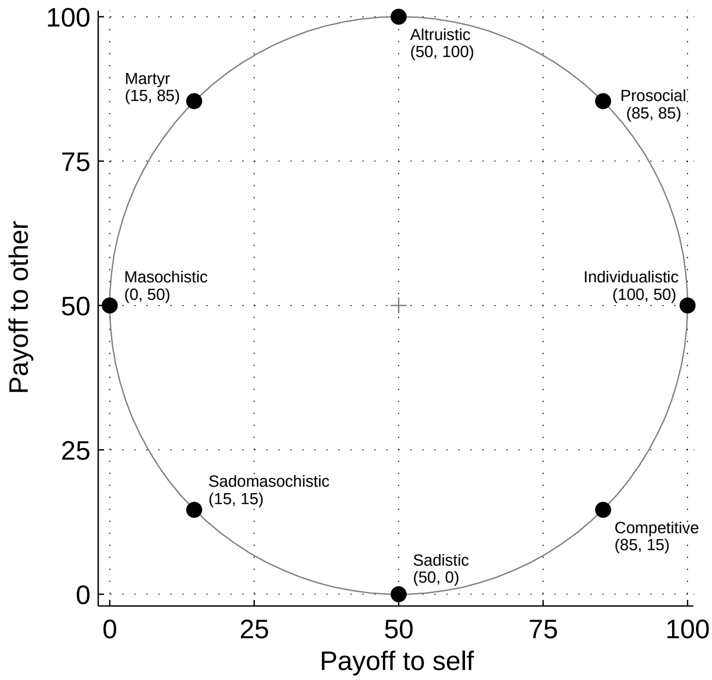


Figure 2.1: A graphical representation of the SVO framework.

A point in the Cartesian plane corresponds to a specific joint outcome. The x-axis corresponds to the value of the DM's individual payoff. The y-axis corresponds to the other person's payoff. Although there are an infinite number of possible joint outcomes, those along the ring, intersecting one of the eight cardinal directions, provide clear and unique exemplars of different joint outcomes that correspond to idealized social preferences. For example, the unique point on the ring that maximizes individual earnings is at $x = 100$ and $y = 50$ (i.e. individualistic or narrow self interest); the point on the ring that maximizes joint earnings is at $x = 85$ and $y = 85$ (i.e. prosocial). These points and their respective archetypical motivations are listed in Table 2.2. The values presented in Table 2.2 are consistent with Figure 2.1 and correspond to different idealized SVOs.[5]

Table 2.2: The archetypal Social Value Orientations.

Self	Other	Orientation	Inferred Motivation	Weight on own outcome	Weight on other's outcome
85	85	Prosocial	Maximize the joint payoff *or* minimize the difference between payoffs	1	1
100	50	Individualistic	Maximize the payoff to self	1	0
85	15	Competitive	Maximize the positive difference between self and the other's payoff	1	-1
50	0	Sadistic	Minimize the other's payoff	0	-1
15	15	Sadomasochistic	Minimize the joint payoff *or* minimize the difference between payoffs	-1	-1
0	50	Masochistic	Minimize the payoff to self	-1	0
15	85	Martyr	Maximize the negative difference between the other's and self payoff	-1	1
50	100	Altruistic	Maximize the other's payoff	0	1

One way to determine an individual's preferences across different joint distributions is to present a DM with a series of allocation decisions and ask her to select her most preferred apportionments (e.g. the binary choice presented in Table 2.1). These resource allocation decisions are sometimes referred to as *decomposed games* (Pruitt, 1967; Messick & McClintock, 1968). As noted before, the term *decomposed games* emerged from seminal work that used simple two-player binary option games (e.g. the Prisoner's Dilemma) to study choice behavior in social contexts. One problem with using a proper game to study intrinsic preferences is that a game is by definition a strategic interaction. Games require DMs to choose not only according to their own preferences, but with those preferences conditioned on their beliefs of what the other player(s) prefer and will choose, knowing that the other player(s) are likely thinking the same thing, and so on. These decisions are complex in that they draw upon personal preferences, beliefs about others' preferences, and beliefs about others' beliefs about preferences, and so on *ad infinitum*. As a method to measure preferences alone, the use of proper games is muddled and confounded by the strategic nature of the social interaction. A solution to this measurement problem is to decouple preference considerations from strategic considerations. This simplified choice task is called a

[5]One thing that is worth noting is that in the economics literature, the term *altruistic* encompasses *any* positive other regarding preferences, whereas in the psychology literature there is a distinction made between *prosocial* and *altruistic* orientations. For this paper we too will maintain the distinction between these two motivations, using *altruistic* to refer to the particular motivation of maximizing another's payoff, indifferent to one's own, and *prosocial* to refer to the preference of maximizing joint gain.

decomposed game. Although it is technically not a game, it does give an unconfounded measure of an individual's preferences for joint outcomes. For example, if a DM chooses Option 2 from the allocation choice presented in Table 2.1, we would infer that her motivation to maximize her own earnings is stronger than her motivation to maximize joint earnings; we would say she has a revealed preference consistent with an individualistic social value orientation.

Any individual choice task where a DM makes a selection among different allocations of resources to himself and others is a decomposed game. Decomposed games have been used extensively to study social preferences (e.g. Messick & McClintock, 1968; McClintock et al., 1973; Kuhlman & Marshello, 1975b,a; Van Lange et al., 1997). The dictator game[6] happens to be a decomposed game too. For clarity, we refer to non-game contexts as *allocation decisions* in order to emphasize the non-strategic nature of the relevant choice tasks.

The framework presented in Figure 2.1 and Table 2.2 provides a taxonomy (see Liebrand, 1984) of revealed social preferences but has several unusual categories that are rarely consistent with real DM's choices. Thus researchers have focused their attention on a subset of social preferences in a particular region of the joint allocation plane. According to Messick & McClintock (1968), a person can fulfill one of three different orientations. A person may be motivated to secure maximal resources for herself, indifferent to how much the other receives (i.e. maximizing own gain). Or a person may prefer to maximize the sum of the outcomes for both self and other (i.e. maximizing joint gain). Or a person may prefer to maximize the difference between her own outcome and the other's outcome (maximizing relative gain). These three motivational orientations have longstanding labels in the psychology literature as individualistic, cooperative (i.e. prosocial), and competitive, respectively (Deutsch, 1960). The most common current measure of social preferences, the 9-Item Triple Dominance Scale (see Van Lange et al., 1997), uses the same three categories. There are a variety of other approaches to the measure of social preferences which range from distinguishing simply between two categories (individualistic and prosocial) to differentiating among up to ten categories as proposed by MacCrimmon & Messick (1976).

2.4 How measures can shape theory: The case of SVO

Theory often precedes measurement. This is certainly true when something – be it an observable natural phenomenon, an inferable underlying force, or a hypothesized latent variable – is measured for the first time. The conceptualization of a measurement naturally requires the conceptualization of the object of measurement. However, once a measurement method exists and is employed, the data it produces has an impact back on how the object of measurement is theorized about. There can be significant interdependencies between theory and measurement since refinements of theories are often driven by data, data depend in part on the measures employed to obtain them, and the employed measures depend on the theories originally proposed. In this vein, the relation between theory and measurement is not exclusively unidirectional or one-way dependent

[6] This is an unfortunate name as it is technically not a game either, as only one DM can influence the joint payoffs.

(see, for instance, Gigerenzer, 1991; Gigerenzer & Sturm, 2007). Furthermore, the instruments we use to assess and process data can have an influence on our thinking in a broader context (see Sturm & Ash, 2005). As John M. Culkin noted, "we shape our tools and thereafter they shape us." We want to make the argument here that the bidirectional relation between theory and measurement has shaped the conceptualization of the SVO construct in ways that impaired progress in measurement quality. In the following section, we want to elaborate on the remarkable anomaly that a valuable continuous construct has commonly been measured on the nominal scale level for decades.

As discussed earlier in this paper, the emergence of the SVO construct was triggered by the observation that most people often do not attempt to maximize the experimenter-defined payoff when interacting with others in strategic situations (McClintock & McNeel, 1966a,c,b, 1967; Messick & Thorngate, 1967). The pattern of results obtained in these studies led Messick & McClintock (1968) to the elemental assumption of three distinct goals that guide behavior in experimental games: maximize own gain (individualistic), maximize relative gain (competitive),[7] and maximize joint gain (cooperative). Hence, an early notion of SVO was categorical. However, Messick & McClintock also tested whether a utility model[8] (see Messick & Thorngate, 1967) that is not restricted to categorical assumptions is useful for describing observed choice patterns. Since they found that a particular utility model poorly fit their data, they abandoned it, and proposed a stochastic choice model instead. This stochastic model was based on the assumption that people are in one of the three suggested motivational states (or indifferent) at a particular point in time according to an individual probability distribution, and that their choices in a particular experimental game would depend on the state adopted at the moment of choice. Therewith, a categorical conceptualization of social preferences won, and in hindsight this had a substantially negative impact on future SVO measurement.

A continuous conceptualization could have emerged however. The thinking behind the use of decomposed games for assessing SVO as originated from Messick & McClintock's seminal work has typically been the following: Discrete options in a decomposed game may dominate each other with respect to certain predefined motivational goals. If an option is chosen by a decision maker, and this option must dominate other available options with respect to a particular motivation, the decision maker's motivation and preferences are revealed. The assumption of three different particular motivations was data driven but still arbitrary, and if the thinking would have been more in line with a utility maximization approach, a continuous conceptualization of SVO could likely have emerged. The three motivational orientations can be represented as three different parameterizations of the same utility function $U(x, y) = x + ay$, with $a = -1$ representing relative gain maximization, $a = 0$ representing own gain maximization, and $a = 1$

[7] It is interesting, in hindsight, to see that the emergence of the SVO construct was mostly driven by the observation that some people appear to maximize relative gain, while this particular preference is the least commonly observed compared to the other two prototypical motives in the vast majority of more recent studies.

[8] The model has the following general form: $U(x, y) = f(x) + g(x - y)$. According to Messick & McClintock (1968, p. 15), "joint gain is ignored in this model [...] also as a result of the data previously reviewed which indicate that relative-gain maximization is a more important choice determinant than joint-gain maximization."

representing joint gain maximization. If such a representation had been salient to Messick & McClintock, it would have been obvious, perhaps, to assume a continuum for $a = [-1, 1]$ and therefore conceptualize SVO as a continuous construct, rather than assuming a rigid categorical typology $a = \{-1, 0, 1\}$. Despite the fact that a continuous conceptualization of social preferences already existed (see Sawyer, 1966) in the late 1960's, the more influential work of Messick & McClintock forged how SVO has been theorized and commonly measured thereafter. Currently the most commonly used SVO measures (the Triple-Dominance Measure and the Ring Measure) produce only categorical output consistent with how Messick and McClintock discussed SVO as informed by decomposed binary games.

We suggest that the persistence of the categorical SVO concept has been promoted by measures commonly employed to assess the SVO construct, and that in this way measurement methods have shaped theory. In order to elaborate on this claim, we focus on the development of the SVO concept and measures thereof following Messick & McClintock (1968). Two traditions of SVO conceptualization have evolved since then. One tradition followed the categorical approach described above, and the other tradition followed a utility model approach. In the utility model tradition, SVO was naturally conceptualized as a continuous, albeit not necessarily unidimensional, construct. The focus of research within this tradition was on postulating and testing different utility functions as representations of social preferences (see Grzelak et al., 1977; Radzicki, 1976; Wyer, 1969). Since parameterization is essential for testing utility models, and parameters are usually not restricted to take on only a very limited number of values (such as only three), a continuous theory of SVO is inherent to this approach. Building on both the work of Messick & McClintock (1968) and Wyer (1969), Griesinger & Livingston (1973) showed how the two conceptualizations relate to each other by employing a geometric approach to represent motivational orientations as vectors in the Cartesian plane with the x-axis corresponding to payoffs to the self and the y-axis corresponding to payoffs to the other. This was a cornerstone in the history of SVO research, since the geometric representation supported visualization of how different motives corresponded to different combinations of weights in a simple utility function. A decision maker's choices could be modeled as if they make tradeoffs between the payoff to the self and the payoff to the other given the utility function $U(x, y) = ax + by$ (see Table 2.2). Hence, Griesinger & Livingston's framework helped clarifying that SVO is a continuous construct. However, this framework paradoxically promoted the categorical conceptualization of SVO thereafter. We claim this because work following and building on Griesinger & Livingston's framework focused on the motivational categories, rather than the underlying continuous motivations (see, for instance, Maki et al., 1979; MacCrimmon & Messick, 1976). Furthermore, the SVO measure that was constructed on the basis of the geometric framework and has become the second most commonly used instrument for assessing SVO, the Ring Measure (Liebrand, 1984; Liebrand & McClintock, 1988), has been proposed and used almost exclusively for categorizing subjects rather than eliciting continuous information.

We see a potential chain of reasoning responsible for why the geometric framework lead to a preference for simple categorical thinking. In the ring framework, the continuous SVO construct is two-dimensional, one dimension referring to the weight a person attaches to her own outcomes, and the other referring to the weight a person attaches to other's outcome. This is inconvenient since multidimensionality hinders the employment of simple statistical tests for evaluating individual differences within the construct and associations or interactions with other variables. However, the two dimensions can be translated into one in terms of an angle. This does not solve the problem since the interpretation of the angle is still not unidimensional. A statement, such as "the higher the angle, the higher the concern for others" does not hold when the full ring is considered. Augmentations in angular degrees beyond both plus and minus 90° imply decreasing concerns for the other, while the opposite is true for angles within this range. Hence, the angle has to be translated back into a corresponding particular motivational category in order to be readily interpretable. This frame of thinking, we speculate, is the reason why the categorical conceptualization finally predominated in SVO research. However, the problem of two-dimensionality could have been solved by simply disregarding one of the two dimensions, namely the dimension corresponding to the weight attached to the own outcome, by assuming that this weight is just equal to one. This assumption appears justifiable since we do not know of any evidence supporting the hypothesis that people ignore (pure altruism or pure aggression) their own payoffs or depreciate (martyrdom, masochism, or sadomasochism) own outcomes either. Evidence suggests that the utility function of own outcomes *per se* is monotonic increasing (Messick & Sentis, 1985), that is – everything else being equal – more is strictly preferred to less. Under this assumption, SVO becomes a unidimensional continuous construct defined as the weight a person attaches to outcomes of others in relation to the own, represented by parameter a in the utility function $U(x, y) = x + ay$. This continuous, unidimensional conceptualization excludes the possibility of particular pathological motives, yet allows for aggression and altruism when letting a approach positive or negative infinity, respectively, and includes competition ($a = -1$), individualism ($a = 0$), and cooperation ($a = 1$) as special archetypical cases.

Although such a continuous conceptualization is at least as old as the categorical one and was once applied for devising a measure of SVO (see Sawyer, 1966), it has apparently been abandoned for decades. Consequently, until the recent advent of a novel, continuous measure (Murphy et al., 2011), for about thirty years SVO has been assessed almost exclusively on the nominal scale. Concurrently, researchers have seemed to be aware that the construct is continuous in principle, but have chosen to apply categorical measures thereof, and then treated the categories as if they constituted the construct as a whole, rather than salient but arbitrary instantiations of an underlying continuum. Typically, in SVO research papers, the construct is introduced as "stable preferences for certain patterns of outcomes for oneself and others" and it is mentioned that "a variety of different social value orientations can be distinguished from a theoretical point of view," but that a "three-category typology" would be applied in the present work.[9] It is usually hard to justify why a continuous construct is categorized or even dichotomized in light of the obvious disadvantages such gross downsampling results in (see Cohen, 1983; Irwin & McClelland, 2003). For example, it would certainly appear as an odd idea – and for very good

[9] The phrases in quotes are taken from Van Lange et al. (1997) which is a well known and widely cited work regarding SVO. The phrases serve to highlight the divergence between measurement and theory that is common in SVO research.

reasons – to measure intelligence with an instrument that only produced a rough categorization of people into the three groups “bright,” “mediocre,” and “dull.” However, the same practice is carried out commonly when considering SVO. It is precisely this circumstance that lead us to claim that in the case of SVO, measurement has fedback and shaped theory. Although the construct is commonly acknowledged as continuous, it has been measured at the nominal level and therefore internally theorized as being categorical, simply because the measures commonly used to assess the construct produced categorical output.

For a surprisingly long time it had been a inauspicious convention in psychology to dichotomize continuous variables for analyses (see MacCallum et al., 2002). The most often cited reasons for such a procedure include convenience and simplicity of ANOVA methods contrasted with multiple regression. However, the adverse effects of discretizing continuous variables have been demonstrated clearly and repeatedly (e.g. Cohen, 1983; Fitzsimons, 2008; Irwin & McClelland, 2003; MacCallum et al., 2002; Royston et al., 2006) and it has also been shown that there are trivial benefits but substantial costs associated with such a practice. Perhaps surprisingly, although the post-hoc degradation of continuous data had been quite common, it apparently has been very uncommon to measure a construct at a lower level of measurement than its theory permits. In fact, it seems to have been standard in psychology to measure constructs, especially personality variables, whenever possible on continuous scales. Apparently this standard has been so strong that hypothesizing about class variables in personality research has required extensive argumentation (for an example, see Gangestad & Snyder, 1985). Interestingly, methodology has been cited as one reason for the ubiquity of a continuous conceptualization of personality dimensions: “[...] Methods for test construction and evaluation generally assume underlying latent continua, and personality researchers adopt, without question, the underlying assumption of dimensionality” (Gangestad & Snyder, 1985, p. 319). In the case of SVO, we argue, methods too have shaped thinking in the same manner as in personality research, but in the opposite direction. The most commonly used SVO measures (the Triple-Dominance Measure and the Ring Measure) produce categorical output, and SVO researchers seem to have adopted, perhaps without deep reflection, a categorical conceptualization. This has resulted in the curious situation that, although in general continuous conceptualizations of individual difference variables have predominated in psychology, SVO has commonly been assessed and thought of as a nominal variable, even though it had been shown to be a continuous construct in principle shortly after its advent by Griesinger & Livingston (1973). To our knowledge, a curious situation of this type is unique in psychology, but it may serve as an important reminder of the need to deliberate on the coherence between theory and measurements methodology from time to time.

The following review on SVO measurement methods is focused on evaluating the methods’ strengths and weaknesses, and in addition is intended to guide the reader through the history of SVO measurement while highlighting how the different measures relate to different theoretical conceptualizations of SVO. The measures are evaluated on the basis of a set of predefined criteria, one of which is output resolution (e.g. scale of measurement). This feature is typically not central in evaluating the quality of a measure. However, in the case of SVO measurement, it is an important issue for the reasons elaborated above, and thus is given considerable weight in the following sections.

2.5 Existing measurements of social preferences

In this section, existing approaches to the measurement of social preferences are described and discussed in approximate chronological order of publication. This provides a general overview of the history of social preference measurements and offers insight into how methods have changed and developed over time. In discussing different approaches to social preference measurement, the current paper focuses on methods that assess people's preferences for certain allocations of resources. These preferences are revealed by eliciting people's judgments or choice behavior when they are presented with options containing different distributions of outcomes for the self and for some other person. Questionnaires or Likert-type scale measures regarding verbally expressed altruistic or prosocial attitudes (e.g. Rushton et al., 1981; Crandall, 1975) are excluded from the present paper. The reason for this exclusion is twofold. First, attitudinal measures are rarely used in SVO research. Second, and more importantly, we think that having people making decisions with real consequences is the right approach to measuring social preferences. For example, we consider the abandonment of real payoffs for the benefit (or detriment) of another person as stronger evidence for social preferences compared to the mere indication of an intention to do so in a hypothetical situation, or the expression of degree of agreement with a qualitative statement. We think that the measurement of real behavior is superior to the measurement of intentions or attitudes given that the object of interest itself is behavior rather than inner processes (see also Baumeister et al., 2007). Hence we will only consider measures that at least allow for having subjects make real decisions with real consequences.

In order to ensure a fair procedure and allow for comparative analyses of the different methods, the SVO measures are evaluated on the basis of a set of five predefined criteria. The first two criteria are standard psychometric ones, validity and reliability. However, we will restrict ourselves to only reporting on predictive validity, convergent validity among SVO measures, and test-retest reliability. The third criterion is output resolution for the reasons explicated earlier in this paper. The fourth criterion is efficiency in terms of the expenditure of time and effort associated with measurement completion and output evaluation. This criterion is included in order to give the reader who seeks the optimal measure for a particular research design some information about pragmatic aspects. The fifth and final criterion is particular advantages, that is, the existence of useful features of a measure that are not commonly shared by other measures. At the end of this review section, the reader is provided with a tabulation of the measures' scores per criterion. Since these scores are based on judgment, they can not be completely objective, and we want to emphasize that we do not purport that they are. Also, it is not the purpose of this review to choose one method as best, but to allow for a comparative evaluation to help selecting the method that best suits a particular research design. Furthermore, the evaluation of the measures is not the only target we aim for. We consider the provision of a historical review on SVO measurement and a statement of how theory and measurement interact as equally important as the evaluation of methods. Therefore, all the methods' individual discussions include, but are not restricted to the criterion-based evaluation.

2.5.1 The Altruism Scale

Early efforts to quantify social preferences can be traced to sociology. Sawyer (1966) devised a method for assessing the degree of concern a DM has for the outcomes for himself and others. He called this method the *Altruism Scale*. However it should be noted that Sawyer's method can assess a range of different orientations including prosocial, individualistic, and competitive motivations, and it would be more accurate to call it as *Social Preference Scale*. Within Sawyer's theoretical framework, the subjective attractiveness of a joint outcome was conceptualized as the linear combination $P_S + wP_O$ where P_S is the payoff for the self and P_O is the payoff for another person. The coefficient w represents how much weight a DM gives to the outcome for the other person, relative to his own outcome (there is an implicit coefficient of 1 in front of the P_S term). If a person is individualistic and therefore interested in only his own welfare, the coefficient w would be zero. If a person is prosocial and cares about both his own, and the other person's welfare, w would be greater than zero. Conversely, if a person is competitive and tries to maximize the difference between the own and the other's payoff, w will be less than zero. Hence, the theoretical conceptualization of social preferences underlying the Altruism Scale is continuous.

The Altruism Scale described

Sawyer's method uses a conjoint measurement technique to estimate an individual's weighting of outcomes for others (w). An index of altruism is computed based on the desirability rankings of own/other outcome combinations. Specifically, participants (college students in this case) were asked to imagine that they would take a seminar with only one other fellow student and that each would receive a grade of A, B or C at the end of the seminar. Participants were then asked to rank their preferences for the allocations of these grade combinations. After all nine of the rankings are made by a DM, the altruism index a can be calculated as follows:

$$a = \frac{\sum \text{ranks in row C} - \sum \text{ranks in row A}}{\sum \text{ranks in column C} - \sum \text{ranks in column A}} \tag{2.1}$$

The index a is a manifest variable and serves as a proxy for the latent variable w. The numerator in computing a corresponds to how much a person cares about the outcome for the other person. If, for example, a person cares about the other's welfare, the DM will assign high ranks to the options where the other person receives A grades and low ranks to the outcomes where the other person receives C grades. In this case, the numerator is positive, indicating positive altruism. Conversely, a negative result indicates competitiveness. If the result is zero, it implies that the person is indifferent to the other student's grade.[10]

[10] Sawyer used a second method to directly assess a by asking the DM to choose 1 out of 21 scale values, which corresponded to values of a ranging from -1 to $+1$ in increments of 0.1. The scale is anchored by statements at the

According to Sawyer, w ranges continuously from -1 to $+1$, i.e. from perfectly competitive ($w = -1$) to perfectly prosocial ($w = +1$), with narrow self interest ($w = 0$) at the midpoint. It is worth noting that the coefficient a can take on values outside of this range given atypical motivations (e.g. masochistic). It is also worth noting that for Sawyer's Altruism scale, a is undefined (perfect altruism implies $w \to \infty$) if a DM provides a purely altruistic ranking (ranks that are consistent with maximizing the grade of the other student and indifferent to the DM's own grade).

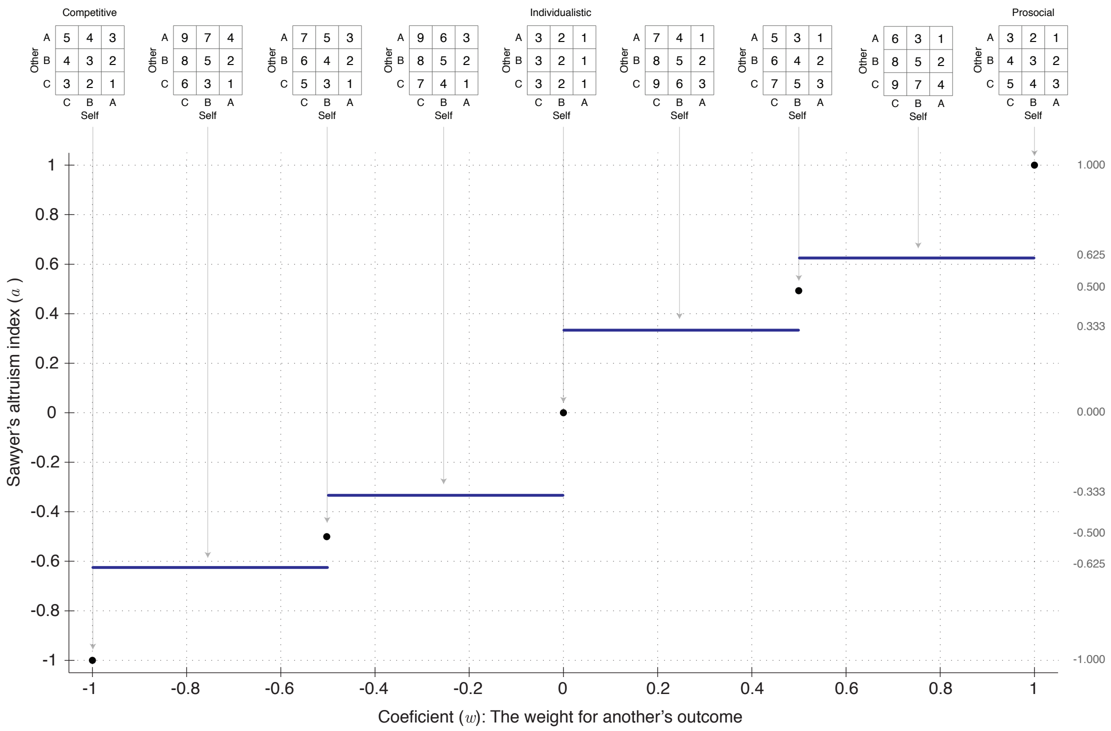


Figure 2.2: Sawyer's Altruism Scale, preference rankings, and the a index. Examples of different preference rankings (the 3×3 matrixes at the top of the figure) are displayed here. Further, the relationship of these rankings to the underlying utility weight (w, shown on the x-axis) for another's outcome, and Sawyer's altruism index a are shown. The ranking of 1 indicates that this allocation of grades is the DM's most preferred joint outcome. As can be seen, ties in ranking outcomes are allowed. Particular rankings are consistent with underlying utility weights, and further each ranking matrix corresponds to an a index. Although w is continuous, the resulting altruism index a is a step function that can take on only one of nine values given w between -1 and 1 inclusive. Some rankings correspond to a single point of w, whereas other rankings are consistent with a range of w values.

values -1, -0.5, 0, $+0.5$ and $+1$. For example, the statement reflecting an a value of $+1$ indicates agreement with the following statement: "I am equally interested in how good his grade is and in how good my grade is," whereas the statement reflecting an a value of 0 indicates agreement with: "I am only interested in how good my grade is; how good or poor his grade is makes no difference to me."

Discussion of the Altruism Scale

The Altruism scale was an early innovation but as a means to measure social preferences it has limitations. First, the metric space of academic grades is not straightforward and further is obviously not amenable to incentive compatible research. This particular choice context may force DMs to take a zero-sum mentality if they are accustomed to curved grading systems or are concerned with their overall class ranking. However, the method per se does not require the employment of school grades as stimuli. Instead, any set of valuable goods containing three elements with transitive and strict preference ordering $A \succ B \succ C$ could be used for eliciting preference indications in principle. Hence, the method could be used for measuring social preferences of individuals who are not experienced with alphanumeric representations, such as children, for instance. Nevertheless, there are other methods, such as utility measurement in general, or the Social Behavior Scale discussed later in this paper, which share this feature.

Second, a procedure for rank ordering preferences that presents participants with all stimuli at the same time runs the risk of yielding unreliable data because people usually are not very skilled at reliably ranking multiple items simultaneously (Saaty, 1980). Hence, as proposed by Sawyer himself, it would probably be beneficial to let participants make sequential pairwise comparisons in order to reduce complexity of the judgment task and therefore yield more accurate rank orders. How one would elicit global rankings based on sequential pairwise rankings is not an issue Sawyer addresses. However, since a rank ordering of nine outcomes can be produced relatively quickly and the computation of output variable a is trivial, the method can be termed efficient with respect to time and effort.

Third, the altruism scale cannot differentiate between the prosocial motivations of joint gain maximization and inequality aversion. It also yields an undefined a index for a DM with a purely altruistic motivational orientation.

Fourth, the process of reducing a set of rank orderings into a single index may be problematic. Each a value at one of the three anchors (-1, 0, +1) has a clear interpretation, whereas values in between are not readily interpretable. Further, the index a is an ordinal variable at best (Stevens, 1946, 1950) and is limited to 9 particular values (see Figure 2.2). Moreover, the mapping from underling utility w to the index a has a “many-to-one” structure which necessarily results in the loss of information.

With respect to psychometric properties of his measure, Sawyer reported weak validity and reliability. For example, the correlation between values obtained by the direct scale estimation measure and the values obtained by the ranking method was only $r = 0.32$, hence challenging the measure’s convergent validity. Sawyer reasoned that the discrepancy between the two measures was probably due to differences in task complexity and to multidimensionality in the rankings (i.e., the direct estimation measure promotes unidimensional judgments whereas the conjoint method allows for more complicated preferences to manifest themselves).

2.5.2 The 9-Item Triple-Dominance Measure

Decomposed games have their roots in two-option two-player games (e.g. the Prisoner's dilemma) that have been the "fruit flies" of social decision research. In psychology, Messick & McClintock (1968) and Pruitt (1967) devised what have been termed decomposed games by deconstructing two-player binary social dilemmas into individual decision problems. The reason for this simplification is to disentangle intrinsic motivations for joint outcomes from strategic concerns. Allocation choices can be constructed that differentiate between archetypal motivations. Messick & McClintock (1968) focused their attention on the three common social preferences, namely prosocial, individualistic, and competitive, ignoring other less common motivations.

For example, the allocation decision presented in Table 2.1 is designed to differentiate between prosocial and individualistic motivations.[11] A prosocial person would choose Option A and an individualist would choose Option B. However, a competitive type would also select Option B, as it has a greater relative difference between the payoffs. So with this particular allocation decision it is not possible to differentiate between individualists and competitors as both types would choose the same option.

There are two general approaches which solve this discrimination problem and can distinguish between the three most common social preferences. First, a researcher can examine the complete set of choices made in a series of two-option, double dominance allocation decisions. The set of choices that pits each of the common social preferences against each of the other common social preference types is necessarily exhaustive and can isolate a DM's primary social motivation. This method would also identify an individual's least preferred social outcome, as well as yield a ranking of preference over the joint options. A second method to differentiate between the three most common social preferences uses a single allocation decision that has three particular options (as proposed by McClintock et al., 1973) such that each option dominates *both* the other two allocations with respect to one particular motivational orientation. These types of items have the property of triple dominance, as they can differentiate between three social orientations. Triple-dominance items were adopted by Kuhlman & Marshello (1975b) who also used other decomposed game classes (double-dominance and single-dominance) for assessing social preferences. Building on this work, the 9-Item Triple-Dominance Measure of SVO (see Van Lange et al., 1997) has evolved and has become a widely-used measurement method for social preferences in social psychology (e.g. applied by, Joireman et al., 2004; Utz, 2004; Utz et al., 2004; van Dijk et al., 2004; Stouten et al., 2005; de Kwaadsteniet et al., 2006; Van Lange et al., 2007; Declerck & Bogaert, 2008; van Prooijen et al., 2008; van den Bos et al., 2009; Haruno & Frith, 2010), in part due to its straight forward structure and ease of use (Van Lange et al., 2007).

[11] In Messick & McClintock's terminology, this type of item is referred to as a *double-dominance* item as either of the two options can dominate the other one with respect to a particular motivation.

The 9-Item Triple-Dominance Scale described

The Triple-Dominance SVO items can be seen in Table 2.3. For each item there is one allocation option that is prosocial, one that is individualistic, and one that is competitive.

Table 2.3: Triple-Dominance items. Note that these values have been standardized to range between 0 and 100 in order to facilitate comparison with the other measures presented in this paper. The original items ranged between 80 and 580 and were presented in the units of points (examples of the original form are shown in Van Lange et al., 1997, 2007).

	Prosocial		Individualistic		Competitive	
Item	Self	Other	Self	Other	Self	Other
1	80	80	92	40	80	0
2	84	84	96	44	84	4
3	88	88	100	48	88	4
4	82	82	96	44	84	4
5	84	84	96	44	82	2
6	84	84	98	44	84	4
7	86	86	96	44	86	6
8	84	84	94	44	84	4
9	82	82	92	44	80	4

The scoring rule for this scale is to count the number of individualistic, prosocial, and competitive options a DM selects. If a DM chooses six or more options from a particular category, then the DM is designated as being that type. If a DM does not choose at least six options from one category, then she is not categorized (e.g. McClintock & Allison, 1989; Platow et al., 1990; Van Lange & Kuhlman, 1994). In other variants of the Triple-Dominance measure, only six items are used and participants are classified when at least five of the six choices are consistent with one of the three social value orientations (e.g. Van Lange, 1999). However, the method of counting choices made in several decomposed games and classifying participants into respective SVO categories according to their choice pattern with respect to a particular consistency criterion is exemplary for the general evaluation procedure used in decomposed game measures. This holds for the 9-Item Triple-Dominance Measure as well as for other variants using mixed dominance classes (e.g. Kuhlman & Marshello, 1975b,a).

Discussion of the 9-Item Triple-Dominance Scale

Although the 9-Item Triple-Dominance Measure is the most commonly used measure of SVO to date, it has some shortcomings. First, it can assign individuals to only one of three categories: prosocial, individualistic, or competitive, and provides no information beyond this basic categorization. This result is only at the nominal scale level (Stevens, 1946, 1950), the lowest level of measurement. However social motivations are conceptualized as a continuous construct (Griesinger & Livingston, 1973; Wyer, 1969) and one practical result of forced categorization is

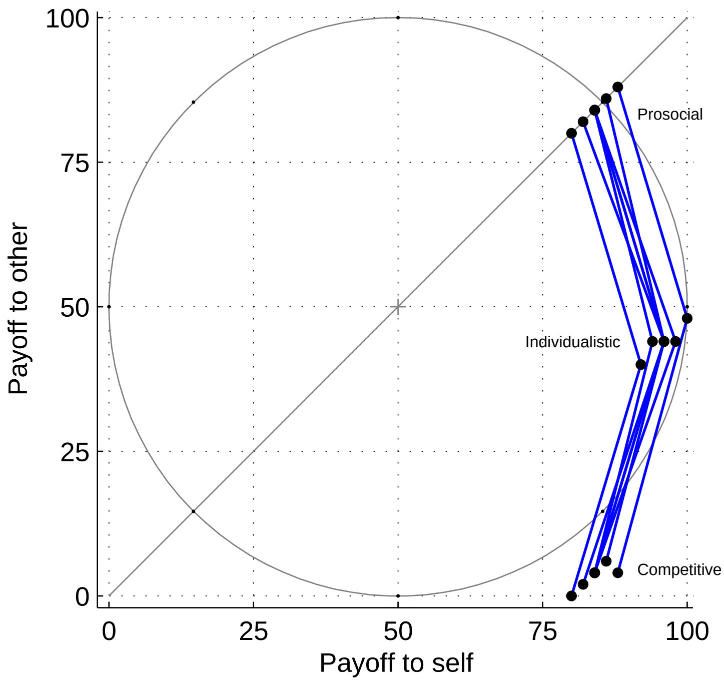


Figure 2.3: The 9-Item Triple-Dominance Scale. Each item is represented in the self/other allocation plane as three points connected by a line. Notice the high degree of similarity among the items. It may be more accurate to say this is a scale with one item repeated nine times.

low statistical power (Cohen, 1983). Attempts have been made to extract continuous information from a set of choices in the Triple-Dominance Measure. For example, the number of cooperative choices has been used as an SVO score (e.g. Hilbig & Zettler, 2009), or the sum of payoffs allocated to the other or the self (see Sheldon, 1999).[12] These scoring methods are similar, since both are based on aggregating prosocial choices, with one method simply being to count the number of prosocial choices, and the other method summing the corresponding payoffs. However, we see several problems with procedures of this kind as they confuse the reliability of a preference with the magnitude of a preference. Although these two things may be related, they are not the same. Hence, an SVO score resulting from a counting procedure is confounded between intensity and reliability and thus its meaning is obfuscated. Consider, for instance, a comparison between a person *A* who chose the cooperative option eight times and a person *B* who chose it nine times in the Triple-Dominance Measure. How much weight do *A* and *B* attach to the outcome of others in relation to the own? This question can not be answered with these choice data. Further we can not determine that *B*'s weight is greater than *A*'s. The 9-Item Triple-Dominance Measure is designed to detect whether a person's choice pattern is more consistent with a weight of 1, 0, or -1, and to categorize a person accordingly given a particular consistency criterion. A more fine-grained

[12]To be precise, Sheldon (1999) used the Kuhlman-Teta Measure, which can be seen as a precursor of the 9-Item Triple-Dominance Measure.

estimation of a person's weight is not possible with this method. SVO scores resulting from a counting procedure thus purport informational richness that is not really there. Moreover, we find evidence that neither of the two counting procedures improves the Triple-Dominance Measure's test-retest reliability, or convergent validity with other SVO measures.[13] Hence, for both conceptual and empirical reasons, we are skeptical that more useful continuous information can be extracted from the Triple-Dominance measure.

Second, the Triple-Dominance Measure cannot discriminate between joint payoff maximization and inequality aversion. All of the prosocial options in this measure happen to both maximize joint outcomes and also minimize inequality. Although these preferences may be related, they are not the same. There is evidence that persons classified as prosocial are concerned with both the maximization of joint gain and equality in outcomes (Van Lange, 1999). However there is conflicting evidence (Eek & Gärling, 2006) that prosocial DMs prefer equal outcomes over maximizing joint outcomes. It is not possible to clarify this issue with the Triple-Dominance Measure.

Third, the 9-Item Triple-Dominance Measure can only establish a DM's first preference, not her lesser preferences. Take for example an individual who has a rank order of preferences as individualistic, prosocial, and competitive. Contrast this individual to someone who ranked preferences of individualistic, competitive, and then prosocial. These individuals may approach the world very differently than each other. Furthermore, knowing what an individual's least preferred allocation is would be informative, as an avoidant personality or *prevention focus* (Higgins, 1997) can serve as a motivational foundation. People who are strongly motivated to avoid their least preferred option, rather than focusing on their most favored option, would make different choices depending on the full ranking of their preferences. It is thus important to know not just a DM's most preferred outcome, but also the entire rank ordering of her social preferences.

Fourth, three-option choice sets are more complicated than binary choices. The simplest choice is between two options and this setting only requires the DM to make one comparison in the process of making a decision. Increasing to three options requires the DM to make three comparisons. The inclusion of one particularly unattractive option has been shown to have an effect on revealed preferences (Huber & Puto, 1983; Simonson, 1989) in surprising ways.

Fifth, it seems to have become quite a common practice that the two categories of individualistic and competitive orientations are merged to form one group which is then compared to DMs in the prosocial category (see e.g. Joireman et al., 2004; Utz, 2004; Stouten et al., 2005; de Kwaadsteniet et al., 2006; Cornelissen et al., 2007; van Prooijen et al., 2008; van den Bos et al., 2009). Obviously

[13] The data from Murphy et al. (2011) allow for a comparison between the different scoring procedures. In order to facilitate a comparison, the categorical data from the normal scoring procedure are treated as ordinal, or dichotomous (combining competitors and individualists). The test-retest reliability of the Triple-Dominance measure is $r_{spearman} = 0.801$, or $r_{phi} = 0.798$, respectively. However, the Pearson correlation between the number of cooperative choices at time one and time two is only $r = 0.692$, and between the sum of payoffs allocated to the other at time one and time two it is $r = 0.621$.

such a procrustean approach sacrifices valuable information. This collapsing across categories is not an intrinsic limitation of the Triple-Dominance SVO measure, but rather a regrettable convention that has evolved with it when the number of participants per category is considered to be too low (see e.g. Van Lange & Liebrand, 1991) to support certain types of statistical analyses.

Sixth, offering the nearly identical choice nine times may induce participants to vary their responses in unexpected ways. In some cases, this variation may be a reflection of their honest preferences. For example we had one participant from pretesting explain during debriefing that he answered about half of the items individualistically and the other half prosocially. His goal, he explained, was to be nice, but not *too nice*. This participant treated the scale holistically and made a set of answers that, when considered in total, were sensible. But this sensible set of responses would have resulted in an uncategorizable result using the standard scoring rule. In other cases, participants may become bored or suspicious of answering the same item repeatedly and thus vary their answers. Ironically, this high degree of redundancy in the Triple-Dominance Measure may undermine its ability to classify participants. For example, in the study conducted by Kuhlman & Marshello (1975b), the percentage of unclassifiable participants was 25% (42 out of 167), Kuhlman et al. (1992) report 29.3% (41 out of 140) unclassifiable participants, and Sheldon (1999), who made use of the Kuhlman and Teta measure, even applied an alternative, and problematic, scoring method after having lost 27% (25 out of 90) of the participants for analysis because they were not classifiable.

The greatest advantage of the Triple-Dominance Measure is probably its high efficiency. The measure focuses on only the three most commonly observed archetypal SVOs, and can be completed by a subject in less than five minutes. Furthermore, data evaluation is straightforward and not time-consuming. Due to these features, the method can be regarded as a quick and simple way to assess SVO.

With respect to the psychometric properties of the measure, results indicate medium quality. The measure shows satisfying test-retest reliability. Usually, about 70-75 per cent of subjects are categorized into the same SVO category at two different points in time (see, for instance, Murphy et al., 2011; Van Lange & Semin-Goossens, 1998). In terms of convergent validity with other SVO measures, results are scarce and less consistent. While Murphy et al. (2011) report satisfying convergent validity (in terms of categorical agreement with the Ring Measure [67%] and the Slider Measure [74%]), Parks (1994) found no association at all between a variant of the Triple-Dominance Measure and the Regression & Clustering approach (discussed later in this paper) by Knight & Dubro (1984). Data on the predictive validity of the 9-Item Triple Dominance Measure are plenty and usually show small to medium effect sizes in a variety of domains (see, for instance, De Cremer & Van Lange, 2001; Van Lange et al., 2007; Van Vugt et al., 1996), although counterexamples of lacking predictive validity exist as well (see Joireman et al., 2004; Parks, 1994). In sum, we regard the psychometric properties of the Triple-Dominance measure as sufficiently strong but with room for improvement (see also Au & Kwong, 2004; Bogaert et al., 2008).

Rank correlation technique with decomposed games

Another measurement technique that relies on decomposed games for assessing SVO was introduced by Iedema & Poppe (1994a,b, 1995). Iedema & Poppe presented DMs with pairwise comparisons of eight (or nine) different own/other payoff allocations, resulting in a total of 28 (or 36, respectively) allocation decisions. Then, ranks were assigned to the payoff allocations for each participant according to how often each of these alternatives had been selected. Prior to this assessment, Iedema & Poppe had compiled ideal rank orders of the alternatives with respect to six different idealized social orientations (individualism, altruism, equality, cooperation, competition, and maximin). The assessed rank orders were then correlated (Spearman's rank order correlation) with each of the six ideal rank orders for each participant, yielding six correlation coefficients per participant, each of which was indicative of the relation between the participant's rank order and the corresponding ideal rank orders of the six social orientations. These coefficients were then transformed into Fisher Z-scores ranging from -3 to $+3$ and participants were classified to a particular SVO category matching their highest Z-score, provided that this score was greater than a predetermined threshold. In one instance (Iedema & Poppe, 1994a,b) a threshold of 0.55 (which corresponds to a correlation coefficient of 0.50) was used. In another instance (Iedema & Poppe, 1995), a threshold of 0.881 (corresponding to a correlation of 0.707, reflecting the threshold of 50% explained variance) was required.

Iedema & Poppe's rank correlation method allows for the detection of particular motives that originally were not part of the SVO concept as proposed by Griesinger & Livingston (1973), but were introduced later by MacCrimmon & Messick (1976), namely inequality aversion (or egalitarianism) and the maximin orientation. However, this is not a unique advantage of this method since other measures, such as Schulz & May's Sphere Measure (discussed later in this paper), can assess these motives as well. One advantage of the measure is that a person's rank order of preferences can be estimated, since Z-scores are obtained for all of the six predefined social motives and can thus be compared to each other. Nevertheless, the method is not very efficient. It employs more than three times as many items as the Triple-Dominance Measure for assessing only twice as many motivational orientations, yet the output is still categorical. Moreover, the data evaluation procedure for computing the categorical output is fairly complicated, which further diminishes overall efficiency. Also, to our knowledge, there are no data available on the method's psychometric properties, which precludes a direct comparison with other methods in terms of measurement quality.

2.5.3 Utility measurement

Utility measurement in general refers to the systematic estimation and mapping of how subjectively valuable payoffs, goods or outcomes are to a DM. Utility is an abstract construct that is inferred from the revealed preferences of DMs as they make choices among available alternatives. These alternatives can include "bundled outcomes," sets of discrete goods that are considered and evaluated as a whole set. Obviously, the utility of these bundled outcomes results from the constituent parts therein. However, the way a DM integrates information about the items and makes tradeoffs between them may not be so obvious. Early studies of these kinds of choices by Thurstone (1931) involved participants making paired comparisons between sets of goods (e.g. [2 hats, 4 pairs of shoes] vs. [3 hats, X pairs of shoes]) where X was varied systematically by the experimenter. This approach yielded an estimated value of X where a DM was indifferent between the sets. Given fungibility and an indifference point, the relative contribution of the discrete items to the bundle's overall utility could be inferred and a personal exchange rate could be estimated between disparate objects.

With respect to social preferences, joint allocations are viewed as "bundled outcomes" that have at least two distinct potential outcomes for a DM – the payoff for the self and the payoff for another. Individual differences emerge because different people place different subjective values upon these sources of utility and make different subjective tradeoffs when evaluating the bundle as a whole. Although the notion of utility is most closely associated with microeconomics, this framework is consistent with functional measurement (Anderson, 1970), specifically in the context of information integration theory (Anderson, 1968). Revealing preferences from finding indifference, and subsequent utility estimation also has a long history in psychology, including Thurstone (1931), Luce & Raiffa (1957), and Kahneman & Tversky (1979).

Utility measurement described

The approach of using utility estimation in the context of own-other-outcome bundles is not a new idea. Francis Edgeworth conjectured that between pure selfishness and pure prosociality (or in his words - *Pure Universalistic*) there are a wide range of middle orientations. Edgeworth (1881, p.16) wrote:

> For between the two extremes Pure Egoistic and Pure Universalistic there may be an indefinite number of impure methods; wherein the happiness of others as compared by the agent (in a calm moment) with his own, neither counts for nothing, nor yet "counts for one," but counts for a fraction.

From this, one can readily see that the idea of utility in social contexts as being affected not only by one's own welfare, but also by the welfare of others, is not new to economics. Edgeworth postulated that the welfare of others does not have the same impact on one's happiness as one's own welfare, but instead has some lesser fraction of that impact. The magnitude of this fraction is an index of prosociality. The coefficients or weights attached to the outcomes of others as specified in utility functions is a modern interpretation of what Edgeworth discussed when using the term *fraction* in this context. Narrow self-interest is just the special case where an individual's coefficient for other's outcomes is equal to exactly 0.

The employment of utility functions for representing social preferences is standard in economics, and a multitude of different other-regarding utility models have been posited to date (e.g. Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Cox et al., 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999; Geanakoplos et al., 1989; Levine, 1998; Loewenstein et al., 1989; Rabin, 1993). However, in economic research, model parameters are estimated or inferred mainly from behavior in strategic situations. This is problematic since behavior in these situations is a function of both preferences and beliefs, and distinguishing between these two factors *ex post* is impossible. Decomposed game techniques were introduced precisely for the reason to overcome this problem by eliminating the possibility of strategic considerations as co-determinants of behavior in interdependent situations. Non-strategic own-other payoff allocation tasks, such as Dictator Games, have also been used for studying social preferences in economics, though. For example, by employing a set of such allocation tasks, Andreoni & Miller (2002) showed that the vast majority of peoples' choice patterns can in principle be represented by a utility function which incorporates payoffs for others, and thus social preferences are rationalizable in a rigorous axiomatic framework.

In psychology, the framework of using joint utility evaluations was chosen by Wyer (1969) who used it to successfully predict choice behavior in particular classes of strategic games. His approach transformed preferences for outcome allocations into utilities. Wyer, as well as Griesinger & Livingston (1973), modeled the utility of joint allocations as a linear combination of the weighted outcomes for the self and for another. Wyer used a utility function with the form

$$u(P_S, P_O) = (\mathbb{1}\, w_1 P_S) + ((1 - \mathbb{1})\, w_2 P_S) + (w_3 P_O) \tag{2.2}$$

where P_S represents the outcome for self, P_O represents the outcome for other, coefficients $w_{1,2,3}$ represent weights of the respective outcomes and $\mathbb{1}$ is an indicator function which yields the value of 1 if $P_S > 0$ and 0 if $P_S \leq 0$. In his experiment, Wyer used a 21-point rating scale to assess the desirability of allocation outcomes. Participants were asked how much they would like, for example, a distribution of 2 points for themselves and −3 points for another. The scale ranged from between −10 and +10 in interval steps. These desirability ratings were then inserted into the above formula as an estimated utility value, conditional on that particular allocation (in this example $P_S = 2$ and $P_O = -3$). After a series of ratings were obtained from a research participant, the weights $w_{1,2,3}$ were estimated by ordinary least-squares fitting.

Different sets of weights indicate different social orientations. Considering instances of positive outcomes, a person with an individualistic orientation would have a high positive w_1 weighting and a w_3 weighting close to zero. Prosocial individuals would have positive values for both w_1 and w_3 that are similar in magnitude. Wyer showed that the competitive orientation would be reflected by a positive w_1 and a negative w_3.

More complex utility models have been posited. For example, second order polynomials have been used to account for joint utilities. Radzicki (1976) used a conjoint measurement technique to identify a best fitting utility function. Participants were asked to make rankings of 25 joint allocations from the most preferred to the least preferred distributions. These rankings were then analyzed and linear programming methods were used to determine the form of the utility function that fit the particular rankings best. This method yielded a particular utility function, along with weighting coefficients, for each of the DMs. Radzicki found that simple linear functions fit 41% of the participants' rankings best, whereas for 8% of the participants' rankings, none of the considered functions fit the ranking data sufficiently. In many cases though, non-linear functions exhibited a significantly better fit to the data than simple linear models. For example, for people with non-linear joint preferences who were concerned with equality in outcomes, the six parameter function which best fit was of the form

$$u(P_S, P_O) = aP_S + bP_O + cP_S^2 + dP_O^2 + e(|P_S - P_O|^q) \tag{2.3}$$

where $0 \leq q \leq 1$ and the other parameters are unrestricted. Although the simplicity of linear models is convenient and in many cases sufficient to describe choice behavior with respect to joint allocations, non-linear models allow for more sophisticated descriptions of choice behavior and are able to account for more complicated patterns in data. Non-linear approaches are suggested by MacCrimmon & Messick (1976), and Wyer (1969) noted that non-linear relations between given outcomes and their utilities could well be possible, if not probable, when payoff amounts exceed a certain range of values under consideration. For example, the increase of a payoff amount from one dollar to two dollars is probably not equally valued as an increase from 500 dollars to 501 dollars with respect to utility. From this point of view, non-linear models are justifiable when the values of outcomes presented to decision makers vary widely in the amounts under consideration.

Discussion of utility measurements

As McClintock & Van Avermaet (1982) noted, the approach of using utility functions for the evaluation of social preferences as performed by Wyer (1969, 1971) or Radzicki (1976) and others (e.g. Messick & Sentis, 1985; Loewenstein et al., 1989) is focused on building simple models that adequately describe the assumed combinatory rules underlying preferences within the framework of own-other-outcome allocations and theoretically allows for an infinite number of possible social value orientations. Therefore, these models do not state particular SVOs *a priori*, but rather let SVOs be inferred from the weight values provided by the resulting fitted utility functions. By contrast, methods following the line of Messick & McClintock (1968) – such as the Triple-Dominance Measure or decomposed games in general as well as the Ring Measure – are more concerned with the "substantive nature" (McClintock & Van Avermaet, 1982, p. 59) of SVO. That is, the existence of a number of predefined SVO categories is assumed *a priori* and the emphasis is placed upon measuring them directly by letting people choose between two or more outcome allocations that are indicative of particular archetypical social orientations.

Nevertheless, there are certain problems with the use of utility measurement for assessing SVO. First, when SVOs for different persons are expressed using different functional forms, each representation potentially containing a different number of parameters, it is difficult to compare or aggregate results. For example, Radzicki's method is likely too flexible to be useful and certainly over-fits rating data that are at least in some part measurement error. However, the problem of interpersonal comparability can be solved by employing only one functional form for model fitting. This way, all subjects can be described in terms of individual best fitting values of the same model parameters, which allows for interpersonal comparisons.

Second, DMs are not making choices in the utility estimation methods described above, but rather judging the attractiveness of different hypothetical allocations. Rating procedures, such as the one applied by Wyer (1969), make strong linear assumptions of the response scale which are likely not met, especially given the scale's structure (a 21-point rating scale!). Eliciting judgments rather than choices introduces a level of abstraction that does not offer clear benefits. Moreover, inducing participants to honestly report their preferences by incentive compatible methods is not easily accommodated with judgment tasks like the ones suggested above. However, this limitation is, of course, not inherent to the method of utility measurement in general.

Since utility measurement is a whole methodology class – which is not restricted to the assessment of social preferences – rather than a particular SVO measurement instrument, its evaluation on the basis of our predefined set of criteria is complicated. The criterion of psychometric properties, in particular, is not readily applicable here. However, the approach can be partially evaluated in terms of the remaining three criteria. With respect to output resolution these methods facilitate continuous, and even multidimensional, data. Nevertheless, the generation of this high resolution output is costly. Model fitting procedures require the use of sophisticated quantitative tools and their application can be quite demanding and time-consuming for researchers. Whether the method can be regarded as more or less efficient depends on the purpose of its use. While it can be considered as highly efficient when employed for the purpose of investigating the nature of social preferences itself, it is quite inefficient when SVO is assessed in an experiment in order to explore its simple linear relationship with other variables. The greatest advantage of utility measurement is its flexibility, which is unique in comparison with the other methods discussed in this paper. All that the method requires is data on preference orderings, that is, choice data, rating data, or data on comparative preference judgments. Hence, the data could be based on option sets involving gains and losses, or tangible objects rather than money. Due to this flexibility, the method can be used in basically any experimental context.

Indifference curves: Measuring SVO graphically

In addition to utility functions as representations of different SVOs, indifference curves too can be used as a representation of preferences for different joint outcomes. Consider a set of curves plotted on a two-dimensional plane defined by the payoff to self on the x-axis and payoff to another on the y-axis. Radzicki (1976) depicted indifference curves resulting from a utility function corresponding to the best fit of participants' rating data.

Another innovative approach along these lines was developed by Harrison (1998) who conducted what he termed an “indifference curve experiment” based on a procedure described by economists MacCrimmon & Toda (1969) and similar to the approach by Thurstone (1931). Harrison requested participants make several pairwise choices between various joint allocations of money. For example, participants chose between an allocation of $10 to themselves and $8 to another, or say $17 for themselves and $15 to another.

The first optional distribution was referred to as the “reference allocation” and for a set of choices was always the same bundle. After multiple choices are made, the researcher can infer an indifference curve running through the reference point and the boundary between those allocations that were preferred over the reference allocation and those allocations which were not preferred over the reference allocation (see Figure 2.4).

By repeating this procedure and using a different reference allocation point, an arbitrary number of indifference curves can be discerned, resulting in a contour map consistent with a DM’s SVO. This measurement procedure can be done to an arbitrary level of precision depending on the number of choice sets presented to a DM. The example depicted in Figure 2.4 shows how one indifference curve can be inferred from 18 distinct pairwise choices; these are stimuli from Harrison (1998).

It is worth noting that such a procedure has some advantages. First, participants are presented with pairwise comparisons rather than multiple comparisons or abstract rating scales. Furthermore, no *a priori* assumption about the existence of a number of predefined social value orientations is needed, while the indifference curve patterns resulting from the procedure allow for interpretations regarding the extent to which they are consistent with respective SVOs. This method can also quickly identify intransitive choice sets or random responding from particular participants, as no indifference curve can be inferred from their allocation decisions. It can also readily accommodate incentive compatible choices.

This approach, however, also has some limitations. First, the resulting indifference curves are identified heuristically and not analytically. This means that a curve is “eye-balled” into place in order to divide the chosen points from the non-chosen points. An undeterminable number of bivariate functions could yield a curve that separates the chosen options from the non-chosen options while intersecting the reference allocation. Identifying the best fitting curve is impossible given the low resolution of the choice data and the heuristic method of curve fitting does not lend itself to parameterization. In order to address this issue, a researcher could specify a functional form for the joint utility equation (similar to Radzicki, 1976), and then roughly estimate an underlying utility function with parameters that are consistent with the choices. Although this would quantify the heuristic indifference curve to some degree, the resulting joint utility function and parameters are not easily comparable between participants given the variety of functional forms that may be used to rationalize the underlying set of binary choices. Another shortcoming with this approach is the relatively large number of choices a DM is required to make in order to infer one indifference curve. In the example shown in Figure 2.4, which is based on stimuli from Harrison (1998), 18 binary choices were used to approximate just one indifference curve. With respect to our predefined set of criteria, the indifference curve approach has the same properties as utility measurement in general.

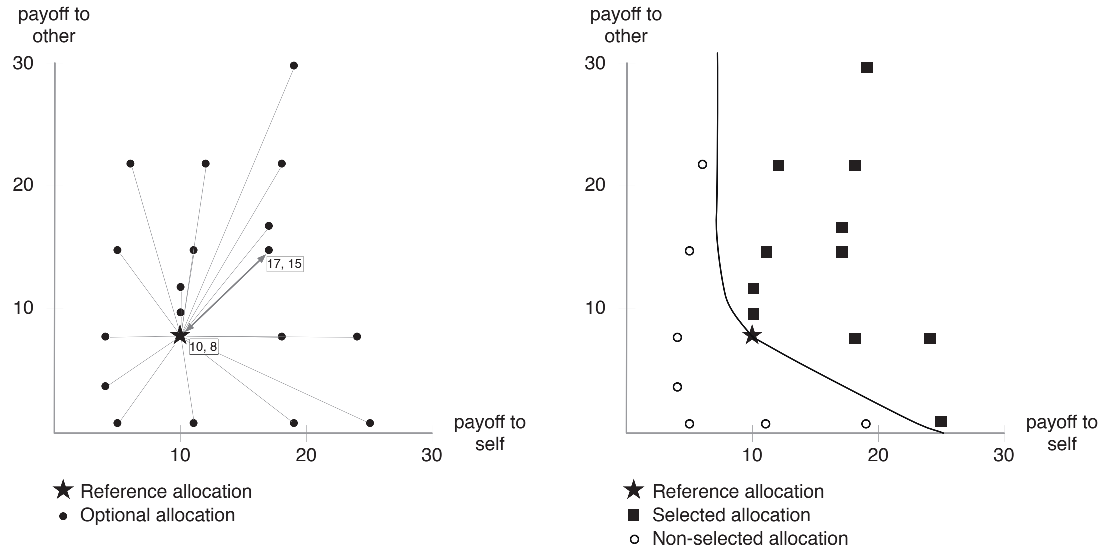


Figure 2.4: An example of the indifference curve method. The left panel shows 18 pairs of potential allocation choices that a participant would consider. Each of them has in common the Reference Allocation. Each of the Optional Allocations is either chosen over the Reference Allocation or not. The resulting pattern of preferred options can be used to identify an indifference curve which necessarily intersects the Reference Allocation point.

2.5.4 The Social Behavior Scale

The Social Behavior Scale comes from developmental psychology and was devised as a measure that controls for individualism by keeping the payoff to the DM constant and varying only the payoff to the other. Consequently, decision makers are not given the opportunity to maximize their own gain by choosing a particular alternative, but only have control over the outcome for another.

The Social Behavior Scale described

The Social Behavior Scale is a choice task with four alternatives as shown in Figure 2.5. The alternatives are: rivalry & superiority; superiority; equality; and altruism & group enhancement. Outcomes for self and other are depicted as small squares and labeled as valuable "chips." This measure was devised by Knight & Kagan (1977) in an effort to study the social behavior of young children from different ethnic groups. In their experimental study, children were told that the more chips they acquired, the more toys they would receive. This "currency" is easily comprehensible and can be presented without any numerical abstraction, ideal for use with children, especially when there may be differences with respect to their formal educational experience. Participants were asked to choose one out of the four alternatives according to their preferences.

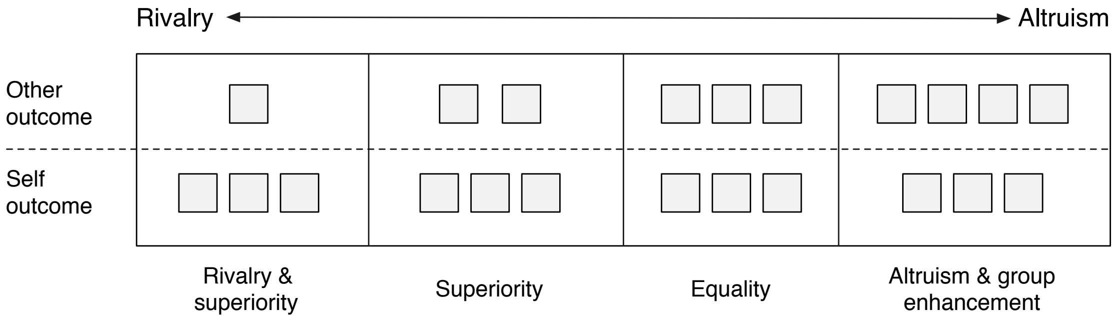


Figure 2.5: The Social Behavior Scale showing potential distributions of valuable items between the DM and some other person.

Discussion of the Social Behavior Scale

As a result of its properties, especially with regard to assessing individualistic orientations, the Social Behavior Scale is too restricted to be an appropriate measure for SVO in general. Of course, one can readily imagine an alternative form of the Social Behavior Scale, where individualism is not strictly controlled. In fact, a variant of the Social Behavior Scale called the Social Orientation Choice Card with a classical triple-dominance structure is available (Knight, 1981). We can imagine numerous alternative forms of that kind which present allocation decisions across a range of different outcomes with simplified stimuli. Such methods can be advantageous when conducting studies with children or populations not accustomed to quantified information, as was the case in Knight & Kagan's research. Since the Social Behavior Scale can be regarded as a non-monetary payoff variant of a decomposed game measure, it receives a similar evaluation as the Triple-Dominance Measure with respect to our predefined criteria. The measure consists of only one item which subjects answer several times, and the subjects are categorized according to their modal choice. Hence, the method is efficient in terms of time required for completion and output computation. However, output is categorical. To date, the measure has been used exclusively to study the development of SVOs in children, such that no data on its predictive validity with respect to other variables are available. However, there is some data on the measure's convergent validity with the Regression and Clustering approach (see Knight & Dubro, 1984, discussed later in this paper) showing 66.7% categorical agreement. Also, Knight & Kagan (1977) report on data hinting at the measure's test-retest reliability. They report a correlation of 0.72 between the total number of chips allocated to other at two points in time separated by 2-5 days[14]. As discussed earlier in this paper, the scoring rule of counting the payoffs allocated to the other is problematic, though, and complicates interpretation of results. From the data available, we can only infer that the measure's psychometric properties are satisfying at best.

[14]The conditions at the two points in time varied slightly. In one condition, the receiver was imaginary, and in the other condition there was a real and visible, but passive receiver.

2.5.5 The Ring Measure

The Ring Measure is a method from social psychology for assessing SVO that uses a series of dichotomous allocation decisions and derives an SVO score from the combined results of the choices. This value is then used for assigning that participant to one of the archetypical SVO categories. The method is based on the notion that joint payoffs can be represented on a Cartesian coordinate system where payoffs to the DM are represented on the x-axis and payoffs to another person are represented on the y-axis (see Figure 2.1). This idea is consistent with the geometrical model which was devised by Griesinger & Livingston (1973), who stated that a person's SVO can be conceptualized as a vector with a certain direction and magnitude in the joint payoff plane. The utility of a particular payoff allocation can then be expressed as the scalar product of the motivational vector with the vector of the given choice, or in other words, the projection of the given choice vector on the motivational vector. Consequently, a person will always choose the payoff allocation with the greatest projection on his or her motivational vector. Further, the angle of the motivational vector is indicative of a person's social preferences. For example, a motivational vector at the angle of $\theta_M = 45°$ represents a prosocial orientation, whereas an individualistic motivation is represented by a vector at $\theta_M = 0°$ (see Figure 2.1). Following this conceptualization, Liebrand (1984) developed and established the Ring Measure as a novel method for categorizing participants into the archetypical SVO classes (see Table 2.2).

The Ring Measure described

The Ring Measure presents DMs with a set of N dichotomous allocation decisions that are defined by N equidistant points on a circle centered at the Cartesian origin $(x = 0, y = 0)$. Each pair of adjacent points (defining a chord on the circle) serves as the two distribution options, and the DM makes a series of choices over these different allocations. Researchers have set the value of N at both 24 (Liebrand & McClintock, 1988) and 16 (Liebrand, 1984).[15] To date, the Ring Measure has generally been implemented by a defining center point at $(0, 0)$, yielding both positive and negative allocation values. However, to facilitate comparison with other measurement methods presented in this paper, the distribution values have been standardized to range between 0 and 100 (see Figure 2.6). This is equivalent to defining a ring with a center at $(50, 50)$ and a radius of 50 units.

After a research participant has made her N allocation choices, a vector is computed by adding her chosen options together, thus yielding two numbers (the sum of money the participant allocated to herself, and the sum of money the participant allocated to the other person). The resulting point can be interpreted as a vector (using the center point of the ring as its origin). The angle of this vector corresponds to a person's SVO and can be computed by

$$\theta = \arctan\left(\frac{(\sum P_O)}{(\sum P_S)}\right) \tag{2.4}$$

[15] Liebrand (1984) employed 16 equally spaced pairs of outcomes on each of two circles (A and B) with radii of \$7.00 (circle A) and \$8.50 (circle B), resulting in a total of 32 outcome pairs as choice allocations.

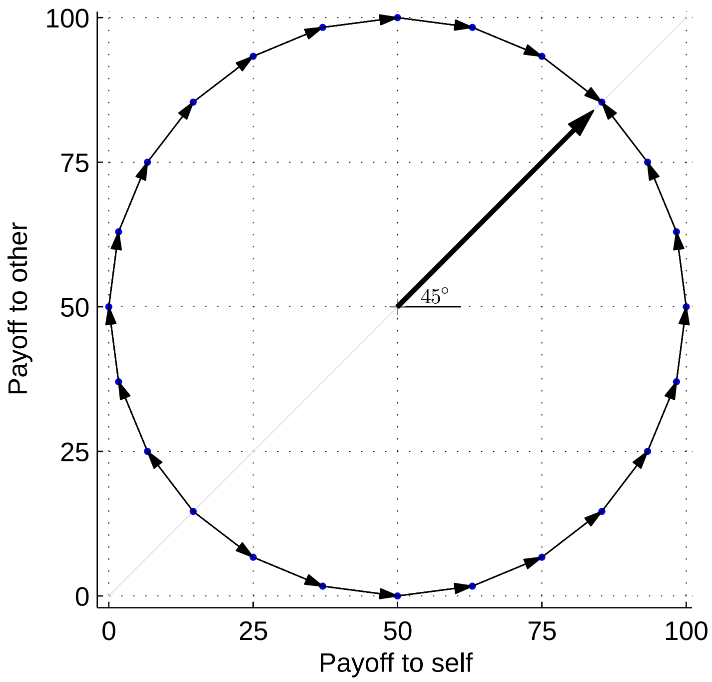


Figure 2.6: Here the 24 allocation choices on a Ring Measure are represented graphically. The smaller arrows correspond to a set of hypothetical choices by a prosocial DM and the points to the joint allocation options. For each set of options, the DM selected the option that maximized joint gain, as indicated by an arrow pointing toward the more preferred joint allocation. This pattern of results reveals an underlying preference vector that is represented by the large arrow. Its angle (in this case +45°) serves as an elegant summary of the DM's social preferences as revealed by her choices of resource allocations.

where $\sum P_O$ is the sum of payoffs selected for the other person and $\sum P_S$ is the sum of payoffs allocated to the self. The length of the vector from the center of the ring indicates the internal consistency of the DM's allocation decisions. If a person makes inconsistent choices, the result is a shorter vector. Perfectly consistent choice sets have the property of having one option being chosen twice (the most preferred distribution in the whole set), one option never being chosen (the least preferred allocation), and the remaining allocations being chosen exactly once (see Figure 2.6 for an example of a perfectly consistent choice pattern). The vector resulting from a perfectly consistent set of choices will have a length equal to twice the radius of the circle used to generate the items, conditional that the center of the circle is fixed at the Cartesian origin. Because of the structure of items in a Ring Measure, the more rigorous property of preference *transitivity* can rarely be evaluated for a participant's set of choices as there is only one possible Hamiltonian cycle in the set of items. Only the weaker condition of *consistency* with a single underlying motivational vector can be evaluated with any fidelity.

When the angle of a person's vector is determined, that person is assigned to one of the eight SVO categories listed in Table 2.2. In order to prevent invalid classifications, DMs were typically only classified if the consistency of their choices was at least 60%. However, there is some variety regarding the standards for establishing a classification. For example, some classifications are made with a 50% consistency level (e.g. Van Lange, 1999) or, in other cases, only if a vector is not shorter than a quarter (e.g. McClintock & Liebrand, 1988) or even a fifth (e.g. Dehue et al., 1993) of the maximum possible vector length.

Discussion of the Ring Measure

Upon first consideration, the Ring Measure's agnostic method of defining items evenly over the complete circle may appear to be a sensible approach. However there is overwhelming evidence demonstrating that SVOs are not uniformly distributed among people and that the vast majority of DMs do not appear to attach negative weight to their own payoffs. Therefore, using items uniformly from the whole space of possible preferences is inefficient. The structure of the Ring Measure assigns equal value to all of the items, including the following two: (1) Is a person more prosocial or more individualistic? (2) Is a person more of a martyr or more of a masochist? Clearly the first question is more useful in trying to understand the motivations of typical DMs. But, because of its blanket approach, the majority of items contained in the Ring Measure provide no useful information about the motivations of the person answering them. The only items that offer any useful diagnostic information are those with slopes that are nearly perpendicular to the underlying motivational vector of the decision maker. This agnostic approach results in a highly inefficient research tool.

Second, the Ring Measure fails to classify a significant number of participants due to inconsistent choice behavior. In their analysis of several studies that applied the Ring Measure, Au & Kwong (2004) reported up to 20% unclassifiable participants and in two experiments performed by Liebrand (1984), the percentage of unclassifiable participants across the two experiments was 15%. In analyzing these percentages, one has to take into consideration that Liebrand (1984) used a 60% consistency criterion, whereas in at least some studies analyzed by Au & Kwong, a 50% consistency criterion was chosen (e.g. Van Lange, 1999). Part of this inconsistency could be the result of asking for people's preferences across such a wide range of potential allocations, some of which the DMs may have only weak preferences about. For example, we have evidence that people show less consistent choice behavior in the items located in quadrants two and three of the Cartesian plane (i.e. the left side of the ring) compared to items located in quadrants one and four (i.e. the right side of the ring). Using the ratio $\frac{(\sum P_O)}{(\sum P_S)}$ as unit of analysis, we found a test-retest reliability of 0.617 for the left half of the ring compared to 0.702 for the right half of the ring.[16] These results indicate that the Ring Measure could be improved by cutting it in half with a vertical line, and only using the items located in quadrants one and four of the Cartesian plane. The resulting Half-Ring Measure has been used in the past (e.g. Balliet, 2007; Joireman, 1996), but with only limited success.

[16] These results are obtained with data from Murphy et al. (2011).

Third, inequality aversion would manifest itself as inconsistency in the Ring Measure. The 45° diagonal line from the origin intersects the ring in two places; the point in the northeast part of the ring corresponds to both minimizing inequality as well as maximizing joint gain, whereas the point in the southwest part corresponds to minimizing inequality but minimizing joint gain. If a DM were sufficiently motivated by inequality aversion, she would produce an inconsistent set of allocations which would result in a shorter vector. The Ring Measure does not address this limitation. And a further complication is that in some studies, the Ring Measure had both positive and negative outcomes. Given the evidence of how losses loom larger than gains (Kahneman & Tversky, 1979), it is possible that DMs make different tradeoffs when considering positive outcomes versus negative (or mixed) outcomes. Lastly, the presence of losses makes it a challenge to implement the Ring Measure in an incentive compatible way as taking money from research participants is generally verboten.

Fourth, although the Ring Measure produces scores in terms of angular degrees, its final output is categorical. As discussed earlier in this paper, one reason for discarding the continuous information may have been that the conceptual interpretation of a Ring Measure angle is two-dimensional, rather than unidimensional. That is, the angle summarizes the weight one attaches to the other one's outcomes as well as the weight one attaches to own outcomes. If only the right half of the ring is used, an angle's interpretation is unidimensional, referring to the weight one attaches to the outcomes of others in relation to the own, such that the angle can be used as a continuous SVO score as employed by Balliet (2007), for instance. Nevertheless, the Ring Measure in its original form predominates, and so does the practice of categorization.

The psychometric properties of the Ring Measure are marginal to weak. In terms of agreement with other SVO measures, Liebrand & van Run (1985, p. 94) report that only 52.54 percent of 236 subjects were categorized into the same SVO category by both the Ring Measure and another decomposed game procedure (see Kuhlman & Marshello, 1975b). Only when altruists and cooperators were combined, the categorical agreement reached a satisfying level (73%). Murphy et al. (2011, p. 775-776) report satisfying categorical agreement of 67% with the Triple-Dominance Measure and 75% with the Slider Measure (discussed later in this paper). In terms of test-retest reliability, Murphy et al. (2011, p. 775) report that the Ring Measure categorized 68% of the subjects into the same SVO category at both of two points in time separated by two weeks. This result is consistent with findings from Dehue et al. (1993, p. 280), who report 70% consistency across a two month period. Although SVO as assessed with the Ring Measure has often been shown to be significantly associated with cooperative behavior in social dilemmas (e.g. Liebrand, 1984; Liebrand & van Run, 1985; Offerman et al., 1996; Smeesters et al., 2003; Sonnemans et al., 1998), effect sizes are rarely reported, which hinders the proper estimation of the method's predictive validity.

Circle-test: A one-item version of the Ring Measure

Sonnemans et al. (2006) conducted a study in economic psychology which required that participants complete an SVO measure four times within the context of an ongoing public goods game. To these ends, the researchers modified the Ring Measure so that participants had to make only one allocation decision to yield an SVO score. They termed this modified Ring Measure the "Circle Test." In the Circle Test, participants were provided with a graphical representation of the SVO Ring on a computer screen (similar to Figure 2.1). Participants were then requested to make their joint allocation decision by clicking on the arc of the circle somewhere. Once a position was tentatively chosen, the corresponding vector appeared on the screen as an arrow. Participants then could, if they wanted to, change the angle of the vector while seeing how these changes affected the payoff allocations for themselves and the other person. Once a participant found her most preferred joint allocation, she confirmed her decision and this completed the measurement.

The Circle Measure is a highly efficient measure of SVO, requiring only one allocation choice be made in order to yield a continuous score for a person (see also Van Winden et al., 2008). But one disadvantage of this brevity is that no information about measurement reliability can be gained. As the circle measure has only one item, it is not possible to check whether the choice is transitive or consistent with respect to other choices. The measure does not provide any possibility to assess the magnitude of measurement error, and at the extremes cannot assess if a participant responded veridically or randomly. Another limitation is that the changes in payoffs that correspond to movements on the arc are non-linear. The visual representation is straightforward but the underlying tradeoffs that occur as a DM moves between different points on the arc are non-intuitive. The arc defining the joint payoffs is necessarily curved (its second derivative is non-zero), thus not only are the joint payoffs changing as a DM adjusts the allocation vector, but the rate of change for each of the payoffs is also changing. DMs may mitigate this complexity by selecting cardinal points on the circle rather than points consistent with their more nuanced actual preferences. Lastly, secondary preferences about different allocation options remain unknown when using the circle measure and inequality aversion remains indistinguishable from joint gain maximization.

In contrast to typical practice, Sonnemans et al. (2006) used the SVO angle as the dependent variable, rather than categorizing subjects according to it. Since 98% of their subjects' angles ranged between $-45°$ and $+45°$, using the angle as a unidimensional continuous scale can be justified and is sensible. To our knowledge, there are no data available on the Circle test's psychometric properties. With respect to the criterion of particular advantages, we acknowledge that the Circle test is the briefest method yielding high-resolution output. However, since no data are available on measurement reliability and validity, it is not possible to estimate the drawbacks associated with the method's high efficiency.

2.5.6 Regression and clustering approach

Consistent with judgment (e.g. Wyer, 1969, 1971) and conjoint measurement techniques (e.g. Luce & Tukey, 1964; Sawyer, 1966; Radzicki, 1976), Knight & Dubro (1984) developed another method for assessing social preferences that applies regression and cluster analysis to a set of well structured preference judgments.

Regression and clustering approach described

To obtaining preference data, Knight & Dubro had participants rate the desirability of joint allocations on a 7-point scale, where the possible allocations were composed of all combinations of payoffs ranging from 0¢ to 6¢ in increments of 1¢, resulting in a total of 49 possible allocations and the same number of ratings. Then, for each person's ratings, a multiple regression equation was used to model the desirability ratings, while three predictors were used in the analyses: own gain (number of cents for self), other's gain (number of cents for the other) and equal gains (difference between the own gain and the other's gain). The resulting regression coefficients were then used in a cluster analysis which yielded six general clusters. These clusters were interpreted as different categories of SVO: *equality*; *group enhancement*; *superiority*; *individualism*; *equality & individualism*; and *individualism & superiority*.

Discussion of regression and clustering approaches

The similarities between the utility measurement approach as proposed by Wyer (1969), and this regression analysis method are clear. In both methods, preference data are elicited and used to compute parameter values by a least squares estimation technique. The weights attached to outcome values in the utility functions are conceptually equivalent to the regression coefficients. The novelty of Knight & Dubro's approach is the use of regression coefficients in a cluster analysis in order to classify people into SVO categories. Given the relatively high median squared multiple correlation coefficients for each of the six clusters, ranging from 0.609 to 0.858, it is clear that participants exhibited substantial consistency in their judgments of the attractiveness of different joint distributions.

One minor drawback with respect to the feasibility of the measure is that Knight & Dubro's procedure, like the utility measure approaches, makes use of more sophisticated statistical tools that may be a barrier to adoption for some researchers, especially if SVO is assessed as only one among several independent variables, for instance. Moreover, the result of this approach is still a categorization of participants into different SVO classes, while the derivation of a unidimensional, continuous scale of SVO which would facilitate analyses is not feasible through this procedure since SVO here is represented by a combination of three parameters. Therefore, the results of this technique (a simple categorization of participants) may not be worth the effort of running regression and clustering analyses. Other approaches, discussed previously in this paper, yield a similarly resolved output by means of much simpler techniques, challenging this method's efficiency. Also, we do not see any particular advantage of this method when compared to others.

With respect to the method's convergent validity with other SVO measures, Knight & Dubro (1984, p. 103) report that 66.7% of the time the subjects' cluster membership was consistent with their choice patterns in the Social Behavior Scale (Knight & Kagan, 1977, discussed earlier in this paper) and its triple-dominance variant called Social Orientation Choice Card (Knight, 1981). To our knowledge, no data on the measure's predictive validity or test-retest reliability are available.

2.5.7 Schulz and May's Sphere Measure

The Sphere Measure described

On the basis of previous work on methods for assessing SVO such as utility measurement (e.g. Wyer, 1969, 1971; Griesinger & Livingston, 1973; Radzicki, 1976), the Ring Measure procedure (Liebrand, 1984), and the regression and clustering approach (Knight & Dubro, 1984), an additional way of determining people's social motivations was devised by Schulz & May (1989). They differentiate between simple linear SVOs (individualism, sacrifice, altruism, aggression, cooperation, competition), non-simple linear SVOs (all possible mixtures of simple linear SVOs), simple conditional linear SVOs (maximin and egalitarianism) as proposed by MacCrimmon & Messick (1976), and non-simple nonlinear SVOs (all possible mixtures of simple conditional and non-conditional linear SVOs). For assessing these different SVO types, Schulz & May applied two measurement methods with the goal of comparing results from each of them. First they used a ranking procedure and second they used a pairwise comparison procedure. Concretely, participants first made pair-wise comparisons between all possible combinations of 15 own-other-payoff distributions, resulting in a total of 105 comparisons per participant. After completing the pair-wise comparison task, participants were asked to rank order the same 15 payoff allocations without ties using a graphic presentation of the allocation options. The data from both methods were then analyzed by using a utility model with the general form

$$u(P_S, P_O) = aP_S + bP_O + c|P_S - P_O| \tag{2.5}$$

which is flexible enough to contain all of the archetypical SVO types as special cases. Roughly speaking, while the Ring Measure uses the parameters a and b for calculating the SVO angle on a two-dimensional plane, Schulz & May extend the model with parameter c, thus yielding a three-dimensional model. The third dimension is useful in accounting for conditional SVOs (e.g. egalitarian or maximin). In order to restrict the model, the authors set the condition such that $a^2 + b^2 + c^2 = 1$, giving the model a spherical geometric representation. Similar to the Ring Measure procedure, participants are then categorized according to their vector directions. In contrast to the Ring Measure, the Sphere Measure vector extends into 3-space and yields a point on the unit sphere rather than a point on a two dimensional circle.

Discussion of the Sphere Measure

Although Schulz & May make use of more sophisticated mathematical tools, and use a more complicated geometric representation than is reflected in previous methods, the measure still yields results at only the nominal scale level. Richer results could be extracted from the data of Schulz & May (e.g. transitivity of individual's choice sets; the angle of the projection of the inferred motivational vector on the self/other plane) but are unfortunately not. Further, this measurement method places substantial demands upon participants, requiring them to make 105 pair-wise decisions about joint payoff allocations, as well as rank order 15 different self/other allocations. Considering the resolution of the results, these demands are hard to justify. Therefore, we judge the method's efficiency as low.

To our knowledge, there are no data available on the Sphere Measure's predictive validity or test-retest reliability. Also, the Sphere Measure's convergence with other SVO measures has not been tested so far. However, Schulz & May (1989, p. 53) report 75.9% agreement between the subjects' categorization as derived from the ranking procedure and the pair comparison procedure. Hence, there is some – albeit limited – evidence in favor of the Sphere Measure's psychometric quality.

2.5.8 The SVO Slider Measure

Murphy et al. (2011) aimed at constructing a measurement method which combines the strengths of existent techniques while avoiding, when possible, some of their weaknesses. Concretely, they posited that a good measure of social preferences should have the following properties: 1) For pragmatic reasons, a measure should be *easy to administer*. Since SVO is often assessed as only one variable among a variety of individual differences, the measurement procedure should be time efficient, straightforward, and the measurement evaluation should not require the application of sophisticated mathematical techniques. 2) An SVO measure should be *efficient*, i.e. able to assess the empirically most relevant SVOs as reliably as possible while neglecting pathological SVOs which are hardly ever observed in the wild (e.g. sadistic, masochistic, sado-masochistic, etc.). 3) A measure should yield a *unidimensional scale of SVO at the ratio level* which facilitates further analyses and manageability. 4) A measure should be highly *sensitive to inter- and intra-individual differences*, which demands high resolution of data. 5) *Inequality aversion* has to be detectable and distinguishable from a preference for joint gain maximization. 6) A measure should allow for checking the consistency of a DM's choices in terms of *detecting intransitive choice patterns* as indicators of random responses. 7) An SVO measure should have good psychometric properties, i.e. *high reliability and validity*.

The SVO Slider Measure described

The SVO Slider Measure can be administered as an online or a paper-pencil assessment (see Figures 2.7 and 2.8, respectively). It consists of six primary and nine optional secondary items, which all have the same general form. That is, each item represents a specific continuum of own-other payoff allocations that can be explored by sliding across the options within the continuum's boundaries. The DM registers her choice by selecting the most preferred outcome. The six primary items reflect the six lines which fully interconnect the coordinates of the empirically most relevant idealized SVO types (altruistic, prosocial, individualistic, and competitive) in the Cartesian SVO framework with the circle having a radius of 50 and its center at 50, 50 as shown in Figure 2.9. This item configuration allows for obtaining a unidimensional SVO score, determining the rank order of social preferences, and checking for transitivity in a DM's responses.

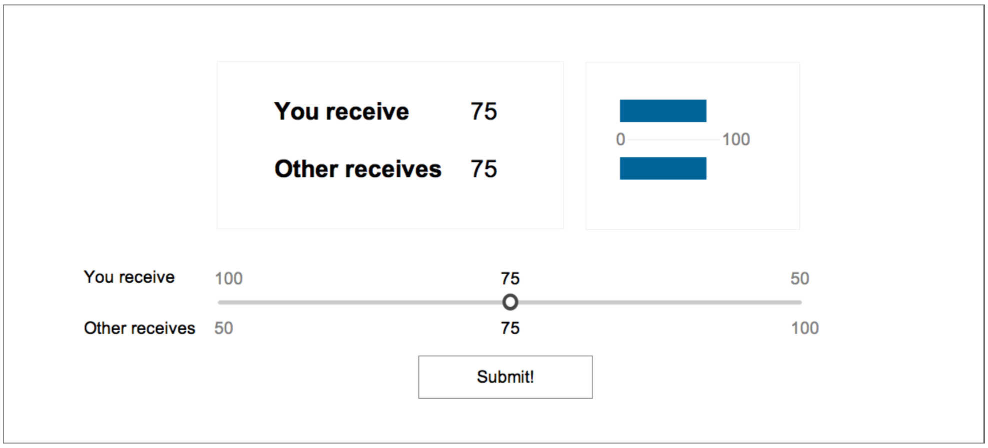


Figure 2.7: A screenshot of one item from the SVO Slider Measure online version. For this item, the DM is choosing between the individualistic distribution on the left and the altruistic distribution on the right. This item is unique in that there is a constant sum (150) that the DM is allocating between himself and the other person. This kind of choice is a dictator game and it is worth noting that it is embedded as part of the Slider Measure.

After a DM has chosen her most preferred payoff allocation in each of the six primary items, her SVO angle can be calculated as follows:

$$\mathrm{SVO}^\circ = \arctan\left(\frac{(\bar{P}_O - 50)}{(\bar{P}_S - 50)}\right) \tag{2.6}$$

where $\bar{P}_S$ is the mean payoff allocated to the self and $\bar{P}_O$ is the mean payoff allocated to the other. The amount of 50 is subtracted from these means in order to shift the center of the ring (50, 50) to the origin of the Cartesian plane such that the inverse tangent of the ratio between $\bar{P}_S$ and $\bar{P}_O$ yields a readably interpretable index, i.e. the individualistic orientation is represented by the angle $\mathrm{SVO}^\circ = 0$. A participant's computed angle is a unidimensional, continuous scale of SVO where higher angular degrees indicate greater concern for the welfare of others, with a lower limit at -16.26° reflecting perfect competitiveness and an upper limit at 61.39° reflecting perfect altruism. If desired, participants' scores can be reduced to one of the four SVO types (altruistic, prosocial, individualistic, or competitive) by means of their SVO angles' values (for the details of this procedure, see Murphy et al., 2011). The categorical output may facilitate comparisons of new results with previous findings, but using the continuous scale is strongly recommended for any other data analysis.

The nine secondary items of the Slider Measure are constructed for the purpose of detecting inequality aversion and distinguishing it from a preference for joint gain maximization. Both are

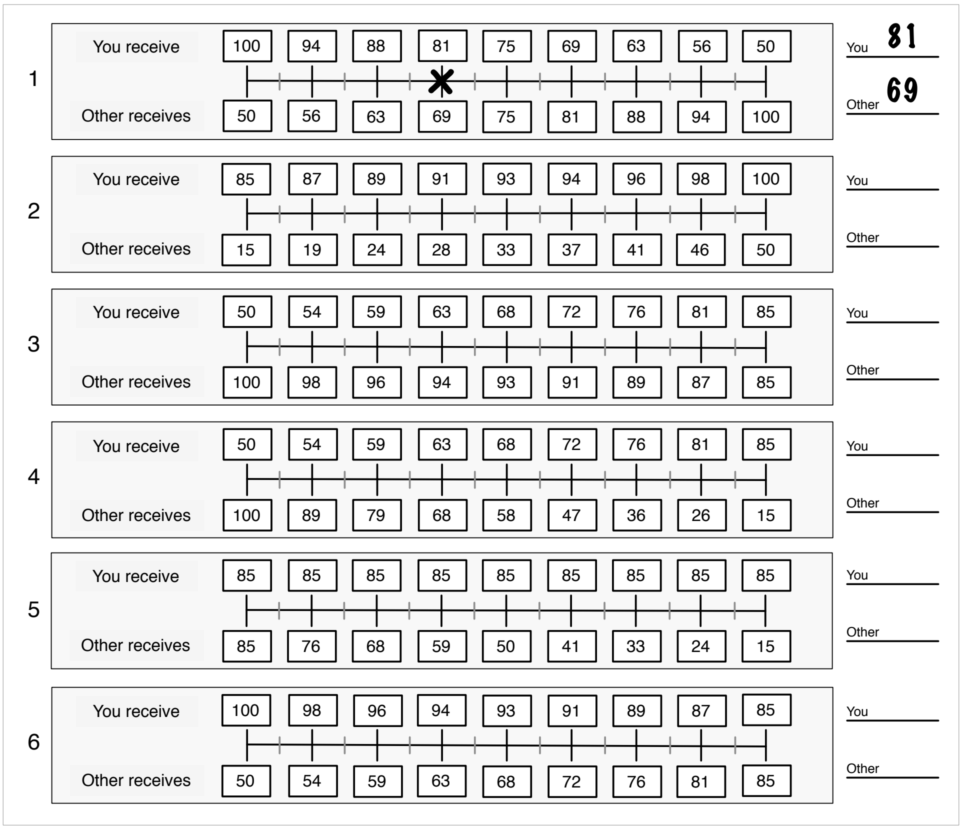


Figure 2.8: The six primary items of the SVO Slider Measure's paper based version as participants saw it.

prosocial preferences but they are different motivations that may represent different goals for a DM. A graphical representation of these items is shown in Figure 2.10. The rationale behind the construction of the secondary items is the idea that inequality aversive participants will choose allocation options close to the 45° line, since these allocations minimize inequality. In contrast, joint gain maximizers will choose the options that maximize the sum of the payoffs; these points are located each at one of the endpoints of the items with a slope other than −45°. Prosocial participants can then be scored along a continuum from perfectly inequality averse to perfectly joint gain maximizing. The results from non-prosocial individuals on the secondary items are not additionally informative and typically confirm their results from the primary items. For example, an individualistic DM will answer the secondary items in such a way as to maximize their own payoff which is neither inequality averse nor joint gain maximizing. The secondary items are maximally informative regarding the more nuanced preferences of prosocial DMs.

Discussion of the SVO Slider Measure

With respect to the SVO Slider Measure's psychometric properties, Murphy et al. (2011) report a test-retest reliability of $r = 0.915$ (or 89% categorical agreement) over a one-week period,

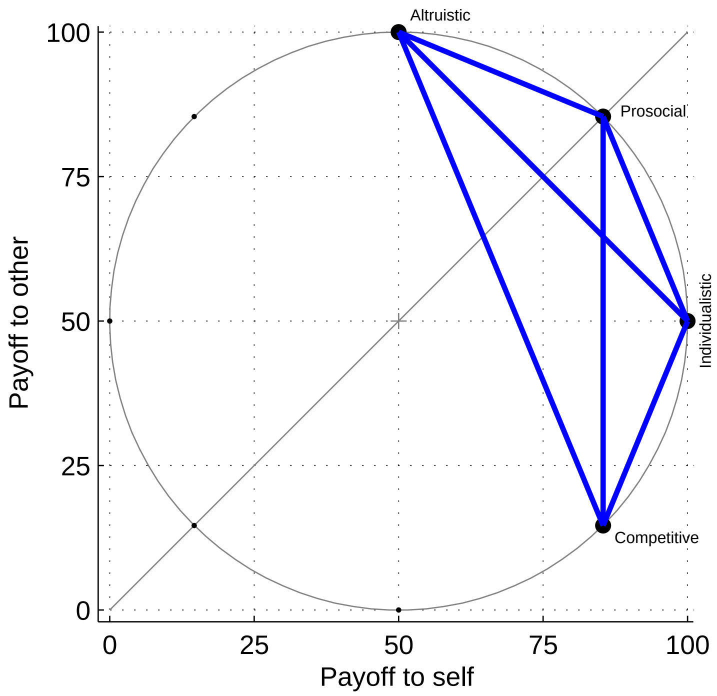


Figure 2.9: A graphical representation of the Slider Measure's six primary items. These items can be scored to yield an index of social preference on a continuous scale ranging from Competitive to Altruistic. The vast majority of people score in the areas of Prosocial and Individualistic but there is pronounced and reliable variance within these categories.

and they could show that the Slider Measure outperformed both the 9-Item Triple-Dominance Measure and the Ring Measure on that metric. Moreover, the Slider Measure exhibited good convergent validity with these two other measures, categorizing the same participants into the same SVO category as did these measures at least 70% of the time. The Slider Measure also shows moderate but significant predictive validity with respect to the binary choices in a Prisoner's Dilemma game (r_{pb} = .24, Murphy et al., 2011) and excellent predictive validity with respect to contributions in a linear Public Goods Game ($r = 0.47$, Murphy & Ackermann, 2014b). Since the Slider Measure requires subjects to complete only six items for computing a continuous score, and because the computation of this score is straightforward, we judge the method as efficient.

An additional feature of the Slider Measure is that the data it yields (for both primary and secondary items) are amenable to mathematical modeling (see Ackermann & Murphy, 2013). Also, the data can be checked for violations of transitivity and rank orderings of SVOs can be computed. Hence, the data produced by the Slider Measure facilitate utility model fitting analyses. Several utility models of other regarding preferences have been developed in behavioral economics (see, for instance, Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002) that include constructs like efficiency maximizing, inequality aversion, fairness, and

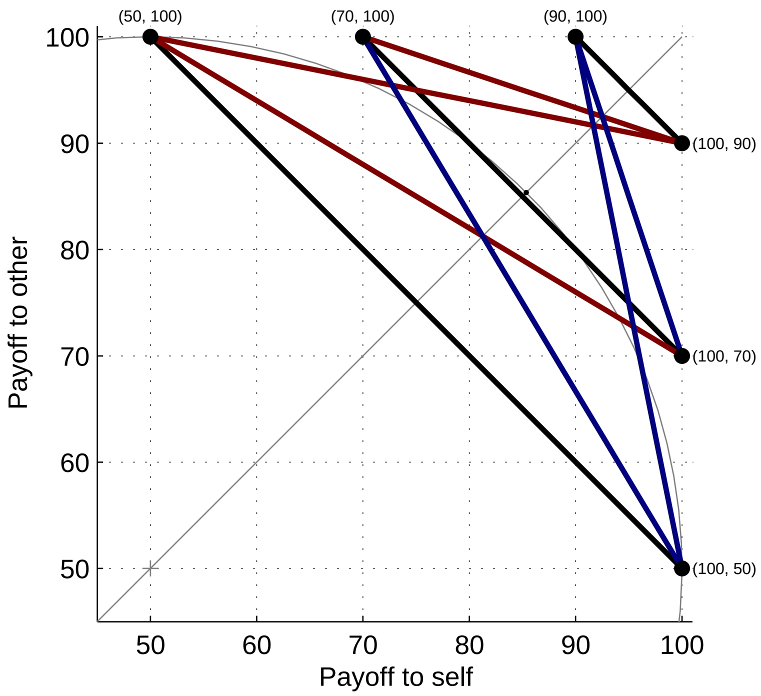


Figure 2.10: A graphical representation of the Slider Measure's nine secondary items. These items are designed explicitly to disentangle the prosocial motivations of *inequality aversion* and *joint gain maximization* and like the primary items yield a score on a continuum between these distinct prosocial motivations.

reciprocity. The psychological literature related to these same issues has developed in parallel but largely done so independently. Perhaps one reason for this schism is the lack of a common measurement method between the two fields. The SVO Slider Measure could act as a bridge to connect these two related but estranged research streams.

One drawback of the SVO Slider Measure is that it does not use a symmetric set of allocation options around the entire ring. As a result the angular boundaries used for determining which SVO category a person is assigned (when reducing data from the ratio level to the nominal level of measurement) are not at intuitive locations. For example, a perfect altruist is represented by an angle of 61.39°, and not 90°. This asymmetry is a consequence of the measure only using a subset of possible items rather than using items allocated symmetrically over the whole ring. However, while efficiency is improved, the measure's validity is unaffected by this asymmetry. It would be possible to extend the Slider Measure in such a way that it would have a symmetric set of items and thus have a rotationally symmetric convex hull of possible scores. This extended measure certainly would be more aesthetically pleasing and would have intuitive angles as boundaries between the categories. However, this extended measure would require about five times as many primary items (6 vs. 28) and would likely not yield significantly better estimations of DM's social preferences. A second related drawback of the new Slider Measure is that it does not accommodate DMs with atypical social preferences (e.g. a masochistic DM– someone with the preference to minimize his own payoff, wholly indifferent to the payoff of the other). In situations where destructive kinds of social preferences (e.g. vengefulness, rage, or spite) are of interest, the SVO Slider Measure in its current form is also likely an inadequate tool. One could imagine an extended version of the Slider Measure that spanned a greater portion of the self-other allocation plane, but such a scale has not been developed nor normed. Furthermore, the Slider Measure can not readily be used for assessing the social preferences of people unexperienced with numeric representations. The method in its original form is therefore likely not suitable for studying the development of social preferences in young children, for instance.

2.5.9 Summary of SVO measure evaluations

Table 2.4: Summary of SVO measure evaluations

SVO Measure	Psychometric Properties	Output Resolution	Efficiency	Special Features
Altruism Scale	–	0	0	No numerical requirements
Triple-Dominance Measure	+	–	+	n/a
Rank correlation technique	n/a	–	–	Rank ordering of SVOs
Utility measurement	n/a	+	–	Flexibility
Indifference curve assessment	n/a	n/a	–	Flexibility
Social Behavior Scale	0	–	+	No numerical requirements
Ring Measure	+	–	–	Assessment of pathological SVOs
Circle test	n/a	+	+	Brevity
Regression and Clustering	0	–	–	n/a
Sphere Measure	n/a	–	–	n/a
Slider Measure	+	+	+	Transitivity check and rank ordering of SVOs

Table 2.4 shows an overview of the different SVO measurement methods discussed in this paper. The overview is supplemented with information about the measures' performance according to our predefined set of criteria. In the table, minus signs (-) indicate unsatisfactory performance, zeros (0) indicate satisfactory or medium performance, and plus signs (+) indicate good performance. If no or insufficient information is available to judge about a measure with respect to a certain criterion, this is indicated by a "not-available / not-applicable" sign (n/a). Regarding special features, the sign indicates a lack of particular or noteworthy comparative advantages. It is also used for evaluating output resolution of the indifference curve assessment technique, since this method produces visual output the quantification of which is possible but would require further complex computation. We are aware that the assignments of performance indications in this table are subjective to a certain degree. However, the information in this table should not be regarded as a substitution of the detailed measure discussions provided throughout this paper. Rather, it is intended to help the reader quickly assess the different measures' relative strengths and weaknesses at one glance. Also, special features of the measures are highlighted to facilitate choosing a method which is best suited for addressing a particular research question or employing a particular experimental design. However, as a general rule, we strongly suggest to use methods that produce continuous output whenever possible. This way, it is not only ensured that the

SVO construct is measured as it is theorized, but also that statistical power is not unnecessarily diminished, undermining evidence for an important individual difference.

2.6 Discussion

The arc of scientific knowledge is bound by our ability to measure things. This paper is about measuring social preferences, a fundamental concept in the social sciences. We have described the concept of SVO and discussed how this construct's theory has been shaped by measurement. Further, we have provided an overview of different ways social preferences have been measured to date across a variety of different disciplines and highlighted the strengths and weaknesses of existent measures. We have also discussed a new measure of social preferences called the SVO Slider Measure that overcomes many of the limitations of previous measures and aims to bridge different research streams by establishing a common language for both theory and testing.

Social preferences are critical to understanding how interrelated DMs allocate scarce resources among themselves and others. The postulate of narrow self interest is a point conjecture, namely that *all* DMs have *exactly zero* interest in the outcomes of other people and only try to maximize their own payoffs. Although this is a useful baseline assumption in that it facilitates tractable models with precise predictions, and in many cases works remarkably well as an "as if" model (Erev & Rapoport, 1998) of decision making, there are numerous examples where it fails to account for, or even roughly approximate, real DM's choice behavior. Real people's preferences are often much richer, more nuanced, and complex than narrow self interest (see, for instance, Camerer & Fehr, 2006). Although simplifying assumptions are useful as starting points for model development, and this conjecture can serve as a very useful starting point, descriptive accuracy and theoretical insight are better supported by the development of empirically accurate descriptions of people's real preferences and motivations. High-resolution measurement methods can serve to provide rich data that can be brought to bear on debates of human motivations, which are fundamental to understanding and predicting behavior in a wide variety of social settings. For example, knowing DM's individual preferences for prosocial outcomes can explain, in part, peoples' willingness to cooperate in social dilemmas (Balliet et al., 2009; Murphy & Ackermann, 2014b; Murphy et al., 2011).

Our review of the literature highlights that social preferences is a rich theoretical construct that can be measured in a variety of different ways. Moreover this construct is of great interest across disciplines in the social sciences. Currently the 9-Item Triple-Dominance Measure is the most popular method for measuring social preferences and it yields at best a nominal level of measurement, often then in practice reduced further to a simple binary result (prosocial vs. individualist). This measurement method constraints thinking and theorizing about social preferences and can hamper the development of better theories to account for how people make tradeoffs when outcomes are interdependent. Paraphrasing Maslow,[17] if the only tool you have is a hammer, then everything looks like a nail. Along the same lines, if the only measurement method one has for social preferences yields a categorical outcome (prosocial or individualistic), then thinking about social preferences veers toward thinking in terms of either/or. This binary approach to contemplating individual differences and preferences is profoundly limiting. First it

[17] Maslow's (1966, p. 15) well known quotation: "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail."

limits statistical power, likely contributing to the file drawer problem (Cohen, 1983; Rosenthal, 1979), which undermines our understanding of the importance of non-selfish preferences in human behavior. Secondly, it limits our ability to work with this valuable theoretical construct in a continuous way. The misfit between the theoretical conceptualization of a continuous individual difference and the predominant measurement method (which in standard practice is dichotomous), yields theories and experiments which tend to be binary when the reality is continuous. Simply put, we all can do better.

Moreover, any static point conjecture about social preferences is inadequate, not only in accounting for different people having different tastes but also in addressing how these preferences change for a person in different situations and contexts, and with the availability of new information. The dynamics of how people's preferences change (Murphy & Ackermann, 2014b), and what factors affect interdependent DMs' willingness to make different tradeoffs (Ackermann, Fleiss, & Murphy, 2014), are of central importance to unraveling the roots of cooperation and conflict (Pennisi, 2005). However, the detection of gradual changes in a person's concern for the well-being of others is impossible with methods that are only able to detect categorical shifts. How social preferences are malleable and reactive is an important and deep question, and efforts to address it empirically require high fidelity measurement methods.

The notion that a DM's utility is not exclusively a function of his own material well being, but is also affected by the well being of others (No man is an island…) is not a new idea. Edgeworth explicitly postulated this notion and anticipated a wide range of social preferences along a continuum. A substantial body of evidence has been built (e.g. Cameron et al., 1998; De Dreu & Boles, 1998; Declerck & Bogaert, 2008; Eisenberger et al., 1992; Joireman et al., 2004; Kanagaretnam et al., 2009; Kuhlman & Marshello, 1975b,a; Roch & Samuelson, 1997; Van Lange et al., 2007; Van Lange & Visser, 1999) showing the pervasiveness and importance of social preferences and the descriptive inadequacy of narrow self interest. A current challenge is to transcend *Homo economicus* by quantifying Edgeworth's fraction by using valid, reliable and efficient methods to measure the degree of entanglement in DM's utilities, and thus constructively expand theories of social decision making that can accommodate the richness and dynamics of real people's social preferences.

3 Measuring Social Value Orientation

Abstract[1]
Narrow self-interest is often used as a simplifying assumption when studying people making decisions in social contexts. Nonetheless, people exhibit a wide range of different motivations when choosing unilaterally among interdependent outcomes. Measuring the magnitude of the concern people have for others, sometimes called Social Value Orientation (SVO), has been an interest of many social scientists for decades and several different measurement methods have been developed so far. Here we introduce a new measure of SVO that has several advantages over existent methods. A detailed description of the new measurement method is presented, along with norming data that provides evidence of its solid psychometric properties. We conclude with a brief discussion of the research streams that would benefit from a more sensitive and higher resolution measure of SVO, and extend an invitation to others to use this new measure which is freely available.

3.1 Introduction

The assumption of narrow self-interest is central to rational choice theory. The postulate is that decision makers (DMs) are concerned about maximizing their own material gain, indifferent to the payoffs of other DMs around them. This is a simplifying assumption that yields a powerful framework to predict and explain human decision making behavior across a wide variety of domains. However there are reliable counterexamples demonstrating that DMs' elicited preferences and choices are often influenced in part by the payoffs of other DMs, thus challenging what some have termed the *selfishness axiom* (Henrich et al., 2005).

Studies on the motivations that underlie interdependent decision behavior have a long history and these motivations have been referred to by a variety of names, including: social preferences,

[1] This chapter is an edited version of the following paper.
Murphy, R.O., Ackermann, K.A., and Handgraaf, M.J.J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771-781.

social motives, other regarding preferences, welfare tradeoff ratios, and Social Value Orientation (SVO). For consistency, we refer to this construct as SVO for the remainder of this paper. Within the SVO framework it is assumed that people vary in their motivations or goals when evaluating different resource allocations between themselves and another person. As examples, a DM may endeavor to maximize her own payoff (individualistic), maximize (competitive) or minimize (inequality averse) the difference between her own and the other person's payoff, or maximize joint payoffs (prosocial). It is worth noting, however, that the assumption of narrow self-interest is itself a particular SVO, namely a perfectly individualistic orientation. Moreover considering a spectrum of different SVOs is not a challenge to rational choice theory *per se*, but rather the extension of a postulate in an effort to increase the theory's psychological realism and descriptive accuracy.

SVO has been found to affect cognitions and account for behavior across a range of interpersonal decision making contexts, specifically in the domain of negotiation settings (De Dreu & Boles, 1998) and resource dilemmas (Roch et al., 2000; Roch & Samuelson, 1997; Samuleson, 1993). SVO has also been identified as a covariate, interacting with different emotional states and influencing the propensity to cooperate (Zeelenberg et al., 2008). SVOs have even been identified in non-human primates (Burkart et al., 2007), indicating that some other species also show intrinsic preferences for prosocial behavior.

In order to use the full explanatory power of SVO as a psychological construct, it is necessary to be able to measure it efficiently, reliably and validly. Several different measurement methods for quantifying variations in SVO across individuals have been developed (for overviews, see McClintock & Van Avermaet, 1982; Au & Kwong, 2004; Murphy & Ackermann, 2014a). Although the use of existent measures has produced a wealth of findings even with categorical approaches (see, for instance De Dreu & Boles, 1998; Kuhlman & Marshello, 1975b,a; Van Lange & Visser, 1999), these measures have substantial limitations. For instance, some measures yield only low-resolution output that lack sensitivity to important individual differences, providing at best a nominal categorization (e.g. the Triple-Dominance Measure, see Van Lange, Otten, De Bruin, & Joireman, 1997). Other measures are highly inefficient and often fail to produce consistent results for a substantial proportion of subjects (e.g. the Ring Measure, see Liebrand, 1984). Yet still other methods require substantial time and effort from a research subject in order to produce a score (e.g. Utility and Conjoint Measurement procedures, or Regression and Clustering techniques, see Wyer, 1969; Radzicki, 1976; Knight & Dubro, 1984, respectively). Moreover, none of these existent measures are explicitly designed to detect more nuanced motivations like inequality aversion. Specifically, disentangling the orientation of joint gain maximization from the motivation to minimize the difference between outcomes has not been explicitly addressed in previous measures. Although these two orientations are related in that they both indicate a deviation from individualism towards prosociality, they are substantially different motivations which should be differentiated both theoretically and operationally.

Furthermore, Social Value Orientation is a continuous construct as it corresponds to the quantity of how much a DM is willing to sacrifice in order to make another DM better off (or perhaps worse off). This quantification of interdependent utilities can be best represented on a continuous scale. Moreover, since the most commonly used SVO measures to date produce only categorical data, a substantial amount of information related to peoples' social preferences is being discarded and ignored. Consequently, the full explanatory power of SVO has not been used because of this unnecessary sacrifice of statistical power (see Cohen, 1983, for a discussion of the unfortunate practice of reducing continuous variables to categories).

In our view, a method for assessing SVO should yield high-resolution output which makes it sensitive to inter- and intra-individual differences and facilitate comparisons thereof, be easy to use, be efficient, be able to detect the most prevalent SVO individual differences, allow for an evaluation of rank orders of social preferences, and yield meaningful results for virtually all subjects. Amongst these criteria, we consider the demand for a high resolution measure which produces data on a continuous scale as crucial.

We introduce here a new measure of SVO which takes this conceptualization into account and allows for greater explanatory potential of SVO through increased statistical power while also meeting the afore mentioned psychometric criteria. This new method is referred to as the *SVO Slider Measure*. A detailed discussion of this new measure is provided, along with norming data and evidence of the new measure's strong psychometric properties.

3.2 The SVO Slider Measure

The SVO Slider Measure can be administered as a paper based choice task or as an online measure. The measure has six primary items with nine secondary (and optional) items. All of the items have the same general form. Each item is a resource allocation choice over a well-defined continuum of joint payoffs. For example, consider a DM choosing a value x between 50 and 100 inclusive. Her payoff would be x, whereas the other's payoff would be $150 - x$. The DM would indicate her allocation choice by marking a line at the point that defines her most preferred joint distribution (see item 5 in Figure 3.1, see also Table A.1). After the DM has marked her most preferred allocation, she would write the corresponding payoffs resulting from her choice to the right of the item. Although this step of writing the values is redundant, it serves to verify that the DM understood the choice task and the resulting allocations.

3.2.1 Primary SVO Slider items

The six primary Slider Measure items are shown in Figure 3.1. These six items were derived from the six lines that fully interconnect the four points corresponding to the most common idealized social orientations reported in the literature (altruistic, prosocial, individualistic, and competitive; see Figure 3.2). A DM evaluates each of the items sequentially and for each one indicates her most preferred joint distribution. The set of responses can then be scored to yield a single score for the DM, the rank order of her social preferences, and additionally contains a check for transitivity in her revealed preferences.

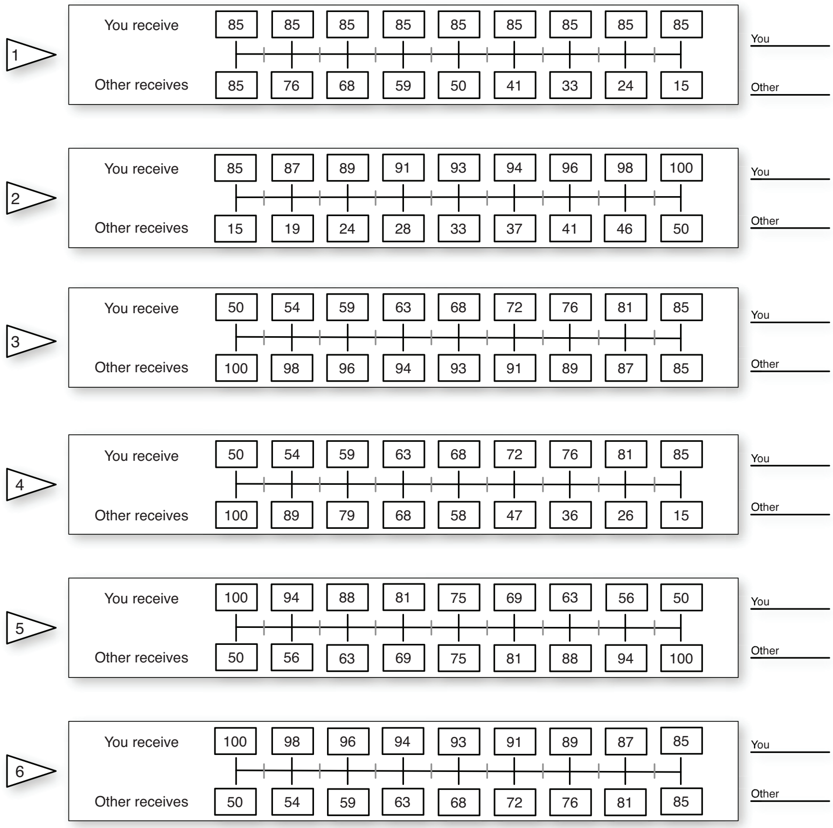


Figure 3.1: This shows the six primary SVO Slider items as seen by the participants.

There are several advantages with the SVO Slider Measure. First, the responses can be evaluated for comprehension (e.g. checking the correspondence between the mark on the distribution line and the written distribution values). Second, the responses can be evaluated for transitivity. Although SVO is a matter of subjective preferences, these preferences should conform to the elemental requirement of transitivity. Random responding on behalf of a subject would likely result in an intransitive set of responses. Third, the responses yield a full ranking of preferences over motivations. Fourth, the measure can be scored in a straightforward manner to yield a single index of SVO as follows. The mean allocation for self ($\bar{A}_s$) is computed as is the mean allocation for the other ($\bar{A}_o$). Then 50 is subtracted from each of these means in order to “shift” the base of the resulting angle to the center of the circle $(50, 50)$ rather than having its base start at the Cartesian origin. Finally, the inverse tangent of the ratio between these means is computed, resulting in a single index of a person’s SVO.

$$\mathrm{SVO}^\circ = \arctan\left(\frac{(\bar{A}_o - 50)}{(\bar{A}_s - 50)}\right) \tag{3.1}$$

This response format is highly sensitive to individual differences and yields an individual score at the ratio level of measurement. Assessing SVO in this way also facilitates parameterization and model assessment that is not possible with other existent measures. Nonetheless, reducing the high-resolution score to a nominal category may be desirable in some cases (e.g. to compare new results to previous studies), and the resulting SVO Slider angles can be transformed into corresponding categories with ease as follows.

If a person would choose the option which maximizes the allocation for the other in each of the six primary items, the resulting angle would be 61.39°, indicating perfect altruism. A prosocial DM with inequality aversion would yield an angle of 37.48°. A prosocial DM who endeavored to maximize joint gain (and is inequality tolerant) would yield an angle between 37.09° and 52.91°. The reason for this range is that this DM would be wholly indifferent across the entire SVO Slider item that has a slope of -1 (i.e. the item with endpoints 100, 50 and 50, 100) as it has a constant sum. A perfectly consistent individualist yields an angle between -7.82° and 7.82°. The reason for this range is that this particular DM would be wholly indifferent across the range of outcomes contained in the SVO Slider item that has an undefined slope (endpoints 85, 85 and 85, 15). A perfectly consistent competitor yields an angle of -16.26°.

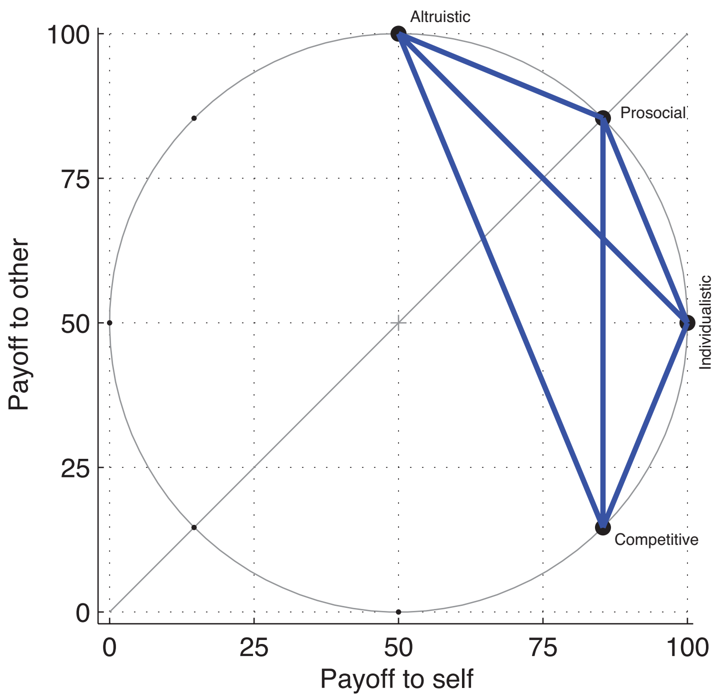


Figure 3.2: This figure shows where in the self/other allocation plane the six primary items are from the Slider Measure.

Given the angles that result from idealized SVO types, proper boundaries between categories can be derived by bisecting the respective adjacent ranges. Altruists would have an angle greater than 57.15°; prosocials would have angles between 22.45° and 57.15°; individualists would have angles between -12.04° and 22.45°; and competitive types would have an angle less than -12.04°.

As it can be seen, these boundaries are not at intuitive locations. The reason for this is that the Slider Measure only uses a subset of possible items from the allocation plane and these items are not symmetrically distributed around the whole of the ring. Because only an asymmetric set of items is used here, the resulting convex hull of possible scores is "squished" to the upper-right, relative to the midpoint of the ring. This characteristic does not adversely affect the validity of the measure.

3.2.2 Secondary SVO Slider items

There are nine secondary SVO Slider Measure items. This set of items is explicitly designed to disentangle the prosocial motivations of *joint maximization* from *inequality aversion*. The items are defined in the prosocial area of the self/other allocation plane and have approximately the same magnitude (ranging between 50 and 100 value units) as the six primary items. One noteworthy feature of these secondary items is that all of the distribution ranges intersect the diagonal line. This is an important feature of the set as points on the diagonal line correspond to perfectly equal allocations, i.e., those distributions that minimize inequality between the DM and the other person. A person motivated to minimize inequality would make allocations on or very near the 45° line. Conversely a person motivated to maximize joint gains would make allocations at the endpoints, as far from the diagonal as possible as it turns out, as these allocations maximize collective earnings. Previous measures of SVO have not been explicitly designed to make a differentiation between these two motivations. The nine items are shown in Figures **??** and A.1. An example of results from these items is discussed in Section 3.3.6.

3.2.3 Web-based SVO Slider Measure

In addition to being administrable as a paper-based measure, the Slider Measure has been programmed as an online research tool which can be freely used by any researcher to easily conduct SVO measurements with their own participants.[2] The online measure and supporting material, as well as the paper based versions of the new measure, can be found at:

http://vlab.ethz.ch/svo/SVO_Slider/

With the online SVO Slider Measure, items are presented in a random order. Subjects record their choices by moving a webpage slider input back and forth, changing the joint allocations until they find their most preferred joint outcome (see Figure 3.3 for a screen shot). The online items are dynamic and display information is updated in real time as the DM moves the slider over the option space. The choice procedure is the same for all of the items. After the subjects

[2]Computing results from the SVO Slider (checking for transitivity, establishing the ranking of preferences, and finding a subject's SVO angle) can be somewhat demanding and thus we have developed an analysis script that automates and simplifies this process. This script is available for download from the SVO website along with a detailed tutorial on its use. We also provide an Excel worksheet for researchers who are interested in quick and basic results.

have participated, the researcher is sent an email with the datafile attached; the datafile contains the subjects' identifying information, date/time stamp, item order, and all of the DMs' allocation choices.

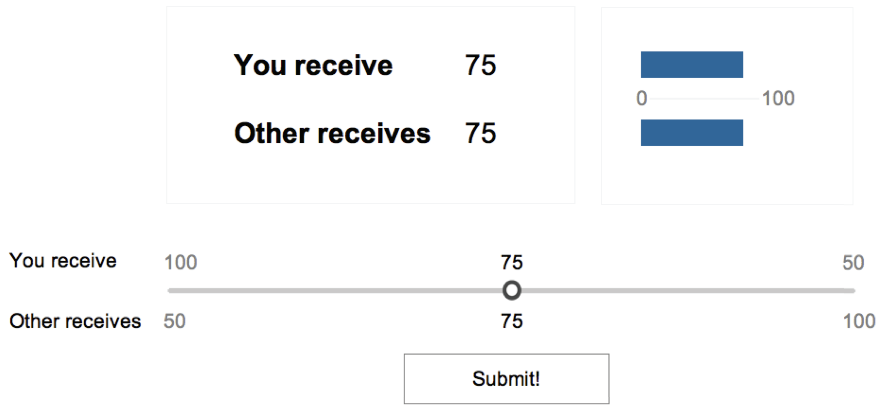


Figure 3.3: Online Slider Measure

3.3 Psychometric properties of the SVO Slider Measure

3.3.1 Slider Measure validation procedure

In order to assess the psychometric properties of the new SVO Slider Measure it was tested in tandem with the established and most commonly used measures of SVO; namely the Triple-Dominance Measure (see Van Lange et al., 1997) and the Ring Measure (Liebrand, 1984). Fifty-six individuals from various majors were recruited to participate in a multi-part "decision making study" at a European university. Participation was voluntary and no deception was used in this research. Participants were guaranteed strict confidentiality for all of their choices. The choices in the experiment were made incentive compatible by means of a lottery– for each experimental session four participants were randomly selected after making their choices and for each selected person one of their allocation decisions was implemented (i.e. their allocation choice was carried out such that they received some chosen payoff, as well as did some other randomly selected person, according to their actual choice). For all research sessions, participants were reminded that their decisions were private and that there was a real chance that their choices would have a pecuniary effect upon themselves and some other person if they happened to be selected by lottery. DMs selected by lottery were paid privately in cash within a week of their participation. Each unit of value in the experiment corresponded to 50 Swiss cents and the average earnings were 81.70 Swiss francs (US$77) per paid participant.

Three research sessions were run, with one week separating the sessions. Each research session required fewer than 15 minutes to conduct and used paper-based methods. In the first session, participants completed the 9-item Triple-Dominance Measure and the 24-item Ring Measure. In the second session, participants completed the Slider Measure and the Triple-Dominance Measure. In the third session, participants completed the Ring Measure and the Slider Measure. All of the measures used standardized values between 0 and 100. This research design allowed us to assess the test-retest reliability of the Triple-Dominance Measure, the Ring Measure, and the Slider Measure. It also allowed us to compute the associations between the different measures and establish norming data and convergent validity for the new SVO Slider Measure.

3.3.2 Results

Table 3.1 shows the percentage of individuals that were assigned to each of the different SVO categories by the different measurement methods, ordered by experimental session. Across all measurement methods there is a clear majority type, namely prosocial, occurring about 59% of the time. Individualist is less common, but found about 35% of the time. Competitive and unclassifiable types complete the remainder of the sample representing about 3-4% each.

Table 3.1: The percentage of individuals that were assigned to each of the different SVO categories by the different measurement methods (TD- Triple Dominance, RM- Ring Measure, SM- Slider Measure), ordered by experimental session.

	Session 1		Session 2		Session 3		
	TD	*RM*	*TD*	*SM*	*RM*	*SM*	*Grand mean*
Prosocial	59	53	61	58	58	64	59
Individualistic	21	45	32	39	36	34	35
Competitive	2	2	3	3	4	2	3
Unclassifiable	18	0	3	0	2	0	4

3.3.3 Reliability

Triple-Dominance Measure test-retest reliability

Table 3.2: A cross tabulation showing the frequency of categorization from test-retest between session 1 (S 1) and session 2 (S 2) for the Triple-Dominance Measure.

		S 2			
		Prosocial	Individualistic	Competitive	Unclassifiable
S 1	Prosocial	23	1	0	1
	Individualistic	2	8	0	0
	Competitive	0	1	0	0
	Unclassifiable	3	5	1	1

Forty-six participants completed both sessions 1 and 2. Of those, 32 $(23 + 8 + 0 + 1)$ were categorized in the same SVO category each time by the Triple-Dominance Measure, yielding a consistency of 70% (Goodman and Kruskal's gamma[3] = 0.391).

Ring Measure test-retest reliability

Forty-four participants completed both sessions 1 and 3. Of those, 30 $(18+12+0+0)$ were categorized into the same SVO category each time by the Ring Measure, yielding a consistency of 68%. Further the correlation between the resulting angles from the test-retest of the ring measure was $r = 0.599$.[4]

Table 3.3: A cross tabulation showing the frequency of categorization from test-retest between session 1 (S 1) and session 3 (S 3) for the Ring Measure.

		S 3			
		Prosocial	Individualistic	Competitive	Unclassifiable
S 1	Prosocial	18	3	0	0
	Individualistic	8	12	1	0
	Competitive	0	1	0	0
	Unclassifiable	0	0	0	0

Slider Measure test-retest reliability

Table 3.4: A cross tabulation showing the frequency of categorization from test-retest between session 2 (S 2) and session 3 (S 3) for the primary SVO Slider items.

		S 3			
		Prosocial	Individualistic	Competitive	Unclassifiable
S 2	Prosocial	25	1	0	0
	Individualistic	3	15	0	0
	Competitive	0	1	1	0
	Unclassifiable	0	0	0	0

Forty-six participants completed both sessions 2 and 3. Of those, 41 were categorized in the same SVO category each time by the Slider Measure, yielding a consistency of 89%. Further the correlation between the resulting angles from the test-retest SVO Slider Measure was $r = 0.915$.

[3] As the Triple-Dominance Measure yields a nominal level variable with more than two categories, a Pearson product moment correlation coefficient is not an appropriate statistic for assessing its reliability, hence the non-parametric Gamma statistic is used as an index of test-retest association.

[4] In order to verify the robustness of these results, non-parametric statistics of association were also conducted in parallel with Pearson's r. The non-parametric statistics yielded the same pattern of results.

3.3.4 Validity

Convergent validity: Categorical agreement

Across research sessions, the Triple-Dominance Measure and the Ring Measure categorized the same subjects into the same SVO category 67% of the time. The Triple-Dominance Measure and the Slider Measure categorized the same subjects in the same SVO category 74% of the time. The Ring Measure and the Slider Measure categorized the same subjects in the same SVO category 75% of the time.

Convergent validity: Correlational agreement

The Ring Measure and Slider Measure both produce continuous results (in the form of angles within the self/other allocation plane), and these results are amenable to computing correlation coefficients across different measures. Table 3.5 displays these correlation coefficients, showing both the test-retest reliability of the Ring Measure ($r = 0.599$) and Slider Measure ($r = 0.915$), as well as the correlations between SVO angles across the different measurement methods.

Table 3.5: The correlation coefficients between the different sessions and methods. These values show both the test-retest reliabilities, as well as the cross method correlations (in gray) which address convergent validity.

	RM-1	RM-3	SM-2	SM-3
RM-1	1	-	-	-
RM-3	0.599	1	-	-
SM-2	0.724	0.536	1	-
SM-3	0.680	0.641	0.915	1

The results show that the Slider Measure correlates as well (if not better) with the Ring Measure as the Ring Measure does with itself across retests. This is strong evidence that the methods are measuring the same thing and further it demonstrates that the Slider Measure is more reliable than the Ring Measure (the mean correlation between the different methods is $r = 0.649$ whereas the test-retest correlation for the Ring Measure is only $r = 0.599$).

Predictive validity

In order to evaluate the Slider Measure's predictive validity, a second study was conducted where different subjects ($N = 100$) first completed the Slider Measure and then played a one-shot anonymous Prisoner's Dilemma game. Identical to the first study, this study used monetary incentives determined by a lottery. We find a moderate and statistically significant point-biserial correlation (r = 0.239) between the subjects' SVO angles and their choices in the Prisoner's Dilemma, indicating a positive relation between SVO angle and cooperation as would be expected. These results are consistent in direction and magnitude with other findings from incentive compatible choice tasks in social dilemmas and measures of SVO (Balliet et al., 2009).

3.3.5 Additional results

As noted before, one advantage of the Slider Measure is its high resolution, as it yields a ratio level of measurement. Previous measures of SVO produce output as a simple categorization, which is a limitation. Conversely, producing a ratio level variable, the distribution of observed SVO angles can be plotted and the density of different orientations can be estimated. Figure 3.4 shows this distribution.

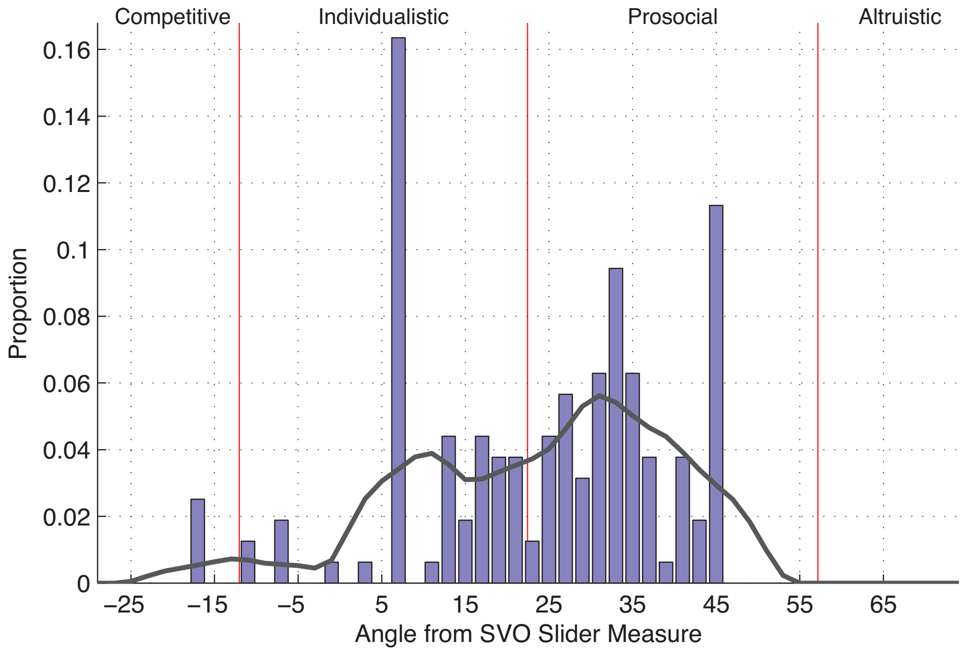


Figure 3.4: The distribution of SVO scores from the Slider Measure as represented by angles. The dark line is a smoothed kernel density estimation.

A LOESS (Cleveland & Devlin, 1988) smoothed kernel density estimation was made of the distribution of SVO scores to provide some general idea of its shape. We find a multimodal distribution of SVO types in our sample. The largest clustering is in the prosocial region shifted slightly to the left (toward individualistic). The second clustering is in the individualistic region and is shifted to the right (toward prosocial). Within this region is the most common SVO score of 7.82° which corresponds to perfectly individualistic choices. The density function trails off to the left, denoting only a few competitive types. As can be seen in the figure, there is substantial variance in the subjects' SVO angles, beyond what a nominal level categorization would indicate. Moreover, the observed variance supports the assertion that a sensitive SVO measure which produces reliable high resolution data on a continuous SVO scale is valuable in that it can capture the rich gradation of social preferences.

As already noted, the transitivity of responses can be assessed with the Slider Measure. We found that 98% of our subjects produced completely transitive sets of social preference choices. This finding stands in stark contrast to the consistency results from the Ring Measure where only 55% of the same subjects produced internally consistent results. This would indicate that almost all subjects have well defined social preferences but that the Ring Measure is not particularly well suited to measure them.

Table 3.6: The full rank orderings of social preferences from the SVO Slider Measure across sessions. Note that 25% of the decision makers were indifferent between Individualistic and Prosocial allocations when their inferred preferences are reduced to ranks.

First preference	Second preference	Third preference	Least preferred	Percent
Prosocial	Individualistic	Altruistic	Competitive	27%
Prosocial	Altruistic	Individualistic	Competitive	25%
Prosocial	Individualistic	Competitive	Altruistic	13%
(Individualistic	Prosocial)	Competitive	Altruistic	25%
Individualistic	Competitive	Prosocial	Altruistic	4%
Individualistic	Prosocial	Altruistic	Competitive	2%
Competitive	Individualistic	Prosocial	Altruistic	4%

As an additional feature, the full ranking of people's social preferences can be obtained from evaluating the six primary items of the SVO Slider Measure (see Table 3.6). These kind of complete ordinal results are not possible with other common SVO measurement methods. Moreover, information about a DM's *least* preferred allocation may be useful to know when measuring individual differences.

3.3.6 Separating the prosocial preferences of inequality aversion and joint maximization

The secondary items from the Slider Measure are designed to differentiate between two different prosocial motivations: inequality aversion and joint maximization. As prosocial behavior can arise from both of these underlying motivations, we demonstrate here how to disentangle these motivations using the secondary SVO Slider items.

In order to identify prosocial DM's underlying motivations, two mean difference scores were computed for each prosocial subject from their allocation choices on the secondary Slider Measure items. The first difference score was defined as the average normalized distance between the subject's allocations and the particular allocations that would maximize equality. For example, if a DM always chose allocations that were on the diagonal line (see Figure A.1), her mean difference score from idealized inequality aversion would be zero, indicating perfect consistency with the preference of inequality aversion. A second difference score was computed for each participant that was defined as the mean distance between her selected allocations and the particular allocations that maximized joint payoffs for that item. If the mean difference for this second index was zero, it indicates that the DM's allocation choices are perfectly consistent with joint maximization. These values can be meaningfully aggregated into a single index by computing the ratio of the first difference score divided by the sum of both difference scores. The result is an index ranging between 0 (indicating allocation choices perfectly consistent with inequality aversion) and 1 (indicating allocation choices perfectly consistent with a preference for joint gain maximization).

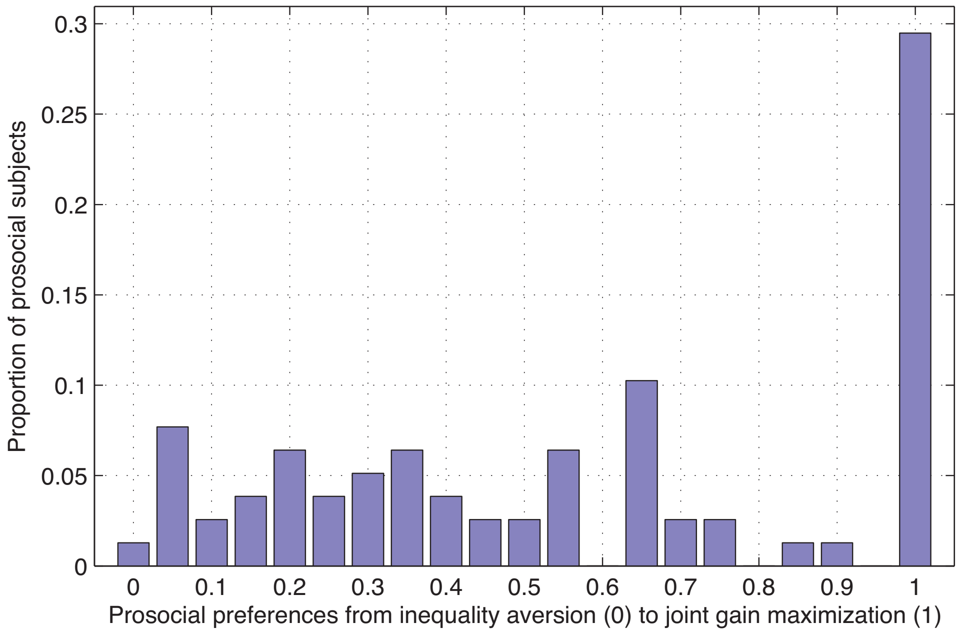


Figure 3.5: The distribution of prosocial preferences, ranging from perfect inequality aversion to perfect joint gain maximization. The most common preference is for joint gain maximization (29%) but there is substantial variance in DM's prosocial preferences.

Results were obtained from the 79 DMs who made consistently prosocial allocations in the primary and secondary items across both studies. The distribution of individuals' inequality aversion / joint gain maximization indices is shown in Figure 3.5. Several results are noteworthy. First, this distribution suggests that prosocial DMs are not homogeneous with respect to their more nuanced prosocial preferences. Some people are striving for maximizing joint gain, whereas others seem to be, at least somewhat, sensitive to equality between payoffs. Second, while the modal preference is for joint gain maximization, a slim majority of DMs are actually closer to inequality aversion. This distribution is both non-uniform and non-skewed with the mean and median at 0.571. Splitting the sample at 0.5, 54% of DMs would be categorized as inequality averse whereas 45% would be better described as joint gain maximizers (one person is exactly at the midpoint of 0.5 and is not categorized). Lastly, the shape of the distribution suggests that there is greater conformity in how joint maximizing DMs make allocation choices compared to how inequality averse DMs allocate resources.

3.4 Discussion

Social preferences are of fundamental importance in understanding interdependent decision making behavior among people. In order to quantify the degree to which people care about outcomes for others, it is necessary to develop reliable measurement methods to assess this construct. Consistent with this goal, we report here on the development of a new measure of SVO and demonstrate that it is quick, efficient, easy to implement, has very good psychometric properties, yields scores for individuals at the ratio level, and facilitates comparison to other measures. The advent of a high-resolution measure of SVO opens opportunities for different research streams to use social preferences as a dependent variable. These types of studies could address questions regarding how context, information, experience, and framing affect peoples' propensities to make tradeoffs in resources between themselves and others. These lines of research could also answer larger questions like under what conditions is the selfishness axiom a good approximation for explaining human behavior, and when is it insufficient, or even grossly inaccurate. This new measurement method can serve as a bridge between perspectives informed by *Homo economicus* and those perspectives which take descriptive accuracy as a starting point.

In a broader sense, we would like to encourage scientists interested in human decision making to develop higher resolution measures. Having more sensitive and reliable measurement methods is critical for the detection of subtle yet important effects that may result from changes in context and information. Therefore, we think that the technique employed with the Slider Measure could also be useful in the development of related methods for assessing other individual differences, such as risk perception (Ganzach, 2000; Ganzach et al., 2008), or temporal discounting (e.g. Stevenson, 1992). In general, we believe that allowing subjects to explore a range of well-ordered and intuitive options facilitates not only the revelation of preferences, but also the discovery and unencumbered expression of those preferences.

4 Modeling Social Value Orientation

Abstract[1]

Over the past decades, the number of utility models representing social preferences has increased substantially. However, no attempt has been made so far to evaluate the most prominent models comparatively on the individual level. The present paper contributes to filling this gap by taking several steps. First, the most prominent models of distributive social preferences are reviewed. Next, the models are fit to empirical data on the individual level by means of three different fitting procedures, one of which is maximum likelihood estimation. Importantly, the data to be fit resulted from subjects completing a measure of social preferences in a non-strategic context, such that confounding effects of beliefs on choice behavior can be ruled out. The different models are evaluated and compared to each other in a horse-race manner according to both the criterion of explanatory power and psychological interpretability. In addition to evaluating the models' explanatory power, we also report on the distribution of individual best fitting parameter values per model. We find substantial differences in the degrees to which the models can account for individual choice data, and in how well best fitting parameter values can be interpreted psychologically.

4.1 Introduction

The validity of the assumption of narrow self-interest as the unique motivation that drives human behavior has been contemplated and debated for decades in the modern social sciences. To date, evidence for deviations from selfishness have accumulated to an extent that can no longer be ignored. As a result, more recent models of human decision making in social contexts often include terms that account for non-selfish preferences. Preferences over own-other resource allocations have been given various labels in different scientific disciplines, such as other-regarding preferences, social preferences, welfare-tradeoff ratios, or social value orientation. For simplicity of abbreviation, we will predominantly refer to the concept as social value orientation

[1]This chapter is an edited version of the following paper.
Ackermann, K.A. and Murphy, R.O. (2013). Modeling social value orientation. Manuscript in preparation.

(SVO) for the remainder of this paper. To date, a variety of different SVO models have been developed, tested, and used for predicting behavior in strategic situations with varying degrees of success. However, we identify three problems with earlier validation procedures in the domain of SVO modeling research. First, models of social preferences are agent based, i.e. utility functions that are supposed to account for how decision makers trade off own material well-being against collective material well-being. Therefore, model parameter values should be estimated on the individual level. Past research, though, has been mainly focused on estimating parameter value distributions on the aggregate level (for qualified counterexamples, see Andreoni & Miller, 2002; Blanco et al., 2011; Daruvala, 2010; Fisman et al., 2007; Graf et al., 2012). That is, often behavior has been observed on the aggregate level and then parameter value distributions that can account for the aggregate behavioral pattern are inferred. This is problematic since the inferred parameter values do not necessarily match the ones obtained when measured directly on the individual level (see, for instance, Blanco et al., 2011). Second, parameter values of simple outcome-based models are often estimated on the basis of behavior observed in strategic situations. This is problematic because behavior in a strategic situation is a function of both preferences and beliefs. Consequently, if parameter values for outcome-based SVO models are estimated from behavior observed in contexts where the decision makers are interdependent, these estimates will necessarily be confounded with the effects the decision makers' beliefs had on the choices they made. Third, in light of the multitude of outcome-based other regarding preference models that have been proposed to date across different scientific disciplines, surprisingly little effort has been dedicated to evaluating the models in a comparative manner.

We address these issues by providing a comparison of the most prominent outcome-based social preference models on the individual level on the basis of behavior observed in a non-strategic situation. Hence, intention-based models that take into account the decision makers' beliefs are excluded from the present analysis. Furthermore, we restrict ourselves to the analysis of the simplest situation possible, where there is only one person making decisions about how to allocate monetary resources between the self and one other person. The different models are compared in terms of their goodness of fit to the data as well as the psychological interpretability of individual best fitting parameter values. The reason why we focus on the simplest of all possible settings where social preferences can be expressed is twofold. First, over the last decades the number of proposed simple outcome-based SVO models has been constantly increasing to an extent which makes it difficult for researchers to overview the literature. Furthermore, the lack of a comparative evaluation of these models makes it difficult for researchers to select one of these models, since there is little information available about which models are most useful. Second, the purpose of studying SVO models is not only to understand human decision making in non-strategic settings, but ultimately also to predict and understand behavior in situations where the decision makers are interdependent. In such situations, decision makers' preferences over own-other outcome allocations are not only a function of the outcome combinations per se, but also of the beliefs (or knowledge) the decision makers have about the past and future behavior and/or intentions of the interdependent other. Such belief- or intention-based models have indeed already been proposed and used (e.g. Levine, 1998; Rabin, 1993; Geanakoplos et al., 1989; Falk & Fischbacher, 2006; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004). However, we think that prior to making further attempts to understand and predict behavior in strategic situations, we have to clarify the question which outcome-based SVO model(s) is (are) most useful for explaining behavior in non-strategic settings. More complex belief- or intention-based models could then build upon the most useful simple outcome-based models. A model supposed to explain behavior in complex situations should pass the test of explaining behavior in simple situations as a minimal requirement. We do not claim that currently used belief-based SVO models are suboptimal or not useful, but we would like to provide an evaluation of simpler models the most successful of which future research could benefit from building upon when modeling behavior in more complex situations.

The paper proceeds as follows. Section two provides a brief overview of the most prominent outcome-based SVO models that have been developed in different scientific disciplines over the last decades. Section three describes the methods applied for comparing the different models. In section four, results from the model fitting procedures are reported and summarized, and section five concludes.

4.2 Existing models of SVO

The assumption that individuals do not only consider their own material well-being, but also take into account the well-being of others when making decisions in social contexts refers back -at least- to 1881, when Edgeworth wrote the following:

> For between the two extremes Pure Egoistic and Pure Universalistic there may be an indefinite number of impure methods; wherein the happiness of others as compared by the agent (in a calm moment) with his own, neither counts for nothing, nor yet 'counts for one,' but counts for a fraction.

The rationale underlying this quote captures at least three ideas: First, people value their own welfare. Second, people also take into account the welfare of others in relation to the own. And third, people vary in the extent to which they value other people's welfare. Taken together, these ideas can be represented by means of a utility function of the following form:

$$U_{(\pi_s,\pi_o)} = \pi_s + \alpha\pi_o \tag{4.1}$$

where π_s denotes the outcome for the self, π_o the outcome for the other, and α the parameter which determines the extent to which the other person's outcome is valued. Since Edgeworth assumed that the extremes of pure egoism and altruism exist, but that for many people the "happiness" of others may neither count for nothing, nor count for one, this would result in a constraint such that $0 \leq a \leq 1$. Although Edgeworth did not explicitly intend to state a model of social preferences, his assumption can easily be translated into one. However, this hypothetical model only considers values of α which are zero or positive, meaning that people always derive positive utility from another person's benefit or none at all. However, one could also think of types of people who do not appreciate the well-being of others, but rather depreciate it. For example, a person may like to be relatively better off than others, hence disapproving other peoples' benefits, which seem to discount the own benefit in their view. A model which incorporates this kind of social preference would have to relax the constraint of parameter α from the former model such that it can also take on negative values. Such a model was stated by Sawyer (1966), who was concerned with finding a way to measure this parameter α. Sawyer's model has the exact same functional form as the model expressed by equation 4.1, but assumes that $-1 \leq \alpha \leq 1$. Sawyer let subjects rank and rate different own-other-outcome combinations for estimating corresponding α values and found that the mean values for 122 subjects were .45 when the other person was labeled as friend, .12 when labeled as stranger, and −.18 when labeled as antagonist.

A further model was then proposed by Messick & Thorngate (1967), who - on the basis of data they had collected - stated that decision makers seem to pay attention to differences in outcomes:

$$U_{(\pi_s,\pi_o)} = f(\pi_s) + g(\pi_s - \pi_o). \tag{4.2}$$

However, the model was discarded shortly after by Messick & McClintock (1968) who found that the majority of their subjects who had made various choices between binary own-other outcome allocations exhibited at least one decision that could not be explained by the stated model. It is interesting, in hindsight, to see that one of the first models of other-regarding preferences already contained a term considering differences in payoffs, while inequality aversion as one type of motivation concerning differences in payoffs has gained excessive attention more recently.

A next step in the history of social preference modeling was taken by Wyer (1969), who assumed that the choices people make when confronted with particular own-other payoff allocations may be predicted by the following linear equation:

$$U_{(\pi_s,\pi_o)} = w_1 \mathbb{1}\pi_s + w_2(1-\mathbb{1})\pi_s + w_3\pi_o. \tag{4.3}$$

Several attributes of this model in comparison with the former ones' are worth emphasizing. In this model, $\mathbb{1}$ is a step function that takes on the value of 1 if the outcome to the self is a gain (i.e. if $\pi_s > 0$) and 0 if the outcome to the self is a loss (i.e. if $\pi_s < 0$). In stating two different parameters w_1 or w_2 attached to π_s, depending on whether π_s is a gain or a loss, Wyer assumed that gains and losses –of the same absolute magnitude– may have different impacts on utility. Ten years later, this foresighted assumption got confirmed such that w_2 turned out to be considerably greater than w_1 in general (Kahneman & Tversky, 1979). Besides the introduction of parameters attached to own outcomes in the first place, and the conditionality of these parameters upon the values of π_s, Wyer's model differs from previously discussed ones in that it does not put a constraint on the values of the parameters w_1, w_2, and w_3. In his study, Wyer presented subjects with several own-other payoff allocations, each of which had to be rated by the subjects in terms of desirability. For fitting the model to the data, a least squares procedure was employed and –averaged over subjects and experimental conditions– the following parameter values were obtained: $w_1 = .708$, $w_2 = .798$, $w_3 = -.334$. The fact that the value of w_2 is greater than the one of w_1 can be interpreted in terms of loss aversion and is in line with what we would expect. What is somewhat surprising about these results is that the average value of parameter w_3 was negative and of considerable magnitude, which indicates that subjects –on average– derive negative utility from the other person's outcome. Later research has shown, though, that motivational orientations such as to decrease the other's payoff or to maximize relative own gain are relatively infrequently observed[2]. However, Wyer's result regarding the mean value of w_3 may be due to the structure of the payoff allocations that subjects rated with respect to desirability, since outcomes also included negative payoffs. If many or even most of the payoff allocations presented to the subjects had a payoff structure similar to a zero-sum situation where one's gain is the other one's loss and vice versa, then a value of w_3 similar to the one obtained by Wyer would be expected.

Quite a different approach to the modeling of SVO was undertaken by Radzicki (1976). He introduced a conjoint measurement technique for the assessment of SVO the result of which is a representation of different SVO types by different utility functions. Radzicki let subjects rank several own-other payoff allocations according to their desirability and applied linear programming methods in order to find the utility functions which best fit these rank orders. The best fit was defined as the maximum Spearman's rank correlation between the empirical rank orders and the rank orders predicted by the respective utility functions. The simplest utility function Radzicki took into consideration was of the following form:

$$U_{(\pi_s,\pi_o)} = \alpha\pi_s + \beta\pi_o. \tag{4.4}$$

It turned out that this simple linear model fit perfectly 41.5% of the subjects. For these cases, Radzicki reported examples of three obtained values of β between $-.12 < \beta < 0$, $0 < \beta < 0.15$, and even $7 < \beta$, where α is held constant at $\alpha = 1$. Only if the simple linear model did not fit the data perfectly, more complex models were taken into consideration. The next simplest - but non-linear - model which fit best another 7.5% of the subject's rankings had the following form:

$$U_{(\pi_s,\pi_o)} = \alpha\pi_s + \beta\pi_o + \gamma\pi_s^2 + \delta\pi_o^2. \tag{4.5}$$

This indicates that subjects for whose rankings equation 4.5 yielded the best fit value the welfare of others in relation to the own according to the proportion $\frac{\beta+2\delta\pi_o}{\alpha+2\gamma\pi_s}$. Yet, a reasonably well fitting function for over 50% of subjects was still not obtained. A further model which fit best an additional 17% of subject's rankings incorporated the idea of inequality aversion, assuming that

[2] In most psychological studies in which SVO was assessed, the frequency of prosocial subjects (i.e. subjects expressing the goal to maximize joint payoffs or efficiency) was by far greater than the frequency of competitive subjects (i.e. subjects expressing the goal to maximize relative gain). As a rule of thumb, the proportion of prosocials ($U_{(\pi_s,\pi_o)} = \pi_s + \pi_o$) to individualists ($U_{(\pi_s,\pi_o)} = \pi_s$) to competitors ($U_{(\pi_s,\pi_o)} = \pi_s - \pi_o$) appears to be roughly 4:2:1 [see, for instance, Au & Kwong (2004, p.74), Van Lange et al. (2007), Van Lange et al. (1997)].

for some subjects, an increase in the difference between the own outcome and the other's outcome is negatively associated with utility:

$$U_{(\pi_s,\pi_o)} = \alpha\pi_s + \beta\pi_o + \gamma\,|\pi_s - \pi_o|\,, \tag{4.6}$$

where it is assumed that $\gamma < 0$. While for 8% of subject's rankings no well fitting function within the models under consideration could be obtained, a more complex non-linear model yielded a reasonably good -albeit not perfect- fit to the remaining 26% of the rankings:

$$U_{(\pi_s,\pi_o)} = \alpha\pi_s + \beta\pi_o + \gamma\pi_s^2 + \delta\pi_o^2 + \epsilon\,|pi_s - pi_o|^\rho\,, \tag{4.7}$$

where it is assumed that $0 < \rho < 1$ in general, indicating diminishing marginal sensitivity with respect to increments in the difference between the own outcome and the other one's outcome. Since Radzicki did not report how well the considered models did when pitted against each other under inclusion of all of the subject's rankings together, it is hard to estimate which of the models is most useful with respect to a compromise between complexity and explanatory power. But since the model expressed by equation 4.4 is included (i.e. nested) in equation 4.6 as a special case (if $\gamma = 0$) and both models together fit the majority of subjects' rankings reasonably well, one is tempted to state that equation 4.6 may be a candidate for the class of most useful models. In line with this appraisal, Grzelak, who was also involved in other investigations on best fitting utility functions in the domain of SVO (Grzelak, Iwinski, & Radzicki, 1977), came to the conclusion "that the simplest but quite satisfactory piecewise linear class of functions" for explaining own-other payoff allocation choices has the form of equation 4.6 (see Grzelak, 1982, p.105). Furthermore, this statement also received support by Schulz & May (1989), who stated that "the majority of empirically recorded preference structures in payoff pairs can be described with the function" represented by equation 4.6. Moreover, they pointed out that this model includes all of the most commonly considered pure SVO types as special cases (see table 4.1).

In their study, Schulz & May used two different procedures for obtaining data. One procedure required subjects to rank 15 own-other payoff allocations according to their desirability, and in the second procedure subjects made pair comparisons between all of these 15 payoff allocations, resulting in 105 decisions per subject. Then, a random utility model technique was applied to find the functions that fit best the rankings produced by the subjects in either of the two measurement procedures. They found that for 42.3% (51.5%)[3] of the subjects, simple orientation models (e.g. models nr. 1-6 in table 4.1) were sufficient to explain subject's choices reasonably well. For another 9.0% (15.3%) non-simple linear models (i.e. functions of the type represented by model nr. 7 in table 4.1) were necessary for a good fit, and another 36.1% (37.5%) of the data was

[3]The first number refers to the data gained by means of the ranking procedure and the second one in parentheses to the data gained by means of the pair comparison procedure.

Table 4.1: SVO types and respective utility functions

Nr.	SVO	Functional form	Constraints on function [4.6]
1.	Individualistic	$U_{(\pi_s,\pi_o)} = \pi_s$	$\alpha = 1, \beta = 0, \gamma = 0$
2.	Prosocial	$U_{(\pi_s,\pi_o)} = \pi_s + \pi_o$	$\alpha = 1, \beta = 1, \gamma = 0$
3.	Competitive	$U_{(\pi_s,\pi_o)} = \pi_s - \pi_o$	$\alpha = 1, \beta = -1, \gamma = 0$
4.	Altruistic	$U_{(\pi_s,\pi_o)} = \pi_o$	$\alpha = 0, \beta = 1, \gamma = 0$
5.	Inequality averse	$U_{(\pi_s,\pi_o)} = -\lvert\pi_s - \pi_o\rvert$	$\alpha = 0, \beta = 0, \gamma = -1$
6.	Maximin	$U_{(\pi_s,\pi_o)} = \pi_s + \pi_o - \lvert\pi_s - \pi_o\rvert$	$\alpha = 1, \beta = 1, \gamma = -1$
7.	Nonsimple linear SVO	$U_{(\pi_s,\pi_o)} = \alpha\pi_s + \beta\pi_o$	$\gamma = 0$

(Table adoptet from Schulz & May, 1989, p.43)

best described by non-simple non-linear (or conditionally linear) models of the type represented by equation 4.6. Surprisingly, when subjects were categorized according to the parameters of the models which best fit the given rankings, Schulz & May found the following distribution of simple SVO types: 48.5% (45.7%) individualistic, 22.4% (21.8%) competitive, 0.7% (0.0%) altruistic, 7.2% (5.6%) egalitarian (inequality averse), 6.5% (9.8%) cooperative (joint gain maximizing), and 15.9% (16.9%) maximin orientations. This is an uncommon distribution since prosocial orientations are usually found to be most frequent, followed by the individualistic orientation, followed by the competitive orientation which is usually the least frequent apart from the altruistic one (see footnote 2). Besides this exceptional distribution, it is surprising that the maximin orientation was the most frequent one among the prosocial (i.e. positively other-regarding) SVOs, which would indicate that roughly 50% of prosocial subjects neither have the goal to maximize joint gains, nor to minimize differences between gains, but to maximize the minimum payoffs.

Further studies regarding models of SVO were conducted by Loewenstein, Bazerman, & Thompson (1989), who found that utility functions in which the payoff to the self, and inequality between payoffs is taken into account yielded a good fit to their data. The novelty of their approach was the incorporation of the distinction between advantageous and disadvantageous inequality into a model[4]. Furthermore, they compared different functional forms of SVOs on the individual level with respect to their goodness of fit, simplicity, and flexibility (i.e. sensitivity to individual differences). The method chosen for the model fitting was a multiple regression technique and the obtained R squares served as goodness of fit criteria. Loewenstein et al. (1989) considered only non-linear utility functions and among those, they found that the model which included terms that account for disadvantageous and advantageous inequality yielded the highest R square value:

$$U_{(\pi_s,\pi_o)} = \alpha\pi_s + \beta(\pi_s - \pi_o) + \gamma(\pi_s - \pi_o)^2 + \delta(\pi_s - \pi_o) + \epsilon(\pi_s - \pi_o)^2, \tag{4.8}$$

[4]A differentiation between advantageous and disadvantageous inequality was also proposed by Conrath & Deci (1969), so the term “novelty" is relativized, but at least Loewenstein et al. (1989) were the first to test such a model explicitly.

where $\{\beta,\gamma\} = 0$ if $(\pi_s - \pi_o) < 0$, indicating disadvantageous inequality, and $\{\delta,\epsilon\} = 0$ if $(\pi_s - \pi_o) > 0$, indicating advantageous inequality. Implicitly, it is assumed that $\{\beta,\gamma,\delta,\epsilon\} \leq 0$. Loewenstein et al. also reported evidence that β and γ are significantly greater (i.e. less negative) than δ and ϵ, which means that disadvantageous inequality has a greater negative impact on utility than advantageous inequality.

The idea of distinguishing between advantageous and disadvantageous inequality was also taken on by Fehr & Schmidt (1999), while the sheer notion of social preferences - often reduced to that of inequality aversion - became more and more acknowledged in economics as a counter model to the selfishness axiom (Henrich et al., 2005), the basic assumption of standard economic theory. For a dyadic situation, Fehr & Schmidt's model (henceforth labeled as FS) is represented by a utility function of the following form:

$$U_{(\pi_s,\pi_o)} = \pi_s - \alpha \max\{\pi_o - \pi_s, 0\} - \beta \max\{\pi_s - \pi_o, 0\}, \tag{4.9}$$

where it is assumed that $\alpha \geq \beta$, indicating a greater impact of disadvantageous inequality on utility compared to advantageous inequality, and $0 \leq \beta < 1$ which means that the possibility of the competitive SVO is ignored.

Fehr & Schmidt inferred parameter value distributions in the population from various experiments on behavior in ultimatum games. In a nutshell, the assumed distribution suggests that 30% of the people are individualists ($\alpha = 0$, $\beta = 0$), and 70% are inequality averse with different shades of this SVO's intensity[5]. To our knowledge, only one study exists which reports data on the empirical estimation of the two parameters in question on the individual level (see Blanco et al., 2011).

A similar model was also proposed by Bolton & Ockenfels (2000) (henceforth labeled as ERC model), which in the dyadic case reduces to:

$$U_{(\pi_s,\pi_o)} = \alpha\pi_s - \frac{\beta}{2}\left(\frac{\pi_s}{\pi_s + \pi_o} - \frac{1}{2}\right)^2, \tag{4.10}$$

where $\alpha \geq 0$ and $\beta > 0$. In this model, no distinction between advantageous and disadvantageous inequality is made, since the term in parentheses only indicates a deviation from an equal share of payoffs, while the information about the direction of a potential deviation is canceled out by the square. Hence, both kinds of inequality aversion have the same impact on utility, while the degree of inequality aversion is determined by parameter β such that a higher β indicates a higher degree of inequality aversion.

[5] However, assumed parameter value distributions vary across papers (see e.g. Fehr & Schmidt, 2004; Fehr et al., 2007).

The three latter models (Loewenstein et al., 1989; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000) have in common that they take into account the own payoff and the difference between the own payoff and the other one's payoff in terms of inequality aversion, but do not include elements reflecting efficiency concerns (joint gain maximization), altruism (maximizing other's gain), or competition (maximizing relative gain), at least not under the constraints placed upon the parameters' ranges. Evidence on whether efficiency concerns or inequality aversion have a larger impact on choice behavior is mixed, though. For instance, Jakiela (2012) found that people seem to be more concerned about equity than efficiency on average. However, Engelmann & Strobel (2004), who also compared the inequality aversion model by Fehr & Schmidt to the one proposed by Bolton & Ockenfels, found that concerns for efficiency (i.e. the goal to maximize joint gains) seem to matter indeed. Moreover, they concluded from their findings that inequality aversion may be a less important motive as previously assumed and that a combination of individualistic, prosocial, and maximin orientations explained their data much better. Consequently, they stated that the model developed by Charness & Rabin (2002) may be sufficient to account for their subjects' behavior in the experiment.

Charness & Rabin's model (henceforth labeled as QM model) of social preferences assumes that people's behavior in games is guided by a (quasi-)maximin orientation and concerns for reciprocity[6]. However, in this paper we do not test models which incorporate beliefs about others or reciprocity concerns, hence we only consider the reciprocity-free quasi-maximin model, which has the following form:

$$U_{(\pi_s,\pi_o)} = [(1 - \mathbb{1}_a\alpha - \mathbb{1}_b\beta)\pi_s] + [(\mathbb{1}_a\alpha + \mathbb{1}_b\beta)\pi_o], \tag{4.11}$$

where $\mathbb{1}_a$ and $\mathbb{1}_b$ are step functions to indicate advantageous inequality (or equality) by setting $\mathbb{1}_a = 0$ and $\mathbb{1}_b = 1$ if $\pi_s \geq \pi_o$, and disadvantageous inequality by setting $\mathbb{1}_a = 1$ and $\mathbb{1}_b = 0$ if $\pi_s < \pi_o$. In other words, the α parameter determines the relative weight a person attaches to the outcome for the other in relation to the own given that $\pi_s < \pi_o$, and the β parameter determines the relative weight a person attaches to the outcome for the other in relation to the own given that $\pi_s \geq \pi_o$. This model can account for various expressions of other-regarding preferences. First, it can account for the expression of social-welfare preferences (represented by patterns of parameter values satisfying $1 \geq \beta \geq \alpha > 0$, including the expression of perfect substitutes given $\alpha = \beta = .5$). Second, it can account for narrow self-interest and competitive SVOs (represented by $\alpha \leq \beta \leq 0$). And third, it can account for inequality aversion (represented by $\alpha < 0 < \beta < 1$). Charness & Rabin did not address the issue of individual differences in SVO, but found that social-welfare concerns (i.e. joint gain maximization) appeared to better account for behavior on the aggregate level compared to other SVO expressions, including inequality aversion. This is consistent with results subsequently obtained on individual choice data (e.g. Engelmann & Strobel, 2004; Daruvala, 2010).

[6] Charness & Rabin (2002) assume that inequality aversion is in fact a combination of concerns for reciprocity and a maximin SVO

In a seminal paper, Andreoni & Miller (2002) could confirm that social preferences are rationalizable and thus can in principle be described in terms of well defined utility functions. For estimating SVO, Andreoni & Miller employed a constant elasticity of substitution (CES) model of the following form:

$$U_{(\pi_s,\pi_o)} = \left(\alpha\pi_s^\beta + (1-\alpha)\pi_o^\beta\right)^{\frac{1}{\beta}}. \tag{4.12}$$

They identified several different SVO types, including narrow self-interest, maximin preferences, and perfect substitutes (joint gain maximization), as well as weaker (i.e. impure) expressions of these prototypes that can be described by the free parameters' values.

Another model of the CES class has been proposed by Cox & Sadiraj (2007), which they labeled egocentric other-regarding preferences model (henceforth referred to as ERP) that has the following functional form:

$$U_{(\pi_s,\pi_o)} = \pi_s^\beta + \alpha\pi_o^\beta, \tag{4.13}$$

where α captures the weight attached to the other person's outcome, and β - as in Andreoni & Miller's model - is the elasticity parameter which regulates the convexity of indifference curves on the plane that spans all possible own-other-outcome combinations. In order to maintain the characteristics of egocentricity, strict positive monotonicity, and strict convexity, the parameter values are constraint such that $\beta < 1$ and $0 < \alpha < 1$. However, Cox & Sadiraj (2007) did not estimate parameter values from empirical data, but showed analytically that their model is better suited for explaining data from Public Goods games compared to Fehr & Schmidt's inequality aversion model.

The list of models discussed above is probably not exhaustive in the sense that all SVO models that have been proposed to date are included. However, we think that it includes the most prominent outcome-based SVO models that are most frequently used in economics and the social sciences today. As can be seen, the models vary widely in terms of complexity, intuitive appeal, and parameterization. In the following section, we will evaluate a selection of the afore discussed models in terms of their goodness of fit to empirical data, but also in terms of the psychological interpretability of parameter values. A model that yields an excellent goodness of fit to data, but can not be interpreted in a meaningful way is not useful for understanding how individuals trade off own material well-being against the well-being of others. In that sense, goodness of fit (GOF) will be treated as a necessary but insufficient criterion for evaluating the usefulness of the different SVO models we test.

Also, we would like to test a further model which we think could be promising, especially in terms of psychological interpretability of parameter values. The model we propose is a slight modification of the model presented in Charness & Rabin (2000), and has the following form:

$$U_{(\pi_s,\pi_o)} = (1-\alpha)\pi_s + \alpha\left(\beta(-|\pi_s - \pi_o|) + [(1-\beta)(\pi_s + \pi_o)]\right), \tag{4.14}$$

This model can represent selfish and competitive SVOs ($\alpha \leq 0, \beta = 1$), social-welfare preferences ($\alpha > 0, \beta = 0$) including perfect substitutes ($\alpha = 1, \beta = 0$), and inequality concerns ($\alpha > 0, \beta = 1$) including pure inequality aversion ($\alpha = 1, \beta = 1$) and pure maximin ($\alpha = 1, \beta = .5$). However, this model rules out pure altruism (i.e. $U_{(\pi_s,\pi_o)} = \pi_o$) and does not differentiate between advantageous and disadvantageous inequality aversion, which could potentially impair the model's ability to fit the data sufficiently well.

4.3 Method

4.3.1 Model selection and parameter specification

The criteria for deciding which of the models proposed in the literature should be tested in this study are the following: 1) Models of pure SVO's, i.e. nonparametric models, should be included since they can serve as benchmarks for the parametric models. Of special interest in this respect will be the comparison between the model reflecting the homo economicus position (i.e. $U_{(\pi_s,\pi_o)} = \pi_s$) and all models including terms reflecting concerns for others. 2) Further, among the parametric models, only those should be included which in principle allow for a psychological interpretation of the parameter values. 3) Third, we will restrict ourselves to the analysis of models that have only two free parameters. We employ this restriction because two-parametric models are tractable, thus facilitating psychological interpretation, and the risk of overfitting the data is reduced when models incorporating three or more parameters are excluded. In accordance with these three criteria, 14 models are selected for testing in this study. The selected models are presented in table 4.2, together with an indication of corresponding parameter ranges that are searched through for determining maximum GOF scores. The models' parameter ranges conform with the constraints put on parameter values as suggested by the corresponding model's authors given that such constraints are indicated, and are determined by us in case no constraints have been stated. The step size for searching through the parameter ranges varies between models depending on how wide the corresponding range (r) ist. For parameter range widths $r \leq 3$, the resolution is .05, for parameter range widths $3 < r$, it is .1, and for the special case where the parameter range width is $r = 50'000$ (ERC model), a step size of 500 is employed.

The highly attentive reader may have noticed that models # 8 (PI) and # 10 (ERC) in table 4.2 are slightly different from their corresponding functional form as represented in section 4.2. In these two models, a parameter has originally been attached to the payoff for the self (expressed as $[\alpha\pi_s]$), which is then not used later on to relativize another term (e.g. expressed as $[1 - \alpha]$). We find it reasonable and justifiable in these cases to assign this parameter a value of 1, thereby fixating it, since the respective second free parameters in these two models are sufficient to represent how much weight is put on terms reflecting non-selfish tendencies relative to the outcome for the self. Consequently, they are tested as one-parametric models.

Table 4.2: Models selected for testing

Nr.	Authors	Label	Functional form	Parameter space under consideration
1	-	Pure individualism (Indi)	$U_{(\pi_s,\pi_o)} = \pi_s$	-
2	-	Pure prosociality (Proso)	$U_{(\pi_s,\pi_o)} = \pi_s + \pi_o$	-
3	-	Pure competitiveness (Compet)	$U_{(\pi_s,\pi_o)} = \pi_s - \pi_o$	-
4	-	Pure altruism (Altr)	$U_{(\pi_s,\pi_o)} = \pi_o$	-
5	-	Pure inequality aversion (Ineq)	$U_{(\pi_s,\pi_o)} = -\lvert\pi_s - \pi_o\rvert$	-
6	-	Maximin	$U_{(\pi_s,\pi_o)} = \pi_s + \pi_o - \lvert\pi_s - \pi_o\rvert$	-
7	Sawyer (1966)	Partial Prosociality (PS)	$U_{(\pi_s,\pi_o)} = \pi_s + \alpha\pi_o$	$-.5 \leq \alpha \leq 1.1$
8	Bolton & Ockenfels (2000)	ERC	$U_{(\pi_s,\pi_o)} = \pi_s - \frac{\alpha}{2}\left(\frac{\pi_s}{\pi_s+\pi_o} - \frac{1}{2}\right)^2$	$0 \leq \alpha \leq 50'000$
9	Radzicki (1976)	Partial Inequality aversion (PI)	$U_{(\pi_s,\pi_o)} = \pi_s + \alpha\pi_o - \beta\lvert\pi_s - \pi_o\rvert$	$[-1.1 \leq \alpha \leq 1.1], [-1.1 \leq \beta \leq 1.1]$
10	Fehr & Schmidt (1999)	FS	$U_{(\pi_s,\pi_o)} = \pi_s - \alpha \max\{\pi_o - \pi_s, 0\} - \beta \max\{\pi_s - \pi_o, 0\}$	$[0 \leq \alpha \leq 4], [0 \leq \beta \leq 1]$
11	Charness & Rabin (2002)	QM	$U_{(\pi_s,\pi_o)} = [(1 - \mathbb{1}_a\alpha - \mathbb{1}_b\beta)\pi_s] + [(\mathbb{1}_a\alpha + \mathbb{1}_b\beta)\pi_o]$	$[-1.1 \leq \alpha \leq 1.1], [-1.1 \leq \beta \leq 1.1]$
12	Andreoni & Miller (2002)	CES	$U_{(\pi_s,\pi_o)} = \left(\alpha\pi_s^\beta + (1-\alpha)\pi_o^\beta\right)^{\frac{1}{\beta}}$	$[-.5 \leq \alpha \leq 1.1], [0 < \beta \leq 1]$
13	Cox & Sadiraj (2007)	ERP	$U_{(\pi_s,\pi_o)} = \pi_s^\beta + \alpha\pi_o^\beta$	$[0 \leq \alpha \leq 1], [0 < \beta \leq 1]$
14	Ackermann & Murphy (2013)	AM	$U_{(\pi_s,\pi_o)} = (1-\alpha)\pi_s + \alpha\left(\beta(-\lvert\pi_s - \pi_o\rvert) + [(1-\beta)(\pi_s + \pi_o)]\right)$	$[-1.1 \leq \alpha \leq 1.1], [-1.1 \leq \beta \leq 1.1]$

4.3.2 Sample and data set

The models are fit to data gained from $N = 269$ subjects who voluntarily participated in decision making experiments at a European university between May 2010 and May 2011. Subjects completed a paper-pencil version of the SVO Slider Measure (Murphy, Ackermann, & Handgraaf, 2011). The SVO Slider Measure consists of 15 items and in each item, subjects are asked to select their most preferred own-other payoff allocation out of 9 options available. Figure 4.1 shows one item of the SVO Slider Measure [for the complete measure, see Appendix B or Murphy et al. (2011)]. The items have different slopes when represented as lines on the own-other outcome plane (see Murphy et al., 2011, p. 779) ranging from -5 to 2.33 (with one item keeping the outcome for the self constant, thus having an undefined slope). Consequently, the cost of deviating from selfishness in order to increase or decrease the other person's outcome varies across items, which is essential for estimating individual parameter values that ultimately inform about an individual's willingness to pay for achieving a goal other than mere own payoff maximization.

All studies were conducted in an incentive compatible manner with a lottery system implemented for payment[7]. The outcomes from the Slider Measure were worth .5 Swiss francs to the subjects. Consequently, those subjects who were randomly selected for payment could earn substantial amounts between 32.5 Swiss francs at the minimum and 100 Swiss francs at the maximum.

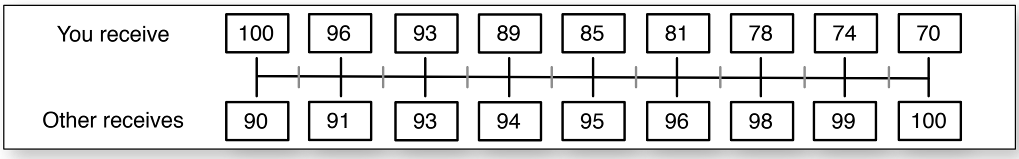


Figure 4.1: One item from the SVO Slider Measure

4.3.3 Model fitting procedure

All models are fit to the choice data of the 269 subjects on the individual level, i.e. GOF scores and best fitting parameter values are obtained for each subject. For fitting the models to the individual data, three different methods are applied. The reason for this is as follows. How good a model fits the data is not only dependent on the model (i.e. the model's functional form and the parameter values considered) and the data, but also on how the GOF criterion is defined (i.e. the method selected for computing the deviation between model predictions and empirical observations), as well as potential interactions between these three factors. Since one major aim of this paper is to present a comparative evaluation of different SVO models, it is of utmost

[7] A subject's payment was not only dependent on his own choices, but also on the choices of one randomly selected other subject. That is, the amount one subject was paid consisted of the payoff he allocated to the self in a randomly selected item and the payoff another randomly selected person allocated to the other in a different randomly selected item.

importance to have confidence in the validity of resulting individual GOF scores and best fitting parameter values. Hence, in order to increase this confidence, we employ three substantially different fitting methods the results of which can then be compared. If the methods agree, we can be confident that a comparative evaluation of the models is feasible and valid. The three methods are explicated below.

Utility percentage fitting (UPF)

We refer to the first fitting method we employ as utility percentage fitting (UPF). This method produces GOF scores as follows. First, for a given fully specified model[8], the utility of each option in the Slider Measure is calculated. Second, the option yielding the highest utility is determined for each of the fifteen items. Third, the sum of the resulting fifteen utility values is computed. This sum is then recorded as the maximum utility that can possibly be obtained across the fifteen Slider Measure items given the corresponding fully specified model. Fourth, the minimum utility possibly obtainable is computed and recorded analogously. Fifth, in case the minimum utility is different from zero, then the minimum and maximum values are shifted by a factor (f_{shift}) such that the minimum utility is equal to zero.
Sixth, for each subject, the utilities of the options actually chosen given the fully specified model are computed, summed up, and shifted by the factor (f_{shift}). Finally, the resulting value is expressed in percentage of the (shifted) maximum utility that can possibly be obtained given the corresponding fully specified model. This percentage is a person's GOF score given a fully specified model and is recorded together with the corresponding parameter values (i.e. the parameter values which are instantiated in the corresponding fully specified model). Hence, a GOF score of 100% means that the options a subject actually chose in the Slider Measure are exactly those options that yield the maximum utility given the fully specified model at hand, thus indicating a perfect fit. Likewise, a GOF score of 0% means that a subject chose exactly those options that minimize utility given the fully specified model at hand, indicating that the subject did exactly the opposite of what the model predicts.
This routine is then repeated for each model specification under consideration[9] (i.e. all parameter values considered), and for each subject the maximum of all resulting GOF scores is stored together with the corresponding parameter values. In case there are several parameter values (or parameter value combinations) that yield the same maximum GOF score, that best fitting parameter value is reported which is closest to zero[10].

[8]We refer to a model as fully specified when particular parameter values are instantiated (i.e. fixing each free parameter at a particular value).

[9]We employ full grid search for all models and model fitting methods rather than optimization algorithms such as simplex or simulated annealing since we found that some error surfaces exhibit many local minima and sometimes a very narrow global minimum, which renders optimization algorithms unreliable.

[10]For parameter value combinations, that best fitting parameter value combination is reported for which the square of the sum of city-block distances between the parameter values and zero is minimized. This procedure is also applied for the two other methods.

Formally, the UPF procedure for obtaining any particular subject's GOF score given a particular fully specified model (M_s) is as follows.

$$U_{\max} = \sum_{j=1}^{15} \left(\max_i U_{(i,j)} \mid M_s \right), \tag{4.15}$$

where $U_{\max}$ is the total maximum utility possibly obtainable given M_s, and $U_{(i,j)}$ is the utility of option i in item j.

$$U_{\min} = \sum_{j=1}^{15} \left(\min_i U_{(i,j)} \mid M_s \right), \tag{4.16}$$

where $U_{\min}$ is the total minimum utility possibly obtainable given M_s.

$$U_{sum} = \sum_{j=1}^{15} \left(U_{(c_j,j)} \mid M_s \right), \tag{4.17}$$

where U_{sum} is the total utility a particular subject obtains given M_s, and $U_{(c_j,j)}$ is the utility of the option chosen by that subject in item j.

$$GOF_{UPF} = \frac{U_{sum} + f_{shift}}{U_{\max} + f_{shift}}, \tag{4.18}$$

where GOF_{UPF} is the goodness of fit of M_s to the choice data of the particular subject as produced by the UPF method, and $f_{shift} = -U_{\min}$.

Inequality count fitting (ICF)

The second model fitting method we employ is referred to as inequality count fitting (ICF) and produces GOF scores as follows. First, for a given fully specified model, the utilities of all options in the Slider Measure are computed. Second, for each subject and item, the utility of the option chosen $\left(U_{(c_j,j)}\right)$ in item j is compared to the utility of each of the remaining 8 options $\left(U_{(i \neq c_j,j)}\right)$ of that item. Each comparison checks whether the inequality $U_{(c_j,j)} \geq U_{(i \neq c_j,j)}$ holds. In case $U_{(c_j,j)} = \max_i U_{(i,j)}$, all 8 inequalities hold, whereas no inequality holds in case $U_{(c_j,j)} = \min_i U_{(i,j)}$. Third, the number of inequalities satisfied is summed up over all items and divided by the total number ($8x15 = 120$) of inequalities that can possibly be satisfied over all items[11]. The resulting percentage then serves as an individual's GOF score given the correspond-

[11] In order to take into account that there may be several options yielding the same utility in a particular item given a particular fully specified model, the number of options yielding the same utility are recorded per item, and subtracted from both the sum of inequalities satisfied and the number of inequalities that can possibly be satisfied (which is 120) prior to the division. This may seem unnecessary, but this results in somewhat more conservative GOF estimates for scores smaller than 100%.

ing fully specified model, where a GOF score of 100% indicates a perfect fit. Given a particular model, this routine is repeated for all instantiations of parameter values under consideration and maximum GOF scores as well as the corresponding best fitting parameter values are recorded for each subject.

Formally, for a particular subject and fully specified model (M_s), the ICF method yields GOF scores as follows.

$$NIS = \sum_{j=1}^{15} \left(\sum_{q=1}^{8} \left(1 - \frac{\max\{U_{(q,j)} - U_{(c_j,j)}, 0\}}{U_{(q,j)} - U_{(c_j,j)}} \right) \middle| \, M_s \right), \tag{4.19}$$

where NIS is the number of inequalities satisfied by the choice pattern of a particular subject given M_s, $U_{(c_j,j)}$ is the utility of the option c_j chosen by the subject in item j, and $U_{(q,j)}$ is the utility of an option *not* chosen by the subject in item j.

$$GOF_{ICF} = \frac{NIS - n_{(ND\,|\,M_s)}}{120 - n_{(ND\,|\,M_s)}}, \tag{4.20}$$

where GOF_{ICF} is the goodness of fit of M_s to a particular subject's choice data as yielded by the ICF method, and n_{ND} is the number of non-discriminant options, i.e. the number of options yielding the same utility in an item minus one, summed over all 15 items.

Maximum likelihood estimation (MLE)

The third fitting method we employ is maximum likelihood estimation (MLE, see e.g. Myung, 2003). For implementing MLE, it is required to introduce a function that assigns choice probabilities to the options in the Slider Measure items. This can be accomplished by applying the Boltzmann softmax function as a generalization of Luce's choice rule (Luce, 1959) to the Slider Measure options' utilitites as given by a particular fully specified model. The assumption introduced thereby is that the option yielding the highest utility in a particular Slider Measure item is not necessarily chosen with certainty, but may also be chosen with a probability $p < 1$. When employing the Boltzmann softmax function the probability $p_{(i,j)}$ that option i is chosen from the set of options (K) in item j from the Slider Measure is given by the following equation.

$$p_{(i,j)} = \frac{\exp\left(\frac{a_{(i,j)}}{T}\right)}{\sum_{k=1}^{9} \exp\left(\frac{a_{(k,j)}}{T}\right)}, \quad \forall \ T > 0, \tag{4.21}$$

where $a_{(i,j)} = \left[\left(U_{(i,j)}|M_s\right) - \min_k \left(U_{(k,j)}|M_s\right)\right] + 1$ expresses the attraction of option i, and $a_{(k,j)} = \left[\left(U_{(k,j)}|M_s\right) - \min_k \left(U_{(k,j)}|M_s\right)\right] + 1$ is the attraction of any of the $k \in K = \{1, 2, ..., 8, 9\}$

options in item j, including option i. This transformation of utilities into attraction scores is employed in order to make sure that only positive values (rather than potential negative utilities) enter the Boltzmann softmax function, and that the option with the lowest utility within an item is assigned a value of 1, i.e. $a_{(m,j)} = 1$, where $U_{(m,j)} = \min_k \left(U_{(k,j)} \mid M_s\right)$. The transformation facilitates the normalization as processed by the Boltzmann softmax function, which ensures that the probabilities with which the options of a particular item are chosen add up to unity:

$$\sum_{k=1}^{9} p_{(k,j)} = 1. \tag{4.22}$$

The parameter T (frequently labeled as the temperature parameter) determines the shape of the probability mass over the options of a given item. With increasing values of T (i.e. with $T \to \infty$), the probability mass approximates a uniform distribution, i.e. all options have the same probability (in our case this would mean $p_{(k,j)} = \frac{1}{9}, \forall\, k$). Conversely, with decreasing values of T (i.e. with $T \to 0$), the probability mass approximates a step function, where the option yielding the highest utility is chosen with probability $p = 1$ (or 1 divided by the number of options yielding the same highest utility), and all other options are chosen with zero probability. However, the Boltzmann softmax function as presented in formula 4.21 is not suited for assuming strict determinism, i.e. $T = 0$. For our purpose, though, we want to allow for the highest utility option being chosen with $p = 1$. Therefore, we implement:

$$\left[p_{(i,j)} \mid (T = 0)\right] = \begin{cases} 1/n_{(max,j)} \\ 0 \end{cases}, \tag{4.23}$$

where $n_{(max,j)}$ is the number of options yielding the maximum utility within item j. If there is only one option uniquely yielding the highest utility, it is assigned a probability of 1. However, if there are two or more options yielding the highest utility, they are assigned equal probabilities. This implementation rules out a perfect fit given that there is at least one item with more than one option yielding the highest utility given a particular fully specified model. Consequently, models that make more than one prediction in at least one item are punished for their predictive flexibility. Given all of the above, we can compute the likelihood[12] of a subject's choice data given a fully specified model and specified temperature parameter (T) by multiplying the probabilites of the options which the subject had actually chosen (i.e. $\prod_{j=1}^{15} p_{(c_j,j)}$, where c_j indicates the option actually chosen by the subject at hand in item j). However, this product can result in extremely small numbers, which is undesirable for technical reasons. Luckily, we can employ a mathematical trick and - rather than multiply the probabilities- sum up the natural logarithm of each of the probabilities. We can then use the resulting joint log-likelihood, which we term lnL

[12]This likelihood must not be confused with the probability of the data given the model. In order to compute the probability rather than the likelihood of the data given the model, this would require a Bayesian approach, where the model is assigned a prior probability.

(log-likelihood) as our GOF score.

$$GOF_{MLE} = \sum_{j=1}^{15} \ln p_{(c_j,j)}. \tag{4.24}$$

Since we allow for $T = 0$, a GOF_{MLE} score can be zero, which would indicate a perfect fit, and lower (i.e. negative) GOF_{MLE} scores indicate worse fits.

As with the UPF and ICF method, we employ full grid search also under the MLE method. That is, for each subject and model, GOF_{MLE} is computed for all possible model instantiations (i.e. parameter values under consideration) and values of the temperature parameter T in the interval $0 \leq T \leq .1$, with a step size of .001. Then, for each subject, the maximum of all GOF_{MLE} scores is recorded together with the corresponding best fitting model parameter values and the corresponding value of T.

4.4 Results

The results are reported according to the following structure. First, we report on the agreement between the three fitting methods we employed in terms of GOF scores. Next, the goodness of fit of each model to the data is reported and the models are compared to each other as well as ranked according to their performance on the basis of their GOF scores as produced by the different methods. Further, we report the results from a hierarchical analysis where we focus on how many subjects' choice data can be perfectly accounted for by particular model classes. Finally, we report on the distributions of best fitting parameter values and discuss the psychological meaning of particular parameter values for the different models.

4.4.1 Agreement between fitting methods

Table 4.3 provides an overview of agreement between the three methods (UPF, ICF, and MLE) per model expressed in terms of Pearson correlations. However, these correlations provide a very conservative estimate of agreement, because the GOF scores as produced by the different methods are not necessarily linearly related. Figure 4.2 shows the relation between GOF scores from the different methods for the PS model (# 7) as an example. The PS model is chosen as the example here because it demonstrates clearly that GOF scores from MLE are not linearly related to both GOF scores from UPF and ICF. As can be seen, a non-linear measure of association would better describe the respective relations between the GOF metrics compared to a linear regression line, hence the agreement is slightly underestimated when represented by simple correlation coefficients. In general, the three methods agree to a very high degree on how good the different models fit the individual choice data, while the mean[13] agreement between ICF and MLE is highest, followed by UPF-ICF-agreement, and is least -but still substantial- between UPF and MLE. Agreement between methods is relatively low, however, for the PI model (# 9). The reason for this will be explicated in the next section.

[13]Mean correlations are computed by taking the mean of the squared correlation coefficients (i.e. turning the correlations into R squares) and then taking the square root of that mean.

Table 4.3: Pearson correlations between GOF scores as produced by the three different methods (rounded to three decimals)

Nr.	Label	UPF x ICF	UPF x MLE	ICF x MLE	Mean correlation
1	Indi	0.981	0.909	0.946	0.946
2	Proso	0.982	0.870	0.915	0.923
3	Compet	0.995	0.934	0.951	0.960
4	Altr	0.987	0.989	1.000	0.992
5	Ineq	0.994	0.933	0.947	0.958
6	Maximin	0.986	0.900	0.951	0.946
7	PS	0.960	0.869	0.911	0.914
8	ERC	0.782	0.705	0.938	0.814
9	PI	0.869	0.618	0.733	0.747
10	FS	0.841	0.708	0.921	0.828
11	QM	0.833	0.709	0.908	0.821
12	CES	0.904	0.804	0.935	0.883
13	ERP	0.890	0.798	0.925	0.873
14	AM	0.864	0.748	0.925	0.849
Mean correlation		0.922	0.828	0.924	0.892

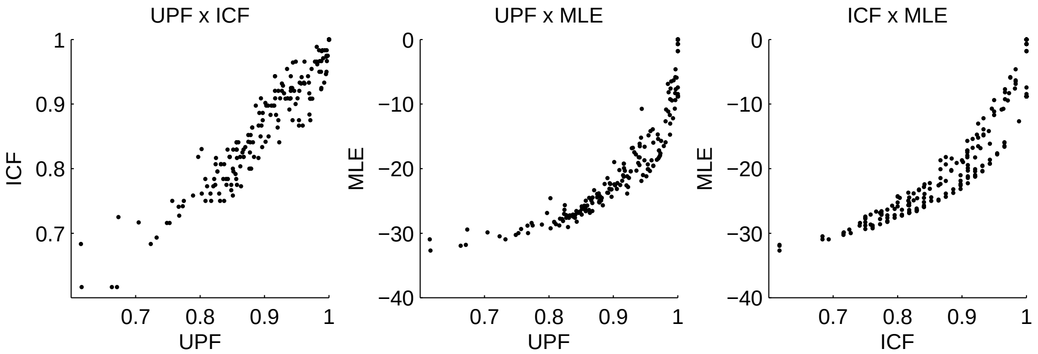


Figure 4.2: Scatter plots of GOF agreement between different fitting methods for the PS model as an example

4.4.2 Goodness of fit analyses

Table 4.4 shows the percentage of subjects perfectly accounted for by the corresponding model (according to UPF), as well as the mean, standard deviation, and median of individual GOF scores as produced by the UPF and ICF methods. The Bayesian information criterion (BIC) is reported as a metric of the models' overall performance as suggested by MLE[14]. Note that a lower BIC indicates higher explanatory power. We are aware that introducing BIC scores complicates cross-method GOF comparisons, since models are punished for the number of parameters they employ when GOF scores are represented as BICs, while GOF scores from UPF and ICF are not transformed in any way. However, we think that for comparing the models rather than the methods, showing BIC scores rather than log-likelihoods facilitates judgment about model performance as assessed by MLE. Table 4.5 provides a ranking of the models according to their ability to account for the data as suggested by different criteria. We compared the models' BICs in terms of Bayesian model weights and Bayes factors, and no two models are equally good according to these information criteria. Hence, in our set, if one model has a lower BIC than another model, it means that there is more than strong evidence in favor of the model with the lower BIC (Bayes factor >> 10). The same holds analogously for comparisons in terms of Bayesian model weights, according to which the PI model shows by far the highest posterior probability (close to one) of having generated the data given the set of models we have tested.

Table 4.4: Overview of the models' explanatory power

Nr.	Label	Perfect fits (%)	UPF GOF		ICF GOF		MLE
			Mean (Std.)	Median	Mean (Std.)	Median	BIC
1	Indi	15.99	0.777 (0.15)	0.750	0.722 (0.18)	0.670	14301.5
2	Proso	14.13	0.782 (0.17)	0.819	0.762 (0.17)	0.784	14774.3
3	Compet	2.23	0.635 (0.19)	0.583	0.617 (0.18)	0.567	17575.7
4	Altr	0.00	0.478 (0.21)	0.529	0.383 (0.18)	0.433	21941.7
5	Ineq	1.12	0.646 (0.22)	0.696	0.640 (0.21)	0.683	17160.1
6	Maximin	1.12	0.734 (0.20)	0.810	0.659 (0.21)	0.700	16344.9
7	PS	33.83	0.926 (0.08)	0.953	0.905 (0.09)	0.920	9806.8
8	ERC	15.99	0.958 (0.05)	0.972	0.875 (0.09)	0.875	11220.3
9	PI	61.34	0.986 (0.03)	1.000	0.966 (0.06)	1.000	7389.9
10	FS	21.19	0.967 (0.04)	0.986	0.904 (0.08)	0.911	10153.5
11	QM	40.89	0.980 (0.04)	0.995	0.949 (0.06)	0.975	7568.3
12	CES	31.60	0.955 (0.06)	0.977	0.915 (0.08)	0.925	9479.9
13	ERP	31.60	0.961 (0.05)	0.976	0.912 (0.08)	0.925	9662.5
14	AM	39.03	0.977 (0.04)	0.992	0.946 (0.07)	0.975	7622.6

Several things are worth noting regarding the information given in table 4.4. First of all, the most simple non-parametric "pure" SVO models (#1-6) can perfectly explain the choice patterns of a full third (33.46%)[15] of all subjects. That is, one third of the subjects made choices that are

[14]The BIC for a particular model M is computed by summing across all $N = 269$ individual model BICs: $BIC_M = \sum_{n=1}^{N} 2 * (-GOF_{(n,MLE|M)}) + k * \ln n_{obs}$, where k is the number of parameters of model M (plus the temperature parameter T), and n_{obs} is the number of observations per subject, which in our case is the number of items in the Slider Measure ($n_{obs} = 15$).

[15]The Ineq (#5) and Maximin (#6) models perfectly explain the choice patterns of the same 3 (1.12%) subjects, since these two models both assign the highest utility to options that yield the same payoff for the self and for the other. The Slider Measure does not include an item where all options yield equal money distributions, but only differ

Table 4.5: Ranking of models from best (1) to worst (14) according to different performance criteria

Nr.	Label	Perfect fits	UPF mean	UPF median	ICF mean	ICF median	MLE BIC
1	Indi	8	10	11	10	12	9
2	Proso	10	9	9	9	9	10
3	Compet	11	13	13	13	13	13
4	Altr	14	14	14	14	14	14
5	Ineq	12	12	12	12	11	12
6	Maximin	12	11	10	11	10	11
7	PS	4	8	8	6	6	6
8	ERC	8	6	7	8	8	8
9	PI	1	1	1	1	1	1
10	FS	7	4	4	7	7	7
11	QM	2	2	2	2	2	2
12	CES	5	7	5	4	4	4
13	ERP	5	5	6	5	4	5
14	AM	3	3	3	3	2	3

perfectly consistent with one of the following simple goals: Maximize the own payoff (15.99%), maximize the sum of payoffs (14.13%), maximize relative gain (2.23%), and either minimize inequality or maximize the minimum payoff (1.12%). No subject consistently chose options that maximize the other person's payoff. It is also noteworthy, that the goal to strictly maximize the sum of payoffs (i.e. increasing social welfare) is almost as common as the goal to strictly maximize the own payoff as narrow self-interest implies. Also, the two models appear to have about the same explanatory power on average, while the Proso model performs slightly better according to UPF and ICF, but slightly worse according to MLE. The finding that social welfare preferences are quite common is consistent with a large and increasing body of evidence (see, for instance, Au & Kwong, 2004; Andreoni & Miller, 2002; Charness & Rabin, 2002; Charness & Grosskopf, 2001; Daruvala, 2010; Engelmann & Strobel, 2004; Choshen-Hillel & Yaniv, 2012; Kritikos & Bolle, 2001), and invalidates the assumption that SVOs are predominantly expressed in terms of a preference for equality in outcomes (for diverging views, see e.g. Fehr, Naef, & Schmidt, 2006; Güth, Kliemt, & Ockenfels, 2003).

Among the one-parametric models, the PS model performs better than the ERC model. The ERC model only shows a higher mean GOF under UPF, but can only perfectly explain the choice patterns of those 15.99% subjects who simply maximize the own payoff, and is outperformed by the PS model both according to ICF and MLE, while the difference in BIC scores is substantial. The reason why the ERC model can not perfectly account for those 3 (1.12%) subjects who consistently chose options that minimize the difference between payoffs is simply that its parameter

in terms of efficiency (i.e. an item with slope one and intercept zero), which potentially -but not necessarily- could differentiate between the two strict goals. Hence, we can not differentiate between those subjects whose strict goal is to minimize inequality, and those whose strict goal is to maximize the minimum payoff. However, the models make different predictions across options which do not yield even payoff distributions, and thus we can analyze which model can better account for the choice patterns of those subjects who deviate from both of the two strict goals.

α is constraint. In order to allow for a perfect fit for those cases, we would have to let $\alpha \rightarrow \infty$, or implement the two-parametric version of the model as expressed in equation 4.10 and assign the parameter attached to the own payoff a value of zero. Needless to say, however, that these 3 subjects obtain a GOF score that is almost indistinguishable from perfect given the ERC model (under UPF, for these subjects the difference to a GOF of 100% is $10^{-4} * .208$). In any case, the ERC model is designed as a model of inequality aversion, and it is thus no surprise that it can not account very well for the choice patterns of subjects who have competitive or social-welfare preferences, while types of the latter preference class appear to be relatively common.
The analysis of GOF scores among the two-parametric models (#9-14) suggests that the PI model (#9) has by far the highest explanatory power. The reason why this model performs so well can be explained by what follows. First, it can perfectly account for all the 94 (34.94%) subjects' choice patterns that can already be perfectly explained by any of the simple non-parametric "pure" SVO models (models #1-6) as well as the one-parametric PS model (# 7)[16]. Furthermore, when we look at best fitting parameter values given the PI model under UPF, it turns out that the best fitting parameter value combination for 92 (34.2%) subjects is $\alpha = -1$ and $\beta = 1$, while for 52 out of these 92 subjects, this parameter value combination results in a perfect fit. The reason for this is that the PI model when fully specified by this particular parameter value combination is $U_{(\pi_s,\pi_o)} = \pi_s - \pi_o - |\pi_s - \pi_o|$, which yields $U_{(\pi_s,\pi_o)} = 0$ for all options where $\pi_s \geq \pi_o$, and $U_{(\pi_s,\pi_o)} = 2 * (\pi_s - \pi_o)$ whenever $\pi_s < \pi_o$, which is necessarily smaller than zero[17]. Consequently, this specification of the model predicts that a person a) chooses in accordance with narrow self-interest among options where she is worse off than the other, b) is indifferent between all options where she receives at least the same amount as the other, and c) will never choose an option that makes her worse off than the other whenever other options are available that yield equal payoffs or a higher payoff for the self. Hence, in addition to various other choice patterns, the model can also perfectly account for the choice patterns of all those 57 (21.19%) subjects who never chose an option that made them worse off than the other, but are neither perfectly competitive (maximizing relative gain), nor perfectly individualistic (maximizing own gain). A choice behavior consistent with the PI model given $\alpha = -1$ and $\beta = 1$ could be seen as the result of a heuristic[18] or preference labeled as "worse-off aversion". We think that it is reasonable to assume that some people are indeed worse-off averse, and simply choose the option that maximizes the own payoff among a set of options where $\pi_s < \pi_o$. On the contrary, it is hard to believe that these people are totally indifferent between all options of a set of options where $\pi_s \geq \pi_o$. This may be an instance where a model performs very well simply because it allows for so many predictions, and the more predictions a model allows for, the higher the likelihood that the data is consistent with one of them. However, the fact that the PI model outperforms all other models considered in our analysis is not just due to the fact that the parameter range

[16]This has to be the case, of course, for the models which are nested in the PI model, i.e. the PS model and all of the simple non-parametric SVO models except "pure" altruism and "pure" inequality aversion.

[17]Given $\alpha = -1$ and $\beta = 1$, the PI model can be rewritten as $U_{(\pi_s,\pi_o)} = \pi_s - \pi_o - \pi_s + \pi_o = 0$ *iff* $\pi_s \geq \pi_o$, and $U_{(\pi_s,\pi_o)} = \pi_s - \pi_o + \pi_s - \pi_o = 2 * (\pi_s - \pi_o)$ *iff* $\pi_s < \pi_o$.

[18]The heuristic would be as follows. If there are only options where you are worse off than the other, choose the one yielding the highest payoff for you. If there are also options that yield a payoff for you that is at least as high as the payoff for the other, choose one of these options at random.

we searched through allows for the particular parameter value combination $\alpha = -1$ and $\beta = 1$. We tested the model also under a different parameter α constraint ($-.5 \leq \alpha$), but the model still performed best according to all criteria we employ except the percentage of perfect fits, which is then slightly lower than the one yielded by the QM model[19]. Furthermore, even when parameter values are not constraint, the parameter combination $\alpha = -1$ and $\beta = 1$ fits best to the choice pattern of only 1 subject according to MLE (scatter plots of individual best fitting parameter value combinations are presented later in this paper), while the model is ranked first according to BIC too. This is evidence that the PI model is not outperforming the other models in terms of GOF just because it allows for a particular parameter value combination that allows for a multitude of predictions. Rather, the model appears to be able to reflect the rich gradation of peoples' social preferences better than the other models under consideration. For parameter value combinations of $-1 < \alpha < 0$ and $0 < \beta < 1$, the model predicts choices that are selfish among options where $\pi_s < \pi_o$, but potentially prosocial among options where $\pi_s \geq \pi_o$, while this sort of conditional prosociality[20] apparently reflects the preferences of 63 (23.42%) subjects according to MLE. By the way, the fact that the MLE method is capable of differentiating more between particular choice patterns than the UPF and ICF methods can is the reason why the agreement between MLE one the one hand and UPF & MLE on the other hand is so low for the PI model (see table 4.3).

It needs to be emphasized, though, that the selection of parameter values that are searched through has a great impact on the performance of models in terms of GOF scores, of course. We tested some models under assumptions of parameter values that are out of the range the corresponding models' authors actually proposed. For instance, we assessed the ERC and FS models' GOF scores also for parameter ranges that include negative values. Due to the fact that negative parameter values for inequality aversion models allow for explaining both competitive as well as social welfare preferences better, overall GOF scores increase substantially. However, extensions of parameter value ranges do hardly affect the ranking of models, with one exception. Allowing for $-.5 \leq \alpha \leq 4$ and $-1 \leq \beta \leq 1$ in the FS model puts the model on rank two or three, depending on the criterion[21]. Like the ERC model, the FS model was designed as a model of inequality aversion, which is a less common social preference as often assumed. Consequently, these models in their original form (i.e. implemented under the parameter constraints suggested by the corresponding authors) are too constraint to account for a considerable amount of choice patterns, which is consistent with previous findings (e.g. Engelmann & Strobel, 2004; Graf et al., 2012). Widening the parameter space thus results in higher GOF scores, but also makes the psychological interpretation of parameter values and parameter value combinations very difficult, because they are then allowed to represent several qualitatively different motivations rather than shades of only one motivation (i.e. inequality aversion) the model was designed to represent.

[19]Restricting α to be greater than -.5 yields 40.15% perfect fits and the following GOF metrics. UPF mean (std.) and median: .983 (.03) and .995; ICF mean (std.) and median: .949 (.06) and .975; MLE BIC: 7455.4.

[20]Such a choice pattern is also consistent with perfect disadvantageous inequality aversion combined with imperfect advantageous inequality aversion.

[21]Under $-.5 \leq \alpha \leq 4$ and $-1 \leq \beta \leq 1$, the FS model yields 110 (40.89%) perfect fits and the following GOF metrics: UPF mean (std.) and median: .984 (.03) and .996; ICF mean (std.) and median: .949 (.06) and .975; MLE BIC: 7594.7.

Hierarchical fitting analysis

Table 4.6 shows the results from a hierarchical analysis (a similar approach was also employed by Radzicki, 1976)[22] focusing on the number of subjects whose choice patterns can be fit perfectly or quite well according to a certain threshold when UPF is employed. The analysis is termed hierarchical since non-parametric models are considered first, and the subjects whose choice patterns can be explained perfectly or quite well according to a certain threshold are recorded and excluded from the sample. Next, the remaining subjects' choice patterns are analyzed according to the one-parametric models, and the subjects whose choice patterns can be explained perfectly or quite well according to a certain threshold are again recorded and then excluded from the sample. Those subjects whose choice patterns could still not be fit perfectly or well enough are then analyzed under the two-parametric models.

Table 4.6 can be read as follows. First, the models with the same number of parameters are grouped together. The first data column shows the absolute number (n) of subjects whose choice patterns can be perfectly (UPF GOF = 100%) explained by the corresponding model. The second column shows the number of subjects whose UPF GOF scores reach or exceed the threshold of 98% given the corresponding model. Columns three and four report analogous information for thresholds of 95%, and 90%, respectively. Data columns five to eight follow the same logic, but report the number of subjects whose choice patterns can *uniquely* be explained by the corresponding model. For example, models # 5 and 6 (Ineq and Maximin) each fit the choice patterns of three subjects perfectly. However, it can be inferred from the information in data column five that both models perfectly account for the choice patterns of the *same* three subjects, since neither of the two models can *uniquely* account for the choice pattern of any subject. The rows labeled "Total unique n (%)" show the absolute number and corresponding percentage of subjects whose choice patterns can be explained by any of the models in the corresponding group (non-parametric, one-parametric, or two-parametric) given a certain GOF threshold. Importantly, the percentages reported in these columns for the one-parametric and two-parametric model groups refer to the number of subjects remaining in the sample after those subjects who had already been explained by any of the models in the previous model group given the corresponding GOF threshold are excluded.

As already mentioned earlier, the non-parametric models can perfectly account for the choice patterns of 33.46% of the sample. This percentage can be compared to the number reported in Andreoni & Miller (2002), who found that 43% of their subjects could be very well accounted for by either the Indi, Proso, or Maximin model. Hence it seems that their sample consisted of more perfectly consistent subjects than our sample. Also, they found quite different proportions of subjects perfectly explained by the Indi (22.7 vs. 15.99%), Proso (6.2 vs 14.13%), or Maximin

[22] Unfortunately, we can not compare our results to the ones obtained by Radzicki (1976), since he fit the models to choice patterns which do not all stem from different people (i.e. some subjects ranked offers twice), and he also noted that he varied how "the other" was characterized to a decision maker. Hence, a comparison of our results to his does not make sense.

(14.2 vs. 1.12%) model. However, the conclusions are the same for both our data and the results from Andreoni & Miller (2002). More than a third of people appear to have very simple and clear SVOs, and the preference for maximizing joint gains is among them.
The choice patterns of only four additional subjects can be perfectly accounted for by the two-parametric models. However, if the GOF criterion is relaxed a bit, the models turn out to be useful in accounting for the data above and beyond what the non-parametric models can explain. Nevertheless, if we focus on perfect fits (UPF GOF = 100%), the choice patterns of the majority of subjects (65.05%) can neither be perfectly accounted for by non-parametric, nor one-parametric models. Two-parametric models are required so that the data from 41.71% of these remaining subjects can be explained perfectly. In total, the data from 62.08% of the subjects can be perfectly explained by at least one of the models we considered in our analysis. If we relax the threshold to (UPF GOF ≥ 90%), this percentage is 98.51%. This appears to be a very high figure, but one has to keep in mind that fitting any model to random data would result in a UPF GOF score of about 50% on average, since a 0% score would result from a choice pattern that is exactly the opposite of what the model predicts. Thus the 90%-threshold has to be evaluated in comparison to a 50% threshold rather than 0%, which relativizes the results obtained when thresholds < 100% are applied.

Table 4.6: Hierarchical fitting analysis of UPF results

Nr.	Label	Subjects perfectly accounted for by criterion				Subjects uniquely perfectly accounted for by criterion			
		1	.98	.95	.9	1	.98	.95	.9
1	Indi	43	47	65	79	43	47	54	26
2	Proso	38	40	49	72	38	40	47	42
3	Compet	6	6	11	53	6	6	0	0
4	Altr	0	0	0	0	0	0	0	0
5	Ineq	3	8	18	31	0	0	1	1
6	Maximin	3	13	26	64	0	5	7	8
Total unique n (%)		90 (33.46)	106 (39.41)	139 (51.67)	186 (69.14)				
7	PS	4	20	22	23	4	16	8	2
8	ERC	0	52	88	69	0	48	74	48
Total unique n (%)		4 (2.23)	68 (41.72)	96 (73.85)	71 (85.54)				
9	PI	71	53	20	8	57	1	0	1
10	FS	10	40	17	7	0	0	0	0
11	QM	16	32	20	6	2	1	0	0
12	CES	0	12	6	1	0	0	0	0
13	ERP	0	14	7	2	0	0	0	0
14	AM	11	25	14	5	0	0	0	0
Total unique n (%)		73 (41.71)	54 (56.84)	20 (58.82)	8 (66.67)				

4.4.3 Analysis of best fitting parameter values

Table 4.7 shows the correlations between all individual best fitting parameter values from all models. Also, the table shows the correlations between these best fitting parameter values and the SVO angle, which is the standard output of the Slider Measure that estimates the weight a person attaches to the outcome for the other in relation to the own[23]. Furthermore, the correlations between best fitting parameter values and the Inequality Aversion (IA) index, ranging from 0 (perfect joint gain maximization) to 1 (perfect inequality aversion), as produced by the Slider Measure are shown[24] (for further details on the Slider Measure's outputs, see Murphy et al., 2011). The two Slider Measure outputs were included in this analysis because they have a clear psychological meaning, and thus may help identify adequate psychological interpretations of the tested models' parameters. As can be seen, the SVO angle is a very good proxy for parameter α from the PS model (#7)[25]. Using the Slider Measure's outputs as standards for interpretation, the ERP (#13) and AM (#14) models' parameters appear to have a straight forward psychological meaning, such that parameter α indicates the weight a person attaches to the outcomes for another person in relation to the own, and parameter β indicates whether a person is concerned more about social welfare maximization or equality in payoffs. For the other two-parametric models, the psychological meaning of a single parameter is less clear. For these models, particular parameter value combinations appear to be the carriers of psychological meaning, which makes the interpretation of one parameter value in isolation of the other parameter's value difficult. For these models, it seems that the meaning of the value of one parameter is dependent on the value of the other parameter. This holds in particular, of course, for the models where one or the other parameter is determining an option's utility, depending on whether the respective option satisfies $\pi_s \geq \pi_o$ or $\pi_s < \pi_o$, as is the case for the FS model (#10) and the QM model (#11). Considering both the models' GOF scores and best fitting parameter values, we find that the AM model (#14) provides a good compromise between the ability to account for the data, and the provision of parameter values that allow for a relatively clear psychological interpretation.

Figure 4.3 displays the distribution of best fitting parameter values for the one-parametric models, and scatter plots of best fitting parameter value combinations for the two-parametric models. The size of a circle in the scatter plots in figure 4.3 is proportional to the number of subjects with the same corresponding best fitting parameter value combination. We will not discuss all the patterns in detail, since this would exceed the scope of this paper. However, one thing that becomes apparent immediately when looking at figure 4.3 is the large heterogeneity in peoples'

[23]The SVO angle is computed as $SVO° = \arctan\left(\frac{(\bar{A}_o - 50)}{(\bar{A}_s - 50)}\right)$, where $\bar{A}_o$ is the mean payoff allocated to the other, and $\bar{A}_s$ is the mean payoff allocated to the self across the six primary items of the Slider Measure.

[24]Notice that the Slider Measure outputs an IA index only for those people who have been categorized as prosocial according to their choices in the measure. The correlations with the IA index are therefore computed on the basis of the subsample of those 152 (56.51%) subjects who consistently made prosocial choices in both the primary and secondary items of the Slider Measure.

[25]This is quite impressive since the SVO angle results from a relatively simple scoring of choices in the six primary items only.

Table 4.7: Pearson correlations between best fitting model parameters as obtained by MLE

Nr.	Model parameters		SVO	IA	PS	ERC	PI		FS		QM		CES		ERP		AM	
			angle	index	α	α	α	β	α	β	α	β	α	β	α	β	α	β
-	SVO angle		1.00	-0.30	0.92	0.60	0.53	0.27	-0.06	0.85	0.19	0.79	-0.93	-0.36	0.92	-0.32	0.89	0.20
-	IA index		-0.30	1.00	-0.12	0.33	-0.51	0.87	0.39	0.10	-0.66	0.64	0.06	-0.48	-0.26	-0.73	-0.26	0.79
7	PS	α	0.92	-0.12	1.00	0.65	0.58	0.25	-0.15	0.90	0.24	0.82	-0.95	-0.36	0.94	-0.31	0.91	0.18
8	ERC	α	0.60	0.33	0.65	1.00	0.39	0.37	-0.13	0.79	0.13	0.68	-0.65	-0.45	0.69	-0.34	0.65	0.20
9	PI	α	0.53	-0.51	0.58	0.39	1.00	-0.46	-0.74	0.53	0.83	0.37	-0.54	0.05	0.62	0.16	0.69	-0.44
		β	0.27	0.87	0.25	0.37	-0.46	1.00	0.61	0.37	-0.76	0.53	-0.31	-0.51	0.24	-0.58	0.18	0.72
10	FS	α	-0.06	0.39	-0.15	-0.13	-0.74	0.61	1.00	-0.13	-0.79	0.02	0.12	-0.22	-0.18	-0.34	-0.22	0.52
		β	0.85	0.10	0.90	0.79	0.53	0.37	-0.13	1.00	0.16	0.81	-0.90	-0.48	0.92	-0.41	0.85	0.22
11	QM	α	0.19	-0.66	0.24	0.13	0.83	-0.76	-0.79	0.16	1.00	-0.01	-0.21	0.26	0.29	0.35	0.38	-0.67
		β	0.79	0.64	0.82	0.68	0.37	0.53	0.02	0.81	-0.01	1.00	-0.84	-0.37	0.76	-0.39	0.83	0.40
12	CES	α	-0.93	0.06	-0.95	-0.65	-0.54	-0.31	0.12	-0.90	-0.21	-0.84	1.00	0.40	-0.95	0.38	-0.91	-0.21
		β	-0.36	-0.48	-0.36	-0.45	0.05	-0.51	-0.22	-0.48	0.26	-0.37	0.40	1.00	-0.41	0.70	-0.29	-0.40
13	ERP	α	0.92	-0.26	0.94	0.69	0.62	0.24	-0.18	0.92	0.29	0.76	-0.95	-0.41	1.00	-0.33	0.92	0.14
		β	-0.32	-0.73	-0.31	-0.34	0.16	-0.58	-0.34	-0.41	0.35	-0.39	0.38	0.70	-0.33	1.00	-0.23	-0.53
14	AM	α	0.89	-0.26	0.91	0.65	0.69	0.18	-0.22	0.85	0.38	0.83	-0.91	-0.29	0.92	-0.23	1.00	0.05
		β	0.20	0.79	0.18	0.20	-0.44	0.72	0.52	0.22	-0.67	0.40	-0.21	-0.40	0.14	-0.53	0.05	1.00

social preferences. There is hardly any clear clustering, but generally widely dispersed individual best fitting parameter value combinations. This is further evidence that a simple dichotomous or trichotomous categorization of people into classes such as competitive, selfish, or prosocial results in far too coarse a representation of peoples' SVOs.

Unfortunately, there is not much data available on individual parameter estimates from past research for most models. However, we can compare our results regarding the parameters from the FS model to previous results from Blanco et al. (2011). In agreement with their findings, our results do not support the assumptions by Fehr & Schmidt (1999) that α and β are positively correlated[26], and that $\alpha \geq \beta$. Furthermore, their scatter plot of best fitting individual parameter value combinations (see Blanco et al., 2011, p .327) looks very similar to ours. However, we find quite different distributions of single parameter values, in particular for parameter α. The comparison is shown in table 4.8. We suspect that the deviations are due to differences in the nature of stimuli used for eliciting parameter value estimates in the study by Blanco et al. (2011) compared to ours. Furthermore, we find quite different joint parameter distributions than originally assumed by Fehr & Schmidt (1999) as shown in table 4.9. The deviations are mainly due, we think, to the fact that the original assumption of $\alpha \geq \beta$ does not seem to hold, as already found by Blanco et al. (2011). We find that for 76.95% of the subjects, $\alpha \leq \beta$ holds, while this is in part due to the many subjects with best fitting $\alpha = 0$.
As for the other models' best fitting parameter values, we do not know of any data from previous studies our results could be directly compared to. We are confident, though, that our reports will serve as reference data for future research.

Table 4.8: Distribution of α and β as assumed in Fehr & Schmidt (1999) [FS] compared to observations from Blanco et al. (2011) [BEN] and our data (AM) in percent

α	FS	BEN	AM
$\alpha < .4$	30	31	72
$.4 \leq \alpha < .92$	30	33	8
$.92 \leq \alpha$	40	36	20
β	FS	BEN	AM
$\beta < .235$	30	29	28
$.235 \leq \beta < .5$	30	15	27
$.5 \leq \beta$	40	56	46

[26]Our results suggest that the correlation -if it is not zero- tends to be negative, which is supportive of the findings from Daruvala (2010).

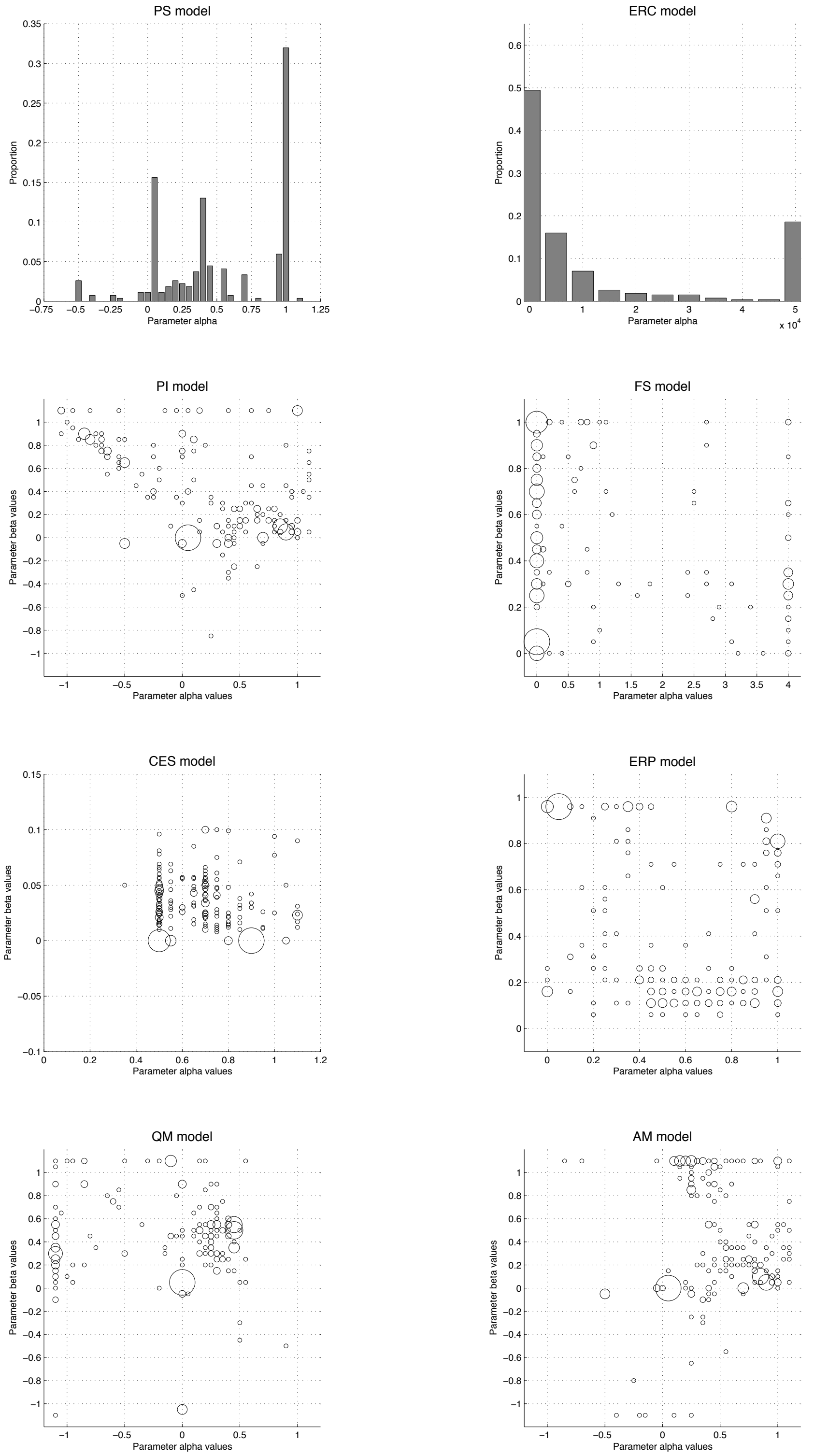


Figure 4.3: Histograms of best fitting parameter value distributions for the one-parametric models, and scatterplots of best fitting parameter value combinations for the two-parametric models as obtained by MLE. Circle radii in the scatterplots correspond to the number of observations.

Table 4.9: Assumed distribution of joint parameter values (in percent) by Fehr & Schmidt (1999) [FS] in comparison to our data [AM]

Joint parameter values	FS	AM
$\alpha < .5$, $\beta < .25$	30	22
$.5 \leq \alpha < 1$, $.25 \leq \beta < .6$	30	2
$1 \leq \alpha$, $.6 \leq \beta$	40	5

(See Fehr & Schmidt, 1999, p.844, p. 864)

4.5 Conclusion

We have compared fourteen different SVO models in terms of their goodness of fit to individual choice data from 269 subjects according to three different fitting methods and reported on resulting individual best fitting parameter values. One major conclusion that we can draw from our findings is that people vary widely in their preferences for own-other outcome allocations. Some people appear to have preferences that can be described by a simple rule, such as maximizing own payoffs, maximizing relative gain, maximizing joint earnings, or minimizing differences in payoffs. Other people appear to have more complicated rules, as expressed, for instance, by a willingness to forgo own gains for the benefit of another person, but only under the condition that oneself is still better off. Yet other people appear to have preferences that cannot be readily verbalized, but rather have to be expressed in terms of particular parameter value constellations in more complicated two-parametric SVO models. Consequently, as we have seen, models that are designed to account for expressions of only one particular SVO are insufficient to adequately describe the choice patterns of a significant proportion of people. Furthermore, we demonstrated that those models that yield the highest explanatory power are not necessarily the ones the parameters of which are most easily psychologically interpretable. We find that a good SVO model should not only fit the data reasonably well, but should also be psychologically interpretable. The next step in constructing SVO models will be to incorporate peoples' beliefs about other peoples' preferences, and people's beliefs about other peoples' beliefs into a given utility function. This will necessarily result in highly complicated models, and if it is already difficult to interpret the basic model a more complicated one is based on, a psychologically meaningful interpretation of such a model's outcomes may become infeasible. We have pointed to a model which yields a relatively good compromise between goodness of fit and psychological interpretability of parameter values, and we suggest that a model of that kind could be useful for building upon in order to develop models which incorporate beliefs about others in a next step.

5 Explaining behavior in public goods games: How preferences and beliefs affect contribution levels

Abstract[1]

There is a large body of evidence showing that a substantial proportion of people contribute positive amounts in public goods games, even if the situation is one-shot and completely anonymous. Clearly, this is in conflict with the prediction of neoclassic economic theory. One of the most promising explanations of why people contribute anything in this context draws upon an interaction between positive social preferences and beliefs about the preferences and anticipated behavior of others. We follow this line of thinking and investigate the predictive power of social preferences and beliefs on contribution levels in both a one-shot and a repeated linear public goods game. We report on the degree to which individual contributions can be explained when individual preferences and beliefs are taken into account, and additionally how preferences and beliefs change in response to the behavior of others.

5.1 Introduction

From the vantage of neoclassic economic theory, cooperation between humans is difficult to explain. Within the normative framework, decision makers (DMs) are conceptualized as narrowly self-interested agents who believe that other DMs are also narrowly self-interested, and that these preferences are common knowledge. These axioms yield powerful, precise, and testable models of behavior in interactive decision contexts. In some cases these axioms yield predictions that are remarkably accurate, whereas in other cases these models are woefully inadequate (for comparative illustrations of both instances, see e.g. Camerer, 2003; Camerer & Fehr, 2006; Camerer & Thaler, 1995). We consider here an instance of the latter case, namely decision

[1] This chapter is an edited version of the following paper.
Murphy, R.O. and Ackermann, K.A. (2014). Explaining behavior in public goods games: How preferences and beliefs affect contribution levels. Manuscript in preparation.

makers' propensity to cooperate in a social dilemma, even when the situation is one-shot and anonymous.

One reason why neoclassic economic theory is unable to explain cooperation in such situations is that it relies on the assumption of complete homogeneity among DMs, i.e. the negation of variance in DMs' tastes and beliefs. We address this shortcoming by extending the normative axioms in three important ways. First, we model the DMs as having a range of possible social (i.e. other-regarding) preferences. This means that a DM's utility is modeled as a joint function and is potentially influenced by the payoffs other DMs receive. This replaces the "selfishness axiom" (Henrich et al., 2005) with a distribution of DM types, where narrow self interest is just one special case of a social preference. Second, we allow DMs to have beliefs about the social preferences of the other decision makers. This replaces the notion that all DMs believe other DMs are perfectly selfish (one part of the common knowledge axiom), with a more complex but still tractable distribution of beliefs about others' preferences. Thirdly, we model DMs as having expectations that other DMs will act in potentially cooperative ways, that is, we allow for heterogeneity in the DMs' beliefs about other DMs' behavior. Again, this conceptualization replaces the normative assumption of complete homogeneity among DMs, and takes into account that DMs can differentiate between the other DMs' intrinsic motivations and their actual (potentially extrinsically motivated) behavior. This allows cooperative DMs to coordinate with other cooperative DMs and achieve more efficient outcomes.

We show that applying these model extensions enables us to explain more than 50% of the variance in real DMs' levels of cooperative behavior in a n-person social dilemma with complete anonymity and no repetition. Furthermore, we investigate how DMs preferences and beliefs affect contribution levels in a repeated interaction, and provide evidence that DMs do not only update their beliefs, but also that DMs' preferences change in response to the observed behavior of other DMs.

5.2 Cooperation in social dilemmas

How people trade off personal benefit against utilitarian outcomes in situations of conflicting interests has been studied extensively across different disciplines (e.g. Dawes, 1980; Fehr & Fischbacher, 2003; Hamilton, 1964; Kollock, 1998; Nowak, 2006; Ostrom, 2000; Sell, Tooby, Cosmides, & Orians, 2009; Tricomi, Rangel, Camerer, & O'Doherty, 2010), yet the puzzle of human cooperation remains an active area of experimental (e.g. Fehr & Gächter, 2000a; Fischbacher & Gächter, 2010; Fischbacher et al., 2001; Frey & Meier, 2004; Gächter & Herrmann, 2009) and theoretical (e.g. Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999; Geanakoplos et al., 1989; Levine, 1998; Rabin, 1993; Cox et al., 2007) research. The best known instance of a social dilemma situation is the Prisoner's Dilemma (PD, see, e.g. Luce & Raiffa, 1957). Typically, substantial cooperation rates (see e.g. Andreoni & Miller, 1993; Cooper et al., 1996; Rapoport & Dale, 1966) are observed when people perform a PD task, even if the situation is one-shot and anonymous (for a general overview, see for instance Dawes, 1980). This also holds for the $n > 2$ extension of the PD, namely the standard linear public goods game (PGG).

Several reasons have been hypothesized as explanations for why people choose strictly dominated strategies in these contexts. Generally, levels of cooperation can depend on exogenous and endogenous factors (for a review on PGGs in this respect, see Ledyard, 1995). For example, the payoff structure of the game, such as the index of cooperation (K) in a PD (see Rapoport, 1967) or the marginal per capita return in a PGG (see Isaac, Walker, & Thomas, 1984) can have a significant effect on cooperation rates in social dilemmas. However, these exogenous factors alone can not explain why people cooperate if they are assumed to be homogeneous with respect to endogenous variables, as is the case in standard economic theory. That is, no matter how small the difference is between the Temptation payoff and the Reward payoff in a PD, or how large the marginal per capita return (MPCR) is in a PGG (as long as MPCR < 1 holds, of course), free riding remains to be the strictly dominant strategy for all DMs. In other words, under the assumption of narrow self-interest and common knowledge of narrow self-interest, altering game parameters such as the K-index in a PD or the MPCR in a PGG should have no effect at all on the DMs' behavior in a social dilemma. Only if DMs are assumed to derive utility not exclusively from their own payoff, but at least in part also from the payoffs others receive (i.e. if DMs have non-selfish distributive social preferences), altering game parameters can be expected to affect behavior. Consider, for instance, a DM who has distributive social preferences that are consistent with efficiency maximizing, i.e. the DM's utility is the sum of the own payoff and the interaction partner's payoff, and who believes that the interaction partner is narrowly self-interested (for simplicity, assume that the DM believes that the interaction partner is going to defect with probability $p = 1$). In a standard symmetric PD given this constellation of the DM's preferences and beliefs, the DM will choose to defect as long as $2P > S + T$, but will choose to cooperate if the game parameters are changed such that $2P < S + T$.[2] Now assume the DM believes (for simplicity again with $p = 1$) that the interaction partner -for whatever reason- is going to cooperate. In this case, the DM will choose to cooperate as well only if $2R > S + T$, but will defect in case the game parameters are such that $2R < S + T$. The example makes clear that changing game parameters can affect behavior only in the presence of social preferences. Under the assumption of narrow self-interest, altering the payoff structure of a social dilemma can not have an effect on behavior, even if DMs are assumed to have diverse beliefs about the intentions and behavior of others, i.e. even if the axiom of common knowledge of narrow self-interest is allowed to be violated.

Hence, exogenous factors can influence choice behavior in social dilemmas only if there are endogenous variables which the exogenous factors can act upon. Consequently, understanding how endogenous variables affect cooperation behavior in social dilemmas is a prerequisite for understanding how exogenous variables do so. Ledyard (1995, p. 143) provides a list of several endogenous (he uses the term "systemic") variables that have a potential impact on cooperation

[2] P, R, S, and T stand for *P*unishment payoff (resulting from mutual defection), *R*eward payoff (resulting from mutual cooperation), *S*ucker payoff (resulting from unilateral cooperation), and *T*emptation payoff (resulting from unilateral defection), respectively.

behavior in PGGs. We will focus on two of these variables, namely 1) social preferences ("altruism, fairness" in Ledyard's terminology) and 2) beliefs as predictors of contribution levels in PGGs for the remainder of this paper.

5.2.1 How social preferences and beliefs affect cooperation behavior in PGGs

Clarifying the concepts

The question how well social preferences and beliefs can account for cooperative behavior in strategic interactions such as linear[3] PGGs has received substantial attention in the past decades. Moreover, apart from social preferences, terms such as reciprocity or conditional cooperation are frequently used in this context. However, we find that these concepts have not been disentangled and defined clearly enough in the past. Therefore we introduce the following definitions for these concepts.

First of all, we follow the line of thinking that individual behavior is a function of personal and situational characteristics (see e.g. Gintis, 2007; Hedström, 2005), while preferences and beliefs constitute the relevant personal characteristics, and constraints as well as opportunities constitute the situational characteristics. Within this framework, we think that the situational characteristics do not primarily affect behavior directly, but do so indirectly by affecting personal preferences and beliefs the combination of which then finally determines behavior. Consequently, if a person's preferences and beliefs in a given situation are known, it should be possible to accurately predict that person's behavior. This is the rationale of the basic framework of thinking we follow throughout this paper.

Furthermore, we make distinctions between the concepts of social preferences, reciprocity, and conditional cooperation. We define social preferences broadly as the degree and direction of care about interaction partners' outcomes in relation to one's own. The most basic type of social preferences is simple baseline distributive preferences, that is, preferences over different allocations of resources between the self and others in a one-shot situation under complete anonymity, such as a standard dictator game, for instance. Simple baseline distributive preferences can be represented by utility functions that exclusively involve terms reflecting the own outcome, others' outcomes, and combinations thereof.

As for reciprocity, we define this concept similar to how it has been defined by Cox (2004, p. 263). Concretely we understand reciprocity as the change in a decision maker's distributive preferences in response to revealed or learned characteristics of the interaction partner. Conditional cooperation, in contrast, is more ambiguous in terms of what it could be supposed to mean.

[3] We will not discuss findings from studies with a focus on step-level PGGs since we think that the introduction of thresholds changes the situation in important ways. In step-level PGGs, contributing can be the payoff maximizing strategy given certain beliefs, while in linear PGGs this is never the case. Hence, thresholds make the situation even more complicated and we find it reasonable to first understand behavior in a simpler situation before turning to more complicated settings.

Fischbacher, Gächter, & Fehr (2001, p. 397) give the following description:

> Conditional cooperation can be considered as a motivation in its own or be a consequence of some fairness preferences like ‘altruism’, ‘warm-glow’, ‘inequity aversion’ or ‘reciprocity’.

On the one hand, we agree with this statement in principle, but on the other hand we find it unsatisfying since it blurs the distinction between preferences and behavior, and therefore also between cause and effect. Obviously, this ambiguity is problematic. There are several reasons why a decision maker may behave in a way that is consistent with conditional cooperation. Firstly, the DM may just have the preference to match the interaction partner’s behavior when it is know, or match the interaction partner’s expected behavior when it is uncertain. In this case, conditional cooperation could indeed be considered a motivation in its own. Secondly, the DM may have distributive preferences consistent with inequality aversion, while in many games (including a linear PGG) inequality in outcomes is minimized if all players choose the same strategy. In this case, conditional cooperation would have to be considered a consequence of simple distributive preferences. Thirdly, the DM’s distributive preferences may change in response to the interaction partner’s (believed) behavior, such that the observation of conditionally cooperative behavior would have to be considered a consequence of reciprocity as defined above. Finally, conditionally cooperative behavior may also be shown by a DM who has prosocial distributive preferences combined with the motivation to avoid being the one who is taken advantage of (sucker aversion). In this case, conditional cooperation would have to be considered a consequence of a combination of different motivations. What complicates the matter even more is that for different decision makers, a conditionally cooperative behavior may be expressed for different reasons. Consequently, from the mere observation of conditional cooperation as a behavior it is not possible to infer its cause. Therefore, we employ a cautious definition of conditional cooperation as a behavior which expresses itself through a positive relation between a decision maker’s level of cooperativeness and the interaction partners’ observed or expected level of cooperativeness.[4]

To summarize, we use the term social preferences to refer to a set of motivations involving consideration of the welfare of others in relation to the own, including simple distributive preferences as well as reciprocity, and use the term conditional cooperation to refer to a specific behavioral pattern defined above.

[4]We are aware that other researchers employ a different definition and consider conditional cooperation as a preference in its own (see, for instance, Chaudhuri, 2011, p. 56).

Knowns and unknowns about the role of social preferences and beliefs in linear PGGs

A note on historical developments. Although social preferences are a prerequisite for potential effects beliefs may have on behavior in PGGs[5], research on the effect of beliefs on contribution levels has been done before light was shed on how social preferences affect contribution levels. In Ledyard's review (1995, p.143), the variable "Beliefs" was assigned a plus sign, indicating a (weak) positive effect on contributions, while the variable "Altruism, fairness" was assigned a question mark, meaning that these variables had not even been measured to that date. Ledyard mentioned one study in which the question of how beliefs affect behavior in a linear PGG had been addressed (Dawes, McTavish, & Shaklee, 1977).[6] However, insights from that study with respect to effects of beliefs on behavior in these contexts were limited to the observation that more cooperative individuals also expect more cooperation from others than uncooperative individuals do. As Dawes (1980) summarized, positive correlations between beliefs about others' cooperativeness and actual cooperation in n-person PDs had also been found by Tyszka & Grzelak (1976) and Marwell & Ames (1979, see p. 1356). This relation between beliefs and cooperation has been replicated several times to date and is consistent with our definition of "conditional cooperation" (e.g. Burlando & Guala, 2005; Fischbacher & Gächter, 2010).

The investigation of how social preferences affect cooperation levels has begun later and has often suffered from methodological flaws. Concretely, researchers have often tried to infer social preferences from behavior in strategic games without taking into account that peoples' behavior in proper games is a function of the decision makers' preferences entangled with their beliefs about the behavior of others and that it is impossible to disentangle these two factors *ex post*, i.e. on the basis of information about peoples' behavior in strategic interactions. In order to assess people's preferences alone, i.e. unconfounded by beliefs, it is necessary to observe their behavior in non-strategic situations. The recognition of this claim has been made explicit in game theoretic research in psychology and has led to a particular conceptualization of distributive social preferences termed *social value orientation* (SVO).

Distributive social preferences. The roots of the SVO construct date back to the late 1960's and developed from the quest to disentangle strategic considerations from intrinsic motivations as determinants of behavior in situations of interdependence, such as a PD, for instance. Messick & McClintock (1968) and Pruitt (1967) independently devised *decomposed games* which can be used for quantifying the degree of peoples' intrinsic concerns for the welfare of others in relation to their own. Concretely, a decomposed game is a decision task requesting a person to decide how to allocate resources between the self and another person by choosing one out of at least two predefined self-other resource allocation options. Hence it is a dictator game,

[5] In the absence of concerns for others, i.e. if all decision makers are perfectly selfish, no decision maker would be expected to contribute in a PGG, no matter what he believes others are going to do.

[6] With respect to our focus on standard linear PGGs, results from this study have to be interpreted with caution. Although no threshold of public goods provision was applied in this study, the setting deviated from a standard PGG in that subjects could only choose to cooperate or defect rather than choosing a contribution level (corresponding to an all-or-nothing contribution decision) and the study also included a loss-frame condition.

but without a constant-sum restriction, i.e. budgets may vary across options. From a person's choices in such decision tasks her social motivation can be determined consistent with the idea of revealed preferences (Samuelson, 1938; for evidence that preferences revealed by choices in dictator game tasks can be rationalized and therefore can be described by utility functions, see Andreoni & Miller, 2002). Consider, for example, a person who is presented with several such decomposed games, and in each of them chooses the option that allocates the highest amount of resources to the self, regardless of what amount of resources is allocated to the other person. Such a choice pattern would be indicative of an individualistic (i.e. selfish) preference, that is, a motivation to maximize own gains. However, another person presented with the same task may choose those options which maximize joint gains, indicating a prosocial motivation, etc. Although theoretically an infinite number of such motivations - reflecting the rich gradation of individual differences in mode and intensity of concerns for others - exist (see, for instance Griesinger & Livingston, 1973; Radzicki, 1976; Wyer, 1969), the most commonly used measures of SVO in psychology produce only very coarse outputs. Concretely, people usually are simply categorized into one out of three (see, e.g. Van Lange et al., 1997) or eight (Liebrand, 1984) motivational classes or types. The limitations inherent to such low-output measures may be one reason why the psychological SVO construct has not gained much attention in adjacent scientific disciplines, such as economics. It is only very recently that SVO has become acknowledged in that field and that instead of inferring distributive social preferences from behavior in strategic situations (e.g. Andreoni, 1995; Bardsley & Moffatt, 2007; Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999; Goeree et al., 2002; Palfrey & Prisbrey, 1997), independent measures of social preferences have been used - albeit rarely - in that field (e.g. Burlando & Guala, 2005; Offerman, Sonnemans, & Schram, 1996; Van Winden, Van Dijk, & Sonnemans, 2008)[7]. Hence, for evaluating the impact of SVO on behavior in PGGs, we will mainly refer to the psychological literature.

In general, SVO has been shown to be a significant predictor of behavior in social dilemmas (e.g. Balliet, Parks, & Joireman, 2009; Murphy, Ackermann, & Handgraaf, 2011). With respect to PGGs, Balliet et al. report an SVO-cooperation effect size of r = .29. They could further demonstrate that when subjects are paid for their decisions, SVO is less -but still- predictive of behavior and they found no evidence that SVO is less predictive in one-shot as compared to iterated social dilemma situations. One reason why SVO may be less predictive of behavior when choices are incentivized could be that information about incentives alter people's beliefs about the other people's behavior - and maybe even beliefs about other peoples' beliefs about peoples' behavior, and so on - which in turn affects their own behavior and weakens the relation between SVO and cooperation decisions. Moreover, there is evidence that SVO and beliefs about other peoples' behavior are not independent from each other. For example, prosocial people are more likely to expect cooperation from others (Kuhlman & Wimberley, 1976; Van Lange & Liebrand, 1991; Van Lange, 1992). Furthermore, prosocial people have been shown to be more likely to reciprocate cooperative behavior than individualistic (a.k.a. selfish) people (De Cremer & Van

[7] Offerman et al. (1996, p. 818) explicitly point out the importance of an independent measure of social preferences for precisely the reasons brought out above.

Lange, 2001; Kanagaretnam et al., 2009; Van Lange, 1999; Van Lange & Semin-Goossens, 1998).

With respect to how SVO relates to beliefs about other people's SVO, evidence is less clear. In general, there exist two competing propositions about the association of SVO with beliefs about SVO: The false consensus hypothesis (Ross, Greene, & House, 1977), and the triangle hypothesis (Kelley & Stahelski, 1970). While the false consensus hypothesis predicts that people expect others to have a similar SVO as they themselves have, the triangle hypothesis assumes that only proself people (individualists and competitors) have expectations in accordance with the false consensus bias, whereas prosocial people would be aware of the large variance in SVO in the population. There is both evidence supporting the false consensus hypothesis (e.g. Iedema & Poppe, 1994b; Kuhlman & Wimberley, 1976; Liebrand et al., 1986; Liebrand, 1984) and the triangle hypothesis (e.g. Maki & McClintock, 1983), and more recent results indicate that both hypotheses are in some sense true, such that people do expect others to have an SVO similar to their own, but that people with an individualistic orientation expect less variance in other people's SVO compared to people with a prosocial or competitive orientation (Aksoy & Weesie, 2012). To date, it remains an open question whether the beliefs people have about other peoples' social preferences affect their behavior in strategic situations, such as a PGG. However, since beliefs about SVO are evidently not independent from SVO, and SVO evidently affects behavior in strategic interactions, it is not unreasonable to assume that beliefs about SVO do so as well.

Beliefs. While -to our knowledge- no study exists that investigated the relationship between beliefs about other peoples' social preferences and one's own behavior in PGGs, some studies have been conducted for investigating the relationship between own behavior and beliefs about other peoples' behavior in PGGs. Although concepts such as reciprocity, or fairness have been discussed in economics for decades (e.g. Rabin, 1993; Sugden, 1984), the hypothesis of conditional cooperation as a specific behavioral pattern in PGGs had been stated relatively recently (Keser & Van Winden, 2000)[8]. In line with this hypothesis, direct evidence that subjects' contributions in a PGG correlate positively with expected contributions of others was found by Croson (2007), Neugebauer et al. (2009), and Fischbacher & Gächter (2010). Croson found that this aggregate result is due to the behavior of allmost all (about 92%; Croson, 2007, p. 207) subjects, indicating that the vast majority of people condition their contribution levels on the (expected) contribution levels of others in a linear way. However, results from other studies suggest that people are heterogeneous with respect to how contribution levels of others affect own contribution amounts in a PGG. For example, Fischbacher et al. (2001) found that only about 50% of subjects condition their contributions on the contribution levels of others in a linear way, while 30% free ride no matter how much others contribute, and about 14% are so called "hump-shaped contributors" who -roughly speaking- match the contribution levels of others for average contribution amounts below 50% of the endowment, and decrease their own contribution

[8]It is worth noting, however, that Weimann (1994) already observed that some people stop cooperating when they see that others do not cooperate sufficiently. He termed the motive guiding this kind of behavior "exploitation aversion" (see Weimann, 1994, p. 198).

levels for average contribution amounts above 50% of the endowment (see Fischbacher et al., 2001, p. 401). Further, they found that those subjects termed “conditional cooperators” do not match the average contribution level of others exactly, but tend to contribute a little less than this average. The authors hypothesize that this may be the reason why contributions decline over time in repeated PGGs - a common and disheartening finding (e.g. Andreoni, 1988, 1995; Burlando & Hey, 1997; Brandts & Schram, 2001; Croson, 1996; Carpenter, 2007; Fehr & Gächter, 2000a; Isaac et al., 1984, 1985; Isaac & Walker, 1988; Keser & Van Winden, 2000; Kim & Walker, 1984; Laury et al., 1995; Weimann, 1994). However, Fischbacher et al. did not elicit peoples' beliefs about other peoples' contributions, but instead used the strategy method (Selten, 1967) and therewith assessed how much people are willing to contribute given specific contribution levels of others. Extending this approach, Fischbacher & Gächter (2010) conducted a similar study, but in addition to assessing subjects' propensity to cooperate conditional on the cooperation of others by applying the strategy method, they also elicited subjects' beliefs about the contribution levels of others. In line with results from other studies (Croson, 2007; Neugebauer et al., 2009), they find that subjects' beliefs about contributions of others are strongly correlated with own contributions on the aggregate level, but also that there is heterogeneity among subjects in this respect, i.e. not all people are conditional cooperators. Also, they could demonstrate that a large proportion of those subjects who do condition their contributions on the contribution levels of others are *imperfect* conditional cooperators. That is, they consistently contribute a little less than they expect others to contribute - a finding which is in accordance with the results reported by Fischbacher et al. (2001). With respect to the formation of beliefs, they found that subjects' beliefs about contribution levels of others in round t can best be modeled as a weighted average of beliefs in round t_{-1} and actual contribution levels in round t_{-1}. From these results Fischbacher & Gächter conclude that the reason why cooperation rates in repeated PGGs decline is because many people are imperfect - i.e. biased - conditional cooperators, who persistently contribute less than they believe others do, which in turn leads to a downward adjustment of beliefs about contribution levels in the next round by others, resulting in a downward spiral culminating in end-stages of almost perfect mutual defection[9]. This explanation for the decline of cooperation in repeated PGGs is also supported by data from Neugebauer et al. (2009), which provides further evidence that many people are selfish-biased conditional cooperators. One issue that has not been clarified to date refers to the question of how well people can predict the contribution levels of others, that is, the congruence between beliefs about contribution levels and actual contribution levels. While Kachelmeier & Shehata (1997) found that peoples' beliefs of other peoples' contributions are too pessimistic, Neugebauer et al. (2009) report results indicating that the opposite is the case. However, optimistic beliefs could not hinder the mechanism that leads to the decline of cooperation in repeated PGGs suggested by Fischbacher & Gächter, but only slow it down. Pessimistic beliefs, on the other hand, would speed it up. Yet, how accurate people's beliefs about other peoples' contributions are remains an open question.

[9]Note that this also means that free-riders do not cause the decline of cooperation per se, but do accelerate it.

5.2.2 Connecting the lines of research

In summary, there seems to be agreement across disciplines that both social preferences and beliefs matter substantially in PGGs. With respect to social preferences, psychologists have first and foremost focused on the SVO construct and have produced a substantial amount of evidence that this construct is predictive of individual contribution levels in PGGs (see the large number of studies investigated in the meta-analysis by Balliet et al., 2009). Economists, on the other hand, have not focused on one particular conceptualization of social preferences, but have investigated a multitude of different models thereof, such as "warm-glow" (Andreoni, 1990), altruism (see Ledyard, 1995), or inequality aversion (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000). While some researchers found support for some of these model, others found support for different ones (see, for instance Ashley et al., 2010; Croson, 2007; Palfrey & Prisbrey, 1997), thus there is no consensus to date about which model does best at explaining data, and one reason for this may be that none of them seems to do satisfactorily well. However, there is clear consensus that *some form of kindness* (Andreoni, 1995, p. 900) is responsible for cooperation in PGGs (e.g. Andreoni, 1990, 1995; Ashley et al., 2010; Bardsley & Moffatt, 2007; Brandts & Schram, 2001; Burlando & Guala, 2005; Croson, 2007; Goeree et al., 2002; Palfrey & Prisbrey, 1997; Weimann, 1994). It is noteworthy at this point what Brandts & Schram (2001) (p. 414) wrote after having considered several economic models as explanations for their data:

> A form of cooperation that is consistent with all of our findings can be found in the classical psychological literature on differences in individual 'value orientation' [...].

However, as mentioned earlier in this paper, the SVO construct has widely been ignored in the economic sciences to date. One reason for this may be that, until very recently, all psychometric measures of SVO that have been available yielded the lowest output resolution possible, i.e. results on the nominal scale level of measurement (Stevens, 1946), hence the measures have probably appeared unattractive for researchers accustomed to parameterized modeling. The necessity of good measures at the individual level has long been acknowledged though[10]. The advent of a high resolution measure (see Murphy et al., 2011) may help to improve interdisciplinary connectivity and encourage the use of the SVO construct by economists.

Evidence strongly suggests that -on top of social preferences- beliefs about the behavior of others are highly associated with own behavior in PGGs (Croson, 2007; Neugebauer et al., 2009; Fischbacher & Gächter, 2010) and has led economists to build a substantial literature on a phenomenon that is supposed to be driven by both beliefs and preferences, has not only been observed in the lab but also in the field (Frey & Meier, 2004), and seems to be global

[10] For example, referring to phenomena observed in PGGs, Palfrey & Prisbrey (1997, p. 843) wrote:

> Given the considerable amount of heterogeneity of behavior across subjects that is known to be characteristic of these experiments, improved measurement at the individual level would seem to be a necessary ingredient to reaching a better understanding of these phenomena.

(Hermann & Thöni, 2009; Kocher et al., 2008): *conditional cooperation* (Fischbacher et al., 2001; Fischbacher & Gächter, 2010; Gächter, 2007; Keser & Van Winden, 2000; Kurzban & Houser, 2005; Neugebauer et al., 2009).

Finally, the question whether beliefs about other peoples' preferences -on top of beliefs about others' behavior - affect peoples' behavior in PGGs remains unanswered to date. However, there is accumulating evidence that people do take into account the intentions of others when making decisions in interdependent situations (Cox, Friedman, & Gjerstad, 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Falk, Fehr, & Fischbacher, 2008; Levine, 1998; Rabin, 1993; Stanca, Bruni, & Corazzini, 2009). For instance, Levine (1998) proposed a model which not only incorporates a parameter a_i, reflecting the weight player i attaches to the outcome of another player j, but also a parameter a_j reflecting the weight player i believes player j is attaching to his own (that is, i's) outcome. The idea behind most models incorporating a term for the interaction partner's intention is that the believed intention of the other may change a decision maker's distributive preferences with respect to this other person. This idea is consistent with our definition of reciprocity as given before. However, in most studies peoples' beliefs about the intentions (i.e. the distributive preferences) of others have not been assessed as an independent variable. We find that gaining further insight into this important issue requires the use of independent measures of both social preferences and beliefs of others' social preferences.

5.2.3 Research questions

The goal of our study is to investigate and disentangle the effects of individual social preferences and individual beliefs on individual choice behavior in a PGG. The disentanglement of preferences and beliefs is achieved by measuring social preferences as an independent measure rather than inferring them from the behavior they are supposed to explain. Further, we directly elicit subjects' beliefs about the behavior of others. With these two variables, we replicate and explain findings from previous studies. Next, we extend previous research by addressing the question whether beliefs about other peoples' social preferences matter on top of beliefs about other peoples' behavior. Again, we avoid inferring variables from behavior these variables are supposed to explain, and therefore use an independent measure of beliefs about social preferences in our study. Finally, we also investigate how preferences and beliefs affect contribution levels in a repeated interaction and the degree to which the dynamics that arise in repeated PGGs affect peoples' social preferences and beliefs.

We conducted two studies for investigating the research questions outlined above. In the first study, the PGG is implemented as a one-shot situation, whereas in study two we employ both a one-shot and a repeated PGG.

5.3 Study 1

5.3.1 Method

Study 1 was conducted in the form of a classroom experiment and data collection took place in May 2011 and April 2012 at a large European university. Subjects were students from a wide range of disciplines. Participation in the experiment was voluntary and no deception was used. Further, all parts of the study were made incentive compatible by the application of a lottery system. That is, 48 subjects were selected at random and paid according to their decisions. The 48 subjects who were selected earned 35 Swiss francs on average. The study consisted of two sessions, with one week separating these two sessions. In the first session, we assessed subjects' SVOs and afterwards let them play a linear anonymous one-shot PGG. In the second session, the same subjects' beliefs about the other subjects' SVOs and the other subjects' contributions in the PGG were assessed. Out of the 330 subjects in total, 227 attended both sessions. However, we had to exclude several subjects due to 1) missing values[11] (17 subjects), 2) intransitive choice patterns in the SVO measure (additional 8 subjects), or 3) incorrect answers to the comprehension check question regarding the PGG in either of the two sessions (another 45 subjects). Therefore, 157 subjects were kept for data analysis. Out of these 157 subjects, 132 (84%) are male and 25 (16%) are female.

5.3.2 Measures

SVO

Subjects' SVOs were assessed with the SVO Slider Measure (Murphy, Ackermann, & Handgraaf, 2011). This measure consists of 6 primary and 9 secondary items. The primary items measure a person's SVO on a continuous scale. One of the primary items of the SVO Slider Measure is shown in figure 5.1. Each item consists of 9 options indicating different divisions of monetary units between the self and another person. When completing a Slider Measure item, a subject considers different payoff allocations and then indicates that own-other payoff allocation she prefers most. When all of the 6 primary items are completed, the measure yields a single index of a person's SVO in terms of an angle, which can range from -16.26° indicating perfect competitiveness up to 61.39° indicating perfect altruism. The Slider Measure has been shown to be highly reliable and valid in terms of both its predictive power and its comparability to other SVO measures (see Murphy et al., 2011). The measure also allows for detecting intransitivity in subjects' choice patterns. As mentioned earlier, violations of transitivity are rare and were found in the choice patterns of only 8 (3.52%) out of the 227 subjects who attended both sessions.

Incentive compatibility was implemented by a lottery system, and subjects were informed about all details of the payoff scheme before completing the task. Concretely, sixteen subjects were

[11] Those subjects who had only missing values in the secondary items of the SVO measure were kept in the sample for further evaluation.

randomly selected for payment. Each of these sixteen subjects was paid according to her choice in one randomly selected item of the Slider Measure plus the choice of another randomly selected subject in an item that was randomly selected for her. More precisely, the sixteen subjects were matched with a procedure we call ring-matching. We employed this procedure in order to negate strategic consideration. Ring-matching was applied as follows: Subject A receives the payoff she allocated to the self in a randomly selected item and in addition receives the payoff subject Z allocated *to the other* in an item that was randomly selected for subject Z. Subject B then receives the payoff she allocated to the self in a randomly selected item and in addition receives the payoff subject A allocated to the other in the item that was selected for her, and so forth, until the ring is closed. This allows for complete incentive compatibility of the choices for the self and for the other in the absence of strategic interaction.

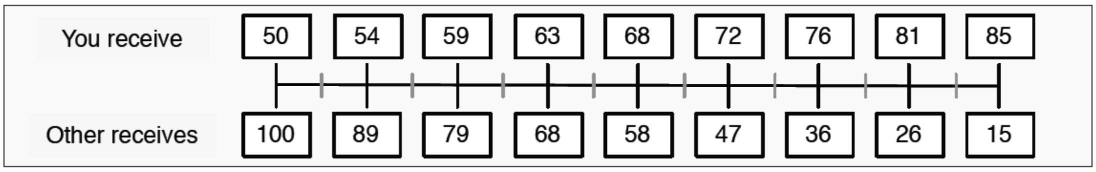


Figure 5.1: Examplary primary item of the Slider Measure

The average SVO angle of subjects as assessed with the Slider Measure was 24.09° ($std.$ = 15.12) with the lowest observation being an angle of $-16.26°$ and the highest observation being an angle of 45.00°. When subjects are assigned to the most commonly used SVO categories, we find that 91 (57.96%) subjects are prosocial, 62 (39.49%) subjects are individualistic, and 4 (2.55%) subjects are competitive. However, for analyses we use the continuous information about people's SVO rather than the categorical one for maximizing statistical power. The full distribution of SVO angles is shown in figure 5.2 and is highly similar to the one reported by Murphy et al. (2011).

Public goods game

Subjects received a single sheet of paper containing detailed instructions about how the public goods game works[12] and how final payoffs depend both on the own contribution and the contribution of others. Subjects were also provided with several examples in order to make it clear for them how final payoffs are realized given the amount of the own contribution and the sum of the group members' contributions. Further, subjects had to answer a comprehension check question so that we could verify that they really understood the task. Finally, they had to indicate how much they wanted to contribute to the public good.

The PGG was a standard linear one with the Voluntary Contributions Mechanism (Isaac et al.,

[12] We did not mention the term "public goods game" in the instructions, of course, but rather used the neutral term "task". Also, we did not use the term "contribute", but "transfer to the group account".

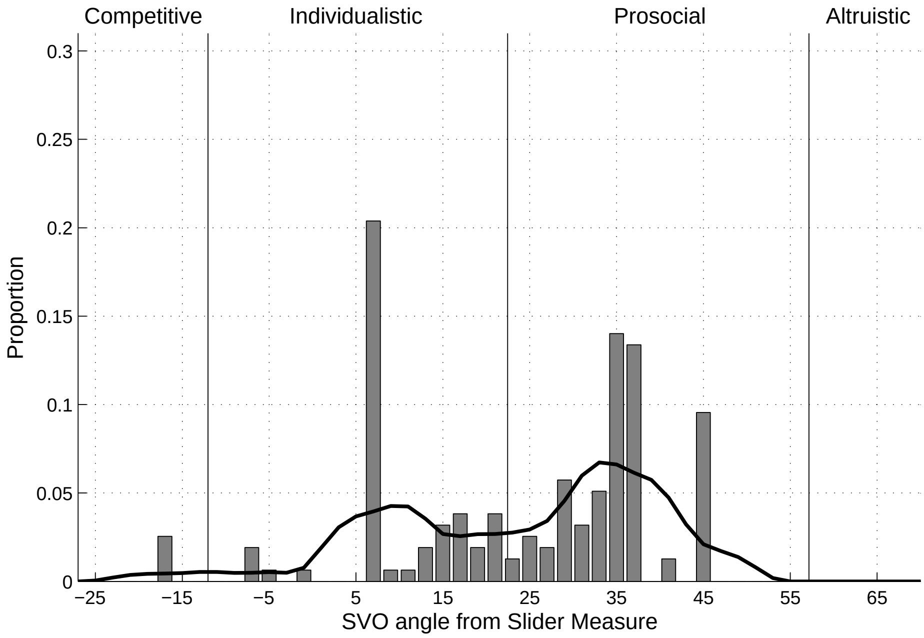


Figure 5.2: Distribution of SVO angles ($n = 157$)

1984) implemented. The endowment was 20 Swiss francs for each subject and group size was 4, with a marginal per capita return (MPCR) of 0.4. That is, the sum of contributions in a group was multiplied by the factor 1.6 and the resulting amount was divided equally among the 4 group members.

Eight subjects were randomly selected for payment according to their decisions in the PGG. These eight subjects were matched in groups of four and final payoffs were computed according to their contributions in the PGG. That is, the contributions were summed up per group, multiplied by 1.6 and the resulting amount was divided equally among them. Accordingly, subjects received the amount resulting from this division plus the amount they had decided not to contribute. Out of the 157 subjects, 34 (21.66%) contributed nothing, and 42 (26.75%) contributed their whole endowment. The full distribution of contributions is shown in figure 5.3.

SVO beliefs

The subjects' beliefs about the other subjects' SVOs were assessed as follows. Subjects were presented with the Slider Measure a second time, but this time, they were asked to complete the measure as they thought most other people would. More precisely, they had to indicate for each item of the Slider Measure which option they thought most other people had chosen in session 1. Further, it was made clear to subjects that they would be paid for accuracy.

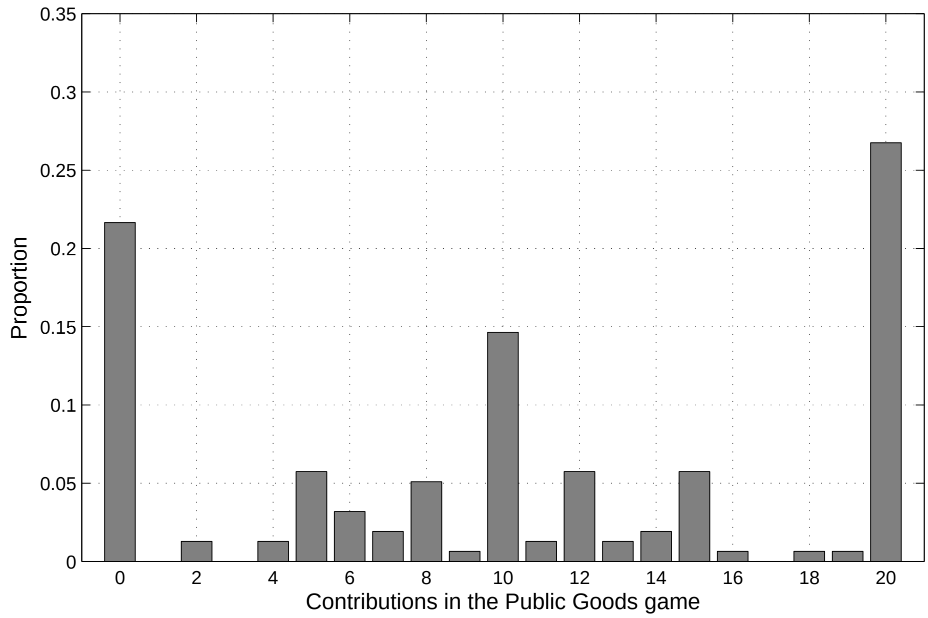


Figure 5.3: Distribution of contributions in the PGG ($n = 157$)

Sixteen subjects were randomly selected for payment and the following scoring rule was applied. Subjects earned 3 Swiss francs if they exactly hit that option which was chosen most often in session 1. A deviation by 1 option still yielded 2 Swiss francs, a deviation by 2 options yielded 1 Swiss franc, and a deviation by 3 or more yielded zero Swiss francs. Since subjects completed both the 6 primary items and the 9 secondary items, the maximum amount they could earn in this task was 15x3 = 45 Swiss francs.

People on average believed that other people would have an SVO angle of 21.58° (min = −7.82°, max = 61.39°). When beliefs are categorized, 70 (44.59%) subjects believe that the majority of others is prosocial, 85 (54.14%) believe that the majority is individualistic, and 2 (1.27%) believe that the majority is altruistic. Figure 5.4 shows the full distribution of SVO beliefs.

PGG beliefs

Similar to the elicitation of beliefs about other peoples' SVOs, the beliefs about other peoples' contributions in the PGG was assessed by presenting subjects again with the same PGG they played in session one, but with different instructions. This time, subjects were asked to indicate how much they believed others had contributed in the PGG in session 1 on average. The reason why we asked subjects to guess the mean rather than the mode was that we wanted to facilitate comparisons between our results and the results from other studies (such as the study by Fischbacher & Gächter, 2010, for instance). Further, it was made clear to the subjects that they

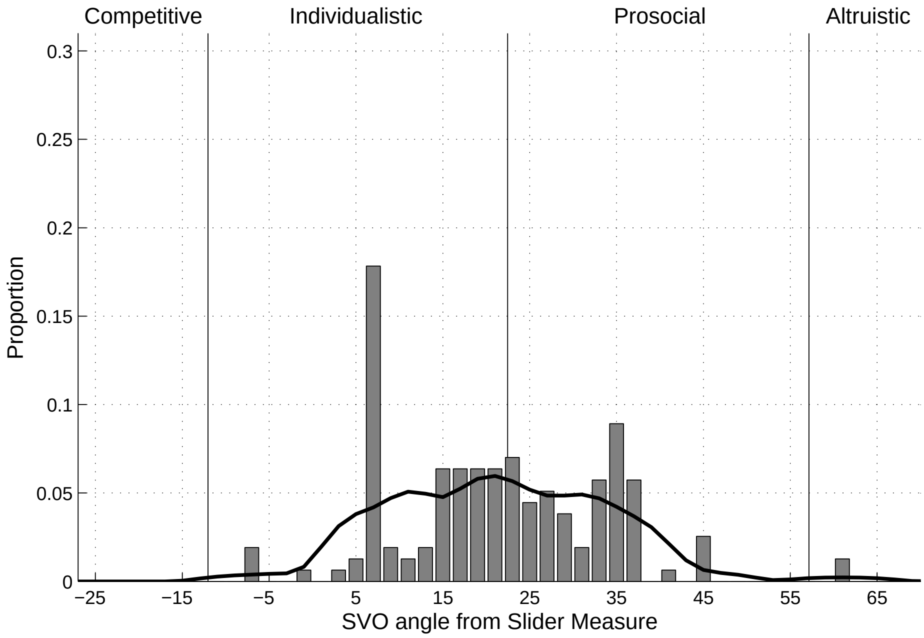


Figure 5.4: Distribution of SVO beliefs ($n = 157$)

would be paid for accuracy.

Again, eight subjects were randomly selected for payment. We applied a scoring rule that yielded subjects 30 Swiss francs if the mean contribution was hit exactly (within rounding error), 20 Swiss francs if they missed it by 1, 10 Swiss francs if they missed it by 2, and zero otherwise.

Subjects on average believed that the mean of the contributions in the PGG was 9.74 Swiss francs, i.e. 48.71% of the endowment. The full distribution of believed mean contributions in the PGG is shown in figure 5.5.

5.3.3 Results

Table 5.1 shows the correlation matrix including all variables assessed in study one. The bivariate results indicate that both SVO and PG belief are highly associated with contributions in the PGG, while SVO belief is not. Furthermore, to our surprise, SVO belief is neither significantly correlated with SVO, nor with PG belief.

Table 5.2 shows the results of an ordinary least squares regression with contributions in the PGG as the dependent variable, and SVO, PG belief, and SVO belief as predictors (Model 1). Only the main effects of SVO and PG belief on contributions in the PG are significant. SVO belief does not explain additional variance in PG contributions above and beyond SVO and

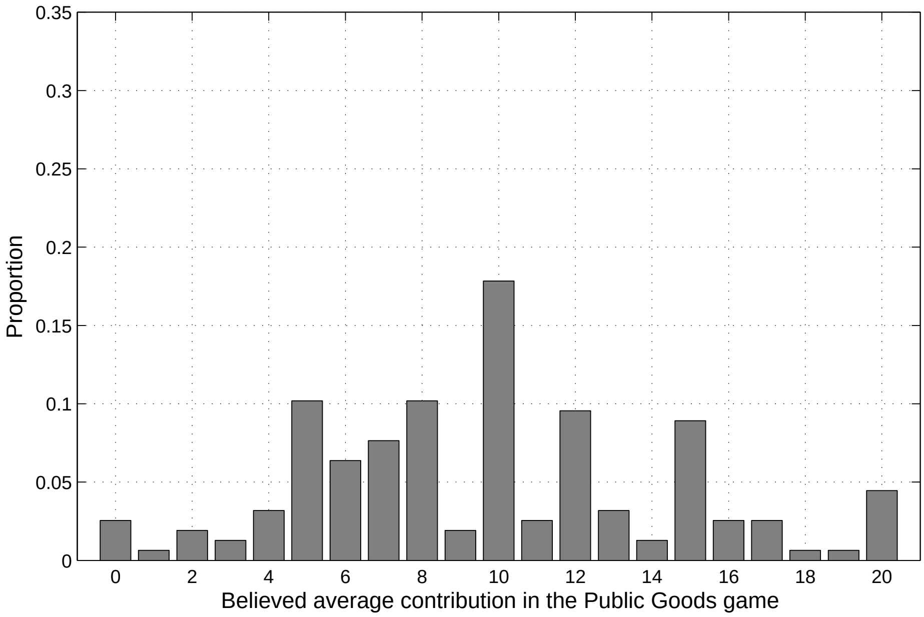


Figure 5.5: Distribution of contribution beliefs ($n = 157$)

PG belief. Nevertheless, the three predictors account for more than a third of the variance in contribution levels in the PGG. The addition of interaction terms as predictors does not improve the explanatory power substantially (see Model 2).

Table 5.1: Bivariate correlations

	PG	SVO	PG belief	SVO belief
PG	-	0.40***	0.51***	-0.07
SVO		-	0.26**	0.12
PG belief			-	0.06
SVO belief				-

$p^{***} < .001$; $p^{**} < .01$

5.3.4 Discussion of study 1

Results from study one suggest that peoples' SVOs and their beliefs about the contributions of others together explain a substantial proportion of variance in contribution levels in a one-shot anonymous PGG. The bivariate correlations as well as the regression coefficients suggest that beliefs about others' contribution levels is the strongest predictor, followed by SVO. Peoples' beliefs about other peoples' SVOs have not been shown to be predictive of contribution levels in study 1. Even though the regression coefficient attached to SVO belief reaches significance in

Table 5.2: OLS main effects and interactions from study 1 ($n = 157$)

	Dependent variable: Contributions in the PGG	
Model	(1)	(2)
SVO	0.30***	0.33***
PG belief	0.44***	0.41***
SVO belief	-0.13	-0.17*
SVO x PG belief		0.09
SVO x SVO belief		-0.04
SVO belief x PG belief		0.11
SVO x SVO belief x PG belief		0.10
R square	*0.35*	*0.38*
Adjusted R square	*0.34*	*0.35*

$p^{***} < .001$; $p^{*} < .05$

model 2, adding SVO belief to SVO and PG belief by a hierarchical regression procedure does not result in a significant increase in explained variance. It may well be, however, that the role of SVO beliefs becomes more relevant in a repeated rather than a one-shot situation. The rationale behind this consideration is as follows. Consider a person who is prosocial and believes that most other people are prosocial as well, but that they will not contribute. The reason why this person thinks that others are not going to contribute although they are believed to be prosocial may be that she either believes others to believe that all people are selfish or believes others to believe that no one will contribute. In a one-shot situation, this constellation of preferences and beliefs may lead this person to contribute zero because she does not want to be the only person contributing (which would be indicative of exploitation aversion as proposed by Weimann, 1994). However, in a repeated situation the same constellation may lead this person to contribute substantially in order to signal to others that there is at least one prosocial person or at least one person contributing in the group, so that others are encouraged to contribute in the next round (for evidence showing that such hard core (persistent) cooperators exist, see e.g. Murphy, Rapoport, & Parco, 2006). Study 2 provides an opportunity to shed more light on the role of SVO beliefs in this respect.

5.4 Study 2

5.4.1 Method

Study 2 took place in January and February 2013 at a decision science laboratory at a large European university in the form of a standard laboratory experiment. The study was made fully incentive compatible and explicitly followed a no-deception policy. The subjects earned 42.05 Swiss francs on average, including a show-up fee of 10 Swiss francs. The tasks subjects performed were programmed in z-Tree (Fischbacher, 2007). We conducted 6 sessions in total with 20 (4 sessions) or 24 (2 sessions) subjects per session, resulting in a total N of 128 subjects. Upon arrival at the laboratory, each subject was randomly assigned a cubicle number and seated at the corresponding computer place. Once the subjects were seated, they were welcomed via loudspeakers and told to follow the instructions displayed on the computer screen. As in study 1, subjects first completed the SVO Slider Measure and then played an anonymous one-shot public goods game in a group of four. Subsequently, they were asked to indicate their beliefs about the average contribution of their group members. Next, subjects' contribution profiles were elicited by use of the strategy method (Selten, 1967) followed by the assessment of their beliefs about other peoples' SVOs. It is important to note here that subjects were not given any information or feedback about the outcome of the one-shot PG or their performance in guessing other peoples' contributions or SVOs at this point in the experiment. Finally, the subjects played a repeated public goods game with fixed matching and feedback for 10 periods and thereafter completed the SVO Slider Measure a second time. As the final part of the experiment, the subjects received detailed feedback about their earnings in the experiment, completed a questionnaire, and were subsequently called to the experimenter window one by one to privately collect their earnings. The duration of the experiment was about 1.5 hours and the exchange rate between experimental currency unit (ECU) and real monetary value was 8 ECUs = 1 Swiss franc.

Incentive compatibility was implemented as follows. Before each session, it was determined at random which parts of the experiment would be relevant for payment. Concretely, it was determined whether the first or second SVO assessment would be relevant for payment, and which item of the SVO Slider Measure would be relevant for payment for both the assessment of SVO and SVO beliefs. Furthermore, one period out of the 10 periods from the repeated PGG was randomly selected to be relevant for payment. At the beginning of the experiment, subjects were informed that the parts relevant for payment had been predetermined, but are unknown to them, and that they should therefore treat every part of the experiment as payoff relevant. They were also informed that the indication of the relevant parts had been written down on a piece of paper in an envelope that was affixed to the wall where it was visible to the subjects during the entire experiment. After the experiment, they were given the opportunity to have a look at that piece of paper to verify that they were truthfully informed about every aspect of the experiment.

Out of the 128 subjects, 4 showed intransitive choice patterns in the SVO Slider Measure and are therefore excluded from the sample, leaving us with a sample n of 124 subjects for analyses. Subjects were mainly university students from a wide range of disciplines, and out of the $n = 124$ sample 76 (61%) subjects are male and 48 (39%) subjects are female.

5.4.2 Measures

SVO and SVO belief

Subject's SVO and SVO beliefs were assessed with the SVO Slider Measure (Murphy et al., 2011) as in study 1. The only difference to the assessment in study 1 was that instead of the paper-based version, the z-Tree implementation of the SVO Slider Measure (Crosetto et al., 2012b) was used. As described above, SVO was assessed two times in study 2, once at the beginning of the experiment (SVO time-1), and once as the final incentive compatible task after the repeated PGG (SVO time-2). When subjects completed the Slider Measure the first time, they were not informed that there will be a second assessment. The instructions for the task were identical at both points in time, with two exceptions. First, the instructions for the second assessment of SVO included the following phrase: "You have done this task before. Now please complete the task one more time. You do not have to be consistent with the answers you gave the first time. Your preferences may or may not have changed during the experiment." Second, subjects were informed that either the first or the second time they complete the task will be relevant for payment. As in study 1, a ring-matching procedure was employed for determining individual payoffs from the SVO assessment, and the scoring rule for payoff determination with respect to the elicitation of SVO beliefs was identical to the one employed in study 1. Subjects were informed that matching would be random. With respect to random matching, the instructions for the second assessment of SVO (i.e. SVO time-2) included an additional phrase: "[...]. Remember that this other person is randomly selected, that is, this person may or may not be someone you interacted with during this experiment."

At time 1, the average SVO angle was 20.43° ($std.$ = 13.58) with a minimum of −16.26° and a maximum of 61.39°. When subjects are categorized, 1 (0.81%) subject is altruistic, 58 (46.77%) subjects are prosocial, 64 (51.61%) subjects are individualistic, and 1 (0.81%) subject is competitive. The distribution of SVO time-1 angles is shown in figure 5.6.

At time 2, the average SVO angle was 18.35° ($std.$ = 12.93) with a minimum of −16.26° and a maximum of 45.00°. When subjects are categorized, no subject is altruistic, 45 (36.29%) subjects are prosocial, 78 (62.90%) subjects are individualistic, and 1 (0.81%) subject is competitive. The distribution of SVO time-2 angles is shown in figure 5.7.

With respect to SVO belief, subjects on average believed that other people would have an SVO angle of 16.62° (min = −7.82°, max = 61.39°). When beliefs are categorized, 1 (0.81%) subject believed that the majority of others is altruistic, 38 (30.65%) subjects believe that the majority of others is prosocial, and 85 (68.55%) believe that the majority is individualistic. Figure 5.8 shows the full distribution of SVO beliefs.

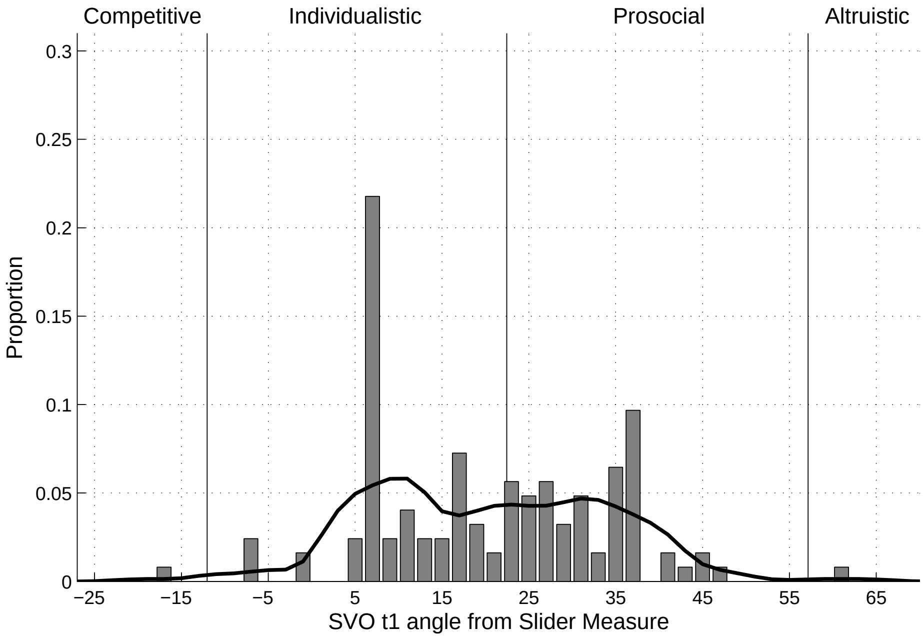


Figure 5.6: Distribution of SVO time-1 angles ($n = 124$)

Public goods game contributions and beliefs

Identical to the one-shot PGG situation in study 1, subjects indicated how much to transfer to the group account before they were asked to indicate their beliefs about the average contribution of others. The PGG implementation was identical to study 1 as well, with an endowment of 20 experimental currency units, an MPCR of .4, and group size of 4, both in the one-shot and the repeated PGG. Also, before indicating their contributions in the one-shot PGG, subjects had to answer two questions to check that they understood the nature of the task[13]. After subjects indicated their contributions and their beliefs about the average contribution of others, they were asked to indicate how much they would want to contribute given any possible integer average contribution of the other group members by means of the strategy method. To maintain incentive compatibility, we followed a procedure that was also applied by Fischbacher et al. (2001). Concretely, we truthfully informed subjects that one member per group would be selected at random to be the “conditional member”, and that their payoff from the (one-shot) PGG would then be calculated as follows. First, in each group, the average contribution of the three “unconditional members” would be computed. Then, the computer would look up what the “conditional member” had indicated to contribute in the strategy method given the actual average

[13]When a subject answered a question incorrectly, a pop-up window appeared on that subject’s computer screen explaining the task once more and giving the subject another opportunity to answer the question. An experimental session continued only if all subjects had answered the questions correctly. Due to this procedure, we could make sure that all subjects understood the task and we did not have to exclude subjects from further analyses due to a lack of understanding of the PGG.

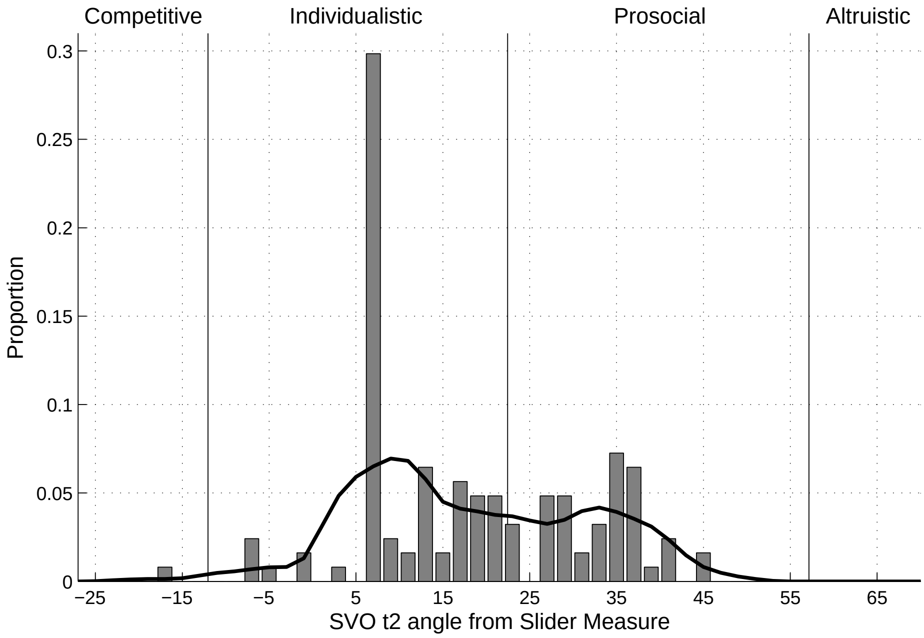


Figure 5.7: Distribution of SVO time-2 angles ($n = 124$)

contribution of the three "unconditional" group members. Finally, based on the contributions of the "unconditional members" and the contribution of the "conditional member", the group contributions are determined and individual payoffs are computed.

The distribution of contributions in the one-shot PGG is shown in figure 5.9, the distribution of subjects' beliefs about the average contribution of others is displayed in figure 5.10, and the distribution of conditional contributions from the strategy method can be seen in figure 5.11.

As mentioned earlier in this paper, subjects began playing the repeated PGG without having received any feedback about the outcomes of the one-shot PGG, nor about their accuracy in predicting the empirical average contribution of other people's SVO. The instructions for the repeated PGG informed subjects that they would make 10 successive decisions about how much to transfer to the group account and to indicate what they believed the other three group members would transfer on average. Further, subjects were informed that group compositions are stable across the 10 periods, such that they are interacting with the same people in all 10 periods. Also, they were informed that they would receive detailed feedback after each period where they would see how much each of the other three group members had transferred to the group account, and the resulting average contribution. We implemented this exhaustive feedback condition to give subjects the opportunity to signal their intentions, such that we could find out whether peoples' beliefs about other people's SVO matter in this situation. Figure 5.12 shows the distribution of contributions across the ten periods on the aggregate, i.e. disregarding group-level information, and figure 5.13 shows the corresponding mean contributions across the ten periods.

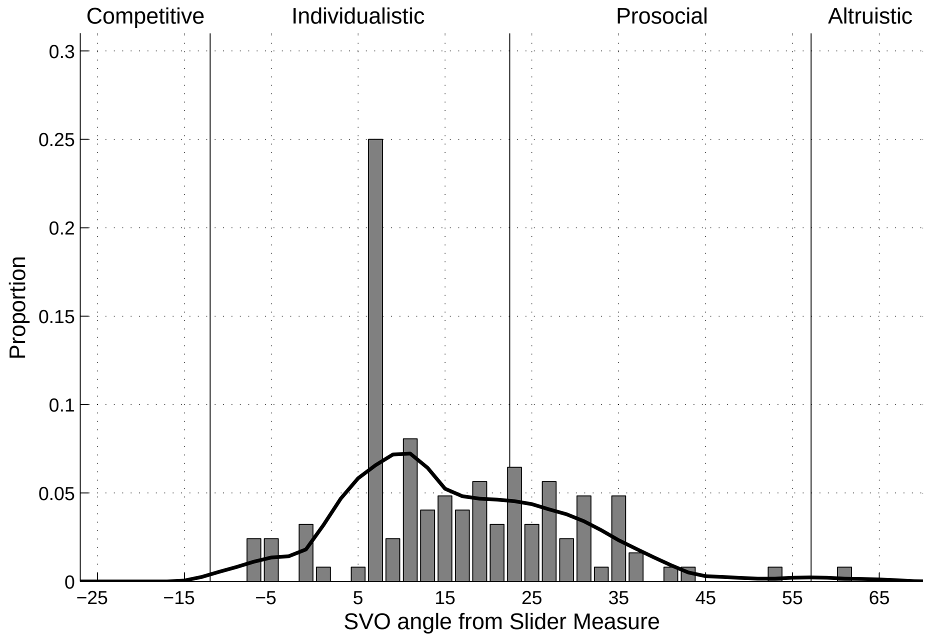


Figure 5.8: Distribution of SVO beliefs ($n = 124$)

The contribution profiles of all groups across the ten periods is shown in figure 5.14.

Questionnaire

After the completion of the SVO Slider Measure at time-2 following the repeated PGG subjects received detailed feedback about their earnings in the experiment and then answered a questionnaire. The questionnaire consisted of three parts. The first part assessed Machiavellianism, a construct that has been shown to be associated with behavior in experimental games in previous research (e.g. Gunnthorsdottir, McCabe, & Smith, 2002), by means of the Mach IV scale (Christie & Geis, 1970). This scale consists of 20 statements, such as, for instance, “Never tell anyone the real reason you did something unless it is useful to do so”, to which subjects indicate the degree to which they agree on a 5-point Likert scale. In our sample, the scale showed low but marginally acceptable internal consistency (Cronbach’s $\alpha = 0.73$). The second part of the questionnaire assessed the Belief in a Just World (BJW) with the 6 items of the General Belief in a Just World Scale (Dalbert et al., 1987), which showed unsatisfying internal consistency (Cronbach’s $\alpha = 0.67$). The third and final part of the questionnaire was the subscale of positive reciprocity (9 items) from the Personal Norm of Reciprocity (PNR) Scale (Perugini et al., 2003), which also showed low but marginally acceptable internal consistency (Cronbach’s $\alpha = 0.72$). The three scales were included in the experiment for exploratory purposes. The questionnaire concluded with the assessment of the sociodemographic variables age, sex, and field of study.

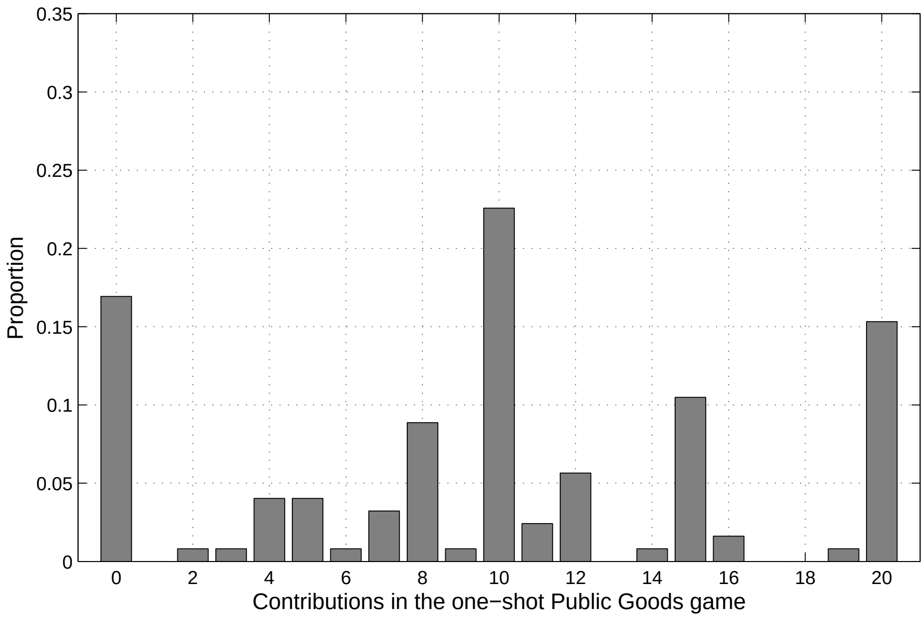


Figure 5.9: Distribution of contributions in the one-shot PGG ($n = 124$)

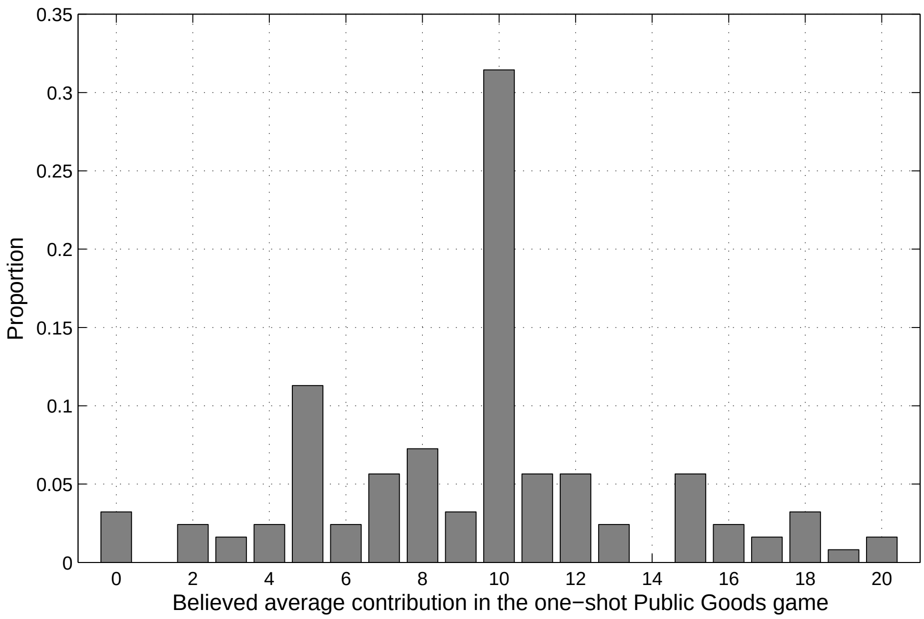


Figure 5.10: Distribution of contribution beliefs from one-shot PGG ($n = 124$)

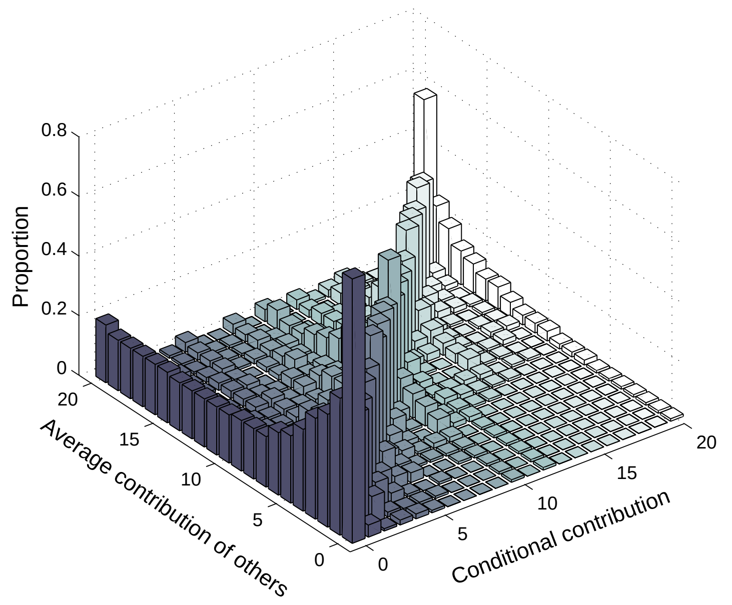


Figure 5.11: Distribution of conditional contributions from strategy method ($n = 124$)

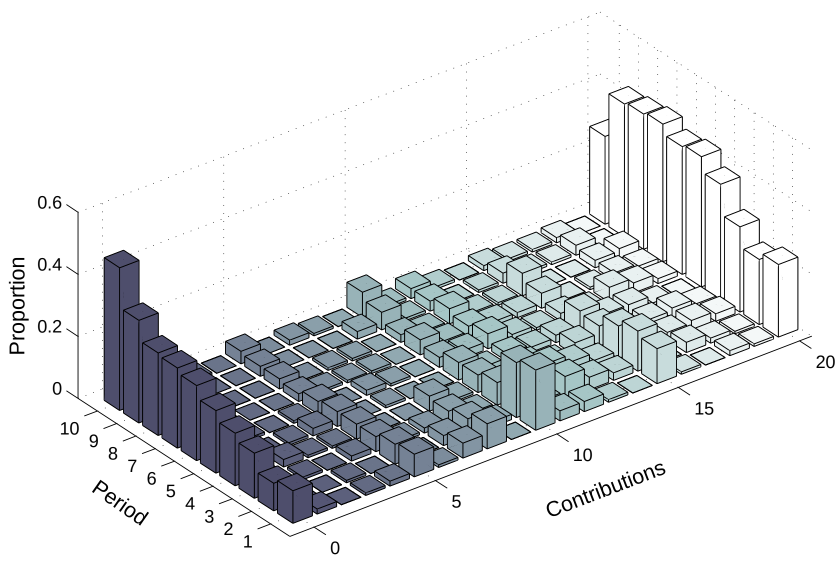


Figure 5.12: Distribution of contributions across the 10 periods in the repeated PGG ($n = 124$)

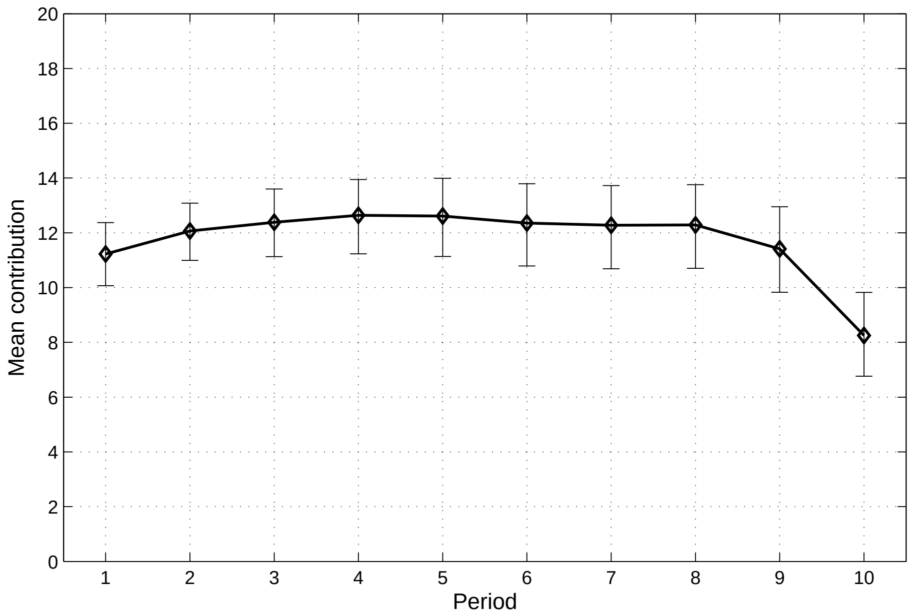


Figure 5.13: Mean contribution per period in the repeated PGG with 95% confidence intervals around the means ($n = 124$)

5.4.3 Results

The bivariate relations between selected variables from study 2 are shown in table 5.3. The variable CCA stands for "Conditional Cooperation Area" and is an index computed from subjects' choices in the strategy method for the purpose of summarizing individual choice patterns in single numbers. Concretely, a subject's CCA is a rescaled approximation of the area under the curve of the subjects' contribution profile from the strategy method. A subject's CCA is 0 if the subject indicated to contribute zero for any possible average contribution of others (free riding), and 1 if the subject indicated to match the average contribution of others exactly (perfect conditional cooperation), and more than 1 if the subject indicated to contribute more than the average contribution of others in total[14].

[14]However, representing a contribution profile with a single number discards information, i.e. a subject's CCA (> 0) is ambiguous. A CCA of 1, for instance, can result from perfect conditional cooperation, but may also be a result of a radically different contribution profile, such as contributing zero for average contributions of others between 0 and 9, matching an average contribution of 10 exactly, and contributing 20 for any average contribution of others strictly above 10. Hence, the CCA has to be interpreted with caution, but it is still a useful approximate representation of the subjects' general tendencies to condition their own contributions on the average contribution of others.

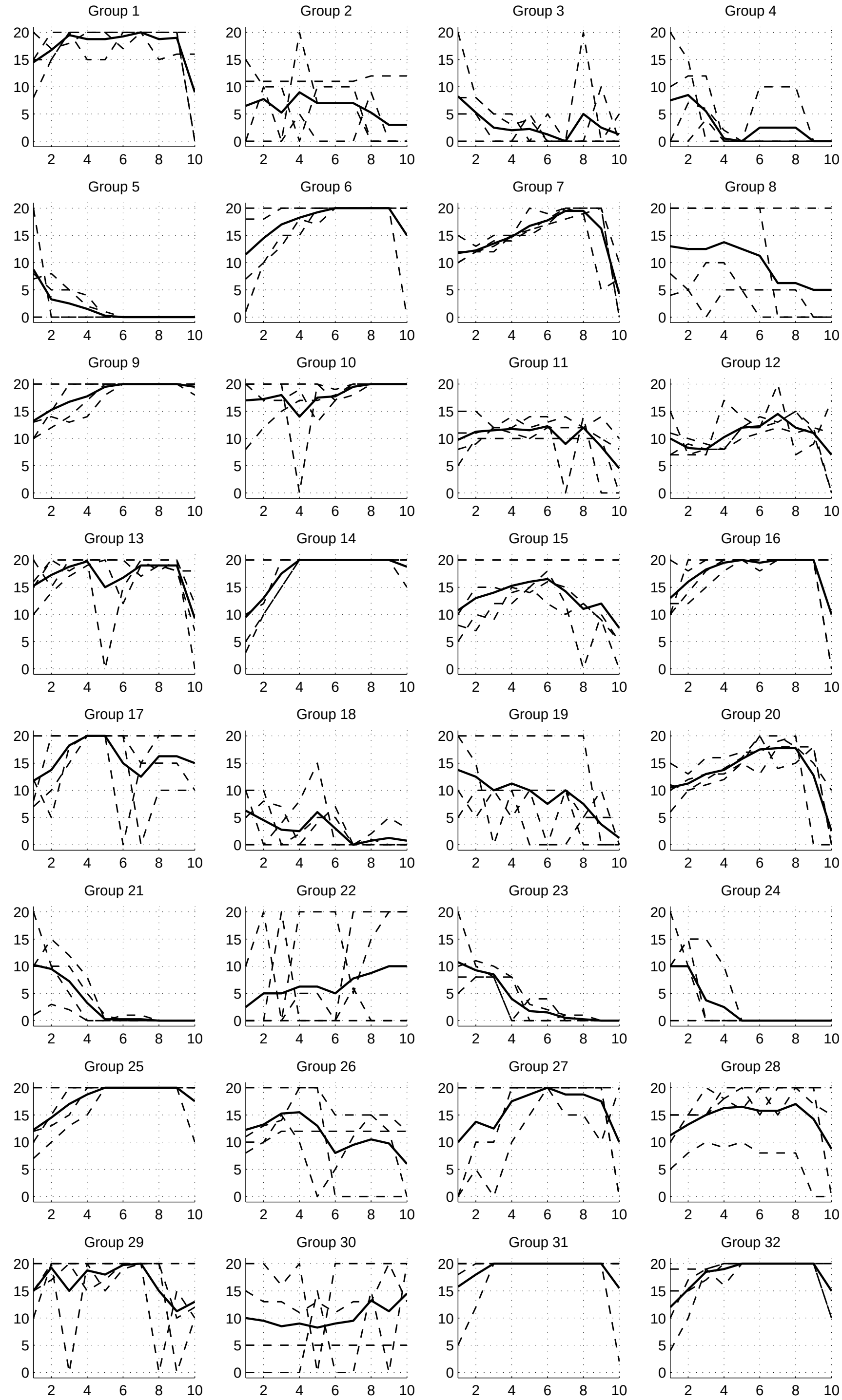


Figure 5.14: Contribution profiles of the 32 groups across the 10 periods of the repeated PGG ($n = 128$). Thick lines indicate group mean contributions, and dashed lines indicate individual contributions.

Table 5.3: Bivariate correlations among preferences, beliefs, and behavior

	PG os	SVO t1	PG os belief	SVO belief	CCA	PG rp p1	PG rp belief p1	PG rp p10	PG rp belief p10	SVO t2
PG os	-									
SVO t1	**0.32**	-								
PG os belief	**0.75**	***0.26***	-							
SVO belief	0.09	**0.63**	0.16	-						
CCA	**0.47**	**0.30**	**0.39**	0.15	-					
PG rp p1	**0.70**	**0.30**	**0.59**	0.09	**0.32**	-				
PG rp belief p1	**0.58**	**0.32**	**0.69**	*0.19*	*0.21*	**0.80**	-			
PG rp p10	***0.27***	***0.26***	*0.20*	0.17	**0.31**	0.15	0.11	-		
PG rp belief p10	***0.27***	0.04	*0.21*	0.05	**0.37**	0.15	0.08	**0.70**	-	
SVO t2	**0.32**	**0.72**	***0.25***	**0.55**	**0.36**	***0.26***	*0.19*	**0.40**	***0.26***	-

Levels of significance are indicated as **bold** at $p < .001$, ***bold and italicized*** at $p < .01$, and *italicized* at $p < .05$. PG os = Contribution in the one-shot PGG; SVO tn = SVO angle at time n; CCA = Conditional cooperation area; PG rp pn = Contribution in period n of the repeated PGG; PG rp belief pn = Belief about the average contribution of others in period n of the repeated PGG.

Explaining behavior in the one-shot PGG

In this section, we consider only the results corresponding to the one-shot public goods game. As can be seen in Table 5.3, there are clear associations of considerable magnitude between SVO, the beliefs about other's behavior (PG os belief), and contribution levels (PG os), while also CCA is significantly correlated with these three variables. However, there is no evidence for a relation between subjects' beliefs about other subjects' SVO and contributions in the PG os. Furthermore, we find a significant and considerable correlation between SVO and beliefs about other people's SVO (SVO belief), suggesting a consensus effect, such that people believe that other people have similar social preferences as they themselves have on the aggregate. However, there is no significant bivariate relation between SVO belief and contribution decisions, or beliefs about contribution levels.

Table 5.4 shows the results of an ordinary least squares regression with contributions in the one-shot PGG as dependent variable. As already indicated by the bivariate correlations, both subjects' SVO and their beliefs about the average contribution of others are clearly predictive of their own contributions in the one-shot PGG. Furthermore, including SVO belief as a predictor explains a significant proportion of variance in PG contributions in addition to the proportion of variance already explained by SVO and PG belief ($p(\Delta R^2|H_0) = 0.01$). Interaction effects are absent. Together, the three predictors SVO, PG belief, and SVO belief explain about 60% of the variance in contribution levels in the one-shot PGG. Furthermore, including CCA as a further predictor again results in a significant R^2 increase (($p(\Delta R^2|H_0) < 0.01$), leading to a total of 0.63% of explained variance in contribution levels in the one-shot PGG, while SVO remains to be a significant predictor. This would suggest that SVO and propensities to condition own contributions on the contribution level of others (as measured by the strategy method) do have differential predictive capacities.

Table 5.4: OLS main effects on contributions in the one-shot PGG (n = 124)

	Dependent variable: Contributions in the one-shot PGG		
Model	(1)	(2)	(3)
SVO	0.26***	0.22**	0.23**
PG belief	0.71***	0.66***	0.54***
SVO belief	-0.19*	-0.18*	-0.19**
CCA		0.17**	
BBCC			0.25**
R square	*0.60*	*0.63*	*0.63*
Adjusted R square	*0.59*	*0.61*	*0.62*

Note: Standardized regression coefficients are reported. Levels of statistical significance are indicated as $p^{***} < .001$; $p^{**} < .01$; $p^{*} < .05$

As an additional test, we replace the general CCA index by the specific belief-based conditional contribution (BBCC) as a predictor of individual contributions in the one-shot PGG.[15] BBCC is computed as follows. First, for each subject, we identify the belief this subject indicated to have about the average contribution of others in the one-shot PGG. Second, in the strategy table completed by this subject, we look up the contribution the subject indicated to make given that average contribution of others that corresponds with the average contribution this subject believed others would contribute in the one-shot PGG, which is then identified as that subject's BBCC. Hence, if a person –given this person's belief about the average contribution level of others– would behave exactely how that person indicated to behave in the strategy method, then his BBCC would be identical to his actual contribution in the one-shot PGG. However, the correlation between subjects' BBCCs and their contributions in the one-shot PGG is not perfect, but rather it is $r = 0.67$ ($p < .001$). Moreover, including BBCC in the regression (see model 3 in Table 5.4) does not at all render the main effects of SVO, PG belief, or SVO belief insignificant, and reduces their relative predictive capacity only slightly. Clearly, the strategy method does reveal useful information about peoples' behavior in a real-time PGG, but the information is far from perfect. Moreover –and surprisingly, perhaps–, the results regarding both CCA and BBCC suggest that data gained by means of the strategy method are not a good substitute for independent measures of distributive social preferences, as they appear to have significant differential predictive power.

What is somewhat puzzling about the results is the role of SVO belief in explaining contribution levels. Subjects' contributions in the one-shot PGG appear to be independent of their beliefs

[15] If we add BCC in addition to CCA in model 2 of the regression, then BBCC takes over all of the variance explained by CCA, thereby rendering the relative effect of CCA insignificant. This is not surprising, since both variables are computed on the basis of the same data, namely subjects' indications of contribution decisions in the strategy method, while BBCC –as it is informed by the subject's elicited beliefs in the one-shot PGG– is by far a more specific indicator of behavior compared to the general CCA index.

about other subjects' SVO when considering the bivariate correlation between the two variables. However, when SVO and PG belief are statistically controlled (and only if *both* are controlled), SVO belief does appear to be predictive of contribution levels, but to our surprise the association is negative. This would indicate that when SVO and PG belief are already taken into account, a subject can be expected to contribute less the more prosocial the majority of others is believed to be. This would be a surprising implication that could perhaps be explained by some sort of diffusion of responsibility, where a person does not feel urged to be the nice guy given the belief that there are so many others who could play this role. However, we can only speculate about the nature of the role of SVO beliefs in explaining contributions in a one-shot PGG. Our data only allow us to conclude that SVO belief does appear to have some explanatory power with respect to contribution levels, but the underlying mechanism has to be investigated in more detail by future research.

Explaining behavior in the repeated PGG

Perhaps surprisingly, the subjects' behavior in the first period of the repeated PGG is very similar to their behavior in the one-shot PGG as evidenced by the high association between individual contributions in the one-shot PGG and in the first period of the repeated PGG (see table 5.3). Although slightly more subjects contribute the full endowment, and also more subjects believe that the others are going to contribute the full endowment on average in the first period of the repeated PGG compared to the one-shot situation, the distribution of contributions in the first period of the repeated PGG does not differ significantly from the distribution of contributions in the one-shot situation.[16] Also, the bivariate relations between SVO, SVO belief, and PG belief on the one hand and contribution levels on the other hand are about the same in the one-shot situation and in the first period of the repeated PGG (see and compare the respective indicators in table 5.3, table 5.4, and figure 5.15). Moreover, the differences between individual contributions in the one-shot PGG and the first period of the repeated PGG can not be explained by SVO beliefs ($r = .01, p = .91$), and there is no significant interaction between SVO and SVO beliefs as predictors of contributions in the first period of the repeated interactions. Thus, we find no evidence of systematic signaling, and differences in contribution levels between the one-shot PGG and the first round of the repeated PGG on the individual level are not accounted for by SVO beliefs. We conclude from this that people enter a repeated PGG as they enter a one-shot PGG on the aggregate, and that the main drivers of contribution levels are just the subjects' individual social preferences and individual beliefs about the contribution levels of others in both situations.

Figure 5.15 shows the results of ten OLS regressions, one per period of the repeated PGG, with contributions in the respective period as dependent variable, and SVO (t_1), SVO belief, and subjects' beliefs about the average contribution of their group members (PG rp belief) in the respective period as independent variables. The figure also shows the proportion of variance explained in contributions by the three independent variables per period.

[16] Both an insignificant Kolmogorov-Smirnov test and overlapping 95% confidence intervals around the means indicate this result.

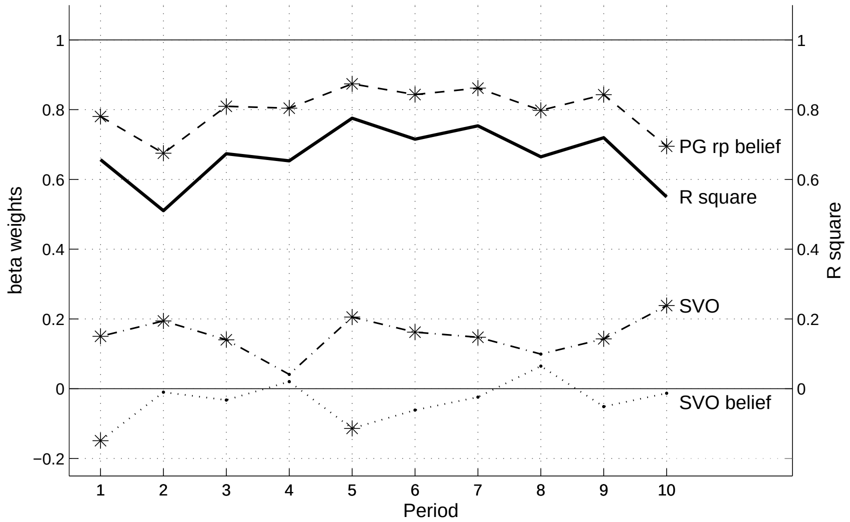


Figure 5.15: Beta weights and R^2 from OLS regression with contributions in the repeated PGG as dependent variable and SVO (t_1), SVO belief, and PG belief (per period) as independent variables across the 10 periods ($n = 124$). Asterisks (*) indicate significance $p < .05$. Obviously beta weights and R^2 values are in different metrics, however this plot shows the relative stability of these components and the overall model over periods of the iterated game.

As can be seen in figure 5.15, the proportion of variance explained in contribution levels by the three predictors is very high across all ten periods, with a minimum of 51.0% explained variance in period 2, and a maximum of 77.6% explained variance in period 5.

Clearly, the predictor that contributes by far the most to the proportion of variance explained in contribution levels is subjects' beliefs about the average contribution of others in a particular period (PG rp belief). However, adding SVO as a predictor to PG belief leads to a significant R^2 increase in 8 out of the 10 periods (not visible in figure 5.15). Subjects' beliefs about other subjects' SVO is only significantly predictive of contribution levels in the first period (and period 5), and thereafter appears to be generally uninformative. Analyses of interaction effects indicated a significant two-way interaction between SVO and SVO belief in periods 2, 3 and 4, indicating that proself subjects contribute more given that they believe that others are prosocial, while prosocials do not respond much to their beliefs about the SVO of others in terms of contribution levels. However, although it appears plausible that such an effect may be there in the first couple of periods, the evidence that this effect is real is too thin to draw conclusions. What seems to be driving subjects' behavior in the PGG after all is primarily and simply their beliefs about the average contribution of others together with their SVO.

Since there is a lot of evidence that beliefs about the behavior of others is affecting subjects' behavior substantially, then the question that arises is to what degree we can predict subjects' beliefs about the average contribution of others. First of all, it is noteworthy that subjects' beliefs about the average contribution of others in the one-shot PGG are not statistically significantly different compared to the first period of the repeated PGG.[17] Among the variables we have assessed that may be predictive of *a priori* PG contribution beliefs, SVO, and CCA show a significant bivariate correlation with PG beliefs both in the one-shot and first period of the repeated PGG. Together, these two variables explain 17.7 % of the variance in beliefs about the average contribution of others in the one-shot situation, and 11.6 % in the first period of the repeated PGG. In terms of a false consensus effect (Ross et al., 1977), it appears plausible to assume that people who are more prosocial and reciprocal are also more likely to believe that others are prosocial (see the positive correlation between SVO and SVO belief in table 5.3) and reciprocal, and consequently also expect higher contribution levels from others compared to less prosocial and reciprocal people.

However, by means of the variables we have assessed, we cannot shed more light on how subjects form their "home grown" beliefs about the behavior of others. But we can now try to shed light on how subjects update their beliefs according to the feedback they receive after each period of the repeated PGG. Figure 5.16 shows the results of 9 nested OLS regressions, one for each period of the repeated PGG starting from the second, with subjects' beliefs about the average contribution of their corresponding group members as dependent variable. The average contribution of group members in period p-1 is inserted first into the regression as independent variable (IV1). As the second independent variable (IV2), subjects' beliefs about the average contribution of the group members in period p-1 is entered into the regression model. As can be seen in figure 5.16, the addition of IV2 into the regression model results in a significant increase in the proportion of variance explained in all but the final period. Finally, subjects' own contributions in period p-1 are inserted as the third independent variable (IV3). The inclusion of IV3 also leads to a further significant increase in R^2 in addition to IV1 and IV2 in most periods, and predominantly in the second and third period, but not in the final period. Clearly, subjects' beliefs about the average contribution of group members can be predicted extremely well by these three variables, especially from period 4 to period 9, where more than 90% of the variance in subjects' beliefs about what the group members are going to contribute in the current period on average can be explained. The prediction of beliefs in the second, third, and last period of the repeated PGG is -albeit still quite accurate- more complicated, presumably because higher-order beliefs are likely to play a significant role at the beginning and the end of a repeated interaction.

Next, we test whether the data obtained by means of the strategy method render the independent measures of preferences and beliefs obsolete. As shown in Figure 5.17, BBCC explains about 30-50% of the variance in contribution levels across the ten periods, but adding subjects' beliefs about the average contribution of others results in a significant -and substantial- increase in R^2 in each period. Also, the addition of SVO results in a significant –albeit not very pronounced– increase in R^2 in six out of the ten periods. It may be worth emphasizing that SVO here is used

[17] Again, both an insignificant Kolmogorov-Smirnov test and overlapping 95% confidence intervals around the means indicate this result.

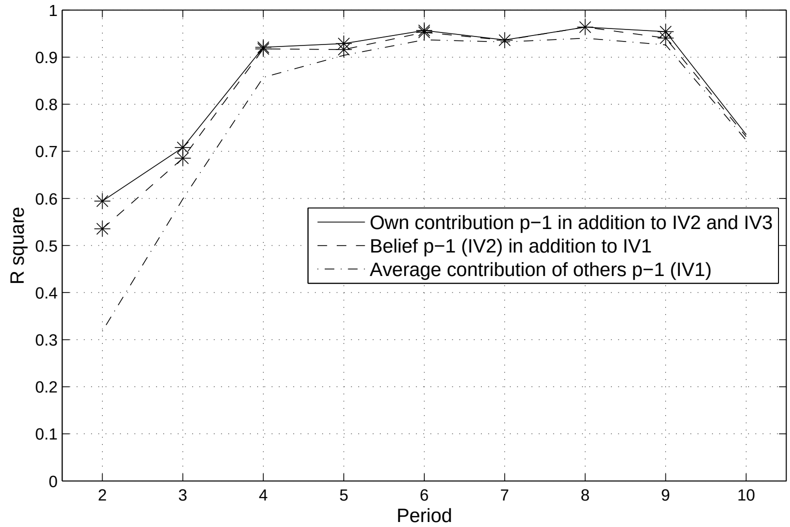


Figure 5.16: R^2s from OLS regressions with beliefs about average contributions of others in the repeated PGG (period 2-10) as dependent variable. For each period, the independent variables *average contribution of others in period p-1*, *PG rp belief in period p-1*, and *own contribution in period p-1* are inserted into the regression in nested blocks in the respective order ($n = 124$). Asterisks (*) indicate a significant increase in R^2 with $p < .05$.

as a static predictor (SVO_{t1}), while beliefs are updated in each period and BBCCs are informed by these updated beliefs.

In combination with the results displayed in figure 5.18 it becomes clear that subjects condition their contributions more on the (believed) behavior of others than they indicated to do in the strategy method, a result consistent with findings from Fischbacher & Gächter (2010). As can be seen in figure 5.18, subjects on average actually contributed more than they indicated they would contribute given the believed average contribution of others in the strategy method in each period except the last, but on average rather matched the believed average contributions of others in all periods except the first and the last. The fact that subjects behave more conditionally cooperative than they indicated to behave in the strategy method could have several causes. First, people may be more conditionally cooperative in a hot situation such as an ongoing PGG interaction compared to a cold situation of mere decision indication. Second, subjects may feel urged to behave more conditionally cooperative because of social desirability effects due to the transparency of their contributions to others. Third, subjects may behave more conditionally cooperative because of strategic considerations. The second and third reasons would both be due to the fact that the subjects completed the strategy method with respect to the completely anonymous one-shot PGG, without even knowing that a repeated interaction would later follow. We can not fully answer the question which reasons may be causing the effect, but we can at least have a look at the difference scores from the one-shot PGG. In the one-shot situation, the mean

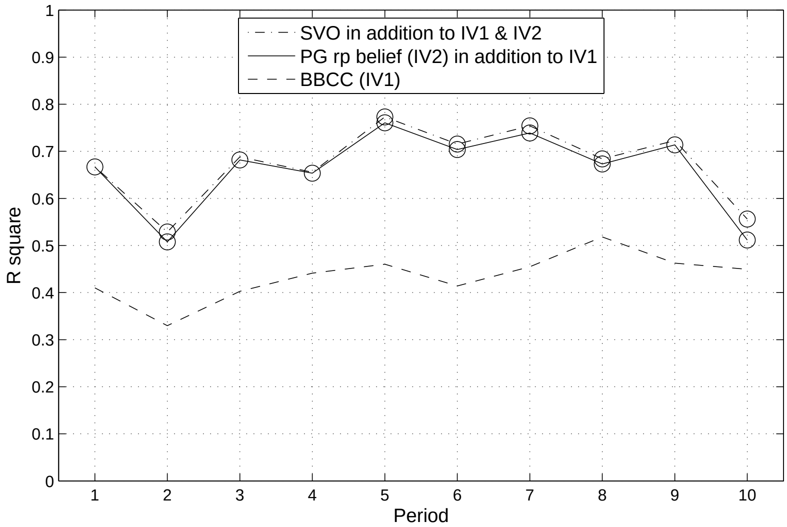


Figure 5.17: R^2s from OLS regressions with individual contributions in the repeated PGG per period as dependent variable. For each period, the independent variables *BBCC*, *PG rp belief* per period, and *SVO* are inserted into the regression in the respective order ($n = 124$). Circles (o) indicate a significant increase in R^2 with $p < .05$.

difference between subjects' contributions and their BBCC is 1.79 (within a 95% confidence interval between 0.97 and 2.74), while the mean difference between subjects' contributions and their beliefs about the average one-shot contribution of others is 0.33 (within a 95% confidence interval between −0.43 and 1.06). Although these figures indicate that in the one-shot situation, too, subjects' contributions on average deviate significantly from their BBCC in the expected direction but do not deviate significantly from the believed average contribution of others, the confidence intervals around the two mean difference scores overlap, meaning that they are not statistically significantly different from each other. Hence, the evidence does not clearly support, nor refute one of the three stated possible explanations for the observation that subjects behave more conditionally cooperative in the repeated PGG than they indicated in the strategy method.

An additional question we can address is which variables on the group level are predictive of group performance. To investigate this issue, we can use the groups' average contribution in the final period of the repeated PGG as a dependent variable, and find it's best predictor on the group level among the variables we have assessed. Since using the group as unit of analysis reduces sample size by a factor of four, and because the subjects we excluded from the sample due to intransitive choice patterns in the SVO Slider Measure are dispersed across groups, we keep all 128 subjects for analyses so the number of groups we can analyze is not further reduced.[18] In our sample of 32 groups, the best predictors of the groups' average contributions

[18]The pattern of results does not change when the groups involving subjects who showed intransitive choice patterns in the Slider Measure are excluded from the analyses.

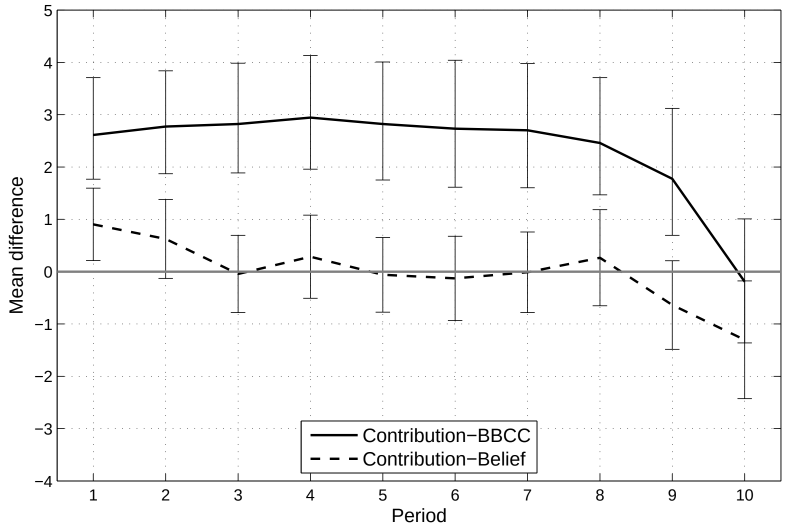


Figure 5.18: Mean differences between actual individual contributions and BBCC on the one hand, and between actual individual contributions and beliefs about the average contribution of others on the other hand per period, with 95% confidence intervals around the mean differences ($n = 124$)

in the final period of the repeated PGG are the groups' average contributions in the first period ($r = 0.44, p < .05$) and the minimum[19] of the groups' members' CCAs ($r = 0.60, p < .001$). These two variables together explain 39.5% of the variance in groups' average contributions in the final period of the repeated PGG.[20] This means that we can predict a group's performance (as operationalized by its average contribution level in the final period) quite well given that we know the group's initial average contribution level and the minimum of the group members' degrees of conditional cooperativeness. This result also holds -and is even more pronounced- if we define group performance as the groups' average contribution levels across all of the 10 periods.[21] The groups' total average contribution levels across periods 2 to 10 depend on the groups' average contribution levels in period 1 ($r = 0.63, p < .001$), and the minimum of the groups' members' CCAs ($r = 0.67, p < .001$). Clearly, the least cooperative player in a group has significant power over the whole group's performance. If group performance is operationalized as the total average group contribution across periods 2 to 10, the two independent variables together explain 58.4% of the variance in group performance.

[19]The mean CCA per group is also quite predictive of a group's average contribution in the final period, but the minimum CCA turns out to show more predictive power than the mean.

[20]If the average contribution in period 9 is used as dependent variable, the two variables together explain almost half (46.6%) of the variance.

[21]Since average contributions in period 1 are used as predictor, we only use the average contributions across periods 2 to 10 as dependent variable in order to avoid artificial boosts of R^2.

Conditional cooperation

Since we have assessed conditional cooperation by means of the strategy method just as it was assessed by Fischbacher & Gächter (2010) and Fischbacher et al. (2001), we can compare some of our findings to previous results. Firstly, we can directly compare our results to the results from Fischbacher et al. (2001). Figure 5.19 shows the total mean conditional contributions of all[22] subjects from the strategy method, and for the subgroups of conditional cooperators, "hump shaped", and free riders. We found that 78.12% of all subjects can be categorized as conditional cooperators, while 5.47% can be termed "hump-shape" contributors, 14.06% free riders, and 2.34% can not be assigned to one of these categories.[23] As for the subjects categorized as conditional cooperators, 24.22% (relative to all 128 subjects) match the average contributions of others perfectly or almost perfectly, 10.94% contribute more than the average contributions of others most of the time (i.e. excessive conditional cooperation), 26.56% predominantly contribute less (i.e. imperfect conditional cooperation), and 16.41% sometimes contribute more and sometimes less than the average contribution of others.

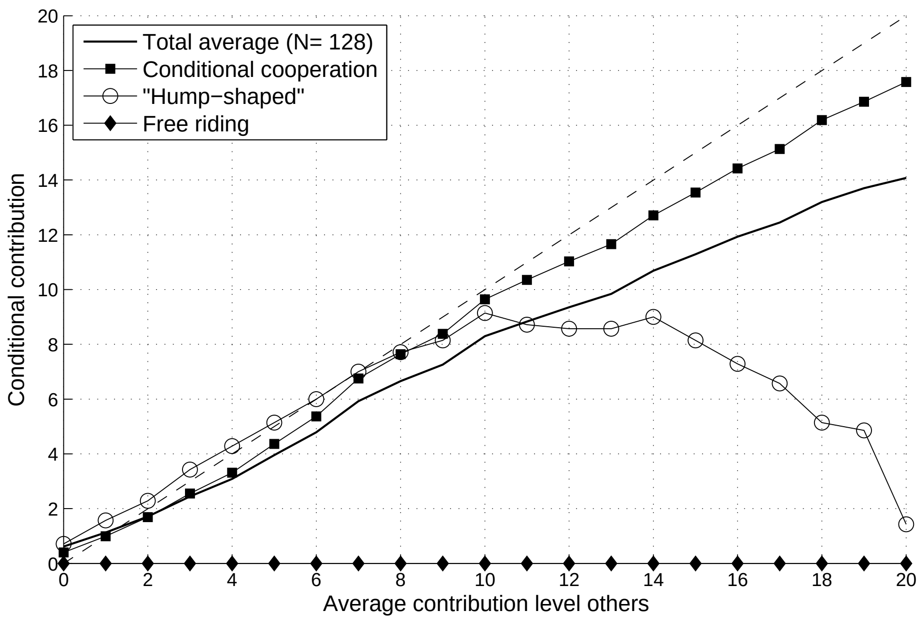


Figure 5.19: Conditional contribution means from the strategy method

Hence, our findings replicate the results from Fischbacher et al. (2001) quite nicely, although subjects' overall willingness to cooperate was clearly much higher in our study.[24] We also

[22]For this analysis, we decided to include all subjects.

[23]For example, one subject indicated to unconditionally contribute the full endowment, and another subject to unconditionally contribute 10.

[24]This conclusion is also supported by the fact that in our sample, the subjects' average contribution in the one-shot PGG was 9.82 tokens (i.e. almost 50% of the endowment), compared to 6.7 tokens (33.5% of the endowment)

find that many subjects can be termed *imperfect conditional cooperators*, resulting in a general -but in our study weak- tendency of subjects to contribute less than others on average. This supports the conclusion stated by Fischbacher et al. (2001) and Fischbacher & Gächter (2010) that imperfect conditional cooperation can cause - or at least contributes substantially to- the decline of cooperation in repeated PGGs that is often observed. However, in our experiment, where subjects received detailed feedback about the contributions of others after each period, and where fixed matching was applied, we do not find a general tendency of subjects to contribute less than the average contribution of their group members in the aggregate. But we can readily imagine that we would have observed such a pattern if the subjects in our experiment had solely received feedback about the average contribution of others in the repeated PGG, rather than receiving a full revelation of each group member's contribution after each period.

We extend previous findings by investigating how SVO relates to cooperator types as assessed by means of the strategy method. Table 5.5 shows the proportion of subjects with a particular SVO value (assigned to the categories low, medium, and high SVO) per choice pattern in the strategy method. Let us first read through the table row-wise, and attend to the corresponding percentages indicated to the right of the diagonal slash in the parentheses. Clearly, the subjects who indicated to free ride completely in the strategy method are predominantly subjects with low or medium SVOs (nearly 80 percent together). However, there are also some subjects with high SVOs who indicated to free ride in the strategy method. Interestingly, among the perfect conditional cooperators, there are predominantly subjects with a medium SVO, followed by those with a high SVO, while still more than a fifth of all perfect conditional cooperators is represented by subjects with low SVOs (i.e. individualists and competitors). In accordance with expectation, the vast majority of excessive conditional cooperators are subjects with high SVOs and medium SVOs. Also, the vast majority of imperfect conditional cooperators are subjects with low-to-medium SVOs, while mixed conditional cooperators are predominantly medium-to-high SVO subjects. Finally, hump shaped cooperators are almost exclusively subjects with intermediate SVO angles.

Let us now read through the table column-wise, and attend to the corresponding percentages indicated to the left of the diagonal slash in the parentheses. Surprisingly, the modal cooperator type among subjects with low SVOs is not "free rider", but "imperfect conditional cooperator", while the percentages of free riders and perfect conditional cooperators are equal among low-SVO subjects. As for subjects with medium SVOs, the distribution is quite flat across the different cooperator types, with the modal type being perfect conditional cooperation followed by imperfect conditional cooperation. One quarter of the subjects with high SVOs are perfect conditional cooperators, while slightly more than a fifth are mixed conditional cooperators and another fifth are excessive conditional cooperators.

In summary, the two concepts of SVO and conditional cooperator types are neither completely unrelated, nor totally redundant, and Table 5.5 reveals some surprising results. First, the strategy to free ride given any possible average contribution of others is not exclusively indicated by subjects with low SVOs, but also to a considerable extent by subjects with medium and high

reported by Fischbacher et al. (2001).

Table 5.5: SVO and conditional cooperation

	Low SVO	Medium SVO	High SVO
Free Rider	7 (19.4 / 38.9)	7 (14.3 / 38.9)	4 (10.3 / 22.2)
Perfect CC	7 (19.4 / 23.3)	13 (26.5 / 43.3)	10 (25.6 / 33.3)
Excessive CC	1 (2.8 / 7.7)	4 (8.2 / 30.8)	8 (20.5 / 61.5)
Imperfect CC	15 (41.7 / 45.5)	11 (22.4 / 33.3)	7 (17.9 / 21.2)
Mixed CC	4 (11.1 / 20.0)	7 (14.3 / 35.0)	9 (23.1 / 45.0)
Hump Shaped	1 (2.8 / 14.3)	6 (12.2 / 85.7)	0 (0.0 / 0.0)
Other	1 (2.8 / 33.3)	1 (2.0 / 33.3)	1 (2.6 / 33.3)

Note: The table shows counts and proportions (in parentheses) of subjects that have SVO angles in a particular range (low, medium, high) per conditional-cooperation-category. The proportions in parentheses that are on the left side of the diagonal slash (/) denote the frequency of subjects per conditional-cooperation category relative to the frequency of subjects within the corresponding SVO category, while the proportions in parentheses that are on the right side of the diagonal slash denote the frequency of subjects per SVO category relative to the frequency of subjects in the corresponding conditional-cooperation category.

SVOs. Second, the modal cooperator type of subjects with low SVOs is not "free rider", but "imperfect conditional cooperator", i.e. the relative majority of individualists and competitors is still inclined to cooperate –up to a certain extent– given that others cooperate as well. Third, perfect conditional cooperators can be found across all ranges of SVO, including the lowest range. Finally, excessive and mixed conditional cooperators are predominantly subjects with high SVO angles as expected. These results show that the tendency to condition own contributions on the contribution levels of others is only mildly associated with SVO, and that conditional cooperation is a behavioral pattern that is shown by subjects who are heterogeneous with respect to their basic distributive social preferences.

ΔSVO

We have assessed SVO twice in study 2, once at the very beginning of the experimental session, and once at the very end right after the repeated PGG, to see whether the dynamics of the repeated interaction (i.e. the increase or decline of cooperation) may cause changes in the subject's distributive preferences. As already shown in table 5.3 the correlation between SVO at time-1 and SVO at time-2 is very high ($r = 0.72$), yet it is lower than the test-retest correlation across a one-week period of $r = 0.92$ we found when validating the SVO Slider Measure (see Murphy et al., 2011).

Figure 5.20 shows the distribution of ΔSVO. About one third of the subjects (38, i.e. 31%) showed no change in their distributive social preferences after the repeated PGG, and among

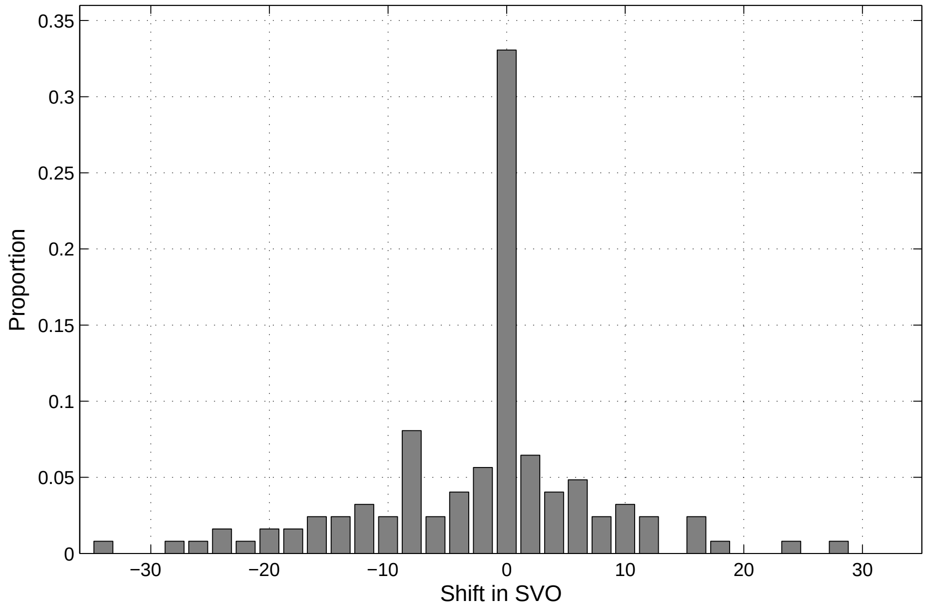


Figure 5.20: Distribution of ΔSVO ($n = 124$)

those two thirds who did show a change in SVO, 35 (28%) showed an increase in SVO angle (i.e. became more prosocial), and 51 (41%) showed a decrease in their SVO angle (i.e. became less prosocial). However, although more people changed their SVO downwards compared to upwards, the distribution is only slightly skewed, indicating that there is no clear general bias in shifts on the aggregate. Also, distributions of SVO angles at time-1 and time-2 are not statistically significantly different from each other as indicated by a Kolmogorov-Smirnov test, suggesting that the PGG interaction by itself did not affect subjects' distributive preferences on the aggregate in a particular direction.

The question then is to what degree we can explain individual changes in SVO by characteristics of the individual PGG interactions. A first natural question we can address is whether group performance is predictive of individual shifts in SVO angles (ΔSVO). Both of the two group performance measures we have used previously are significantly associated with ΔSVO, namely the group mean contribution levels across all periods ($r = 0.36, p < .001$) and likewise the mean contribution levels in the final period ($r = 0.36, p < .001$). It is interesting to see that what happens in the very last period appears to be as important as what happened across all ten periods on average. However, the best single predictor of ΔSVO is the average contribution of the subjects' corresponding group members in the final period with a correlation of $r = 0.38$ ($p < .001$). This may be due to a recency effect (see e.g. Ebbinghaus, 1913; Atkinson & Shiffrin, 1968) or that contributions in the final period are perceived as revelations of others' intentions unconfounded by the shadow of the future (Axelrod, 1984).

Furthermore, the difference between subjects' beliefs about the average contribution of others and the actual average contribution of others in the first half of the repeated PGG -especially in the very first period- is predictive of ΔSVO. In particular, the difference between what subjects believed their group members would contribute in the very first period, and what the group members then actually contributed in that first period shows a correlation of $r = 0.26$ ($p < .01$) with ΔSVO. This means that shifts in SVO are not independent from the experience of positive surprise or disappointment about the average contribution of group members in the first period. Finally, also the constellation of a group's members' CCAs is informative with respect to the group's members' ΔSVO. Concretely, ΔSVO is significantly correlated with the average CCA of the three other group members ($r = 0.19, p < .05$), and more importantly, with the minimum of the other three group members' CCAs ($r = 0.28, p < .01$).

Clearly, what happens in a group during a repeated PGG has an effect on the majority of corresponding interaction partners' social preferences, which then affects their behavior. Contribution levels in the final period of the repeated PGG are far better predicted by SVO at time 2 ($r = .40, p < .001$) than by SVO at time 1 ($r = .26, p < .01$). In conclusion, the analyses regarding ΔSVO suggest that people do not only update their beliefs about the behavior of others during a repeated PGG interaction, but in addition update their social preferences. Hence, for understanding the dynamics arising from peoples' behavior in a repeated PGG interaction, both the process of belief updating, and the process of preference updating have to be taken into account.

Questionnaire data

In general the questionnaire data are not very informative with respect to subjects' behavior in the one-shot PGG, nor with respect to their behavior in the first and last period of the repeated PGG. The only scale that shows some predictive power with respect to contribution levels in the repeated PGG is Machiavellianism, which is significantly negatively associated with contributions in periods 2, 3 and 5 (r ranging between .18 and -.27), and marginally significantly negatively associated with contributions in periods 4, 6, and 7. On average, the correlation between Machiavellianism and contribution levels in the repeated PGG across all periods is $r = -0.20$ ($p < .05$). The negative association -albeit rather weak- between Machiavellianism and contribution levels in the repeated PGG makes sense, since Machiavellianism is also significantly negatively associated with both SVO ($r = -.23, p < .05$) and CCA ($r = -.26, p < .01$), which themselves are positively associated with contribution levels. Furthermore, Machiavellianism is the only variable among all the variables we have assessed or computed which is significantly (negatively) associated with the other two questionnaire scales we have assessed, namely the Belief in a Just World (BJW), and Positive Reciprocity (PR)[25].

[25]PR shows a significantly positive relation with contribution levels in the final period of the repeated PGG ($r = .18, p < .05$), though, but since this is the only instance where PR shows a significant correlation with any of the variables we have assessed or computed (except Machiavellianism) - including several measures of reciprocity, such as CCA, mean difference between own contribution and believed average contribution of others, and further variables of this kind- we do not pay attention to this one single correlation which may have reached significance just by chance.

5.4.4 Discussion of study 2

The results of study 2 confirmed the findings from study 1 that both social preferences and beliefs about the behavior of others are important determinants of contributions in an anonymous one-shot public goods game, which together can explain up to more than 50% of the variance in individual contribution levels. Furthermore, peoples' beliefs about the social preferences of others turn out to be predictive of contribution levels as well, but only when peoples' own social preferences and beliefs about the contribution levels of others are statistically controlled. Furthermore, the relation between peoples' beliefs about the SVO of others and their own contribution levels appears to be negative, which is surprising. However, since the general pattern of results in this respect is the same both in study 1 and study 2, the effect does not seem to be a measurement error. Our expectation that SVO beliefs are more informative of contribution levels in a repeated interaction compared to a one-shot situation is not supported by the data, though.

To our surprise, contribution levels and beliefs about the contribution levels of others did not differ noticeably between the one-shot PGG and the first period of the repeated PGG. From this we can conclude that only very few people took the opportunity to signal their willingness to cooperate at the beginning of the repeated interaction to induce cooperation by others. Rather, contribution levels can best be predicted by peoples' beliefs about the contribution level of others together with their social preferences in both situations. Furthermore, we found that the vast majority of people have a clear tendency to behave conditionally cooperative, and that this tendency is likely underestimated when measured by means of the strategy method. Our data clearly supports the finding by Fischbacher & Gächter (2010) that people behave more conditionally cooperative in an ongoing PGG interaction than they indicated they will when asked in the strategy method.

Regarding the dynamics of the repeated PGG interaction, we found that once people receive feedback about the outcomes of a PGG period, they update their beliefs on the basis of the observed average contribution of others, their own contribution level, and the previous belief they had about the average contribution of others. Due to the predominance of the tendency to condition own contributions on the contributions of others, the updated beliefs are then excellent predictors of individual contributions in the following period. However, as it turns out the accuracy of the prediction can be significantly increased when individual social preferences (i.e. individual SVO) are taken into account.

A further conclusion from study 2 is that a group's performance in the repeated PGG can be predicted quite well by the group members' average contribution in the first period and the constellation of the group members' tendencies to behave conditionally cooperative. Concretely, the higher the contribution level a group starts with, and the higher the tendency of group members to behave conditionally cooperative, the more likely it is that the group can sustain a high contribution level on average, resulting in higher group performance. Likewise, the lower the contribution level a group starts with, and the lower the group members' tendencies to behave conditionally cooperative, the more likely it is that the group will end up with a low average contribution level. Since we found no interaction between initial average contribution level and

CCA, the effects of the two factors seem to be additive. This means that even if a group starts with a rather high average contribution level, it can end up performing poorly given that one or more group members have a very low tendency to behave conditionally cooperative (imperfect conditional cooperators) or free ride completely. Imagine, for instance, a group consisting of one free rider and three perfect conditional cooperators as an extreme case. Even if the three perfect conditional cooperators contribute the full endowment in the first period, the decline of cooperation would unfold due to the distribution of the group members' CCAs (see Fischbacher & Gächter, 2010), and the group will end up with an average contribution level of 1.12 tokens in the final period. On the contrary, a group starting with a rather low level of cooperation can end up performing very well given that one or more group members have a very high CCA, and sometimes contribute more than the believed average of others (i.e. excessive cooperators).

The finding that a group's performance is quite strongly dependent on the group's initial average contribution level, and that initial individual contribution levels depend strongly on individual beliefs about the contributions of others leads to an interesting implication, namely that peoples' beliefs about the contribution levels of others can turn into self-fulfilling prophecies. Consider, for instance, a group with very prosocial subjects who have a quite high -but not perfect- tendency to behave conditionally cooperative. Assume now that the group members all believe that the other group members are going to contribute very little in the first period, so they will also contribute very little since they are conditional cooperators. Once they get feedback about the outcomes of the first period, their beliefs are confirmed and due to their tendency to imperfectly conditionally cooperate, they are very likely to end up performing very poorly, even though they would all have wished and been willing to achieve very high contribution levels.

A final and new finding from study 2 is that peoples' distributive social preferences (i.e. SVO) can change as a function of what happens during the repeated PGG interaction. Concretely, ΔSVO is associated in the expected direction with both group performance and "surprise" (positive or negative) about the average contribution of others in the first period(s). This is direct evidence for reciprocity according to the definition of reciprocity we have given earlier in this paper. Hence, in the process of a repeated PGG interaction, people do not only update their beliefs about the behavior of others, but do also update their own distributive social preferences in response to the behavior of others. This result extends previous findings and shows that people may not only alter their social preferences towards a specific other person (Sonnemans et al., 2006), but that the behavior of interaction partners can change peoples' social preferences more generally. It is an open question, however, to which degree the shifts in peoples' social preferences are decoupled from specific others, i.e. for how long peoples' distributive preferences remain altered, and to which extent the subjects' altered distributive preferences would carry over towards a completely different group consisting of people subjects know for sure they have not interacted with.

5.5 General discussion

We have conducted two studies to investigate how preferences and beliefs affect contribution levels in a one-shot and repeated linear public goods game (PGG). The evidence is clear that people enter a PGG situation with very diverse preferences and beliefs, resulting in tremendous heterogeneity among players. However, the data indicate that beliefs are not independent from preferences, and that people generally belief that others have similar preferences as they have and will show a behavior similar to their own. Once we know peoples' distributive social preferences (SVO), their tendency to condition their behavior on the behavior of others, and their beliefs about others' preferences and behavior, we can predict pretty well how much they are going to contribute in a one-shot PGG and also in the first period of a repeated PGG. And once we know a group's distribution of preferences, beliefs, and behavior in the first period of a repeated interaction, we can predict quite well how the interaction will continue and how the group will perform. We have seen that especially peoples' beliefs about the behavior of others are critical to their own behavior, and that due to people's strong tendency to cooperate conditionally, beliefs are the best single predictor of a person's future behavior in a repeated PGG. However, we have also seen that the impact of peoples' social preferences on their contribution levels never really fades out or becomes completely overridden by the effect beliefs have on behavior. Both preferences and beliefs are important determinants of behavior, and we could clearly show that people do not only update their beliefs about others during a repeated interaction, but may also change their distributive social preferences in response to the behavior of others - a phenomenon known as reciprocity and quantified unambiguously here.

6 Reciprocity as an individual difference

Abstract[1]

There is accumulating evidence that decision makers are sensitive to the distribution of resources among themselves and others, beyond what is expected from the predictions of narrow self-interest. These social preferences are typically conceptualized as being static and existing independently of information about the other people influenced by a decision maker's allocation choices. In this paper we consider the reactivity of a decision maker's social preferences in response to information about the intentions or past behavior of the person to be affected by the decision maker's allocation choices (i.e., how do social preferences change in relation to the other's type). This paper offers a conceptual framework for characterizing the link between distributive preferences and reciprocity, and reports on experiments in which these two constructs are disentangled and the relation between the two is characterized.

6.1 Introduction

It has been shown in many different studies that decision makers (DMs) generally take into account other peoples' payoffs when making decisions in situations of interdependence, and thus typically do not behave according to the predictions of neoclassic economic theory. For instance, there is a large body of evidence showing that people voluntarily forgo some of their own gains for the benefit (or detriment) of others in experimental games, even if the situation is one-shot and anonymous (for reviews, see for instance, Andreoni, Harbaugh, & Vesterlund, 2008; Camerer & Thaler, 1995; Camerer, 2003; Fehr & Fischbacher, 2003). Behavior deviating from simple own payoff maximization has been attributed in part to motivations referred to as social preferences (a.k.a. other-regarding preferences, social value orientation [SVO], altruism,

[1] This chapter is an edited version of the following paper.
Ackermann, K.A., Fleiß, J., and Murphy, R.O. (2014). Reciprocity as an individual difference. *Journal of Conflict Resolution, In press*.

or welfare tradeoff ratios). Primarily, social preferences have been understood in terms of simple distributive concerns, where a DM's utility is modeled as a joint function of his own outcome, and also of others' outcomes (see Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999; Grzelak, 1982; Loewenstein et al., 1989; Messick & Thorngate, 1967; Radzicki, 1976; Wyer, 1969). That is, a DM with social preferences derives non-zero utility from benefitting or harming another person, even when information about this other person is absent, i.e., under complete anonymity and lack of information about any of this other person's characteristics, past actions, or type (e.g., is the other person altruistic, prosocial, individualistic, or competitive).

The measurement and investigation of this *unconditional* type of social preferences has a long history in psychology, where this kind of motivation is referred to as Social Value Orientation (for reviews of SVO, see e.g. Au & Kwong, 2004; Bogaert, Boone, & Declerck, 2008; McClintock & Van Avermaet, 1982; Murphy & Ackermann, 2014a). A person's SVO can be understood as the general motivational goal a person has when allocating resources between herself and another person. Different types of SVOs are commonly considered and discussed in the literature. For instance, a person may be motivated to simply maximize his own payoff (individualism) as standard economic theory would predict, maximize the sum of all payoffs (prosociality), minimize the difference between payoffs (another kind of prosociality called inequality aversion), maximize the other's payoff indifferent to his own (altruism), or maximize his own relative gain (competitiveness), etc. These archetypes are useful when first considering social preferences, but SVO is a continuous construct and can be defined generally by the weights a DM attaches to their own outcomes and to other person's outcomes. These kinds of social preferences can be represented by a joint utility function, i.e., a utility function that involves separate terms corresponding to outcomes for the self, the other, and arithmetic combinations thereof.[2]

Empirical work has shown the reliability of SVO as an individual difference (e.g. Au & Kwong, 2004; Murphy, Ackermann, & Handgraaf, 2011); the distribution of SVOs across individuals has been estimated, revealing substantial reliable differences across individuals and a bimodal pattern of preferences as well. Moreover, SVO has been shown to be a valid predictor of choice behavior in strategic situations, such as social dilemmas (see Balliet, Parks, & Joireman, 2009; Murphy & Ackermann, 2014b). However these examples use SVO as a static construct, not attending to the systematic reactivity of preferences and specifically how DMs' SVOs change in response to information about other people's SVOs and behavior.

There is support for the notion that DMs do sometimes condition their preferences and choices on characteristics of their interaction partners in experimental situations. Early research on this issue has shown that people exhibit dramatically different SVOs depending on whether the recipient in an own-other resource allocation task is described as a "friend", "unknown", or as an "antagonist" (Sawyer, 1966). These results suggest that DMs' distributive preferences, with

[2] The following is an example of a joint utility function that can account for social preferences:
$U(\pi_s, \pi_o) = \pi_s + \alpha * \pi_o - \beta * |\pi_s - \pi_o|$,
where π_s is the payoff for the self, π_o is the payoff for the other, and α and β are parameters (weights) attached to the respective terms.

respect to how resources are allocated between themselves and another person, are sensitive to revealed characteristics of the corresponding other person. Moreover, it has been shown that people may change their distributive preferences depending on the (potential) interaction partners' observed past behavior (e.g. Murphy & Ackermann, 2014b; Sonnemans, van Dijk, & van Winden, 2006). That is, people may become less prosocial, or even hostile (for evidence on costly punishing behavior, see e.g. Fehr & Gächter, 2000b), after observing non-cooperative behavior by their interaction partner (this pattern is sometimes referred to as *behavioral assimilation*, a term coined by Kelley & Stahelski, 1970), or DMs may become (more) prosocial after observing a cooperative move from the interaction partner (e.g. Berg, Dickhaut, & McCabe, 1995; Diekmann, 2004). These two kinds of behavioral responses are often referred to as negative, and positive reciprocity, respectively (see Fehr & Gächter, 2000c). In addition to observed actions taken by the interaction partner, behavior can also be shaped by more complicated expectations of, and beliefs about, the interaction partner. For example, Schubert & Lambsdorff (2013) performed Ultimatum Game experiments in which Palestinians where confronted with offers from either Palestinian or Israeli proposers. The experimental results showed that Palestinian responders indeed responded differently to Israelis, as shown by the significantly higher rejection rate as compared to the situation where they are matched with Palestinian proposers.

Results suggest that what people respond to when exhibiting positive or negative reciprocity is not merely the outcomes resulting from the interaction partners' behavior (i.e., its consequences), but the interaction partners' *intentions* as revealed by the interaction partners' behavior (see e.g. Charness & Rabin, 2002; Cox, Friedman, & Gjerstad, 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Falk, Fehr, & Fischbacher, 2008; Levine, 1998; Rabin, 1993; Stanca, Bruni, & Corazzini, 2009). It is important to notice that – at least in the context of experimental games – what has been termed the interaction partners' "intentions" is simply a synonym for the interaction partners' "unconditional social preferences", their baseline SVO, or their type (e.g., altruistic, prosocial, individualistic, competitive). That is, the basic idea behind intention-based reciprocity models has been that DMs' SVOs may change depending on the interaction partners' assumed or observed SVO. In other words, people condition their social preferences upon what they believe the other person is like.

We follow this line of reasoning and apply a definition of reciprocity similar to the one provided by Cox (2004, p. 263) to address these changes in SVO. Concretely, we define reciprocity as *the change in a decision maker's distributive social preferences (i.e., SVO) in response to information about the interaction partner, compared to the decision maker's unconditional distributive social preferences* that are expressed when information about an interaction partner is absent. It is important to note that this definition of reciprocity is a difference score: it is the difference between unconditional (i.e., baseline) SVO and conditional SVO. Operationally then, a researcher would need to measure social preferences twice on the same DM in order to gage the magnitude of reciprocity an individual exhibits. This reactivity, or the dynamics of social preferences, is what concerns us here in this paper.

In the present study, we investigate how information about an interaction partner's type – i.e., the interaction partner's revealed SVO – changes people's distributive social preferences (the DMs' SVOs) in a non-strategic situation. This design allows us to disentangle strategic concerns (anticipated benefits or reprisals) and isolate the effect of changes in SVO. For a similar argument in the context of the sequential prisoner's dilemma, see Clark & Sefton (2001, p. 55) and Dufwenberg & Kirchsteiger (2004). Hence, we can measure reciprocity alone, without the potential confounding effects beliefs may have on a change in behavior.

There is already clear evidence that people exhibit reciprocity in one-shot situations[3] by showing a willingness to forgo their own gains in order to respond prosocially to observed prosocial behavior in a sequential dictator game where no subsequent interaction will take place (Diekmann, 2004). Diekmann also showed that the degree of kindness in reciprocating depended on the degree of kindness of the behavior being reciprocated, at the aggregate level. Our experimental design allows us to qualitatively replicate these findings and extend them in several important ways. First, rather than use a between-subjects design, we employed a within-subjects design and elicited complete reciprocity profiles of individuals across a set of different others. Second, subjects made a series of decisions in dictator games with varying tradeoff slopes (i.e., a mix of constant-sum and non-constant-sum dictator games with varying rates of marginal substitution), which allows a more fine-grained assessment of the willingness to pay for the benefit or detriment of the person whose previously observed behavior is being responded to. And third, our experimental design rules out the possibility that the previously observed behavior being responded to might be perceived as a mere strategic gambit. The subjects responding to previous dictator game decisions will be truthfully informed that the person who had made the dictator game decision was not informed that her decision would be revealed to the DM. This ensures that subjects are responding to –i.e., the object of reciprocation is– the interaction partner's honest revealed social preferences. This means that the DM does not have to try to glean strategic considerations on behalf of the other, such as acting prosocially only for the sake of inducing positive reciprocity.

The second goal of this paper is to investigate the extent to which SVO is related to reciprocity. Studies have previously shown that people with different social value orientations follow different reasoning when engaged in experimental games. For example, Boone, Declerck, & Kiyonari (2010) found that cooperative behavior of subjects with an individualistic value orientation tends to depend on external incentives, while the cooperation of prosocial subjects tends to depend on trust. There exists a fair amount of evidence suggesting that people with prosocial distributive preferences are more likely to reciprocate a corresponding interaction partner's prosocial choices compared to people with individualistic (i.e., goal to maximize own gain) or competitive (i.e., goal to maximize relative gain) distributive preferences (e.g. De Cremer & Van Lange, 2001; Kanagaretnam, Mestelman, Nainar, & Shehata, 2009; Van Lange, 1999; Van Lange & Semin-Goossens, 1998), although the latter two SVO types have also been shown to reciprocate significantly under certain conditions as well (e.g. Parks & Rumble, 2001; Sheldon, 1999). There is also evidence from researchers applying questionnaire methods, which supports the conclusion

[3] This form of reciprocity that is expressed in one-shot situations, where no future interactions are taking place and thus beliefs are inconsequential is sometimes referred to as "altruistic reciprocity", and is in accordance with our definition of reciprocity given above.

of a positive relationship between SVO and reciprocity (Perugini & Gallucci, 2001; Perugini et al., 2003), although results are somewhat mixed in these studies.

In any case, the existent literature regarding the relation between SVO and reciprocity leaves room for improvement. First, the measures that have been used to assess SVO in these studies yield categorical data, i.e., data on the lowest scale level of measurement (Stevens, 1946), which results in low statistical power due to restricted variance. Concretely, subjects have typically been categorized as prosocial, individualistic, or competitive, or even dichotomized as prosocial vs. proself. This severely limits statistical power and may yield null results erroneously. Second, some of the cited studies asked subjects to make hypothetical choices rather than decisions with real consequences, which complicates the interpretation of results since a person's reported intention of how to behave in a particular situation is not necessarily consistent with that person's real behavior in the respective situation (see, e.g. Ajzen, Brown, & Carvajal, 2004; Sheeran, 2002). Third, and most importantly, no study so far has assessed reciprocity profiles at the individual level. That is to say that no study to our knowledge has assessed how individuals differing in SVO responded differently to a set of distinct previously observed choices from different others. The present study addresses these three issues by *a)* measuring SVO on a continuous scale by means of the SVO Slider Measure (Murphy, Ackermann, & Handgraaf, 2011), *b)* implementing complete incentive compatibility, and *c)* requiring that each subject responds to a set of empirically observed previous decisions made by matched real interaction partners. Consequently, the data obtained by the current study allow for a fine-grained and comprehensive assessment of reciprocity as an individual difference variable, and support a more detailed analysis of its relation with SVO compared to previous studies.

The conjecture that SVO is to some extent dependent on situational factors raises an important question, namely whether SVO is a responsive motivational state rather than a purely stable personality trait, as has often been tacitly assumed in previous SVO research. The answer to this question is fundamentally connected to a long-standing discussion in personality and social psychology research, namely the person-situation debate. It has long been acknowledged in psychology that some personality factors can be conceptualized as both state and trait. The most prominent exemplars are anxiety (see, e.g. Catell, 1966; Spielberger, 2010), and anger (e.g. Spielberger et al., 1999), but other personality factors show substantial within-person variability as well (Fleeson, 2001). The apparent contradiction that the same personality variable can be both stable over time and variable within a person has evoked a considerable and sometimes heated debate in psychology. Some researchers went so far as to deny the usefulness of a state-trait distinction (Allen & Potkay, 1981), while others decidedly expressed a contrary opinion (e.g. Zuckerman, 1983). However, many researchers appear to have adopted the position that the distinction between state and trait is not arbitrary, and that the two concepts are to some extent dynamically intertwined, and often dependent on each other (e.g. Chaplin et al., 1988; Endler & Kocovski, 2001; Fridhandler, 1986; Mischel & Shoda, 1998; Steyer et al., 1992). To quote Fleeson (2004, p. 83), "The person-situation debate is coming to an end because both sides of the debate have turned out to be right." With respect to SVO, previous results show a high test-retest reliability of $r = 0.915$ over a one-week period (Murphy et al., 2011). This may lead some to conclude that SVO is only a static trait, but we would not support this simplistic conclusion. As we show in the remainder of this paper, there is substantial systematic variability in how people's social preferences change in regards to updated information about the other person, and we posit that it is worth paying attention to the situation (i.e., state) side of the SVO construct as well as the person (i.e., trait) side.

6.2 Method

In order to address the research questions outlined above, we use an experimental setup that consists of two parts, A and B, where the data collected in experiment A serve as stimuli for experiment B. Concretely, experiment A allows us to conduct experiment B without using deception, such that subjects in both experiments A and B make real decisions with real monetary consequences, resulting in a fully incentive compatible experimental design. The procedure we employed is explained in the next sections.

6.2.1 Experiment A: Collecting stimuli for experiment B

Experiment A was conducted in terms of a paper-pen choice task, where a total of 148 subjects from various disciplines made just a single decision, namely to choose one out of four options (A, B, C, or D) of how to distribute money between themselves and a mutually anonymous other person. These four distribution options are shown in Table 6.1. Each option dominates[4] the other three options with respect to a particular motivational goal, namely the goal to maximize relative gain (option A dominates), the goal to maximize the own payoff (option B dominates), the goal to maximize the sum of payoffs (option C dominates), and the goal to maximize the other's payoff (option D dominates). These four goals represent four prototypical SVOs: competitiveness, individualism, prosociality, and altruism, respectively. Hence, when a person chooses one of the four options, her "type" is revealed and her primary social preference can be identified. The purpose of experiment A was to obtain choice results that would serve as stimuli for experiment B. The reason for this two-part design is to avoid using deception as a research practice and moreover to maintain incentive compatibility for the participants.

Data collection for experiment A was as follows. The experimenter was introduced to the subjects by a university instructor who then announced that a brief decision making task would be handed out. The experimenter then thanked the group for their willingness to participate in the experiment and told the subjects that participation is voluntary and participants would be compensated. Each subject was then handed the decision sheet with the four distribution options and an envelope. Furthermore, subjects were informed that decisions of the same kind would be made by other people in the future, and that they will be randomly matched in pairs with one of these future

[4] In the terminology of Messick & McClintock (1968), the item is a quadruple dominance decomposed game.

decision makers, such that final payoffs would be determined. Importantly, the subjects were not informed, however, that their decisions would be revealed to the future decision makers they could then be matched with. After all the subjects had made their decisions and put their marked decision sheet in their envelope, the experimenter collected the envelopes. Also an email list was distributed so that subjects could be contacted for payment once the future decision makers had made their choices. Concretely, the subjects' student id numbers were used to match subjects with their corresponding payment that they then received once experiment B (see below) was completed. On average the pen and paper task took 10 minutes to complete and subjects earned an average amount of 4.1 Euros (min = 2.5, max = 4.8). The choices subjects made in experiment A are reported in Table 6.1. To be clear again, experiment A was conducted for the purpose of generating stimuli for experiment B.

Table 6.1: Distribution options in the pen and paper task with corresponding choice frequencies

SVO Type	Decision options: Option label	Payoff for: Self	Payoff for: Other	Choice frequency: #	Choice frequency: %
Competitive	A	85	15	6	4.1
Individualistic	B	100	50	25	16.9
Prosocial	C	85	85	109	73.6
Altruistic	D	50	100	8	5.4

6.2.2 Experiment B: Assessing reciprocity profiles

Experiment B was carried out over 12 experimental sessions with a total of 148 subjects (same sample size as in experiment A, but different subjects) conducted at the Max Jung laboratory at the University of Graz. The experiment was programmed using z-Tree (Fischbacher, 2007) and subjects were recruited using ORSEE (Greiner, 2004) which ensured that each subject only participated in the experiment once. In the experiment, subjects made decisions in terms of allocating points which were then exchanged at a conversion rate of 100 Points = 2.50 Euros. Subjects were paid according to their decisions and the decisions of a corresponding interaction partner in one randomly selected Slider Measure item from the baseline condition and also one randomly selected Slider Measure item from one randomly chosen reciprocity condition (explicated below). In the latter case, the interaction partner was a subject from experiment A. This remuneration scheme was common knowledge and is incentive compatible. On average, subjects earned 11.10 Euros (min = 8.50, max = 12.80) including a show up fee of 3 Euros. The average duration of a session was about 60 minutes.

Phase 1: Measuring baseline SVO

Upon arrival to the research laboratory, subjects were welcomed by the experimenter and each participant drew shuffled cards with numbers. These numbers corresponded to the workstation numbers inside the laboratory. Subjects were then led into the laboratory where they first read the instructions regarding phase 1 of the experiment and thereafter were given the opportunity to ask questions. The instructions informed subjects that in phase 1 of the experiment they will be making a series of 15 decisions about how to allocate monetary resources between themselves and an unspecified anonymous other person who would remain unknown to them. After all subjects had read and indicated they understood the instructions, phase 1 of the experiment began. All subjects then completed the 15 items of the SVO Slider Measure (Murphy et al., 2011) in order to assess their individual SVO. This was implemented in terms of a z-Tree SVO module (Crosetto, Weisel, & Winter, 2012a). Phase 1 therefore served as the baseline condition, where the subjects' unconditional distributive preferences were assessed.

The SVO Slider Measure consists of 6 primary and 9 secondary items. The 6 primary items allow for the assessment of a person's general SVO on a continuous scale in terms of an angle. An SVO angle of 0° indicates perfect selfishness, while a positive angle indicates the degree of positive concern about the payoff for another person (i.e., increasing prosociality), and a negative angle indicates the degree of negative concern about the payoff for the another person (i.e., increasing competitiveness).[5] The secondary items of the Slider Measure (items 7-15) allow for distinguishing inequality aversion from joint gain maximization among prosocial decision makers. For further details about the SVO Slider Measure, see Murphy et al. (2011) and Ackermann & Murphy (2012).

Phase 2: Measuring conditional SVO

When all subjects had completed the baseline condition (i.e., phase 1) they were given information regarding phase 2 of the experiment (i.e., the reciprocity conditions) on their computer screens. Subjects were informed that in phase 2 they would be presented with the choices of four different people who had decided previously (i.e., in experiment A) how to allocate monetary resources between themselves and an anonymous other person by choosing one out of the four options indicated in Table 6.1. The subjects were then instructed to complete the 15 items of the Slider Measure with respect to *each* of these four specific and different people. Subjects were informed that they will be randomly matched with one of these four people and that the corresponding decisions made by the subject and the matched other person will be realized for payment.[6]

[5] In essence, the SVO angle is a trigonometric function of parameter α in the utility function $U(\pi_s, \pi_o) = \pi_s + \alpha\pi_o$, where π_s is the payoff for the self, and π_o is the payoff for the other person.

[6] The relative frequencies of how many times each of the four different options had been chosen in experiment A were taken as the probabilities that a subject would be matched with a person who made a corresponding choice for determining final payoffs. However, subjects in experiment B were neither informed about these relative frequencies

Subjects were further informed that the people they are responding to had only been informed that they would be matched with someone for the determination of final payoffs, but that they had not known that their decisions would be revealed to this other person they would be matched with. We informed subjects about this in order to make clear to them that the choices revealed to them are direct indicators of the other persons' baseline preferences (or their *type*), rather than potentially the result of strategic thinking or misrepresentation. Furthermore, before subjects began to complete the Slider Measure in phase 2 with respect to each of the four others, the experimenter walked around the laboratory showing the subjects the stack of decision sheets from experiment A to assure them that they were going to respond to real decisions made previously by real people. However, subjects were not given information about the distribution of "types" observed in experiment A. They were only informed that "one decision concerning one of these four other persons will be randomly drawn at the end of the experiment" and that this "randomly drawn decision will determine their own payoff and the payoff of this specific other person".

In this second phase of experiment B, each subject was presented with each of the four possible choices people from experiment A had made. Hence, the experiment is a within-subjects design where the subjects were sequentially responding to one person who had chosen option A (competitive condition), to one person who had chosen option B (individualistic condition), to one person who had chosen option C (prosocial condition), and to one person who had chosen option D (altruistic condition) in experiment A. The sequential order of the four conditions was randomized across subjects in order to control for order effects. In each of the four conditions, subjects were first presented with the four options indicated in Table 6.1 and the corresponding choice made by the other person in experiment A (competitive, individualistic, prosocial, or altruistic), and were then asked to indicate in words what they think of this other person and the choice this other person made in an open ended text box. In addition, subjects had to indicate how much the other person had allocated to themselves and to the other, and could only proceed in the experiment when the response was verified as correct. This additional comprehension check was implemented to mitigate any possible confusion between the amount allocated to themselves and the amount allocated to the other. The written statements were elicited from subjects for two reasons. First, the statements allow for verifying whether the decision made by the "other" was interpreted in a sensible way by the subject. If, for example, a subject confronted with the altruistic choice indicated a statement such as "This person was obviously motivated to minimize the payoff the other person –in this case me– would receive", this would indicate that the subject misinterpreted the altruistic choice by confusing it with a competitive choice; these confused responses can be flagged and removed from subsequent analyses. Second, the elicited statements allow for a qualitative analysis of the cognitive and emotional reactions subjects express when confronted with other peoples' revealed social preferences. After the subjects indicated their opinion about the interaction partner and the interaction partner's decision, they were then asked to complete the 15 items of the SVO Slider Measure in response to this particular interaction partner whose distributive social preferences have been revealed. This procedure was

nor about the corresponding details of the matching procedure.

the same under each of the four conditions all subjects went through during the experiment. The procedure of presenting each subject with each possible decision made by people in experiment A is analogous to applying the strategy method developed by Selten (1967). As in other cases, this data collection method yields a rich dataset for analysis and modeling.

Subjects then filled out the HEXACO personality questionnaire (Ashton & Lee, 2009; Ashton et al., 2004) and answered some sociodemographic questions. The HEXACO measure of personality was chosen because it is claimed to allow for a more fine-grained analysis of prosocial behavior (Hilbig et al., 2013). We used the 60 item German HEXACO version (Moshagen et al., forthcoming). Subjects who had finished were asked to step outside the laboratory and wait until all subjects had finished answering the questions. Finally, the experimenter asked subjects to step into the laboratory one by one to privately collect their payment.

In the following results section one can note that sample size varies slightly across different conditions and analyses. The reason for this is that we include subjects for a particular analysis only if they satisfy both of the following two conditions. First, a subject must show transitive choice patterns in the Slider Measure under the experimental conditions that are analyzed. Intransitive choice patterns in the Slider Measure are indicators of random responding (see Murphy et al., 2011) and were observed here only rarely (i.e. 2.7% in the baseline condition). Second, there must not be written statements collected under the reciprocity conditions indicating that a subject misinterpreted the choice made by the corresponding other person. If, for example, a subject had written a statement such as "The other person made a choice that maximized his own payoff" in the altruism condition, this subject would be excluded from all analyses involving the altruism condition. Because of these exclusion criteria the sample sizes are not constant across analyses. An overview on the respective sample sizes in the different conditions can be found in Table 6.2. Less stringent exclusion criteria were also considered, and when implemented the pattern of results remained consistent with what is reported here. Raw data are publicly available for download.

6.3 Results

6.3.1 Primary results

A summary of the main results is shown in Figure 6.1. The distributions of SVO angles as produced in response to the choices made by the competitive, individualistic, prosocial, and altruistic "others" each differ significantly from the baseline SVO angle distribution as indicated by Kolmogorov-Smirnov (K-S) tests.[7] This is clear evidence that a sufficient number of subjects

[7] The K-S statistics from the SVO reciprocity distributions in comparison to the baseline SVO distribution are $D = 0.30, p < .001$ for the competitiveness condition, $D = 0.17, p = 0.022$ for the individualism condition, $D = 0.25, p < .001$ for the prosociality condition, and $D = 0.27, p < .001$ for the altruism condition. Wilcoxon signed rank tests

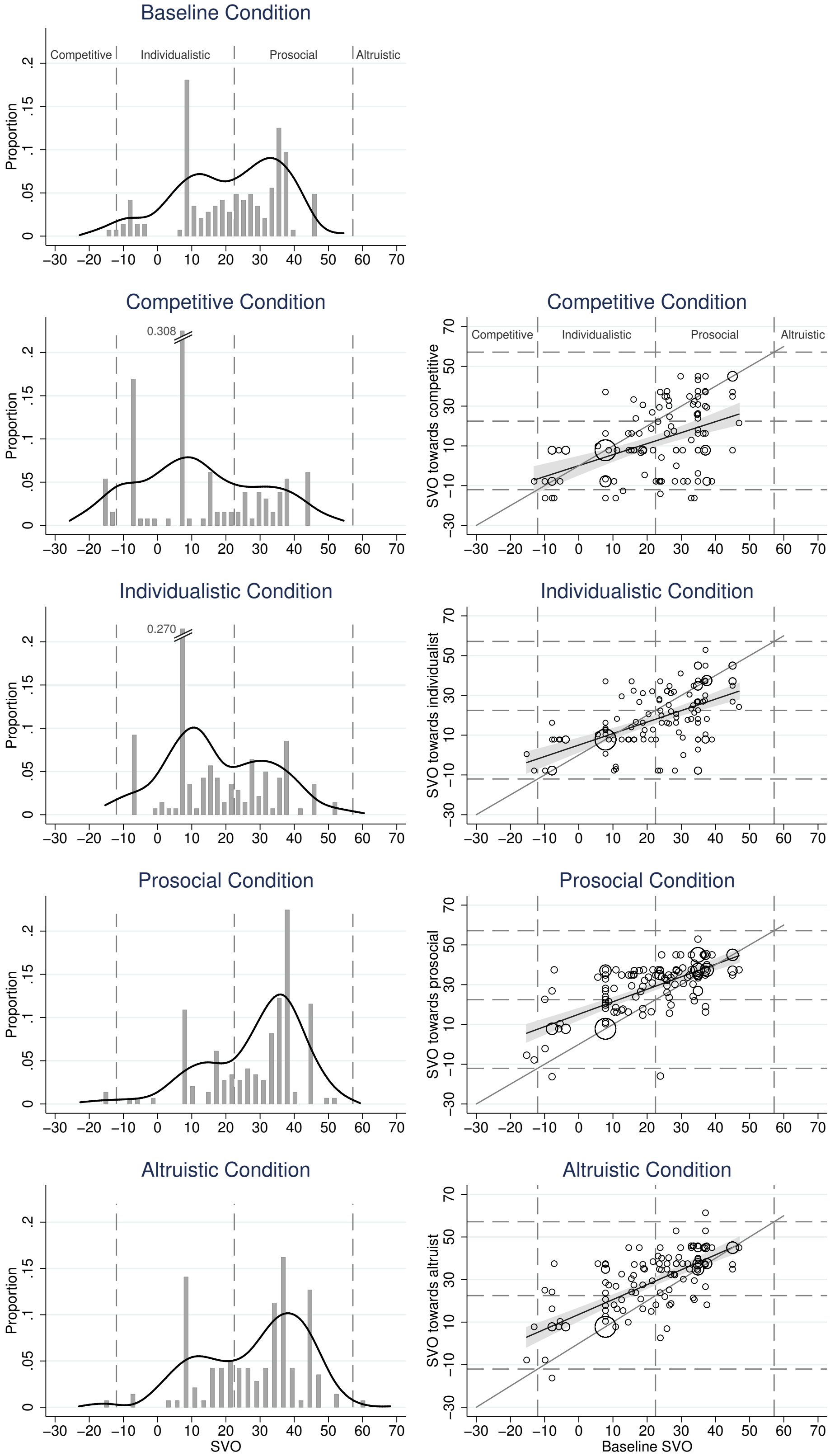


Figure 6.1: SVO and Reciprocity

altered their SVO angles in response to the choices made by the corresponding “others" *in particular directions*. Furthermore, the comparisons[8] among the SVO reciprocity distributions indicate that they are all significantly different from each other as well, except for the comparison between SVO in response to prosociality and in response to altruism (Wilcoxon signed rank test $p = 0.89$; K-S test $D = 0.08, p = 0.66$). This means that subjects respond differently to competitiveness than to individualism, prosociality, and altruism, and differently to individualism than to prosociality and altruism, but respond to prosociality and altruism in generally the same way. To have an impression of what the reciprocal reactions mean on the aggregate level in monetary terms, we can compute an approximation of subjects’ willingness to pay for an increase of one monetary unit on the side of the interaction partner by taking the tangent of subjects’ SVO angles. In the baseline measurement, subjects are on average willing to pay 0.42 monetary units to increase the “other’s” payoff by one unit. However, when the “other” is known to have made a competitive choice, the average willingness to pay drops by 40% to 0.25 monetary units. In response to an individualistic person, the average willingness to pay drops by 17% to 0.35. In contrast, when the “other” is known to be prosocial, the average willingness to pay for a one-unit increase in the payoff for the “other” increases by 38% to 0.58 compared to the baseline willingness to pay. Finally, when the “other” is known to have made an altruistic choice, the average willingness to pay increases by 41% to 0.59 monetary units compared to the baseline.

The scatterplots in Figure 6.1 summarize the entirety of experiment B. They show how subjects’ SVO angles from the baseline condition relate to their SVO angles as produced in response to each of the four other types. Observations on the diagonal line indicate no shift in SVO (i.e., no reciprocity), while observations above the diagonal indicate upward shifts (i.e., increase in SVO angle indicating positive reciprocity) and observations below the diagonal indicate downward shifts (i.e., decrease in SVO angle indicating negative reciprocity). Table 6.2 shows the mean shifts in SVO angles as well as the percentage of subjects shifting and not shifting in a particular direction for each condition on the aggregate level. As can be seen in Figure 6.1, shifts in SVO on the individual level predominantly occur when the interaction partner’s SVO diverges from the decision maker’s own SVO. A competitive interaction partner evokes negative reciprocity predominantly among subjects who tend to be prosocial, while subjects who themselves tend to be competitive do not show much of a shift in their SVO angles. In contrast, a prosocial or even altruistic interaction partner evokes positive reciprocity predominantly among subjects who tend to be individualistic or competitive, while subjects who themselves tend to be prosocial do not show much of a shift, but rather just express their baseline prosociality towards the prosocial interaction partner. It is interesting to see that even subjects who tend to be competitive, as indicated by negative baseline SVO angles, can show considerable positive reciprocity in response to a prosocial or altruistic interaction partner.

---

corroborate these results, indicating that subjects’ SVO angles changed in each reciprocity condition compared to the baseline condition.

[8] Again, both K-S tests and Wilcoxon signed rank tests were used and indicated the same results.

Table 6.2: Descriptive statistics on SVO and changes in SVO per condition

Condition	Mean SVO	N_{SVO}	Mean ΔSVO	$N_{\Delta SVO}$	Negative shift (%)	No shift (%)	Positive shift (%)
Baseline	21.1	144					
Competitive	12.4	130	-9.4	127	59.8	16.5	23.7
Individualistic	17.7	141	-4.0	137	47.5	22.6	29.9
Prosocial	28.7	147	7.3	144	15.3	15.3	69.4
Altruistic	28.5	142	7.3	139	15.8	18.7	65.5

Note: The column *Mean SVO* indicates the average SVO angle as obtained by the SVO Slider Measure per condition, while the subsequent column (N_{SVO}) provides the corresponding sample sizes. The column *Mean ΔSVO* indicates the average difference between the SVO angles as obtained under the respective condition and those obtained under the baseline condition, while the subsequent column ($N_{\Delta SVO}$) provides the corresponding sample sizes. The numbers in column $N_{\Delta SVO}$ are smaller than the values in column N_{SVO} because subjects had to show a transitive choice pattern in both the baseline condition and the corresponding reciprocity condition in order to be included for the analysis of difference scores (i.e. ΔSVO). The percentages of subjects showing a negative, zero, or positive shift in their SVO angle per condition are reported in columns 6, 7, and 8, respectively.

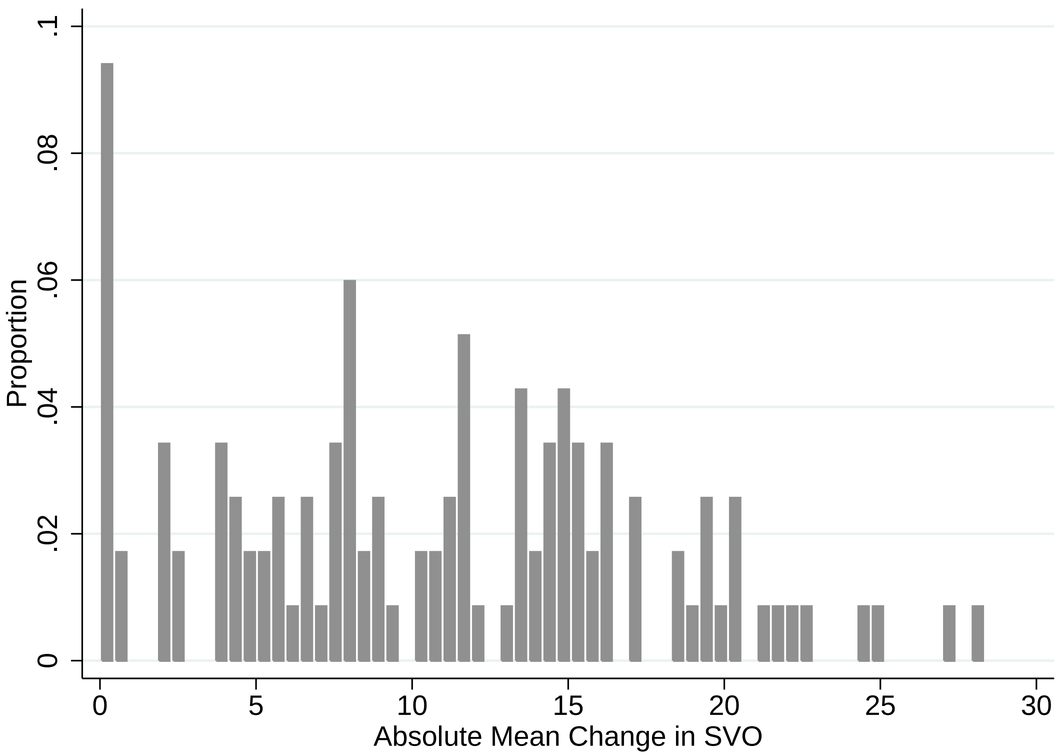


Figure 6.2: Aggregate absolute reciprocity (*n = 117* as indicated by the distribution of average absolute shifts in SVO (i.e., $|\Delta SVO|$) across the four conditions.)

However, there are large individual differences in the degree to which individuals are reciprocal. About ten percent of subjects are not reciprocal at all, as evidenced by identical SVOs in response to the revealed actions of others as compared to their baseline SVO.[9] However, the vast majority of subjects did show at least some degree of reciprocity, and many of them show considerable degrees of reciprocity. Figure 6.2 shows the distribution of the absolute average shifts in SVO angle. One thing worth noting here, however, is that the degree of reciprocity when computed as the absolute average shift in SVO angle does not take into account that some subjects showed negative reciprocity towards prosociality and/or altruism, some other subjects showed positive reciprocity towards individualism and/or competitiveness, and yet some other subjects showed consistent positive or negative reciprocity across all conditions. Although these patterns are difficult to rationalize, we are hesitant to exclude these cases from the analysis since we can not rule out that these patterns are consequences of the subjects' real preferences rather than just noise. Figure 6.3 shows examples of eight different general patterns of reciprocity profiles, each with an indication of the proportion of subjects showing the corresponding type of pattern. Subject 14 is an example of those 9.4% of subjects who do not react in any way to the revealed preferences of "others", while subject 54 is an example of subjects who show negative reciprocity towards competitiveness and individualism, and positive reciprocity towards prosociality and altruism. This pattern of both positive reciprocity when matched with a prosocial and altruistic other and negative reciprocity when matched with a competitive and individualistic other is the most common pattern (23.9%). Subject 113 is an example of an individualistic subject who does not react to competitiveness nor individualism, but shows positive reciprocity towards prosociality and altruism. Subject 111 is an example of a prosocial subject who does not react to prosociality nor altruism, but shows negative reciprocity towards competitiveness and individualism only. Note that these patterns are exhibited by only 4.3% and 1.7% of subjects, respectively. In addition, there are a number of subjects (22.2%) who only show partial negative and/or positive reciprocity. An exemplar of this response type is subject 109 who only reacts with a decrease in SVO when matched with a competitive interaction partner and an increase in SVO when matched with an altruistic interaction partner.

Subjects 90 and 71 are examples of subjects who show a reciprocal reaction to others but do not differentiate between these revealed types, resulting in a general decrease (7.7%) or increase (16.2%) in SVO across conditions. As mentioned earlier, patterns of that kind are difficult to explain. Perhaps the simplest explanation for such patterns would be that these subjects are not reciprocal at all, but made a mistake in one or a few items in the Slider Measure under the baseline condition which they then corrected under the reciprocity conditions. Another explanation might be that these subjects really have different distributive social preferences when the other is a specified particular other person compared to an unspecified "someone" who will be randomly

[9] The actual share of subjects who do not show significant reciprocity is likely slightly underestimated due to measurement error. There may be subjects whose SVO angles changed slightly across conditions not because they reacted to the others in some way, but because they did not reproduce their choice pattern exactly between conditions. However, due to the Slider Measure's very high reliability (Test-retest $r = 0.915$, see Murphy et al., 2011), measurement errors are expected to be fairly small.

selected after the decisions have been made. However, we can not shed further light on the rationale behind patterns of that kind by means of the data available to us, but can only speculate about potential causes. Other patterns that can not be assigned to one of the afore mentioned categories are exhibited by 14.5% of the subjects. These examples make clear that there is substantial heterogeneity in the patterns of how subjects react to various others when information about them is revealed.

One question would be to consider if different types of DMs are more reactive than others. For example it could be conjectured that people with prosocial SVOs are more reactive than people with individualistic motivations. However, the bivariate correlation between baseline SVO and degree of reciprocity (i.e., absolute shifts in SVO across the four conditions) is not significantly different from zero, indicating that there is no significant relation between SVO and reciprocity (Figure C1 of the Appendix shows the corresponding scatterplot). The absolute degree of reactivity in people's social preferences is not conditioned on their baseline preferences.

6.3.2 Secondary results

SVO and its relation to positive and negative reciprocity

The mean positive change in SVO angles over all four conditions captures the strength of subjects' positive reciprocal reactions, while the absolute mean negative change indicates the strength of their negative reciprocal reactions. On the aggregate level, subjects showed an average positive change in SVO of 5.3 degrees and an average negative change in SVO of 5.4 degrees over all four conditions.[10] We observe that 26.5% of all subjects exhibit no positive reciprocity, and 35.0% do not show negative reciprocity. Overall, we do not find a significant difference between the two distributions (K-S test, $D = 0.11$, $p = .407$), indicating that the effects of positive and negative reciprocity are about the same on the aggregate.

Figure 6.4a shows a scatterplot indicating how positive and negative reciprocity are jointly distributed on the individual level. We observe some cases near the diagonal that exhibit a similar degree of positive and negative reciprocity. Cases below the diagonal show stronger negative than positive reciprocity including cases that exhibit negative reciprocity only. The opposite is true for cases above the diagonal where either less negative reciprocity as compared to positive reciprocity is observed or where negative reciprocity is completely absent. We can identify several factors that are partially responsible for the pattern shown in Figure 6.4a. First of all, we have seen that individual reciprocity profiles are dependent on SVO in that subjects with higher SVO angles predominantly show negative reciprocal reactions towards competitive and individualistic interaction partners while showing no or only little positive reciprocity towards prosociality and altruism, while the opposite holds for subjects with lower SVO angles. Furthermore, we have seen that some subjects appear to show only positive or only negative reciprocity across all conditions.

[10] The difference is not significant (Wilcoxon signed rank test, $p = 0.985$). This difference is increased if we exclude cases that show no positive or negative reciprocity (7.2 and 8.3) but remains insignificant (Wilcoxon signed rank test, $p = 0.171$). Figure C2 of the Appendix shows the distributions of positive and negative reciprocity separately.

Figure 6.3: Examples of individual reciprocity profiles (n = 117)

And finally, there are subjects -predominantly with intermediate baseline SVO angles- who show as much negative reciprocity towards competitiveness and individualism as they show positive reciprocity towards prosociality and altruism. As mentioned earlier, Figure 6.3 shows examples for each of these patterns. The combination of these factors together, we argue, shape the triangle form of observations shown in Figure 6.4a, and lead to a negative correlation of $r = -0.367$ ($p < .001$) between positive and negative reciprocity. This interpretation of the results is also supported by the observation of a negative correlation of $r = -0.474$ ($p < .001$) between SVO (baseline) and positive reciprocity, and a positive correlation of $r = 0.479$ ($p < .001$) between SVO (baseline) and negative reciprocity as visualized in Figure 6.4b and 6.4c, respectively. Since there are two separate effects in opposing directions for positive and negative reciprocity, this explains why we observe no correlation between baseline SVO and overall reciprocity.

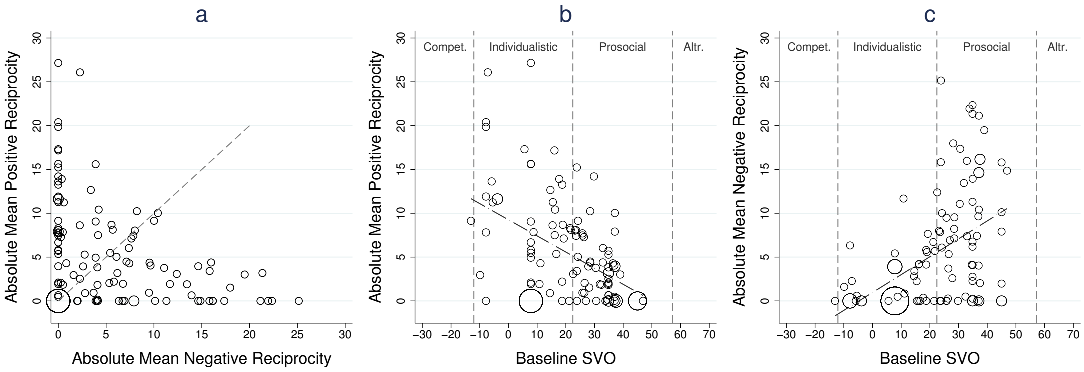


Figure 6.4: Relations between positive and negative reciprocity and baseline SVO (n = 117)

Hexaco

All factors of the Hexaco questionnaire show acceptable internal reliability.[11] The correlation between the Honesty-Humility factor and baseline SVO is $r = 0.25$ ($p < 0.01$). The direction and magnitude of this relation is in accordance with previous findings (e.g. Hilbig & Zettler, 2009; Hilbig, Zettler, Moshagen, & Heydasch, 2012). Also, this relation is the only one between baseline SVO and Hexaco scales that remains significant after Bonferroni correction. Without Bonferroni correction, the relation between baseline SVO and the Hexaco scale "Openness" is significant as well with $r = 0.17$ ($p < .05$).

We find no relationship of any of the HEXACO scales with overall reciprocity. However, when SVO (baseline) is statistically controlled, we find a significant positive relation between the Honest-Humility factor and average *positive* reciprocity ($r_{partial} = .23, p < .05$), and a significant negative relation between Honesty-Humility and average absolute *negative* reciprocity ($r_{partial} = -.20, p < .05$).

[11] Cronbach's alphas for the factors are as follows. Honesty-Humilty: 0.77; Emotionality: 0.84; Extraversion: 0.77; Agreeableness: 0.76; Conscientiousness: 0.81; Openness: 0.78.

Honesty-Humility is described as a "tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation" (Ashton & Lee, 2007, p.156). So it makes sense that people who exhibit this trait show less negative and more positive reciprocity than those who do not. These findings add further support to the argument of separating Honesty-Humility from other personality traits (Hilbig et al., 2013), since it appears to have separate effects on cooperative behavior in different situations. Nevertheless, the overall pattern of our results also suggests that tendencies towards positive and negative reciprocity may indeed be considered relatively independent motivational inclinations rather than collinear ones, which would be in support of other findings (e.g. Eisenberger et al., 2004; Yamagishi et al., 2012; Egloff et al., 2013). Importantly, however, our results are exploratory rather than confirmatory in this respect, and a new experiment with a potentially larger sample size would be required to judge the robustness of the effects concerning the relationship between personality factors and reciprocity that we found.

Inequality aversion

An alternative explanation for shifts in SVO angles may be that people do not change their baseline distributive social preferences, but simply always express the same distributive social preference, namely inequality aversion, when responding to advance payoff allocations. If a person wanted to minimize the difference between *final* payoffs for both decision makers, then this person would be expected to show a behavioral pattern that –by itself– is indistinguishable from perfect reciprocity. That is, such a person would respond competitively to a competitor, individualistically to an individualist, prosocially to a prosocial, and altruistically to an altruist, since these responses would guarantee that both decision makers receive equal payoffs in the end. We can test whether this alternative explanation holds. The secondary items of the Slider Measure allow for the assessment of the degree of inequality aversion (IA) among prosocial people (i.e. subjects with an SVO angle between 22.45° and 57.15°) given that their choices in the secondary items, too, are consistent with a prosocial orientation (i.e. inequality aversion or joint gain maximization) rather than an individualistic or altruistic orientation.[12] Concretely, for subjects who meet these requirements, we can compute an IA index ranging from zero to one, where an index of zero indicates that a person is perfectly inequality averse (i.e., showing a choice pattern in the secondary items of the Slider Measure that is in perfect accordance with minimizing differences in payoffs), and an index of one indicates that the person is perfectly joint gain maximizing (i.e., showing a choice pattern in the secondary items of the Slider Measure that is in perfect accordance with maximizing the sum of payoffs). Among the 117 subjects who showed transitive choice patterns in the baseline SVO measurement and correct interpretation of the interaction partner's type across all conditions, there are 51 prosocial subjects (out of 64 prosocials in total) for which an IA index can be computed. Among these 51 prosocial subjects, 33 (i.e., 28.21% of all 117 subjects under consideration) tend to be inequality averse as indicated by an IA-index of less than 0.5. Only one out of the 117 subjects under consideration is perfectly

[12] For details about the IA index, see Murphy et al. (2011) and Ackermann & Murphy (2012).

inequality averse as indicated by an IA-index of exactly zero (Figure C3 of the Appendix shows the distribution of the IA-index). If the alternative explanation for the shifts in SVO angles holds, then we would expect that the vast majority of these 33 subjects who tend to be inequality averse respond competitively to a competitive interaction partner. However, this is not what we find. In fact, only one out of these 33 inequality averters responded competitively to a competitor, thereby equalizing final payoffs (the distribution of these 33 subjects' SVO angles when matched with a competitive other is shown in Figure C4 of the Appendix). The modal response was individualistic with an SVO angle of -7.82°, which results from a perfectly individualistic choice pattern across all items where benefiting or harming the "other" is costly, and a competitive choice in the item where harming the "other" is free. Clearly, such a choice pattern does not serve the purpose of equalizing final payoffs, but punishing the other when it is free, and being concerned with one's own payoff when punishment is costly. We therefore refute this alternative explanation of inequality aversion as the main driver of shifts in SVO angles.

A further question regarding inequality aversion is whether this particular distributive social preference can be expressed in terms of a reciprocal reaction. Table 6.3 shows for each condition the percentage of subjects that can be categorized as prosocial (i.e., subjects with an SVO angle between 22.45° and 57.15°) subdivided into three categories: Joint Gain Maximization (JointMax, i.e., prosocial subjects with an IA index ≥ .5), Inequality Aversion (IneqAvers, i.e., prosocial subjects with an IA index < .5), and prosocials whose choice patterns do not allow for the computation of an IA index (n/a, i.e. prosocial subjects who showed a choice pattern in the secondary items that is neither consistent with joint gain maximization, nor inequality aversion). The fifth row shows the median IA index across all subjects for which an IA index can be computed in the corresponding condition. In order to facilitate comparisons across conditions, only those $n = 117$ subjects are considered who show transitive choice patterns in the Slider Measure and no misperception of the "other's" types across all conditions. As can be seen in table 6.3, the percentage of subjects showing an inequality averse choice pattern in the Slider Measure varies substantially across conditions in expected ways. Regarding our question, the comparison between the percentage of inequality averse subjects in the baseline condition and the percentage of inequality averters in response to prosociality is most interesting. There are 7.7 percent more inequality averters among the subjects when they are responding to prosociality compared to the baseline condition. Table 6.4 informs about how this 7.7 percent increase in total is realized. While 9 subjects who showed inequality aversion in the baseline condition changed their distributive social preferences when responding to prosociality, 18 subjects who did not show concern about equality in the baseline condition responded to prosociality with an inequality averse choice pattern. Hence, there are 18 observations of inequality aversion as a response to prosociality. This is clear evidence that inequality aversion can be expressed in terms of a reciprocal reaction.

Table 6.3: Proportion of inequality averse, joint max, and other subjects as well as the median IA index per condition

		Conditions				
		Baseline	Compet	Indi	Proso	Altr
Other types		45.3	71.8	64.1	31.6	33.3
	IneqAvers	28.2	13.7	15.4	35.9	27.4
Prosocial types	JointMax	15.4	9.4	12.8	23.9	30.8
	n/a	11.1	5.1	7.7	8.5	8.5
IA median		.35	.33	.38	.34	.53

Note: Numbers in row one through row four are percentages of subjects categorized in the respective class per condition. These proportions are based on n = *117* subjects. The numbers in row five are the median values of the IA index from those subjects who are categorized as prosocial and show a choice pattern in the Slider Measure's secondary items consistent with a prosocial orientation. The IA index can range between 0 (i.e., perfect inequality aversion) and 1 (i.e., perfect joint gain maximizing).

Table 6.4: Type comparison between baseline and prosocial condition

		Prosocial condition			
		Other types	IneqAvers	JointMax	n/a
Baseline	Other types	31	12	4	6
	IneqAvers	2	24	5	2
	JointMax	1	3	14	0
	n/a	3	3	5	2

Note: Numbers in the table are counts of subjects (n = 117).

6.4 Discussion

We have defined reciprocity as a change in the distributive social preferences of an individual in response to information about an interaction partner's characteristics. Thus we have conceptualized reciprocity as an individual difference that can be measured and used as a dependent variable. We show that the vast majority of people do change their distributive social preferences towards an interaction partner when the interaction partner's "type" is revealed, thereby showing reciprocity. Furthermore, we show that there is considerable heterogeneity in the degree to which people are reciprocal and how degrees of reciprocity relate to SVO. These findings have significant implications regarding the study and modeling of preferences, beliefs, and behavior in experimental games that involve the revelation of information about interaction partners. It has been common practice in the economic sciences to infer distributive social preferences from responder behavior in experimental settings such as ultimatum games, for instance. It has been neglected, however, that the responder behavior may not reveal the responder's baseline distributive social preferences, but the responder's reciprocal reaction to revealed information about the proposer's characteristics, namely the proposer's behavior. In situations of that kind, it has often been assumed that the expression of one particular distributive social preference, inequality aversion, is responsible for the decision to accept an equal split and refute offers that deviate too much from it in ultimatum games, for instance. We have shown that there is another explanation for the behavioral pattern often observed in these types of experimental games. Namely that distributive social preferences may change in response to information about the interaction partner, such that decision makers may assign a different weight – positive, zero, or negative – to the outcome for the interaction partner once information about the interaction partner is revealed compared to when no information is available. Consequently, both punishment and reward behavior can result from reciprocity as an individual difference and social preferences show a reactivity to information about the other.

Our results clearly support the validity of recently developed models of behavior in experimental games that take into account the significance of information about the types of interaction partners (e.g. Charness & Rabin, 2002; Cox et al., 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Falk et al., 2008; Levine, 1998; Rabin, 1993; Stanca et al., 2009). However, there is large heterogeneity in both peoples' baseline distributive social preferences and the degree to which people are reciprocal. Our results suggest that apart from people's beliefs, both individual differences in distributive social preferences and individual differences in the degree to which these preferences can change in response to information about the environment – predominantly about interaction partners – are key to understanding behavior in many situations of interdependence and strategic choice.

Lastly, this paper highlights the general usefulness and broad potential of measuring SVO as a dependent variable. One can readily imagine a host of similarly structured experiments where baseline SVO is first measured and then contrasted against another measure of SVO in regards to a specified other, or in regards to the same other in a different context or with updated information. For example, one could consider to what extent SVO changes when the other person is part of an in-group/out-group? Or how does the reactivity of social preferences relate to the decision maker's own, and the other's, demographic or other characteristics. Are there interactions between the decision maker's and the other's characteristics in affecting reciprocal reactions (i.e., homophily effects)? How does the status of the other (earned or otherwise bestowed arbitrarily) influence SVO? In general, to what degree are decision maker's willing to make costly tradeoffs when the features (attributes, identity, or past behavior) about the other are made explicit? And to what extent are these preference dynamics dependent on social context? Our findings demonstrate the flexibility of social preferences, but understanding the general structure of these contingencies at the individual level is an open question and one of great interest across many disciplines in the social sciences.

7 Discussion and conclusion

7.1 Lessons learned

Research on SVO has produced a tremendous wealth of findings over the last forty years and a substantial body of literature has been built on the basis of this important construct. Also, the thinking about how to conceptualize and measure this construct has evolved over time from a categorical account to a continuous one, which opens up new opportunities. Let us recapitulate what we have learned so far from past SVO research, and what questions SVO research may address in the future.

Since its advent in the late 1960s until very recently the SVO construct has been used almost exclusively as an independent variable. Concretely, SVO research has focused on predicting behavior by the SVO construct or finding relationships between the construct and other variables. For instance, SVO has been shown to be predictive of various types of behavior, such as helping behavior (McClintock & Allison, 1989), donation behavior (Van Lange et al., 2007), proenvironmental behavior (Van Vugt et al., 1996), negotiation behavior (De Dreu & Van Lange, 1995), and cooperation in social dilemmas (Balliet et al., 2009). Furthermore, it has been shown that SVO is associated with attitudinal or personality variables, such as Machiavellianism (e.g. Murphy & Ackermann, 2014b, see chapter five of this dissertation), Social Mindfulness (Van Doesum et al., 2013), Honesty-Humility[1] (Hilbig & Zettler, 2009; Hilbig et al., 2014), and Agreeableness[2] (Hilbig et al., 2014). Also, other branches of research have revealed that a person's SVO is not independent from her belief about other people's SVOs (e.g. Kuhlman & Wimberley, 1976; Aksoy & Weesie, 2012), or her evaluation of cooperative versus uncooperative behavior (e.g. Kuhlman et al., 1992; Liebrand et al., 1986; Van Lange & Liebrand, 1991), for instance. Yet other studies have investigated the ontogenetic development of SVO (e.g. Van Lange et al., 1997) or the construct's association with socioeconomic and cultural factors (e.g. Knight, 1982; Knight

[1] Honesty-Humility refers to a personality factor of the HEXACO scale (see Ashton & Lee, 2007).

[2] Agreeableness here refers to the corresponding factor of the classical Five-Factor model (see, for instance, Costa & McCrae, 2009).

& Kagan, 1977). Clearly, SVO is an extraordinarily well and broadly studied construct, but researchers have used it primarily as an independent variable.

As we have discussed in chapter two, one reason for why SVO has rarely been used as a dependent variable is rooted in the way the construct has predominantly been conceptualized and measured in the past. Originally, the construct has been thought of as a trichotomous variable (see Messick & McClintock, 1968), and has accordingly been assessed on the nominal scale level of measurement (see Liebrand & McClintock, 1988; Van Lange et al., 1997). It was assumed that a person may either have the motivation to maximize the own payoff (individualistic), or to maximize the relative payoff (competitive), or to maximize the sum of payoffs (cooperative) at a given point in time. However, Griesinger & Livingston (1973) showed that these three motivational types are just representations of particular values of an underlying continuum reflecting the weight a person attaches to the outcome of others in relation to the own. Unfortunately, this clarification had not led to a general acceptance of SVO as a continuous variable for reasons that are discussed in chapter two. Instead, researchers often just increased the number of motivational categories people could theoretically be assigned to (see, for instance, MacCrimmon & Messick, 1976; Liebrand, 1984). In another tradition of SVO research, the continuous nature of the construct was naturally acknowledged, though. Concretely, researchers from this tradition have used utility functions for describing preferences over own-other resource allocations in terms of best fitting parameter value combinations (e.g. Grzelak et al., 1977; Radzicki, 1976; Wyer, 1969). This approach of accounting for social preferences is also particularly widespread in economics (e.g. Bolton & Ockenfels, 2000; Breitmoser, 2013; Charness & Rabin, 2002; Cox et al., 2007; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999; Fisman et al., 2007; Geanakoplos et al., 1989; Levine, 1998; Loewenstein et al., 1989; Rabin, 1993). In chapter four of this thesis, several of these utility models were described, ways to fit these models to empirical data were discussed, and corresponding results were reported. In addition to the conceptual clarification provided by Griesinger & Livingston (1973), the substantial variance usually found in individual best fitting parameter values as obtained by utility measurement procedures (discussed in both chapters two and four) serves as further evidence that a categorical account of the SVO construct is unnecessarily limiting and theoretically as well as empirically unwarranted.

In chapter three, a novel instrument for measuring SVO was introduced that accommodates the construct's continuous conceptualization. Concretely, the SVO Slider Measure (Murphy, Ackermann, & Handgraaf, 2011) assesses the weight a person attaches to the payoff for the "other" in terms of a continuous score, allows for distinguishing between inequality aversion and joint gain maximization, and in addition provides a check for transitivity in choice patterns. Also, the measure has been shown to have very strong psychometric properties. Due to the measure's ability to produce continuous scores, it opens up the opportunity to use SVO not only as an independent variable with higher resolution, but also as a dependent variable. That is, the measure makes it possible to study the extent to which SVO can change in response to changes in the environment, such as changes in the information provided about interaction partners, for instance. With categorical measures that produce output on the nominal scale level of measurement, the detection of potentially subtle yet important changes in a person's concern for others is not possible.

Chapter five then reported on a study we conducted (Murphy & Ackermann, 2014b) in which SVO as assessed by the Slider Measure was used both as independent and dependent variable. We could show that SVO is an important predictor of behavior in a public goods game in that it explains a significant proportion of variance in contribution levels in addition to the variance explained by people's beliefs about other people's behavior. Nevertheless, a person's belief about the behavior of others is the best single predictor of that person's contribution in both a one-shot and repeated game. The reason for this is that the vast majority of people who tend to be prosocial are conditional cooperators. That is, they are motivated to contribute but only under the condition –and to the extent– that others contribute as well. Colloquially speaking, prosocials are nice but not stupid. Also, we found that people (i.e. both prosocials and proselfs) are generally more conditionally cooperative in an ongoing game than they indicated to be when asked in a hypothetical situation. Clearly, the use of the Slider Measure substantially facilitated the investigation of the relative impact of SVO (as an independent variable) and beliefs on behavior in a public goods game due to its high-resolution output.

Furthermore, we could show that individual SVO scores can change in response to individual experiences made during a repeated public goods game. Concretely, we found that people who were part of a relatively cooperative group were likely to show higher SVO scores (i.e. shifts towards prosociality) after the game was finished as compared to their baseline score that was measured before the game started, whereas people who were part of a relatively uncooperative group were likely to show lower SVO scores (i.e. shifts towards individualism/competitiveness) after the game. Hence, we could provide one of the first pieces of evidence in the history of SVO research showing that the weight people attach to other people's payoffs can change in response to observations of how other people behave. That is, SVO is malleable to some extent.

In following up on these findings we then conducted an experiment that allowed for an explicit investigation of the extent to which SVO can change on the individual level, which was reported in chapter six. Concretely, we define reciprocity as a change in SVO in response to revealed information about the interaction partner. In order to assess reciprocity we then let subjects complete the SVO Slider Measure in response to each of a set of different interaction partners whose SVOs were unilaterally revealed to the subjects. The resulting SVO scores were then compared to the subjects' baseline SVO scores, which had been assessed beforehand. We could clearly show that the majority of subjects behaved reciprocally in expected ways. On the aggregate, subjects behaved more prosocially when responding to a person who knowingly has a prosocial SVO, and behave less prosocially -sometimes even competitively- when responding to a person who knowingly has an individualistic or competitive SVO. In total, the results clearly indicate that the majority of people are reciprocal, and that a person's SVO can change substantially in response to revealed information about an interaction partner.

7.2 SVO: State or Trait?

The finding that SVO is to some extent dependent on situational factors raises an important question, namely whether SVO is a variable motivational state rather than a stable personality trait as has often been silently assumed in past SVO research. The answer to this question is fundamentally connected to an old debate in personality research, namely the person-situation debate. It has long been acknowledged in psychology that some personality factors can be conceptualized as both state and trait. The most prominent exemplars are anxiety (see, e.g. Catell, 1966; Spielberger, 2010), and anger (e.g. Spielberger et al., 1999). The apparent contradiction that the same personality variable can be both stable over time and variable within a person has evoked a considerably heated debate in psychology. Some researchers went so far as to deny the usefulness of a state-trait distinction (Allen & Potkay, 1981), while others decidedly expressed a contrary opinion (e.g. Zuckerman, 1983). However, many researchers appear to have adopted the position that the distinction between state and trait is not arbitrary, but that the two concepts are to some extent intertwined or at least dependent on each other (e.g. Chaplin et al., 1988; Endler & Kocovski, 2001; Fridhandler, 1986; Steyer et al., 1992). It has been found that even the most widely accepted and used concept of stable personality in psychology –the Big Five (see Costa & McCrae, 2009)– shows considerable within-person variability (Fleeson, 2001). With respect to SVO, we found a test-retest reliability of $r = .92$ over a one-week period (Murphy et al., 2011), but could show that SVO can change dramatically in response to changes in the environment (Ackermann et al., 2014). How can this apparent contradiction be reconciled?

Personally, I very much agree with Fleeson (2004). He argues that traits can be understood as density distributions of states. That is, a person's score on a personality variable may vary considerably on a day-to-day basis, but the average of these scores is highly stable and characteristic of a person. Roughly speaking, the aggregate value of state scores reflects the trait score. For me, the best analogy to the state-trait relationship is the weather-climate relationship. Two regions on earth may have a dramatically different climate, but may at the same time show very similar weather conditions on a particular day depending on environmental factors. Analogously, two people may differ from each other dramatically in how extraverted they generally are, but depending on situational factors, they may show very similar degrees of extraversion in particular situations. However, in a situation that evokes higher degrees of extraversion in both individuals (e.g. a party situation with good friends), the person who has a higher (baseline-) score on extraversion in general may still be more extraverted than the other person in that particular situation. That is, they are likely not only different from each other on average, but also on the extremes. Also, a person may have quite high a general (baseline-) score of extraversion, but may behave quite introvertedly when interacting with a particular interaction partner in the vicinity of which the person does not feel comfortable –i.e. when situational factors exert influence. It is in this sense that I compare the state-trait relation to the climate-weather relation. It makes sense to distinguish between climate and weather, and the distinction is not arbitrary. One region may have a dramatically different climate than another region on earth in that the weather in these regions is dramatically different on the aggregate, but that does not mean that the weather does not vary, or that on certain occasions the weather in these two regions can be very similar depending on the environmental conditions that exert influence at a given point in time. Both weather and climate are important concepts that are intertwined but nevertheless should be distinguished, and so are the psychological concepts of state and trait. With respect to SVO, this means that this construct, too, can be considered a state when it is measured at a time when external conditions do likely exert influence (e.g. when information about the interaction partner is provided), or it can be considered a trait when it is measured at a time when external conditions are unlikely to exert influence (i.e. in the absence of experimental manipulation).

Another question in this respect is whether different measures should be used to assess a trait compared to a state. I argue that the same measure can be used for assessing state and trait, and again I draw upon the climate-weather analogy. A climate has certain attributes, and so does personality. One attribute of climate is temperature, for instance, and one attribute of personality is SVO –I argue. If we want to assess the temperature that is representative of the climate in a particular region, we can either measure it many times over the course of the year and aggregate results for a precise assessment, or measure it on one day where no extraordinary or unusual environmental factors exert influence for a proximate assessment. However, we can also measure the temperature on a day when extraordinary environmental factors do exert influence, thereby focusing on the weather. My point is that in both situations one would use the same instrument, namely a thermometer, because this is the instrument that provides us with the information that we are interested in. Likewise, we can measure SVO in the laboratory where no information about interaction partners is provided and extraordinary environmental influences are unlikely exerting influence on the measurement for assessing SVO-trait, or we can measure SVO when the conditions are very different for assessing SVO-state. But we would use the same instrument for both measurements. And I argue that this holds true for any other personality variable. The nice thing about this approach is that we can then measure a person's score on a variable under manipulated conditions and compare it to the baseline score of that person so we can quantify the impact the manipulation had on the variable for that person. Hence, I fully agree with Fleeson (2004) in stating that within-person variability is an opportunity for personality research, and not a threat. And we should acknowledge that it makes sense to distinguish between the concept of state and the concept of trait, but at the same time accept that they are fundamentally intertwined, just like the concepts of weather and climate.

7.3 Outlook and Conclusion

Forty years of SVO research have resulted in an admirable wealth of findings and many interesting insights about the nature of social preferences have accumulated to date. Yet the study of SVO is far from finished, and new tools provide new opportunities for innovative investigations. Clearly, experimental designs that focus on the malleability of SVO by using it both as independent and dependent variable hold a great potential for examining the dynamics of social preferences. Concretely, it is now possible to quantify the impact situational factors have on the weight people attach to the welfare of others. Research questions that can be addressed in this respect are plentiful. For instance, it can be tested to what extent SVO changes in response to changes in status, group membership, or priming. Also, it can be assessed what kind of information about an interaction partner can evoke changes in a decision maker's SVO. For example, it could be studied to what extent SVO changes in response to revealed information about the interaction partner's political attitudes, religious beliefs, moral values, or ethnical background – just to mention a few. The idea of using high-resolution measures to assess baseline preferences and deviations from it in manipulated contexts may also be applied in a broader context of preference dynamics research. Clearly, SVO is just one particular type of preferences this methodology can be applied to. However, in the case of SVO, the detection of variables a decision maker's concerns for others is dependent on may reveal powerful pathways for enhancing human cooperation in situations of interdependence.

Another opportunity for future research arises from the fact that findings about whether SVO is nurture rather than nature are mixed. While evidence from early research suggests that SVO is mainly formed by socialization (see Au & Kwong, 2004), findings from more recent studies relativize this conclusion. For example, results from molecular genetic studies indicate a relationship between social preferences and vasopressin (Avinun et al., 2011; Knafo et al., 2008), dopamine (Bachner-Melman et al., 2005; Reuter et al., 2011), and oxytocin (Israel et al., 2009) receptor gene polymorphisms (for reviews, see Donaldson & Young, 2008; Israel et al., 2008). Studies on population genetics –i.e. twin-studies– also found that genetic factors explain a significant proportion of variance in social preferences and cooperation behavior (Cesarini et al., 2008, 2009; Knafo & Plomin, 2006a,b; Wallace et al., 2007). Clearly, future research will have to show the degree to which baseline SVO can be formed by socialization, since this may have important implications for private and scholarly education of children in terms of fostering cooperation. Furthermore, another approach for addressing the nature vs. nurture question is to conduct a large-scale study on cross-cultural differences in SVO with the goal to create a world map of social preferences. A project that aims at doing exactly this was recently initialized. This undertaking may also reveal the extent to which different political and societal systems foster or hinder the development and expression of prosociality.

Yet another issue that future research on SVO can address in more detail concerns the neuropsychology of social preferences. For instance, it is unclear to date whether the expression of prosociality requires self-control or is expressed automatically. Several studies have found evidence in favor of the former hypothesis, thereby stating that SVO is reflexive. Concretely, it has been shown that disruption of the right dorsolateral prefrontal cortex (rDLPFC) results in higher acceptance rates in ultimatum games (Knoch et al., 2006) and lower ability to resist the temptation of immediate payoffs in a trust game (Knoch et al., 2009). Consistent with these findings, the assumption that the expression of prosociality does require self-control is also supported by experiments in which ego-depletion tasks were employed (Balliet & Joireman, 2010; Seeley & Gardner, 2003). However, other experiments addressing the same research question indicated evidence that prosociality is expressed automatically (Haruno & Frith, 2010; Rand et al., 2012; Roch et al., 2000) or yielded inconclusive results (Cornelissen et al., 2011; Kinnunen & Windmann, 2013). Clearly, more research on this issue is necessary to reconcile divergent streams of evidence and embedding them in a coherent theoretical framework.

The preceding listing of opportunities for future SVO research is certainly not exhaustive, and new research questions will surely emerge in the process of answering the present ones. However, due to the availability of new tools and improved technology, the breadth of questions that can possibly be addressed has never been greater than today. For instance, the combination of a continuous measure of SVO with brain imaging or non-invasive brain stimulation techniques offers powerful and unprecedent possibilities for understanding the neurobiology of human prosociality. Hence, the time has never been better to explore the genetic and neurological bases of SVO, to examine aspects of its ontogenetic and phylogenetic development, and to investigate its psychological, social, and economic correlates. The new tools now available to us can be considered amplifiers of the light that can be shed on the nature of our concerns for others, and thus the future shines very bright for continued research on one of the most important interdisciplinary constructs in psychology, sociology, and economics.

A Appendix

The items from the SVO Slider Measure

Table A.1: SVO item endpoints and subsequent slopes that define each of the SVO Slider Measure items.

	Endpoint 1		Endpoint 2		Descriptive information	
Item	Self	Other	Self	Other	Slope	Equation
1	85	85	85	15	Undefined	$y \in [15, 85],\ x = 85$
2	85	15	100	50	2.33	$y = \frac{7}{3} \cdot x - \frac{550}{3}$
3	50	100	85	85	-0.43	$y = -\frac{3}{7} \cdot x + \frac{850}{7}$
4	50	100	85	15	-2.43	$y = -\frac{17}{7} \cdot x + \frac{1550}{7}$
5	100	50	50	100	-1.00	$y = -1 \cdot x + 150$
6	100	50	85	85	-2.33	$y = -\frac{7}{3} \cdot x + \frac{850}{3}$
7	100	50	70	100	-1.67	$y = -\frac{5}{3} \cdot x + \frac{650}{3}$
8	90	100	100	90	-1.00	$y = -1 \cdot x + 190$
9	100	70	50	100	-0.60	$y = -\frac{3}{5} \cdot x + 130$
10	100	70	90	100	-3.00	$y = -3 \cdot x + 370$
11	70	100	100	70	-1.00	$y = -1 \cdot x + 170$
12	50	100	100	90	-0.20	$y = -\frac{1}{5} \cdot x + 110$
13	50	100	100	50	-1.00	$y = -1 \cdot x + 150$
14	100	90	70	100	-0.33	$y = -\frac{1}{3} \cdot x + \frac{370}{3}$
15	90	100	100	50	-5.00	$y = -5 \cdot x + 550$

Note: The SVO Slider Measure has embedded in it several items which are also dictator games. Item number 5 from the primary set has this structure. From the secondary set, items 8, 11, 13 also have a slope of -1, giving them the same tradeoff rate between the payoff for self and other but over different ranges. Results from these items can be analyzed separately.

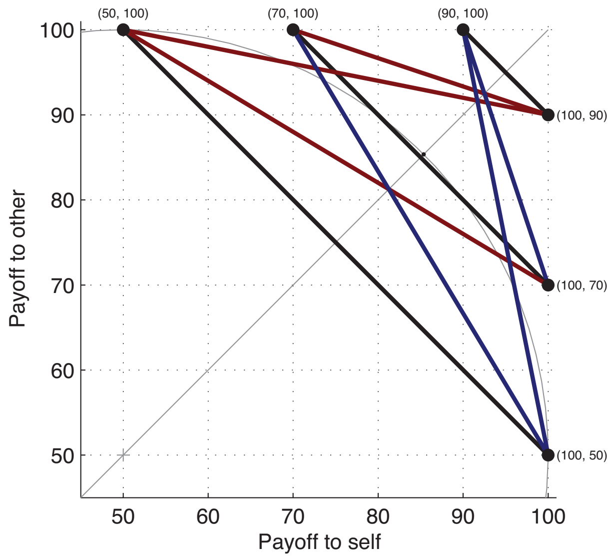


Figure A.1: This figure shows the location of the nine secondary items of the Slider Measure in the self/other allocation plane.

SVO angle calculation

1. Calculate the mean of the payoffs a subject allocated to herself across the six primary items ($\bar{A}_s$).
2. Calculate the mean of the payoffs a subject allocated to the other person across the six primary items ($\bar{A}_o$).
3. Subtract 50 from both means: $\bar{A}_s - 50$ and $\bar{A}_o - 50$.
4. In order to compute the SVO angle, calculate the inverse tangent of the ratio of the mean of the payoffs allocated to the other minus 50 and the mean of the payoffs allocated to the self minus 50:

$$\mathrm{SVO}^\circ = \arctan\left(\frac{(\bar{A}_o - 50)}{(\bar{A}_s - 50)}\right)$$

5. It is not recommended, but if for some reason categorical results are preferred to ratio level results, individual subjects' scores can be diminished to the categorical level following this scheme:
   - Altruism: SVO° > 57.15°
   - Prosociality: 22.45° < SVO° < 57.15°
   - Individualism: -12.04° < SVO° < 22.45°
   - Competitiveness: SVO° < -12.04°

Explication of boundary determination

The boundaries between categories were derived as follows:

If a subject would choose the option which maximizes the other one's payoff in each of the six primary items, the resulting angle would be 61.39°, indicating perfect altruism (see Table A.2). Likewise, if a person would choose the option which maximizes the difference between the own and the other one's payoff in each of the six primary items, the resulting angle would be -16.26°, indicating perfect competitiveness (see Table A.5). For prosocial subjects, there are two ways in which they could answer the six primary items perfectly consistent (see Table A.3). First, if a subject would choose the option which minimizes the difference between payoffs in each of the six items, the resulting angle would be 37.48°. Second, if a subject would choose the option which maximizes joint gain in each of the six items, the resulting angle would be between 37.09° and 52.91°. The reason for this range is that this DM would be wholly indifferent across the entire SVO Slider item that has a slope of -1 (i.e. the item with endpoints 100, 50 and 50, 100) as it has a constant sum. For the domain of individualism, if a subject would consistently choose the option which maximizes the own payoff in each of the six items, this would yield and angle between -7.82° and 7.82° (see Table A.4). The reason for this range is that this particular DM would be wholly indifferent across the entire SVO Slider item that has an undefined slope (endpoints 85, 85 and 85, 15).

Table A.2: Derivation of the SVO angle that would result if a person would consistently choose the altruistic options

	Endpoint 1		Endpoint 2		Altruistic Choice	
Item	Self	Other	Self	Other	Self	Other
1	85	85	85	15	85	85
2	85	15	100	50	100	50
3	50	100	85	85	50	100
4	50	100	85	15	50	100
5	100	50	50	100	50	100
6	100	50	85	85	85	85
Resulting means:					70	86.7
Resulting means - 50:					20	36.7
Resulting angle:					61.39°	

Table A.3: Derivation of the SVO angle that would result if a person would consistently choose the prosocial options

	Endpoint 1		Endpoint 2		Prosocial Choice	
Item	Self	Other	Self	Other	Self	Other
1	85	85	85	15	85	85
2	85	15	100	50	100	50
3	50	100	85	85	85	85
4	50	100	85	15	50	100
5	100	50	50	100	100 ↔ 50	50 ↔ 100
6	100	50	85	85	85	85
Resulting means:					84.2 ↔ 75.8	75.8 ↔ 84.2
Resulting means - 50:					34.2 ↔ 25.8	25.8 ↔ 34.2
Resulting angle:					37.09° ↔ 52.91°	

Table A.4: Derivation of the SVO angle that would result if a person would consistently choose the individualistic options

	Endpoint 1		Endpoint 2		Individualistic Choice	
Item	Self	Other	Self	Other	Self	Other
1	85	85	85	15	85	85 ↔ 15
2	85	15	100	50	100	50
3	50	100	85	85	85	85
4	50	100	85	15	85	15
5	100	50	50	100	100	50
6	100	50	85	85	100	50
Resulting means:					92.5	55.8 ↔ 44.2
Resulting means - 50:					42.5	5.8 ↔ -5.8
Resulting angle:					-7.82° ↔ 7.82°	

Table A.5: Derivation of the SVO angle that would result if a person would consistently choose the competitive options

	Endpoint 1		Endpoint 2		Competitive Choice	
Item	Self	Other	Self	Other	Self	Other
1	85	85	85	15	85	15
2	85	15	100	50	85	15
3	50	100	85	85	85	85
4	50	100	85	15	85	15
5	100	50	50	100	100	50
6	100	50	85	85	100	50
Resulting means:					90	38.3
Resulting means - 50:					40	-11.7
Resulting angle:					-16.26°	

The boundaries according to which subjects can be categorized were derived by bisecting the ranges between the angles that are produced when a subject with one of the four classical motivational orientations answers the Slider Measure perfectly consistent. When there is a range of angles which can be produced by perfectly consistent choice behavior (as is the case for individualistic and prosocial subjects), the maximum / minimum values are used for computing the boundaries. Concretely, the boundaries were calculated as follows:

- Boundary between altruism and prosociality:

$$\frac{61.39° + 52.91°}{2} = 57.15°$$

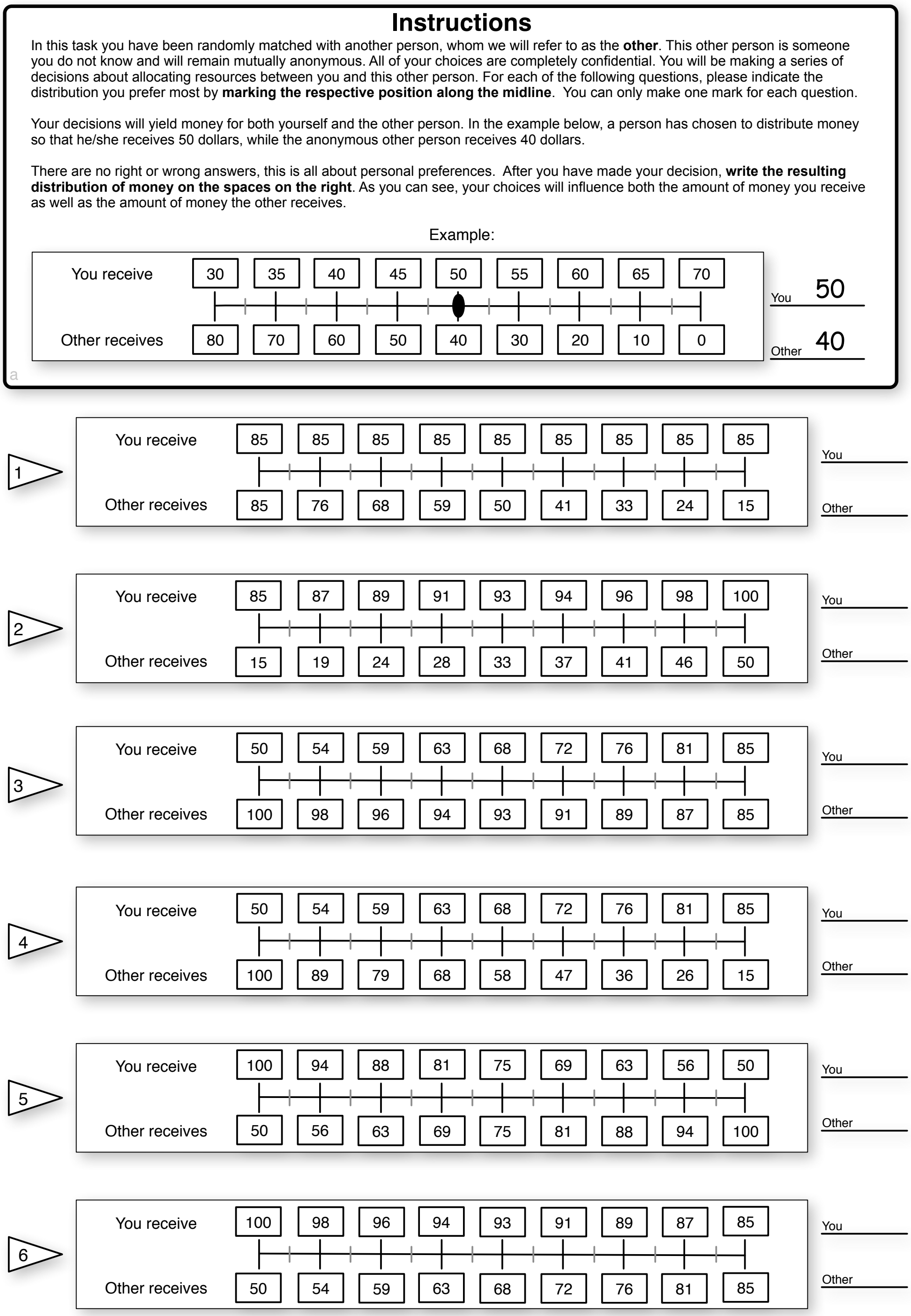
- Boundary between prosociality and individualism:

$$\frac{37.09° + 7.82°}{2} = 22.45°$$

- Boundary between individualism and competitiveness:

$$\frac{-7.82° + -16.26°}{2} = -12.04°$$

B Appendix



Instructions

In this task you have been randomly matched with another person, whom we will refer to as the **other**. This other person is someone you do not know and will remain mutually anonymous. All of your choices are completely confidential. You will be making a series of decisions about allocating resources between you and this other person. For each of the following questions, please indicate the distribution you prefer most by **marking the respective position along the midline**. You can only make one mark for each question.

Your decisions will yield money for both yourself and the other person. In the example below, a person has chosen to distribute money so that he/she receives 50 dollars, while the anonymous other person receives 40 dollars.

There are no right or wrong answers, this is all about personal preferences. After you have made your decision, **write the resulting distribution of money on the spaces on the right**. As you can see, your choices will influence both the amount of money you receive as well as the amount of money the other receives.

Example:

You receive	30	35	40	45	50	55	60	65	70
Other receives	80	70	60	50	40	30	20	10	0

You 50
Other 40

a

1

You receive	85	85	85	85	85	85	85	85	85
Other receives	85	76	68	59	50	41	33	24	15

You ______
Other ______

2

You receive	85	87	89	91	93	94	96	98	100
Other receives	15	19	24	28	33	37	41	46	50

You ______
Other ______

3

You receive	50	54	59	63	68	72	76	81	85
Other receives	100	98	96	94	93	91	89	87	85

You ______
Other ______

4

You receive	50	54	59	63	68	72	76	81	85
Other receives	100	89	79	68	58	47	36	26	15

You ______
Other ______

5

You receive	100	94	88	81	75	69	63	56	50
Other receives	50	56	63	69	75	81	88	94	100

You ______
Other ______

6

You receive	100	98	96	94	93	91	89	87	85
Other receives	50	54	59	63	68	72	76	81	85

You ______
Other ______

Figure B.1: Instructions and the Slider Measure's six primary items

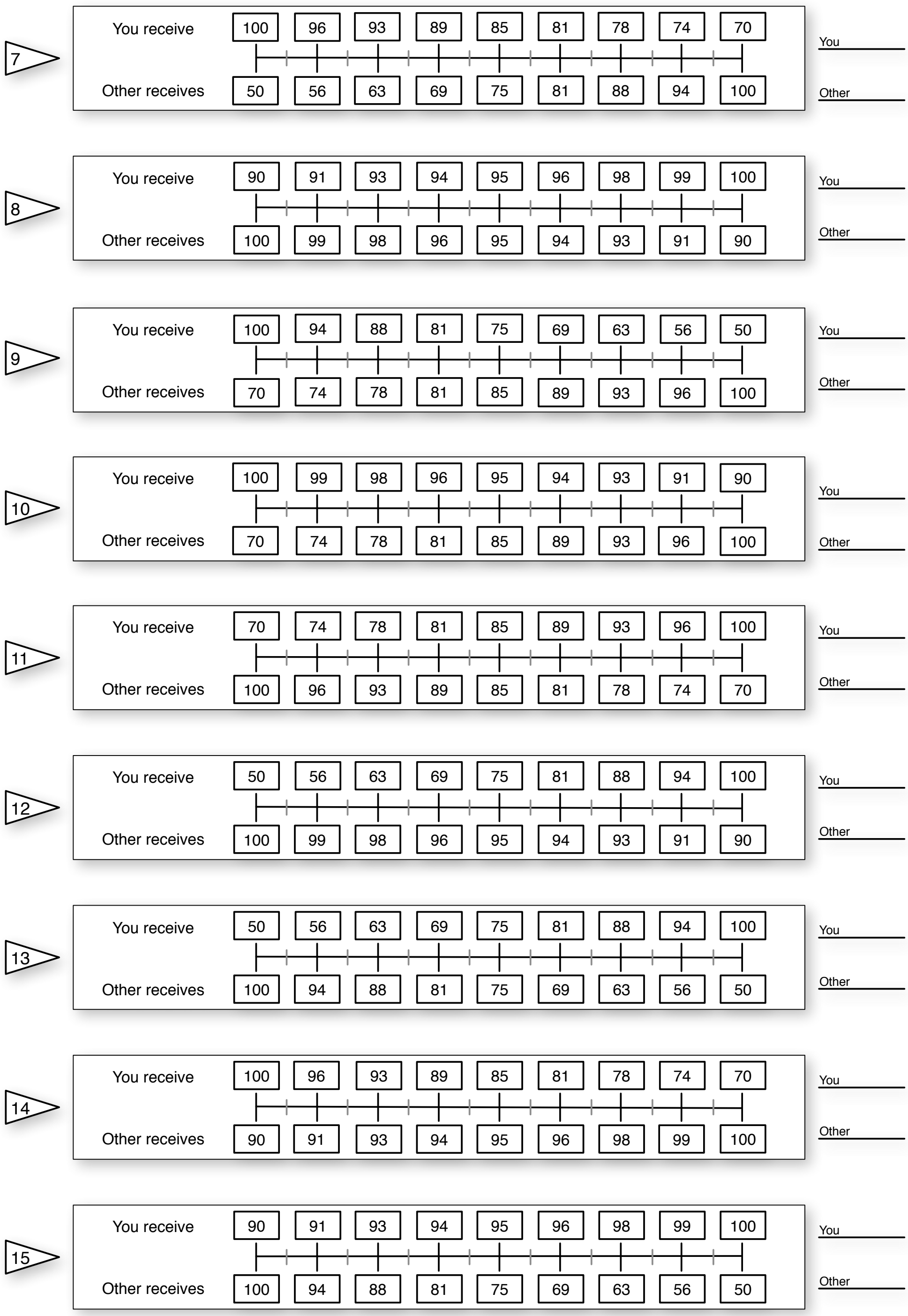


Figure B.2: The Slider Measure’s nine secondary items

C Appendix

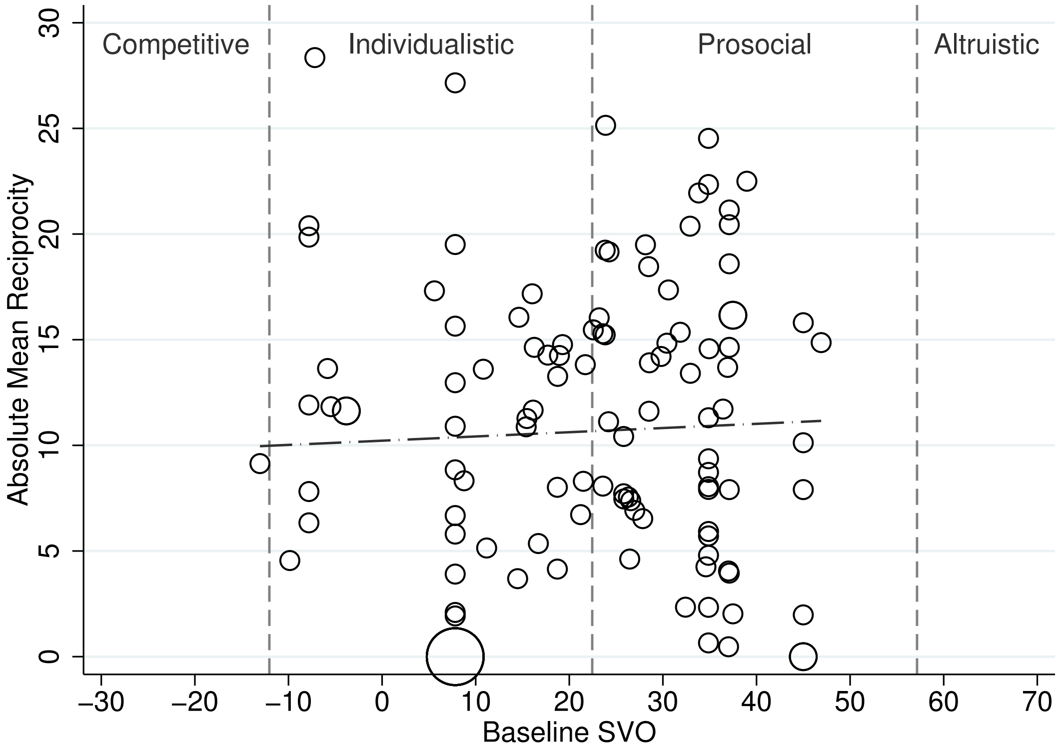


Figure C.1: Relation between SVO and aggregate absolute reciprocity (*n = 117*)

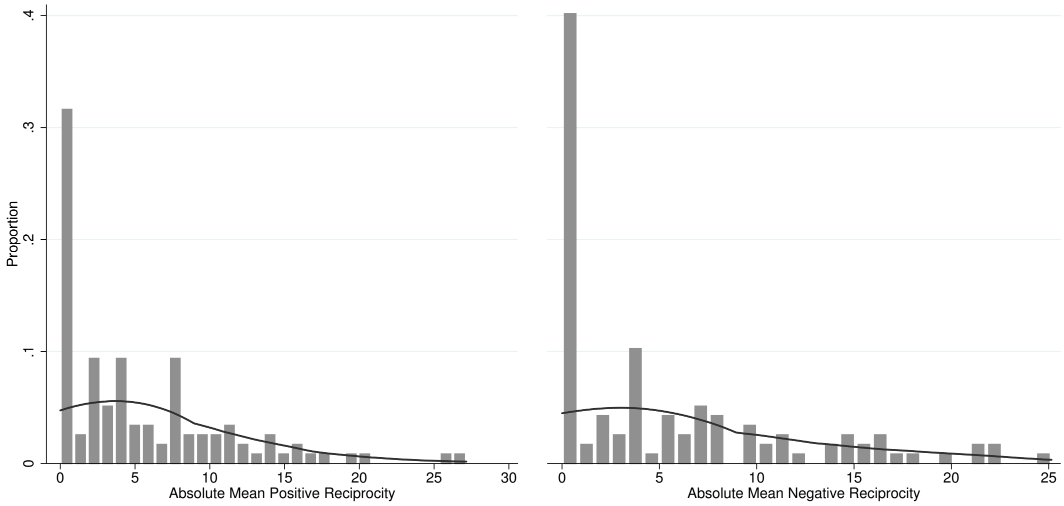


Figure C.2: Aggregate positive and negative reciprocity (n = 117)

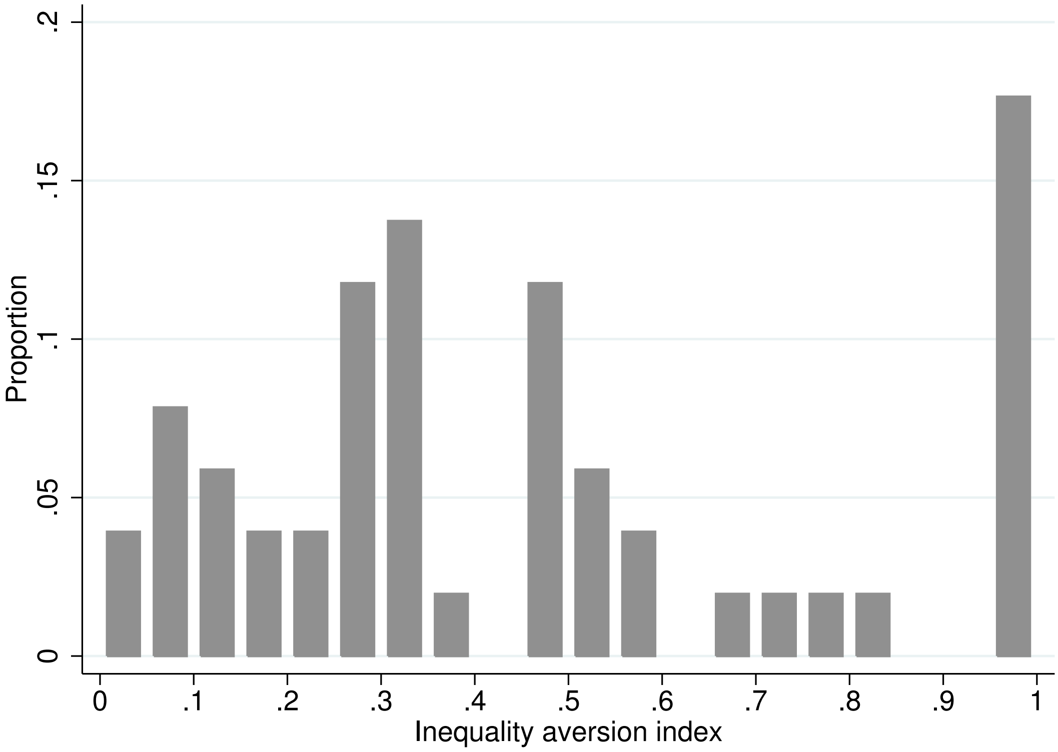


Figure C.3: IA index of prosocial subjects from baseline condition (*n* = *51*)

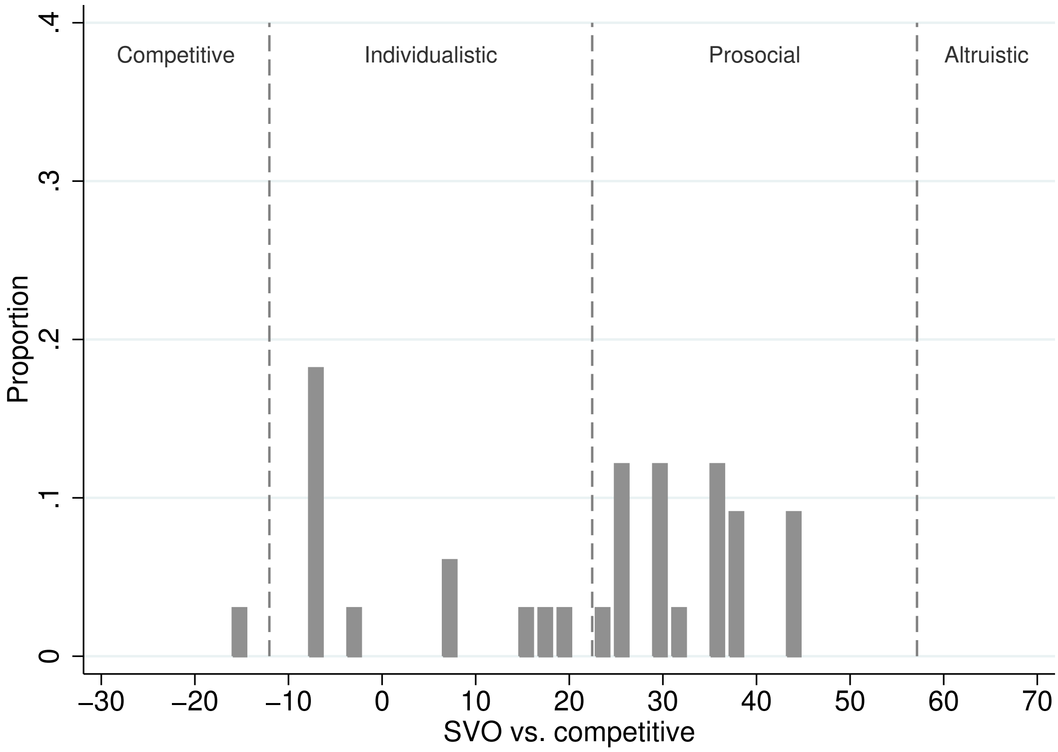


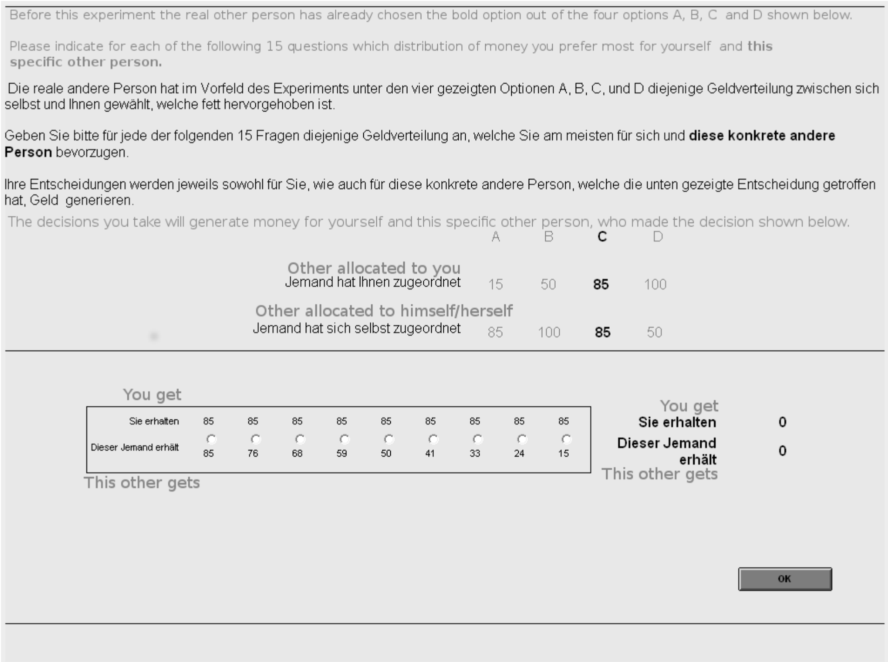
Figure C.4: SVO angles of inequality averse subjects when responding to competitive interaction partner (*n = 33*)

Table C.1: Results of the pen and paper task

	Field of Study					
SVO Type	Other Field		Social Work		Total	
	#	%	#	%	#	%
Competitive	5	8.1	1	1.2	6	4.1
Individualistic	16	25.8	9	10.5	25	16.9
Prosocial	39	62.9	70	81.4	109	73.6
Altrustic	2	3.2	6	7.0	8	5.4
Total	62	100.0	86	100.0	148	100.0

Note: A comparison of the choices made by students of social work with those made by students of other fields shows that the distributions are significantly different from each other (Fishers exact test, $p < .01$), i.e. the proportion of prosocials and altruists is larger among students of social work as compared to students of other fields.

Figure C.5: Decision Screen for SVO Slider Measure with specific other



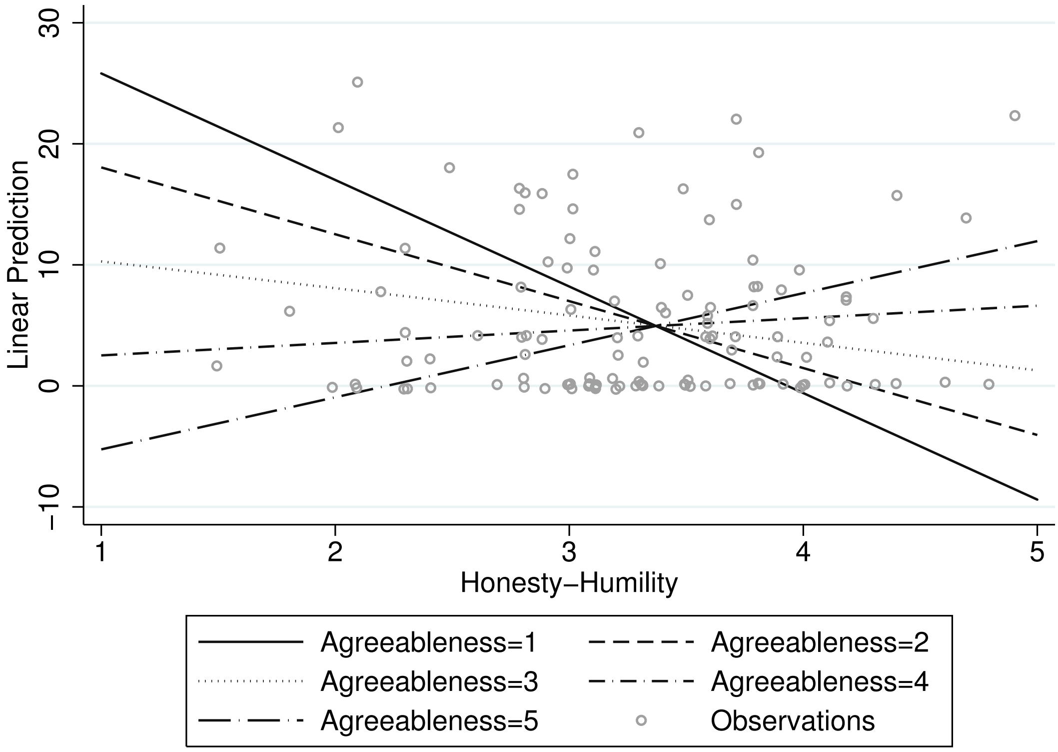


Figure C.6: Negative reciprocity predictions for different values of Agreeableness and Honesty-Humility

Table C.2: OLS regressions on positive and negative reciprocity

	(1) Overall Reciprocity	(2) Positive Reciprocity	(3) Negative Reciprocity
Baseline SVO	0.014	-0.214**	0.227**
Honesty-Hum.	-5.250	6.829*	-12.079*
Emotionality	0.232	-0.363	0.595
Extraversion	-0.446	-1.310	0.865
Agreeablen.	-6.319	4.723	-11.042*
Conscienti.	-1.594	-1.117	-0.477
Openness	1.147	0.601	0.546
Honesty-Hum. *Agreeablen.	1.673	-1.603	3.276*
Constant	32.413	-3.119	35.533*
Observations	117	117	117
R-squared	0.050	0.319	0.334

Note: Significance levels are ** $p < 0.01$, * $p < 0.05$. Unstandardized coefficients are reported. Robust standard errors were used. The interaction term was computed without centering the corresponding independent variables. If centering is employed, the main effect of Agreeableness on Negative Reciprocity is not significant. A new experiment with a larger sample size would be required to make a conclusive judgment about the robustness of the effects found in this exploratory analysis.

Bibliography

Ackermann, K. A., Fleiss, J., & Murphy, R. O. (2014). Reciprocity as an individual difference. *Journal of Conflic Resolution*, *In press*.

Ackermann, K. A., & Murphy, R. O. (2012). Tutorial on how to evaluate the SVO Slider Measure's secondary items. *Available online: http://vlab.ethz.ch/svo/SVO_Slider/SVO_Slider_scoring.html*.

Ackermann, K. A., & Murphy, R. O. (2013). Modeling social value orientation. Manuscript in preparation.

Ajzen, I., Brown, T., & Carvajal, F. (2004). Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation. *Personality and Social Psychology Bulletin*, *30*(9), 1108–1121.

Aksoy, O., & Weesie, J. (2012). Beliefs about the social orientations of others: A parametric test of the triangle, false consensus, and cone hypotheses. *Journal of Experimental Social Psychology*, *48*(1), 45–54.

Allen, B. P., & Potkay, C. R. (1981). On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology*, *41*(5), 916–928.

Allison, S. T., & Goethals, G. R. (2010). *Heroes: What they do and why we need them*. New York: Oxford University Press.

Anderson, N. H. (1968). A simple model for information integration. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.) *Theories of Cognitive Consistency: A Sourcebook*. Chicago: Rand McNally.

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, *77*(3), 153–170.

Andreoni, J. (1988). Why free ride? *Journal of Public Economics*, *37*(3), 291–304.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, *100*(401), 464–477.

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

Andreoni, J., Harbaugh, W. T., & Vesterlund, L. (2008). Altruism in experiments. In Durlauf, S. and Blume, L. (Ed.) *New Palgrave Dictionary in Economics*. Macmillan, 2 ed.

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, *103*(418), 570–585.

Andreoni, J., & Miller, J. H. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Ashley, R., Ball, S., & Eckel, C. (2010). Motives for giving: A reanalysis of two classic public goods experiments. *Southern Econonomic Journal*, *77*(1), 15–26.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*, 150–166.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence, & J. Spence (Eds.) *The psychology of learning and motivation: Advances in research and theory*, vol. 2, (pp. 89–195). New York: Academic Press.

Au, W. T., & Kwong, J. Y. Y. (2004). Measurements and effects of social-value orientation in social dilemmas: A review. In R. Suleiman, D. V. Budescu, I. Fischer, & D. M. Messick (Eds.) *Contemporary Psychological Research on Social Dilemmas*, (pp. 71–98). New York: Cambridge University Press.

Avinun, R., Israel, S., Shalev, I., Gritsenko, I., Bornstein, G., Ebstein, R. P., & Knafo, A. (2011). Avpr1a variant associated with preschoolers' lower altruistic behavior. *PLoS ONE*, *6*(9), e25274.

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Bachner-Melman, R., Gritsenko, I., Nemanov, L., Zohar, A. H., Dina, C., & Ebstein, R. P. (2005). Dopaminergic polymorphisms associated with self-report measures of human altruism: A fresh phenotype for the dopamine d4 receptor. *Molecular Psychiatry*, *10*, 333–335.

Balliet, D. (2007). *A matter of time: Does the impact of Social Value Orientation and self-efficacy on contributions to Public Goods depend on the temporal framing of the dilemma?*. Doctoral dissertation, Washington State University.

Balliet, D., & Joireman, J. (2010). Ego depletion reduces proselfs' concern with the well-being of others. *Group Processes & Intergroup Relations*, *13*(2), 227–239.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533–547.

Bardsley, N., & Moffatt, P. G. (2007). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision*, *62*, 161–193.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements. *Perspectives on Psychological Science*, *2*(4), 396–403.

Becker, S. W., & Eagly, A. H. (2004). The heroism of women and men. *American Psychologist*, *59*(3), 163–178.

Bennett, R. (2006). Factors underlying the inclination to donate to particular types of charity. *International Journal of Nonprofit and Voluntary Sector Marketing*, *8*(1), 12–29.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338.

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, *47*(3), 453–480.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, *90*(1), 166–193.

Boone, C., Declerck, C., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: The moderating role of incentives and trust. *Journal of Conflict Resolution*, *54*(5), 799–824.

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

Brandts, J., & Schram, A. (2001). Cooperation and noise in public goods experiments: Applying the contribution function approach. *Journal of Public Economics*, *79*, 399–427.

Breitmoser, Y. (2013). Estimation of social preferences in generalized dictator games. *Economics Letters*, *121*(2), 192–197.

Bryan, J. H., & Test, M. A. (1967). Models and helping: Naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology*, *6*(4), 400–407.

Burkart, J. M., Fehr, E., Efferson, C., & van Schaik, C. P. (2007). Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proceedings of the National Academy of Sciences*, *104*(50), 19762–19766.

Burlando, R., & Hey, J. D. (1997). Do anglo-saxons free ride more? *Journal of Public Economics*, *64*, 41–60.

Burlando, R. M., & Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics*, *8*(1), 35–54.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., & Fehr, E. (2006). When does "economic man" dominate social behavior? *Science*, *311*(5757), 47–52.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, *9*(2), 209–219.

Cameron, L. D., Brown, P. M., & Chapman, J. G. (1998). Social value orientations and decisions to take proenvironmental action. *Journal of Applied Social Psychology*, *28*(8), 675–697.

Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, *60*, 31–51.

Catell, R. B. (1966). Anxiety and motivation: Theory and crucial experiments. *Anxiety and Behavior*, *1*, 23–62.

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences*, *105*(10), 3721–3726.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics*, *124*(2), 809–842.

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541–557.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization*, *45*, 301–328.

Charness, G., & Rabin, M. (2000). Social preferences: Some simple tests and a new model. Unpublished manuscript.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*(3), 817–869.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, *14*, 47–83.

Choshen-Hillel, S., & Yaniv, I. (2012). Social preferences shaped by conflicting motives: When enhancing social welfare creates unfavorable comparisons for the self. *Judgment and Decision Making*, *7*(5), 618–627.

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.

Clark, K., & Sefton, M. (2001). The sequential prisoner's dilemma: Evidence on reciprocation. *The Economic Journal*, *111*, 51–68.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*(403), 596–610.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

Conrath, D. W., & Deci, E. L. (1969). The determination and scaling of a bivariate utility function. *Behavioral Science*, *14*(4), 316–327.

Cooper, R., De Jong, D. V., Forsythe, R., & Ross, T. W. (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior*, *12*, 187–218.

Cornelissen, G., Dewitte, S., & Warlop, L. (2007). Social value orientation as a moral intuition: Decision-making in the dictator game. Unpublished paper, available at SSRN: http://ssrn.com/abstract=978469.

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin*, *37*(8), 1080–1090.

Costa, P. T., & McCrae, R. R. (2009). The five-factor model and the neo inventories. In J. N. Butcher (Ed.) *Oxford handbook of personality assessment*, (pp. 299–322). Oxford University Press.

Cox, J. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*, 260–281.

Cox, J. C., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, *59*, 17–45.

Cox, J. C., & Sadiraj, V. (2007). On modeling voluntary contributions to public goods. *Public Finance Review*, *35*(2), 311–332.

Crandall, J. E. (1975). A scale for social interest. *Journal of Individual Psychology*, *31*(2), 187–195.

Crosetto, P., Weisel, O., & Winter, F. (2012a). A flexible z-Tree implementation of the social value orientation slider measure (Murphy et al. 2011) – manual –.

Crosetto, P., Weisel, O., & Winter, F. (2012b). A flexible z-Tree implementation of the Social Value Orientation slider measure (Murphy et al. 2011) - Manual. Jena Economic Research Paper, 62.

Croson, R. T. (2007). Theories of commitment, altruism, and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, *45*(2), 199–216.

Croson, R. T. A. (1996). Partners and strangers revisited. *Economic Letters*, *53*, 25–32.

Dalbert, C., Montada, L., & Schmitt, M. (1987). Belief in a just world: Validity correlates of two scales. *Psychologische Beiträge*, *29*, 596–615.

Daruvala, D. (2010). Would the right social preference model please stand up. *Journal of Economic Behavior & Organization*, *73*(2), 199–208.

Dawes, R., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, *35*(1), 1–11.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*(1), 169–193.

De Cremer, D., & Van Lange, P. A. M. (2001). Why prosocials exhibit greater cooperation than proselfs: The roles of social responsibility and reciprocity. *European Journal of Personality*, *15*, 5–18.

De Dreu, C. K. W., & Boles, T. L. (1998). Share and share alike or winner take all?: The influence of social value orientation upon choice and recall of negotiation heuristics. *Organizational Behavior and Human Decision Processes*, *76*(3), 253–276.

De Dreu, C. K. W., & Van Lange, P. A. M. (1995). The impact of social value orientations on negotiator cognition and behavior. *Personality and Social Psychology Bulletin*, *21*(11), 1178–1188.

de Kwaadsteniet, E. W., van Dijk, E., Wit, A., & de Cremer, D. (2006). Social dilemmas as strong versus weak situations: Social value orientations and tacit coordination under resource size uncertainty. *Journal of Experimental Social Psychology*, *42*(4), 509–516.

Declerck, C. H., & Bogaert, S. (2008). Social value orientation: Related to empathy and the ability to read the mind in the eyes. *The Journal of Social Psychology*, *148*(6), 711–726.

Dehue, F. M. J., McClintock, C. G., & Liebrand, W. B. G. (1993). Social value related response latencies: Unobtrusive evidence for individual differences in information processing. *European Journal of Social Psychology*, *23*, 273–293.

Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations*, *2*, 129–151.

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, *13*(2), 123–139.

Diekmann, A. (2004). The power of reciprocity: Fairness, reciprocity, and stakes in variants of the dictator game. *Journal of Conflict Resolution*, *48*(4), 487–505.

Donaldson, Z. R., & Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science*, *322*, 900–904.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Ebbinghaus, H. (1913). *On memory: A contribution to experimental psychology*. New York: Teachers College.

Edgeworth, F. Y. (1881). Mathematical psychics. (L.S.E. Reprint 1934).

Eek, D., & Gärling, T. (2006). Prosocials prefer equal outcomes to maximizing joint outcomes. *British Journal of Social Psychology*, *45*, 321–337.

Egloff, B., Richter, D., & Schmukle, S. C. (2013). Need for conclusive evidence that positive and negative reciprocity are unrelated. *Proceedings of the National Academy of Sciences*, *110*(9), E786.

Eisenberger, R., Kuhlman, D. M., & Cotterell, N. (1992). Effects of social values, effort training, and goal structure on task persistence. *Journal of Research in Personality*, *26*(3), 258–272.

Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who takes most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin*, *30*(6), 787–799.

Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Anxiety Disorders*, *15*, 231–245.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, *94*(4), 857–869.

Erev, I., & Rapoport, A. (1998). Coordination, “magic”, and reinforcement learning in a market entry game. *Games and Economic Behavior*, *23*, 146–175.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness - Intentions matter. *Games and Economic Behavior*, *62*, 287–303.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.

Fehr, E., & Gächter, S. (2000a). Cooperation and punishment in public goods experiments. *The American Economic Review*, *90*(4), 980–994.

Fehr, E., & Gächter, S. (2000b). Cooperation and punishment in public goods experiments. *The American Economic Review*, *90*(4), 980–994.

Fehr, E., & Gächter, S. (2000c). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*, 159–181.

Fehr, E., Klein, A., & M., S. K. (2007). Fairness and contract design. *Econometrica*, *75*(1), 121–154.

Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American Economic Review*, *96*(5), 1912–1917.

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. (2004). Fairness and incentives in a multi–task principal–agent model. *Scandinavian Journal of Economics*, *106*(3), 453–474.

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*, 272–284.

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

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review*, *100*(1), 541–556.

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

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*, 1858–1877.

Fitzsimons, G. (2008). Editorial: Death to dichotomizing. *Journal of Consumer Research*, *35*(1), 5–8.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distribution of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*(2), 83–87.

Franco, Z. E., & Zimbardo, P. G. (2011). Heroism: A conceptual analysis and differentiation between heroic action and altruism. *Review of General Psychology*, *15*(2), 99–113.

Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing ”conditional cooperation” in a field experiment. *The American Economic Review*, *94*(5), 1717–1722.

Fridhandler, B. M. (1986). Conceptual note on state, trait, and the state-trait distinction. *Journal of Personality and Social Psychology*, *50*(1), 169–174.

Gächter, S. (2007). Conditional cooperation. Behavioral regularities from the lab and the field and their policy implications. In B. Frey, & A. Stutzer (Eds.) *Economics and psychology. A promising new cross-disciplinary field.*. Cambridge: MIT Press.

Gächter, S., & Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1518), 791–806.

Gangestad, S. W., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review*, *92*, 317–349.

Ganzach, Y. (2000). Judging risk and return of financial assets. *Organizational Behavior and Human Decision Processes*, *83*(2), 353–370.

Ganzach, Y., Ellis, S., Pazy, A., & Ricci-Siag, T. (2008). On the perception and operationalization of risk perception. *Judgment and Decision Making*, *3*(4), 317–324.

Geanakoplos, J., Pearce, D., & Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, *1*(1), 60–79.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*(2), 254–267.

Gigerenzer, G., & Sturm, T. (2007). Tools = theories = data? on some circular dynamics in cognitive science. In M. Ash, & T. Sturm (Eds.) *Psychology's territories: Historical and contemporary perspectives form different disciplines*, chap. 14, (pp. 305–342). London: Lawrence Erlbaum Associates.

Gillespie, T. W., & Hillyer, C. D. (2002). Blood donors and factors impacting the blood donation decision. *Transfusion Medicine Reviews*, *16*(2), 115–130.

Gintis, H. (2003). Solving the puzzle of prosociality. *Rationality and Society*, *15*(2), 155–187.

Gintis, H. (2007). A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences*, *30*(1), 1–61.

Goeree, J. K., Holt, C. A., & Laury, S. K. (2002). Private costs and public benefits: Unraveling the effects of altruism and noisy behavior. *Journal of Public Economics*, *83*, 255–276.

Graf, C., Vetschera, R., & Zhang, Y. (2012). Parameters of social preference functions: measurement and external validity. *Theory and Decision*, (pp. 1–26).

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer, & V. Macho (Eds.) *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, (pp. 79–93). Göttingen: Ges. für Wiss. Datenverarbeitung.

Griesinger, D. W., & Livingston, J. W. (1973). Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, *18*(3), 173–188.

Grzelak, J. L. (1982). Preferences and cognitive processes in interdependence situations: A theoretical analysis of cooperation. In V. J. Derlega, & J. Grzelak (Eds.) *Cooperation and Helping Behavior*, (pp. 95–122). Academic Press, New York.

Grzelak, J. L., Iwinski, T. B., & Radzicki, J. J. (1977). "Motivational" components of utility. In H. Jungermann, & G. de Zeeuw (Eds.) *Decision Making and Change in Human Affairs*. Reidel, Dordrecht.

Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, *23*(1), 49–66.

Güth, W., Kliemt, H., & Ockenfels, A. (2003). Fairness versus efficiency. an experimental study of (mutual) gift giving. *Journal of Economic Behavior & Organization*, *50*, 465–475.

Hamilton, W. D. (1964). The genetical evolution of social behavior. I. *Journal of Theoretical Biology*, *7*(1), 1–16.

Harrison, M. (1998). The bioeconomics of altruism and rivalry. Doctoral Dissertation, Department of Economics, University of Texas at Austin.

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161.

Hedström, P. (2005). *Dissecting the social: On the principles of analytic sociology*. Cambridge: Cambridge University Press.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., & Tracer, D. (2005). “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(06), 795–855.

Hermann, B., & Thöni, C. (2009). Measuring conditional cooperation: A replication study in russia. *Experimental Economics*, *12*(1), 87–92.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*, 1280–1300.

Hilbig, B. E., Glöckner, A., & Zettler, I. (2014). Personality and pro-social behavior: Linking basic traits and social value orientations. *Journal of Personality and Social Psychology*, (p. in press).

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty-humility, social value orientations, and economic behavior. *Journal of Research in Personality*, *43*(3), 516–519.

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–humility and agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598 – 603.

Hilbig, B. E., Zettler, I., Moshagen, M., & Heydasch, T. (2012). Tracing the path from personality -via cooperativeness- to conservation. *European Journal of Personality*.

Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, Brace & World.

Huber, J., & Puto, C. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *The Journal of Consumer Research*, *10*(1), 31–44.

Iedema, J., & Poppe, M. (1994a). The effect of self-presentation on social value orientation. *The Journal of Social Psychology*, *134*(6), 771–782.

Iedema, J., & Poppe, M. (1994b). Effects of social value orientation on expecting and learning others' orientations. *European Journal of Social Psychology*, *24*, 565–579.

Iedema, J., & Poppe, M. (1995). Perceived consensus of one's social value orientation in different populations in public and private circumstances. *European Journal of Social Psychology*, *25*, 497–507.

Irwin, J., & McClelland, G. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, *40*, 366–371.

Isaac, R. M., McCue, K. F., & Plott, C. R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, *26*(1), 51–74.

Isaac, R. M., Walker, J., & Thomas, S. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, *43*(1), 113–149.

Isaac, R. M., & Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, *103*(1), 179–199.

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Reibold, M., Bachner-Melman, R., Granot, R., Bornstein, G., Knafo, A., Yirmiya, N., & Ebstein, R. (2008). Molecular genetic studies of the arginine vasopressin 1a receptor (avpr1a) and the oxytocin receptor (oxtr) in human behaviour: from autism to altruism with some notes in between. *Progress in Brain Research*, *170*, 435–449.

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Riebold, M., Laiba, E., Bachner-Melman, R., Maril, A., Bornstein, G., Knafo, A., & Ebstein, R. P. (2009). The oxytocin receptor (oxtr) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS ONE*, *4*(5), e5535.

Jakiela, P. (2012). Equity vs. efficiency vs. self-interest: On the use of dictator games to measure distributional preferences. *Experimental Economics*, *15*(3).

Jevons, W. S. (1875). *Money and the Mechanism of Exchange*. Appleton and Co.

Joireman, J. A. (1996). *Differing perceptions of the interdependence structure underlying commuting decisions: The role of social value orientation, trust, and travel related concerns.*. Doctoral dissertation, University of Delaware.

Joireman, J. A., Van Lange, P. A. M., & Van Vugt, M. (2004). Who cares about the environmental impact of cars?: Those with an eye toward the future. *Environment and Behavior*, *36*(2), 187–206.

Kachelmeier, S. T., & Shehata, M. (1997). Internal auditing and voluntary cooperation in firms: A cross-cultural experiment. *The Accounting Review*, *72*(3), 407–431.

Kahneman, D., & Tversky, A. (1979). Prospect Theory - An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology*, *30*(3), 368–380.

Kelley, H. H., & Stahelski, A. J. (1970). Social interaction basis of cooperators and competitors beliefs about others. *Journal of Personality and Social Psychology*, *16*(1), 66–91.

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal Relations: A Theory of Interdependence*. Wiley, New York.

Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics*, *102*(1), 23–39.

Kim, O., & Walker, M. (1984). The free rider problem: Experimental evidence. *Public Choice*, *43*(1), 3–24.

Kinnunen, S. P., & Windmann, S. (2013). Dual-processing altruism. *Frontiers in Psychology*, *4*, 1–8.

Knafo, A., Israel, S., Darvasi, A., Bachner-Melman, R., Uzefovsky, F., Cohen, L., Feldman, E., Lerer, E., Laiba, E., Raz, Y., Nemanov, L., Gritsenko, I., Dina, C., Agam, G., Dean, B., Bornstein, G., & Ebstein, R. P. (2008). Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor rs3 promoter region and correlation between rs3 length and hippocampal mrna. *Genes, Brain and Behavior*, *7*, 266–275.

Knafo, A., & Plomin, R. (2006a). Parental discipline and affection and children's prosocial behavior: Genetic and environmental link. *Journal of Personality and Social Psychology*, *90*(1), 147–164.

Knafo, A., & Plomin, R. (2006b). Prosocial behavior from early to middle childhood: Genetic and environmental influences on stability and change. *Developmental Psychology*, *42*(5), 771–786.

Knight, G. P. (1981). Behavioral and sociometric methods of identifying cooperators, competitors, and individualists: Support for the validity of the social orientation construct. *Developmental Psychology*, *17*(4), 430–433.

Knight, G. P. (1982). Cooperative-competitive social orientation: Interactions of birth order with sex and economic class. *Child Development*, *53*, 664–667.

Knight, G. P., & Dubro, A. F. (1984). Cooperative, competitive, and individualistic social values - an individualized regression and clustering approach. *Journal of Personality and Social Psychology*, *46*(1), 98–105.

Knight, G. P., & Kagan, S. (1977). Development of prosocial and competitive bahaviors in anglo-american and mexican-american children. *Child Development*, *48*(4), 1385–1394.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Dminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–832.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Sciences*, *106*(49), 20895–20899.

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economic Letters*, *101*(3), 175–178.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*, 183–214.

Kritikos, A., & Bolle, F. (2001). Distributional concerns: Equity- or efficiency oriented? *Economic Letters*, *73*(3), 333–338.

Kuhlman, D. M., Brown, C., & Teta, P. (1992). Judgements of cooperation and defection in social dilemmas: The moderating role of judges's social orientation. In W. B. Liebrand, D. M. Messick, & H. A. Wilke (Eds.) *Social dilemmas: Theoretical issues and research findings*, (pp. 111–132). Oxford, England: Pergamon Press.

Kuhlman, D. M., & Marshello, A. (1975a). Individual-differences in game motives of own, relative, and joint gain. *Journal of Research in Personality*, *9*(3), 240–251.

Kuhlman, D. M., & Marshello, A. F. J. (1975b). Individual-differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *Journal of Personality and Social Psychology*, *32*(5), 922–931.

Kuhlman, D. M., & Wimberley, D. C. (1976). Expectations of choice behavior held by cooperators, competitors, and individualists across four classes of experimental games. *Journal of Personality and Social Psychology*, *34*(1), 69–81.

Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *PNAS*, *102*(5), 1803–1807.

Laury, S. K., Walker, J. M., & Williams, A. W. (1995). Anonymity and the voluntary provision of public goods. *Journal of Economic Behavior and Organization*, *27*(3), 365–380.

Ledyard, J. (1995). Public goods: a survey of experimental research. In J. Kagel, & A. Roth (Eds.) *Handbook of Experimental Economics*. Princeton University Press, Princeton.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*, 593–622.

Lewin, K. (1936). *Principles of Topological Psychology*. New York: McGraw-Hill.

Liebrand, W. B. G. (1984). The effect of social motives, communication and group-size on behavior in an n-person multi-stage mixed-motive game. *European Journal of Social Psychology*, *14*(3), 239–264.

Liebrand, W. B. G., Jansen, R. W. T. L., Rijken, V. M., & Suhre, C. J. M. (1986). Might over morality: Social values and the perception of other players in experimental games. *Journal of Experimental Social Psychology*, *22*(3), 203–215.

Liebrand, W. B. G., & McClintock, C. G. (1988). The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European Journal of Personality*, *2*(3), 217–230.

Liebrand, W. B. G., & van Run, G. (1985). The effects of social motives on behavior in social dilemmas in two cultures. *Journal of Experimental Social Psychology*, *21*, 86–102.

Loewenstein, G. F., Bazerman, M. H., & Thompson, L. (1989). Social utility and decision-making in interpersonal contexts. *Journal of Personality and Social Psychology*, *57*(3), 426–441.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Luce, R. D., & Raiffa, H. (1957). *Games and Decisions. Introduction and Critical Survey*. New York: Wiley.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*(1), 1–27.

MacCallum, R., Zhang, S., Preacher, K., & Rucker, D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40.

MacCrimmon, K. R., & Messick, D. M. (1976). A framework for social motives. *Behavioral Science*, *21*(2), 86–100.

MacCrimmon, K. R., & Toda, M. (1969). The experimental determination of indifference curves. *The Review of Economic Studies*, *36*(4), 433–451.

Maki, J. E., & McClintock, C. G. (1983). The accuracy of social value prediction: Actor and observer influences. *Journal of Personality and Social Psychology*, *45*(4), 829–838.

Maki, J. E., Thorngate, W. B., & McClintock, C. G. (1979). Prediction and perception of social motives. *Journal of Personality and Social Psychology*, *37*(2), 203–220.

Marwell, G., & Ames, R. E. (1979). Experiments on the provision of public goods. Resources, interest, group size, and the free-rider problem. *American Journal of Sociology*, *84*(6), 1335–1360.

Maslow, A. H. (1966). *The Psychology of Science: A Reconnaissance*. New York: Harper and Row.

McClintock, C. G. (1972). Social motivation: A set of propositions. *Behavioral Science*, *17*(5), 438–454.

McClintock, C. G. (1988). Evolution, systems of interdependence, and social values. *Behavioral science*, *33*(1), 59–76.

McClintock, C. G., & Allison, S. (1989). Social value orientation and helping behavior. *Journal of Applied Social Psychology*, *19*, 353–362.

McClintock, C. G., & Liebrand, W. B. G. (1988). Role of interdependence structure, individual value orientation, and another's strategy in social decision making: A transformational analysis. *Journal of Personality and Social Psychology*, *55*(3), 396–409.

McClintock, C. G., & McNeel, S. P. (1966a). Cross-cultural comparison of interpersonal motives. *Sociometry*, *29*, 406–427.

McClintock, C. G., & McNeel, S. P. (1966b). Reward and score feedback as determinants of cooperative game behavior. *Journal of Personality and Social Psychology*, *4*, 606–613.

McClintock, C. G., & McNeel, S. P. (1966c). Societal membership, score status, and game behavior. a phenomenological analysis. *International Journal of Psychology*, *1*, 263–279.

McClintock, C. G., & McNeel, S. P. (1967). Prior dyadic experience and monetary reward as determinants of cooperative and competitive game behavior. *Journal of Personality and Social Psychology*, *5*(3), 282–294.

McClintock, C. G., Messick, D. M., Kuhlman, D. M., & Campos, F. T. (1973). Motivational bases of choice in three-choice decomposed games. *Journal of Experimental Social Psychology*, *9*(6), 572–590.

McClintock, C. G., & Van Avermaet, E. (1982). Social values and rules of fairness: A theoretical perspective. In V. J. Derlega, & J. Grzelak (Eds.) *Cooperation and Helping Behavior*, (pp. 43–71). Academic Press, New York.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1–25.

Messick, D. M., & Sentis, K. P. (1985). Estimating social and nonsocial utility functions from ordinal data. *European Journal of Social Psychology*, *15*(4), 389–399.

Messick, D. M., & Thorngate, W. (1967). Relative gain maximization in experimental games. *Journal of Experimental Social Psychology*, *3*, 85–101.

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, *49*(1), 229–258.

Moshagen, M., Hilbig, B. E., & Zettler, I. (forthcoming). Faktorenstruktur, psychometrische eigenschaften und messinvarianz der deutschsprachigen version des 60-item hexaco persönlichkeitsinventars. *Diagnostica*.

Murphy, R. O., & Ackermann, K. A. (2014a). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41.

Murphy, R. O., & Ackermann, K. A. (2014b). Explaining behavior in public goods games: How preferences and beliefs affect contribution levels. Manuscript in preparation.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781.

Murphy, R. O., Rapoport, A., & Parco, J. E. (2006). The breakdown of cooperation in iterative real-time trust dilemmas. *Experimental Economics*, *9*(2), 147–166.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Neugebauer, T., Perote, J., Schmidt, U., & Loos, M. (2009). Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology*, *30*(1), 52–60.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563.

Offerman, T., Sonnemans, J., & Schram, A. (1996). Value orientations, expectations, and voluntary contributions in public goods. *The Economic Journal*, *106*(437), 817–845.

Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, *14*(3), 137–158.

Palfrey, T. R., & Prisbrey, J. E. (1997). Anomalous behavior in public goods experiments: How much and why? *The American Economic Review*, *87*(5), 829–846.

Parks, C. D. (1994). The predictive ability of social values in resource dilemmas and public goods games. *Personality and Social Psychology Bulletin*, *20*, 431–438.

Parks, C. D., & Rumble, A. C. (2001). Elements of reciprocity and social value orientation. *Personality and Social Psychology Bulletin*, *27*, 1301–1309.

Pennisi, E. (2005). How did cooperative behavior evolve? *Science*, *309*(5731), 93.

Perugini, M., & Gallucci, M. (2001). Individual differences and social norms: The distinction between reciprocators and prosocials. *European Journal of Personality*, *15*, 19–35.

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*, 251–283.

Platow, M. J., McClintock, C. G., & Liebrand, W. B. G. (1990). Predicting intergroup fairness and ingroup bias in the minimal group paradigm. *European Journal of Social Psychology*, *20*, 221–239.

Pruitt, D. G. (1967). Reward structure and cooperation: The decomposed prisoners dilemma game. *Journal of Personality and Social Psychology*, *7*(1), 21–27.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, *28*(1), 363–392.

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

Radzicki, J. (1976). Technique of conjoint measurement of subjective value of own and other's gains. *Polish Psychological Bulletin*, *7*(3), 179–186.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.

Rapoport, A. (1967). A note on index of cooperation for prisoners dilemma. *Journal of Conflict Resolution*, *11*(1), 100–103.

Rapoport, A., & Dale, P. S. (1966). The "end" and "start" effects in iterated prisoner's dilemma. *Journal of Conflict Resolution*, *10*(3), 363–366.

Reuter, M., Frenzel, C., Walter, N. T., Markett, S., & Montag, C. (2011). Investigating the genetic basis of altruism: the role of the comt val158met polymorphism. *Social Cognitive and Affective Neuroscience*, *6*(5), 662–668.

Roch, S. G., Lane, J. A. S., Samuelson, C. D., Allison, S. T., & Dent, J. L. (2000). Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups. *Organizational Behavior and Human Decision Processes*, *83*(2), 185–212.

Roch, S. G., & Samuelson, C. D. (1997). Effects of environmental uncertainty and social value orientation in resource dilemmas. *Organizational Behavior and Human Decision Processes*, *70*(3), 221–235.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279–301.

Royston, P., Altman, D., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*, 127–141.

Rushton, J. P., Chrisjohn, R. D., & Fekken, G. C. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, *2*(4), 293–302.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, (pp. 61–71).

Samuleson, C. D. (1993). A multiattribute evaluation approach to structural change in resource dilemmas. *Organizational Behavior and Human Decision Processes*, *55*, 298–324.

Sawyer, J. (1966). Altruism scale - A measure of co-operative, individualistic, and competitive interpersonal orientation. *American Journal of Sociology*, *71*(4), 407–416.

Schubert, M., & Lambsdorff, J. G. (2013). Negative reciprocity in an environment of violent conflict: Experimental evidence from the occupied Palestinian territories. *Journal of Conflict Resolution. DOI: 10.1177/0022002713478792*.

Schulz, U., & May, T. (1989). The recording of social orientations with ranking and pair comparison procedures. *European Journal of Social Psychology*, *19*, 41–59.

Seeley, E. A., & Gardner, W. L. (2003). The "selfless" and self-regulation: The role of chronic other-orientation in averting self-regulatory depletion. *Self and Identity*, *2*(2), 103–117.

Sell, A., Tooby, J., Cosmides, L., & Orians, G. H. (2009). Formidability and the logic of human anger. *PNAS*, *106*(35), 15073–15078.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkten rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In H. Sauermann (Ed.) *Beiträge zur Experimentellen Wirtschaftsforschung*, (pp. 136–168). J.C.B. Mohr: Tübingen.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, *12*(1), 1–36.

Sheldon, K. M. (1999). Learning the lessons of tit-for-tat: Even competitors can get the message. *Journal of Personality and Social Psychology*, *77*(6), 1245–1253.

Shellenbarger, S. (2012). Are you are hero or a bystander? Available online: http://online.wsj.com/article/SB10000872396390443989204577603341710975650.html.

Shelley, G. P., Page, M., & Kuhlman, D. M. (2010). Semantic differential judgments of cooperators, competitors, and individualists at zero-acquaintance. *Letters on Evolutionary Behavioral Science*, *1*(1), 23–26.

Shelley, G. P., Page, M. J., Rives, P., Yeagley, E., & Kuhlman, D. M. (2009). Nonverbal communication and detection of individual differences in social value orientation. In R. Kramer, M. Bazerman, & A. Tenbrunsel (Eds.) *Social decision making: Social dilemmas, social values, and ethical judgments*, (pp. 147–169). New York: Psychology Press.

Shug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*(2), 87–94.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174.

Slade, R. M. (2004). *Software Forensics*, chap. 2: The players – Hackers, crackers, phreaks, and other doodz. McGraw-Hill.

Smeesters, D., Warlop, L., Van Avermaet, E., Corneille, O., & Yzerbyt, V. (2003). Do not prime hawks with doves: The interplay of construct activation and consistency of social value orientation on cooperative behavior. *Journal of Personality and Social Psychology*, *84*(5), 972–987.

Smith, D. H. (1994). Determinants of voluntary association participation and volunteering: A literature review. *Nonprofit and Voluntary Sector Quarterly*, *23*(3), 243–263.

Sonnemans, J., Schram, A., & Offerman, T. (1998). Public good provision and public bad prevention: The effect of framing. *Journal of Economic Behavior & Organization*, *34*, 143–161.

Sonnemans, J., van Dijk, F., & van Winden, F. (2006). On the dynamics of social ties structures in groups. *Journal of Economic Psychology*, *27*(2), 187–204.

Spielberger, C. D. (2010). State-trait anxiety inventory. In *The Corsini Encyclopedia of Psychology*. John Wiley & Sons.

Spielberger, C. D., Sydeman, S. J., Owen, A. E., & Marsh, B. J. (1999). Measuring anxiety and anger with the state-trait anxiety inventory (STAI) and the state-trait anger expression inventory (STAXI). In M. E. Maruish (Ed.) *The use of psychological testing for treatment planning and outcomes assessment*, (pp. 993–1021). Lawrence Erlbaum Associates Publishers, 2 ed.

Stanca, L., Bruni, L., & Corazzini, L. (2009). Testing theories of reciprocity: Do motivations matter? *Journal of Economic Behavior and Organization*, *71*(2), 233–245.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.

Stevens, S. S. (1950). Mathematics, measurement, and psychophysics. In S. Stevens (Ed.) *Handbook of Experimental Psychology*, (pp. 1–49). Wiley, New York.

Stevenson, M. K. (1992). The impact of temporal context and risk on the judged value of future outcomes. *Organizational Behavior and Human Decision Processes*, *52*, 455–491.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8*(2), 79–98.

Stouten, J., De Cremer, D., & Van Dijk, E. (2005). All is well that ends well, at least for proselfs: Emotional reactions to equality violation as a function of social value orientation. *European Journal of Social Psychology*, *35*(6), 767–783.

Sturm, T., & Ash, M. G. (2005). Roles of instruments in psychological research. *History of Psychology*, *8*(1), 3–34.

Sugden, R. (1984). Reciprocity: The supply of public goods through voluntary contributions. *The Economic Journal*, *94*(376), 772–787.

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.

Thomson, W. (1889). *Popular Lectures and Addresses*, vol. 1. London: Macmillan and co.

Thurstone, L. L. (1931). The indifference function. *Journal of Social Psychology*, *2*, 139–167.

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*(7284), 1089–1091.

Tyszka, T., & Grzelak, J. L. (1976). Criteria of choice in non-constant zero-sum games. *Journal of Conflict Resolution*, *20*(2), 357–376.

Utz, S. (2004). Self-activation is a two-edged sword: The effects of "I" primes on cooperation. *Journal of Experimental Social Psychology*, *40*(6), 769–776.

Utz, S., Ouwerkerk, J. W., & Van Lange, P. A. M. (2004). What is smart in a social dilemma? Differential effects of priming competence on cooperation. *European Journal of Social Psychology*, *34*(3), 317–332.

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the trust game. *Social Cognitive and Affective Neuroscience*, *4*(3), 294–304.

van Dijk, E., De Cremer, D., & Handgraaf, M. J. J. (2004). Social value orientations and the strategic use of fairness in ultimatum bargaining. *Journal of Experimental Social Psychology*, *40*(6), 697–707.

Van Doesum, N. J., Van Lange, D. A., & Van Lange, P. A. M. (2013). Social mindfulness: Skill and will to navigate the social world. *Journal of Personality and Social Psychology*, *105*(1), 86–103.

Van Lange, P. A. M. (1992). Confidence in expectations: A test of the triangle hypothesis. *European Journal of Personality*, *6*, 371–379.

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*(2), 337–349.

Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M., & Van Vugt, M. (2007). From games to giving: Social value orientation predicts donations to noble causes. *Basic and Applied Social Psychology*, *29*(4), 375–384.

Van Lange, P. A. M., De Cremer, D., van Dijk, E., & Van Vugt, M. (2007). Self-interest and beyond: Basic principles of social interaction. In A. W. Kruglanski, & E. T. Higgins (Eds.) *Social Psychology: Handbook of Basic Principles*, (pp. 540–561). New York: Guilford Press.

Van Lange, P. A. M., & Kuhlman, D. M. (1994). Social value orientations and impressions of partner's honesty and intelligence: A test of the might versus morality effect. *Journal of Personality and Social Psychology*, *67*, 126–126.

Van Lange, P. A. M., & Liebrand, W. B. G. (1991). Social value orientation and intelligence: A test of the goal prescribes rationality principle. *European Journal of Social Psychology*, *21*, 273–292.

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*(4), 733–746.

Van Lange, P. A. M., & Semin-Goossens, A. (1998). The boundaries of reciprocal cooperation. *European Journal of Social Psychology*, *28*(5), 847–854.

Van Lange, P. A. M., & Visser, K. (1999). Locomotion in social dilemmas: How people adapt to cooperative, tit-for-tat, and noncooperative partners. *Journal of Personality and Social Psychology*, *77*(4), 762–773.

van Prooijen, J. W., De Cremer, D., van Beest, I., Ståhl, T., van Dijke, M., & Van Lange, P. A. M. (2008). The egocentric nature of procedural justice: Social value orientation as moderator of reactions to decision-making procedures. *Journal of Experimental Social Psychology*, *44*(5), 1303–1315.

Van Vugt, M., Van Lange, M., P. A, & Meertens, R. M. (1996). Commuting by car or public transportation? a social dilemma analysis of travel mode judgements. *European Journal of Social Psychology*, *26*, 373–395.

Van Winden, F., Van Dijk, F., & Sonnemans, J. (2008). Intrinsic motivation in a public good environment. In C. Plott, & V. Smith (Eds.) *Handbook of Experimental Economics Results*, vol. 1. North-Holland: Amsterdam.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wallace, B., Cesarini, D., Lichtenstein, P., & Johannesson, M. (2007). Heritability of ultimatum game responder behavior. *Proceedings of the National Academy of Sciences*, *104*(40), 15631–15634.

Weimann, J. (1994). Individual behavior in a free riding experiment. *Journal of Public Economics*, *54*(2), 185–200.

Wyer, R. S. (1969). Prediction of behavior in 2-person games. *Journal of Personality and Social Psychology*, *13*(3), 222–238.

Wyer, R. S. (1971). Effects of outcome matrix and partner's behavior in two-person games. *Journal of Experimental Social Psychology*, *7*, 190–210.

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., & Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences*, *109*(50), 20364–20368.

Zeelenberg, M., Nelissen, R. M. A., Breugelmans, S. M., & Pieters, R. (2008). On emotion specificity in decision making: Why feeling is for doing. *Judgment and Decision Making*, *3*(1), 18–27.

Zuckerman, M. (1983). The distinction between trait and state scales is *not* arbitrary: Comment on allen and potkay's "on the arbitrary distinction between traits and states". *Journal of Personality and Social Psychology*, *44*(5), 1083–1086.