ECONOMIC STUDIES

DEPARTMENT OF ECONOMICS
SCHOOL OF BUSINESS, ECONOMICS AND LAW
UNIVERSITY OF GOTHENBURG
187

Self-Control and Altruism

Conny Wollbrant

UNIVERSITY OF GOTHENBURG

ISBN 978-91-85169-47-4
ISSN 1651-4289 print
ISSN 1651-4297 online

Printed in Sweden,
Geson Hylte Tryck 2010

To my grandfather, Bent Sand-Petersen

Table of Contents

Chapter 5. Conditional Cooperation and Social Group – Experimental Results from Colombia

Chapter 6. The Role of Beliefs, Trust, and Risk Preferences in Contributions to a Public Good

Acknowledgments

Though I have been told that finishing the thesis is just the beginning, I feel this also marks the end of a long journey and a home coming. As my friends, and in particular my family know all too well, I left home at 17 to study abroad; first at the United World College of the Atlantic, then at the University of York and finally at the London School of Economics. As time went by, I became more and more torn between returning to my family and pursuing other things that were perhaps not so easy to come by in my home town.

Already in my first year in York I started looking into possibilities of pursuing a Ph.D. in Göteborg. I remember writing Arne Bigsten e-mails inquiring about the program and visiting public lectures as often as I could, which typically meant Easter time when we had a few weeks off between two trimesters. I decided to move home, started planning the application, and when I was finally admitted to the Ph.D. program, around March 2006, I was overwhelmed with joy.

In getting accepted I believe I owe much to my family, and as such I believe my gratitude should go to them first. When thinking of my mother, her husband Lennart and my two brothers, Peter and Mikael, I am always astounded by the support I have had. I was often spending the night on Peter's sofa in Johanneberg after having stayed up too late on a week day. I loved sleeping on that sofa and, ironically enough, I got to inherit it. During my time on the British Isles, my father came visiting me in Wales, York and London which was very much appreciated. Peter accompanied him to both York and London for my graduations and on both occasions I was told how proud they both were, which was much appreciated. I should add to this the many phone calls I have gotten during the years, before the time of broadband internet.

Among the callers, my grandmother is perhaps at the top of the list. Always curious to hear what was going on, I could call from a payphone only to hear "Is it Conny, I'll call you back!" I spent a lot of time with my grandparents as a child and though I moved so far away I never felt that we lost touch or that their care was in any way diminished. And so my first set of bed linen was a gift from my grandparents along with a Swedish cheese slicer and a pair of gloves my grandmother knitted, along with numerous other items. In fact, being a longstanding fan of Donald Duck, my grandmother started a subscription for me and posted the magazine every week.

While still on the British Isles, I want to thank those who made my time there pleasant and enjoyable. I am sad to say that one of you, Benjamin Oabona, just recently passed away. My sympathy goes out to your wife and children. I have a feeling you'd say we'll meet again. But not yet. You are dearly missed. Along with Ben, I shared my room with Sebastian Hayes, who was always up for a bit of fun. I also met my first teacher of Economics, Andi Kumalo, whom I much admired. Jason Fairbourne I got to know during my time in London and he has always been helpful both before and after I began my research efforts.

A lesson in determination might be had from my uncle Mikael. He suffered a severe injury early in life leaving him disabled and he now lives in my grandparents' home. Still he now walks every day, at an extraordinary pace, for at least an hour, paints and shovels large volumes of snow. His frequent inscription on his crafted plaques *tappa inte sugen*, roughly instructs us not to get discouraged. Whenever I was abroad for study, I always kept one beside my bed. Recently, I attended his art exhibition and was glad to see that he is painting in bright colors.

Of course, getting in to a Ph.D program is one thing but getting through it is quite another, and though my family have been equally important these last few years as before, I now have a larger group of people that needs to be added to the list.

First of all, and without doubt most importantly I would like to thank Clara. She has been alongside me through all of it these past few years. With regard to the academic work, she knows it as well as I know it myself. This is because I talk about it a lot and remarkably enough, she has just kept listening. I can honestly say that I never felt alone with this project and I can only hope I have managed to return some of her favors. Academic work aside, I hardly have to say that she is far more important to me on a personal level, compared to which this whole enterprise fades.

Since she is basically inseparable from them, I will take this opportunity to thank Clara's family and friends along with her. It is a curious thing, visiting Colombia, when you are from the frosty north. On my first trip, I confess, I had trouble remembering all the names of her family members and close friends. Not due to disinterest on my part, it is just that they were so many! I was astounded at how well received I was everywhere and I always felt very welcome. I have spent many days in *Las Tías* (*the aunts'*, actually including a few uncles) place meeting Carlos, Clarita, Marta, Alcira, Gonzalo, Alberto Elias, including the late Leonor and Adela; Clara's close family: Luz, Antonio, Sergio and Gloria, Alejando and Clemencia, Juan Camilo and Paula (who also hosted me on my visit to Tucson). Together you made me feel very welcome. The Colombian experience also included many outings and adventures with a bunch of Clara's friends: Johnny and Paula, Paula, Liana, Vivi, Tata, Clifford and Pily, and Angela. I would like to especially thank Clara's father, Antonio, for being so helpful and supporting me with the logistics involved in the experiments; everything from printing instructions, giving me a ride back and forth to venues where the experiments were conducted, going to the petrol stations in the late evenings to exchange notes for coins, and always with good humor.

Second on this list is a dear friend, Kristian. He was actually the one encouraging me to send the e-mail to Arne, read my application for the LSE and Göteborg and during my time here, I have had almost daily contact with him. We first met during my time in Wales where we were boarded in the same corridor. Much of the work in this thesis is in part the result of our friendship and complementary skills, but being slightly more senior, he has also acted as a quasi-supervisor, for which I am very grateful.

Third on this list are a few other dear friends: Linus Nordén, Marcin Szklarski, and Anders Gustafsson. I shared my apartment with both Anders and Marcin during different periods and we shared many concerns. I also had quite number of meals prepared for me by both, and Anders sometimes drove me to the office at Södra Allégatan, very early in the mornings before he had to go for work himself, often in the opposite direction. I am very grateful for all your support and hope I will be given the opportunity to repay some day. Linus lived very close to me and for about a year or so, we met early every morning to go for walks around a nearby lake, Härlanda tjärn, before we set off for work. During our walks I know I sometimes talked excessively about my work and he always listened. And, came back the next morning. Thank you for your patience and attention, you have always been, and continue being a great friend.

I also had two formal supervisors, Peter Martinsson and Katarina Nordblom, and I can attest to this being an excellent mix. When I first came to visit Peter in his office I was immediately encouraged by his enthusiasm. He has always set ambitious deadlines for me and I have been trying to keep up. We have several interesting projects at the moment and I hope it continues in the same fashion. Katarina officially became my supervisor at a much later date, though I had already informally visited her office several times. She has given me invaluable input, but most of all she made me feel confident about what I was doing. When deadlines drew near, they were both very comforting and I think this is simply a testament to what nice people they are.

I have also had several excellent teachers. Among these I would like to extend a special thanks to Olof Johansson-Stenman, Johan Stennek and Måns Söderbom. Olof has always kept an open door and I have had many interesting discussions with him both about ideas pertaining to my own work but also on other topics. Johan enthusiastically designed a graduate level reading course in game theory and introduced me to the subject. Following that, we have had many interesting conversations on widely disparate topics, which I have very much appreciated. He also provided very constructive advice on the introduction to the thesis which, I believe, improved it a great deal. Måns taught a brilliant course in econometrics here at the school which I concluded just shortly before running my first experiment. The familiarity with econometrics and Stata that I gained during the course has been very valuable to me and made my work much easier.

I would also like to extend my gratitude to my teachers in the first year, Ola Olsson, Fredrik Carlsson, Olof Johansson-Stenman, Lennart Hjalmarsson, Renato Aguilar, Lennart Flood, Roger, Rune Stenbacka, Elias Tsakas and Dick Durevall. Among them, I especially remember the lectures of Lennart Flood which were not only excellent but also very entertaining. Renato Aguilar’s lectures got me curious about mathematics and without this (and Katarina’s guidance) I don’t think I would have dared trying anything theoretical. In the department in general, I think we maintain a very friendly and constructive atmosphere which I am truly happy about participating in. I have had the opportunity to get to know a lot of nice and interesting people. In my own cohort, I was lucky to share the courses with Måns Nerman, Andreas Kotsadam, Eyerusalem Siba, Pham Khanh Nam, Haoran He, Yonas Alem,

Clara Inés Villegas Palacio and Kofi Vondolia. I would also like to thank the colleagues in the department, past and present. Especially I want to thank, Elias Tsakas, Andreea Mitrut, Macela Ibañes, Niklas Jakobsson, and Amrish Patel and Michele Valsecchi who both read parts of the thesis before submission, and Jesper Stage who guided me to material on resource use. I also received very helpful comments at different stages of the dissertation from Martin Dufwenberg and Martin Kocher (who also hosted me in Munich) that were very valuable both in determining the direction of the work but also improved its quality. Thank you both! I would also like to thank the administrators, and among them, Eva-Lena Neth Johansson and Gerd Georgsson. Both of you have been very helpful during my time as a PhD student, especially during the time which the experiments were conducted. Thank you very much!

Finally, and most importantly, both personally and professionally, I would like to thank my grandfather to whom this thesis is dedicated. When I first got involved with the application process for the UWC schools, he drove me to the interviews, and later on, he and my uncle Mikael drove me all the way to the gates of the school in Wales. He was a strong larger-than-life personality who from a very early age instilled in me a sense of wonder and curiosity with the world in general and with learning and knowledge in particular. I held him in very high esteem and always tried to emulate him. As far as role models go, I cannot praise him enough and wish I could have handed over a copy.

I think it is fair to say that in my family expressions of love are indirect and subtle: “dress warm,” “get enough sleep,” drive carefully.” Direct and vocal expressions are rare, so in keeping with this tradition, I will stop here.

Göteborg, April 2010 *Conny Wollbrant*

Abstract

This thesis consists of two theoretical papers on self-control (Chapters 1-2) and four empirical papers (3-6) of which two (Chapter 5 and 6) explore determinants of pro-social behavior, while the other two (Chapters 3-4) examine the relationship between self-control and pro-social behavior.

(1) “A Theory of Self-Control-Conflict: The Pyrrhic Motions of Reason and Passion.” (with Kristian Ove R. Myrseth).

We model self-control conflict as a struggle between an agent and a visceral influence, which impels the agent to act against her better interest. The agent holds pre-commitment technology to avoid the conflict altogether, though at a cost. The agent’s decision to face down temptation, to pre-commit, or to succumb without resisting is determined by three factors: (1) the payoff from the goal, (2) the strength of the temptation, and (3) willpower. We consider implications from the agent (1) underestimating the anticipated visceral influence and (2) overestimating her stock of willpower. Underestimating the anticipated visceral influence may lead the agent to exaggerate the expected value of resisting temptation, and so mistakenly forego pre-commitment. Overestimating her stock of willpower may lead to a similar result. Finally, a welfare analysis yields the counterintuitive prediction that higher willpower under certain circumstances reduces welfare.

(2) “Self-Control in Games.”

People are often tempted to deviate from their optimal strategies. A situation reflecting such interference by temptation is defined as a self-control game where each player consists of two cognition types. One type generates biases in decision making by producing visceral influences. In contrast, another cognition type can ameliorate visceral influences by exercising self-control. The set of outcomes reflecting perfect self-control are called "self-control equilibria" and is equal to the set of subgame perfect Nash equilibria. In contrast, the set of "temptation equilibria" reflects imperfect self-control and is a superset if will-power is "high enough." We explore implications for several instances of social interaction when players are altruists tempted to be greedy.

(3) "Reconciling Pro-Social vs. Selfish Behavior: Evidence for the Role of Self-Control." (with Peter Martinsson and Kristian Ove R. Myrseth)

We test the proposition that individuals may experience a self-control conflict between short-term temptation to be selfish and better judgment to act pro-socially. Using a public goods game and a dictator game, we manipulated the likelihood that individuals identified self-control conflict, and we measured their trait ability to implement self-control strategies. Consistent with our hypothesis, we find that trait self-control exhibits a positive and significant correlation with pro-social behavior in the treatment that raises likelihood of conflict identification, but not in the treatment that reduces likelihood of conflict identification.

(4) "Conditional Cooperation and Self-Control." (with Peter Martinsson and Kristian Ove R. Myrseth**)**

When facing the opportunity to act either in self-interest or in the interest of others, individuals may experience a self-control conflict between pro-social preferences and urges to act selfishly. We explore the domain of conditional contribution, and we test the hypothesis that an increase in an individual's belief about others' average contribution increases contributions more when her willpower is high than when it is low. We employ a subtle framing technique and the strategy method in a public goods experiment. Consistent with our hypothesis, we find that conditionally cooperative behavior is stronger when beliefs of high contributions are accompanied by high rather than low levels of self-control.

(5) "Conditional Cooperation and Social Group – Experimental Results from Colombia." (with Peter Martinsson and Clara Inés Villegas Palacio)

In contrast to previous studies on cross-group comparisons of conditional cooperation, this study keeps cross- and within-country dimensions constant. The results reveal significantly different cooperation behavior between social groups in the same location.

(6) "The Role of Beliefs, Trust, and Risk Preferences in Contributions to a Public Good." (with Martin Kocher, Peter Martinsson, and Dominik Matzat)

This paper experimentally investigates the role of beliefs, trust, and risk preferences in shaping cooperative behavior. By using a linear public goods game and the strategy method for revealing conditional contribution schedules, we categorize subjects into different types of contributors. Our results support the notion that beliefs about others' behavior and trust are positively associated with cooperation while risk preferences do not seem to matter.

Keywords: Self-Control, Temptation, Game Theory, Experiment, Pro-Social Behavior, Conditional Cooperation, Altruism, Public Goods, Trust, Risk

JEL Classification: C91, C79, D01, D03, D64, D69, D70, D90, H41

Introduction

- Self-Control and Altruism

Conny Wollbrant, *University of Gothenburg*

> It is clear, then, that a human being is more of a political animal than is any bee… and humans are the only animals who possess reasoned speech
>
> *Aristotle*

At the very moment of biting into a delicious cookie, one is aware that this conflicts with one's goal of maintaining a slim figure and better health. Similarly, when on occasion, we find ourselves angered in quarrels with our colleagues, friends or loved ones, we know we have little, if anything to gain by responding in a rude manner. But, irritable after insufficient sleep or perhaps increasingly impatient after an extended wait, we sometimes do. Moreover, even at the moment of taking an unjustified large share of the pie for ourselves, we know we are being unfair.

The feeling of trying to adhere to a diet, contain one's anger, or act in congruence with social norms or moral codes, while at the same time feel an urge not to, is an experience familiar to most of us; we try to resist, while our sweet tooth join forces with our angry and egoistic impulses to seduce us with promises of cookies, retaliation and larger shares of the pie. Similar experiences permeate our daily lives, manifesting themselves in a range of situations and as a consequence, we regularly act against our better judgment by over-eating, lashing out and acting selfishly, even while fully aware that we are doing so.

Reason and passion

Ancient philosophers understood the discrepancy between one's perceived best interest and one's urges in terms of a conflict between our reason and passion (e.g., Plato, 2000).

Roughly, the human soul was thought to consist of a rational self (reason) and an emotional (passion) as well as a third part which acted as the enforcer of reason, commonly referred to as our will. Passion corresponds to older parts of our brains which we share with other animals (the limbic system), while System 2, by contrast, is a later development (the brain's prefrontal cortex) responsible for abstract thought processes such as planning and language (the brain's prefrontal cortex) (Lowenstein, 1996; 2000).[1]

The properties of each self are particularly revealing as our emotional self was primarily developed in order to secure certain vital functions. As such it is designed to operate quickly and almost in the absence of any deliberation. Moreover, it is often accompanied by powerful emotions (*passions*), which motivate us to engage in one or other specific behaviors.[2] For example, when hungry one feels the urge to eat and when angry one is more likely to engage in aggressive behaviors, such as being rude, raising one's voice or fight. Similarly, selfish behavior may be understood partly as a result of greed (O'Donoghue and Loewenstein, 2007).

Our rational self, by contrast, operates at a much slower pace than the emotional self and relies on deliberation, also referred to as our *reason*. It is responsible for highly abstract exercises such as planning and strategizing, both of which rely on higher cognitive functions. When reasoning, we often realize that overeating will cause problems in the future, that fighting might not be the best course of action (for example, if we are sure to lose) and that our quarrels often are unproductive. Similarly, we also realize that we ought not always behave in a selfish manner but instead think about others around us. Considerations such as planning, thinking about the well-being of others or what constitutes a fair share, are all the product of abstract though processes, represented by our rational self.[3]

It is thus by virtue of our reason that we deem our diets, rude responses and selfishness inappropriate while our passions such as, hunger, anger and greed, would have us indulging in behaviors that sometimes conflict with our better judgment. This ability to reason is something humans possess above our instinctual and emotional facilities. It is for this reason that a human being is indeed more of a political animal than any bee, thinking, planning,

[1] Reason and Passion correspond to dual process theories in psychology where passion is referred to as System 1 and Reason System 2. This terminology reflects the evolutionary sequencing in humans as passion developed first (Kahneman, 2003; Sloman, 1996).

[2] Classical writers talked about "the passions" with the implication that these are a kind of suffering. Anger, in the Greek *Menis* was mainly appropriate for the Gods and consequently represented a kind of cosmic force which we suffer passively. Hence, we also speak of "the passion of the Christ." In fact, we talk about the passions in precisely these terms: "we get angry," "we get hungry," and "we get thirsty" and so on.

[3] Consistent with this idea, Pronin et al. (2008) find that decisions about others resemble decisions about "future selves," both classes of which contrast to decisions about less abstract "present selves."

strategizing and debating, as is suggested by Aristotle (1981) in the epigraph at the beginning of the introduction.

Self-control conflict

Self-control problems arise when the operation of the rational and emotional selves (reason and passion) result in conflicting prescriptions for behavior. For example, when the rational self would like to diet to become slim and healthy, contain one's ill-temper or be a fair person, while the emotional self would like to indulge in a delicious cookie, retaliate on a slight or act selfishly.

To avoid such problems, many of us try to outsmart ourselves by constructing elaborate strategies that will lead us off the path to temptation. Like in game of poker, however, our strategies are relatively impotent when our opponents know about them and for the same reason, it is hard to play against ourselves. Nevertheless, we can try to avoid games where we are sure to lose by picking our opponents well. Dieters are often advised to do their shopping shortly after they have had a large meal in the hope that being full will stop them from buying high calorific foods they might otherwise have bought, had they entered the shop hungry. Other common tricks include not bringing sweets or alcohol into our homes, buying cigarettes by the pack or placing the alarm clock far from our beds. When our demons come to haunt us with cravings for chocolate, having another smoke or snoozing just one more time, these options are unavailable to us. Like Ulysses in Homer's *The Iliad*, we in a sense tie our hands to the mast by excluding the possibility of acting in ways we deem undesirable by using some form of pre-commitment device.

In Chapter 1, we address the self-control problem from a theoretical perspective.[4] An interesting psychological insight is that self-control effort expended by the rational self is non-monotonic in temptation strength, following the path of an inverted U, initially rising, reaching a maximum and finally falling towards zero when temptation strength becomes insurmountable. That is to say, an individual might decide not to expend much effort in resisting a cookie for two reasons: either the cookie is not too tasty and hence it does not significantly tease her sweet tooth, or, it is so tasty that there is no point in resisting.

[4] Other theoretical work includes Ainslie (1975), Thaler and Shefrin (1981), Schelling (1984), Laibson (1997), Gul and Pesendorfer (2001), Fudenberg and Levine (2006). Among these, Fudenberg and Levine (2006) present the model closest to ours. For an overview of work on time preference, see Fredericks et al. (2003).

In the context of pre-commitment, our model yields particularly interesting implications for behavior when the individual overestimates her willpower or underestimate the strength of temptation. In either case, increasing willpower will sometimes lead to greater welfare losses. When, for example, we deciding whether or not to pre-commit by avoiding to bring chocolates into our homes, we inevitably have to consider our chances of resisting the temptation of eating them once they are there, right in front of us on our kitchen tables, or lurking in our cupboards.

If we do not assess our chances correctly, we might choose to bring the chocolates home when we ought not to. Since very strong individuals are more likely to think they have good chances of resisting, overconfidence and underestimating temptation is more likely to cause them more harm in terms of welfare than weaker individuals. In fact, this turns out to be the case even if they are only slightly over confident. In effect, it is much worse to think that one is twice as strong when one is actually strong than when one is weak. In the former case, one's perceived strength is far more incorrect and will therefore lead to bad choices more often. A very strong individual, thinking she is a little bit stronger, might incorrectly believe she can manage having chocolates in her home without eating them. In the latter case, incorrectly assessing one's chances will not have such dramatic effects. A very weak individual thinking she is a little bit stronger still would not think she is strong enough to bring the tempting chocolates home. In short, being stronger is not always better.

Self-control and social interaction

Viewing behavior as an outcome of a conflict between reason and passion provides a useful framework for thinking about the link between self-control and altruism. In chapter 3, we explore the hypothesis that the problem of pro-social vs. selfish behavior may represent one of self-control in an experimental setting. Just like consuming a small chocolate just once will not be detrimental to one's diet, being selfish just once does not render an individual anti-social. In contrast, if the small chocolate becomes many small chocolates and one repeatedly behaves selfishly, one will soon find oneself both overweight and anti-social.

Our experiments made use of a public goods game and a standard dictator game. In the public goods game, each individual belonged to a group of four individuals. Each individual was then given an amount of money and had to decide how much to put in her own pocket, and how much to give to the group. The money that was given to the group automatically grew, and was then divided between all members of the group. In this game, it is best for the group if everyone gives their money to the group, but for each individual, it is best if everyone else gives money to the group while she does not. In the dictator game, individuals are given a sum of money and have to decide how much to put in their pocket and how much to give to the Red Cross in Colombia.

We then manipulated the likelihood that individuals viewed their decisions either as an isolated event or part of a larger pattern of behavior, and we measured the participants' trait self-control using the Rosenbaum Self-Control Schedule (Rosenbaum, 1980).[5] We hypothesized that if individuals think of their decisions as an isolated event, they might be more selfish in both games and give less than they would if they thought of this decision, not as an isolated event, but rather as one of many similar decisions taking place in the future. Consistent with our hypothesis, we find that the self-control measure is positively correlated with pro-social behavior when individuals viewed their decisions as part of a larger pattern of behavior. [6]

Whenever we are tempted to be selfish rather than pursuing our pro-social preferences, we must recruit our willpower to avoid yielding to the tempting selfish behavior. [7] The fact that individuals seem to condition their pro-social behavior on the expected behavior of others, seems to suggest some abstractly structured judgment of fairness, reflecting the operation of our rational self. Moreover, there is now neurological evidence demonstrating that brain regions pertaining to our rational self (prefrontal cortex) is more active when the individual considers issues of fairness, cooperation and trust (Lieberman, 2010).

[5] The framing manipulation is adopted from Myrseth and Fishbach (2010).

[6] That is not to say that people do not engage in altruistic behavior in the absence of abstract cognitive functions. On the contrary, individuals may for example experience strong sympathetic urges to give to another, or even give more than what one considers fair (O'Donoghue and Loewenstein, 2007). For example, one might suspect that a beggar in the street is a "con" seeking "easy money", but one cannot help yielding to the sorry gestures. In such as case, one might say that one is tempted to be generous while knowing full well that one ought not to. In other instances, one might experience the same sympathetic urge while simultaneously thinking that generosity is justified. In such as case, sympathetic urges and generosity do not conflict and while experiencing the sympathetic urges, one cannot claim to be acting against better judgment.

[7] In addition, the passions (also known as visceral influences) tend to undermine altruism in general as they tend to take the form of aversive unpleasant sensations. For example, anger, pain and hunger are all unpleasant encouraging the individual to end her suffering and as a result, individuals tend to focus inwards. [7] From an evolutionary perspective, this serves us well since it promotes survival, for example by motivating us to eat, but it also trumps the efforts of the rational self for altruistic behavior (Damasio, 1994; Loewenstein, 1996).

In Chapter 4 we explore the conceptually related hypothesis that the problem of *conditional cooperation* vs. selfish behavior may too represent one of self-control in a public goods game using the strategy vector method. In this game, people do not simply give an amount to the group as before in Chapter 3, but instead have to decide how much to give conditional on every possible average donation of the other group members. They are asked, for example, how "much would you like to give if the others in your group decide to give 5 on average?," or 6, or 7, and so on. Roughly speaking, conditional cooperators are those who contribute when they believe others contribute, as such their contributions increase as their expectations of other contributions increase (see e.g. Fishbacher et al. 2001).

Employing the same manipulation as in Chapter 3, we find that the interaction between an individual's self-control and her belief about others' contribution is positively correlated with contributions to the public good when individuals are more likely to view their decision as one of many similar decisions, rather as an isolated event. When a conditional cooperator thinks her group members will contribute 50%, she too would like to contribute something close to it, say 40%. Of course, if she is tempted to be selfish and contribute nothing, she would experience a self-control conflict between contributing and acting selfishly. A higher willpower would then lead to higher contributions as long as she takes a long term view of the decisions. Taken together, chapters 3 and 4 suggest that cheap framing manipulations may be used in order for individuals to take a long term view of their behavior and then activate their own cognitive resources in the service of the public good.

Self-control and strategy

The model in Chapter 1 states that a necessary condition for the problem of altruistic vs. selfish behavior to represent one of self-control is our *reason* would like to be pro-social, while our *passions* dictate behavior that conflict with this preference. The list of such potentially conflicting passions is not confined to greed, but is rather long. We have already mentioned the cases of hunger and anger but one can equally well imagine, for example, jealousy, fear and contempt. The fact that many emotions have social characteristics - that is to say, they involve or take as their object another person - suggests that self-control problems might be even more pervasive in instances of social interaction than in individual decision making.[8]

To see that many emotions are very often social in this sense, we only need to consider the fact that our anger, jealousy, fear and contempt are often about someone else. For example, we are *angry at* someone, *jealous of* another, we have fear *of someone* and in contempt *we look down on* someone. In Chapter 2, we elaborate on this idea by first constructing a general theoretical framework that allows for analyzing self-control problems in strategic interaction where such social emotions are likely to present themselves. The model is applied to several instances of social interaction, yielding novel predictions. For example, when thinking about whether to cooperate or not, one ought to take into account the partners capacity to control her urges for personal gain. Similarly, before trusting someone, it is prudent to assess the partner's ability not to stray from agreements and commitments. It would not be wise to enter a partnership, romantic, financial or otherwise, with someone who would jump at any other opportunity as soon as it presents itself. In fact, this is likely to be an important aspect of trust as the analysis from chapter 2 verifies.

Pro-social preferences

Pro-social preferences, when understood as preferences incorporating some form of concern for the welfare of others, seem to be widespread and much attention has been devoted to the exploration of its different forms (for an overview, see Fehr and Schmidt, 2006). Examples include pure altruism, reciprocity, fairness and the aforementioned conditional cooperation.

Chapters 5 and 6 further explore conditional cooperation using experimental methods. We find that poorer socio-economic groups contribute more than richer groups when the expectation of others' contribution is the same. For example, when partners' contributions are expected to be 50%, individuals in poorer groups might want to contribute 40% while individuals in the richer group contribute less. A possible explanation for this finding is that poorer individuals rely more on networks for insurance purposes than do richer individuals.

[8] In fact, emotions such as jealousy and contempt are impossible in a world consisting of a single individual, while this is not the case for hunger or thirst.

While the suggestion that the problem of altruistic vs. selfish behavior may represent one of self-control is the main proposition of this thesis, self-control problems are not limited to the presence of an urge to behave selfishly. For example, should a strategic opponent get sufficiently angry, she might fail to control her urges for retaliation and engage in punishment that is costly to both. When devising strategies, one thus ought to take into account the potential emotions the expectation of one's strategy might stir up in the opponent, as well as her preferences. As the list of emotions that could be of strategic interest is long, the role of self-control in social interaction is potentially very powerful.

References

Ainslie, George. 1975. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin,* 82(4): 463-96.

Aristotle. 1981. "The Politics." Rev. Trans. Saunders, Trevor J. London, England: The Penguin Group. (Orig. work, 350 B.C.)

Damasio, Antonio, R. 1994. "Descartes Error: Emotion, Reason, and the Human Brain." New York: Putman.

Fehr, Ernst, and Klaus M. Schmidt. 2006. "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories." In *Handbook on the Economics of Giving, Reciprocity and Altruism,* Kolm, Serge-Christophe, and Jean Mercier Ythier (Eds.)., 615-91. Amsterdam: Elsevier.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters,* 71, 397-404.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature,* 40(2): 351-401.

Fudenberg, Drew, and David Levine. 2006. "A Dual-Self Model of Impulse Control." *The American Economic Review,* 96(5): 1449-76.

Gul, Faruk and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica,* 69(6): 1403-35.

Kahneman, Daniel. 2003. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist,* 58, 697-792.

Laibson, David I. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics,* 112(2): 443–77.

Lieberman, Matthew D. 2010. "Social Cognitive Neuroscience." In *Handbook of Social Psychology,* (5th ed.), Fiske, Susan T., Daniel T. Gilbert, and Gardner Lindzey (Eds.)., 143-93. New York: McGraw-Hill.

Loewenstein, George. 1996. “Out of control: Visceral Influences on Behavior.” *Organizational Behavior and Human Decision Processes*, 65(3), 272–92.

Loewenstein, George. 2000. “Preferences, behavior and welfare: Emotions in Economic Theory and Economic Behavior.” *The American Economic Review*, 90(2): 426-32.

Myrseth, Kristian O. R., and Ayelet Fishbach. 2010. “Seeing Self-Control Conflict: The Problem of Isolated versus Interrelated Temptations.” Unpublished.

O’Donoghue, Ted, and George Loewenstein. 2007. “The Heat of the Moment: Modeling Interactions Between Affect and Deliberation.” http//:www.arts.cornell.edu/econ/edo1/heat.pdf

Plato. 2000. “The Republic.” Ed. Ferrari, Giovanni G.R., *Cambridge Texts in the History of Political Thought,* Trans. Griffith, Tom. (Orig. pub. 380 B.C.)

Pronin, Emily, Cristopher Y. Olivola, and Kathleen A. Kennedy. 2008. “Doing Unto Future Selves as You Would Do Unto Others: Psychological Distance and Decision Making.” *Personality and Social Psychology Bulletin,* 34(2): 224 – 36.

Rosenbaum, Michael. 1980. “A Schedule for Assessing Self-Control Behaviors: Preliminary Findings.” *Behavior Therapy ,* 11(1): 109–21.

Schelling, Thomas. 1984. “Self-Command in Practice, in Policy, and in a Theory of Choice.” *The American Economic Review,* 74(2): 1-11.

Sloman, Steven. A. 1996. “The Empirical Case for Two Systems of Reasoning.” *Psychological Bulletin,* 119(1): 3-22.

Thaler, Richard H. and Hersh M. Shefrin. 1981. “An Economic Theory of Self-Control.” *The Journal of Political Economy,* 89(2): 392-406.

Chapter 1

A Theory of Self-Control Conflict: The Pyrrhic Motions of Reason and Passion[1]

Kristian Ove R. Myrseth, ESMT European School of Management and Technology, Germany[2]

Conny Wollbrant, University of Gothenburg, Sweden[3]

Abstract

We model self-control conflict as a struggle between an agent and a visceral influence, which impels the agent to act against her better interest. The agent holds pre-commitment technology to avoid the conflict altogether, though at a cost. The agent's decision to face down temptation, to pre-commit, or to succumb without resisting is determined by three factors: (1) the payoff from the goal, (2) the strength of the temptation, and (3) willpower. We consider implications from the agent (1) underestimating the anticipated visceral influence and (2) overestimating her stock of willpower. Underestimating the anticipated visceral influence may lead the agent to exaggerate the expected value of resisting temptation, and so mistakenly forego pre-commitment. Overestimating her stock of willpower may lead to a similar result. Finally, a welfare analysis yields the counterintuitive prediction that higher willpower under certain circumstances reduces welfare.

JEL Classification: D01, D03, D69, D90

Keywords: self-control, temptation, inter-temporal choice

---

[1] Financial support from the Swedish Research Council (Vetenskapsrådet) is gratefully acknowledged. We are grateful for helpful comment and suggestions on this project from Olof Johansson-Stenman, Peter Martinsson, Katarina Nordblom, Clara Inés Villegas Palacio and Robert Östling

[2] ESMT European School of Management and Technology, Schlossplatz 1, 10178 Berlin, Germany; Ph +49 (0) 30 212 31 1529; Fax +49 (0) 30 212 31 1281 ; E.mail myrseth@esmt.org

[3] Department of Economics, School of Business, Economics and Law, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 26 15; Fax +46 31 786 10 43; E-mail conny.wollbrant@economics.gu.se

1. Introduction

We introduce a simple model of self-control that reduces to a one-period maximization problem for an agent of bounded rationality. We conceptualize the self-control conflict as a costly struggle between the agent and a conflicting visceral influence that acts on the agent like a force. The agent may anticipate the influence, and she holds technology to pre-commit to avoiding the influence (as in Thaler & Shefrin, 1981), though at a cost. The agent may decide to avoid the influence, to struggle against the influence, or to succumb to the influence without struggle. Critically, we examine the effect of underestimating anticipated visceral influence. We also consider the effect of overestimating willpower. We find that underestimating anticipated visceral influence may lead the agent to exaggerate the expected value of resisting the influence, thereby causing her to mistakenly forego pre-commitment. We find similar results for overestimating willpower. Furthermore, and perhaps surprisingly, our welfare analysis suggests that higher levels of willpower under certain circumstances reduce welfare.

Our model of self-control, as most others, is suitably illustrated by antiquity's story of Ulysses and the song of the Sirens.. Upon hearing their seductive song, mariners were said to leap into the sea. Ulysses was curious to hear it, and so he ordered his crew to tie him to the mast and to stuff their ears with beeswax. They were to leave him tied while within range of the song, no matter how much he begged. Having thus prevented himself from leaping overboard, Ulysses heard the song of the Sirens as he and his crew sailed past their island.

Since Ulysses anticipated that the song would cause him to act against his better judgment, he constrained his future choice set by eliminating the possibility of acting on the temptation. While the canonical case of successful self-control by pre-commitment (Elster, 1977; Schelling, 1984; Thaler & Shefrin, 1981), the story begs the question, what would Ulysses have done had he believed that the song's influence was weaker, that his desire to live was stronger, or that his power of will was mightier? We explore these questions, but we also explore the welfare implications of changes in these parameters. Suppose Ulysses, just as individuals of the 3rd millennium, also tended to underestimate the influence of anticipated visceral influences. How would he then benefit from having a larger stock of willpower? Similarly, how would overconfidence in his own willpower affect his decisions and hence his welfare?

A key insight from our model is its prediction that the agent's effort in resisting temptation does not simply increase monotonically with the strength of the visceral influence, as might be expected from intuition. Rather, since effort is costly, effort as a function of the strength of visceral influence follows an inverted U, rising at first, reaching a maximum, and eventually dropping to zero. That is, when the visceral influence is too strong, it is not worthwhile for the agent to expend effort in the struggle. This has important implications for the agent's decision of whether or not to pre-commit, especially if she underestimates the influence of anticipated temptations on behavior. In such cases, the agent will exaggerate the expected value of trying to resist temptation, and hence mistakenly forego pre-commitment. That is, had Ulysses underestimated the influence of the song of the Sirens, he may have mistakenly thought it feasible to resist the temptation and thus not necessary to have himself tied to the mast. Finally, a welfare analysis of the model yields the surprising result that more willpower under certain circumstances reduces welfare. Suppose Ulysses underestimated the degree to which the song exerted influence over him, but that his faulty judgment, due to his low stock of willpower, nevertheless led him to the correct decision of tying himself to the mast. A higher stock of willpower, then, could have precipitated the wrong conclusion that he was strong enough to successfully face the song without trying himself to the mast. Overestimating his stock of willpower could have lead to the same mistake.Although based on somewhat different conceptual foundations, our model is closely related to the reduced form of Fudenberg and Levine's (2006) dual-self model. Their model reduces to a maximization problem, and it is slightly different to the axiomatic model proposed by Gul and Pesendorfer (2001, 2004). We discuss the relationship of our model to both of these models in the final section.

2. Psychological foundations

We ground our model on the idea that self-control conflict can be understood as a battle between will and passion (Loewenstein, 1996; Metcalfe & Mischel, 1999; Baumeister, Heatherton, & Tice, 1994; for exceptions, see Fishbach and Shah, 2006). While past models of self-control conflict posit a game between multiple "selves," wherein a far-sighted "self" strategizes against a more myopic "self" (e.g., Thaler & Shefrin, 1981; Schelling, 1984; Benabeu & Pycia, 2002; Fudenberg and Levine, 2006), our model does not. Rather, and in line with

Loewenstein's (1996) paper on visceral influences, we conceptualize the self-control conflict as a maximization problem for an agent faced with temptation that impels the agent to act against her better judgment. This also is consistent with recent dual-process models, which classify cognition into one of two distinct processing modes (e.g., see Kahneman, 2003; Sloman; 1996):

- System 1: effortless, parallel processing; associative reasoning; "hot," emotional influences; uncontrolled

- System 2: effortful, serial processing; rule-based reasoning; "cool" thinking; controlled

Psychologically, we treat self-control conflict as a maximization problem for an individual's System 2 cognition (i.e., the agent) in the face of conflicting System 1 cognition (i.e., the visceral influences of temptation). That is, the agent must determine an optimal course of action when facing or anticipating a force that impels behavior against her better judgment.

The agent in our model can choose either to fight conflicting temptation by exercising willpower or simply to yield without struggle, thereby succumbing to temptation. The choice between fighting and yielding is important because fighting temptation requires cognitive resources that are limited (e.g., Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven & Baumeister, 2000). In a dynamic context, therefore, fighting is costly due to resource depletion. Moreover, when having decided to fight, the agent must determine the optimal level of effort to be invested in the struggle. Naturally, the more cognitive effort invested, the higher the probability of winning the contest.

Similar to Loewenstein (1996), we conceptualize temptation as a System 1 visceral influence on the agent. This visceral factor may be thought of as a drive-state; it acts like a force on the agent, impelling the agent to act in a specific manner. Typical examples of such drive-states include hunger, sex, and pain-relief, but also more complex emotions such as fear, anger, and greed. Temptations vary in strength according to their inherent properties (fresh fruit is more tempting than rotten fruit) and physical and temporal proximity (e.g., thinking about eating chocolate today makes me salivate, while thinking of eating chocolate next year does so to a lesser extent), and to recent exposure, for example due to the mechanisms of satiation and addiction (Loewenstein, 1996; 2000). We assume that the stronger the temptation (i.e., the stronger the force acting on the agent), the larger the effort required by the agent to resist the temptation.

While individuals are equipped with dual processing modes, occasionally yielding conflicting prescriptions for behavior, the two modes need not conflict. In fact, quite often they are in line. The self-control conflict in a given situation arises only when the agent identifies that there is a conflict between her goals and the impulses acting on her. Identification of self-control conflict often is not trivial (Myrseth & Fishbach, 2009). However, our present model only applies to the case where the agent has identified self-control conflict.

3. A model of self-control conflict

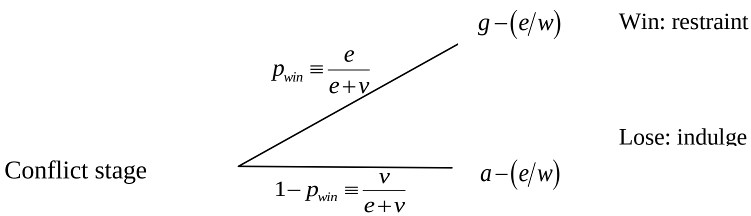
When an individual faces a temptation that conflicts with her better judgment, we say that she experiences self-control conflict. Experimental psychologists have defined this as a conflict between a "higher-order" goal with larger and often delayed benefits and a "lower-order" temptation with more immediate benefits (Schelling, 1984; Metcalfe & Mischel, 1991; Loewenstein, 1996; Trope & Fishbach, 2000; Myrseth & Fishbach, 2009). For example, upon hearing the song of the Sirens, Ulysses was strongly tempted to leap overboard, even knowing that this conflicts with his more important preference for self-preservation. A more mundane case is the dieter who upon seeing a cookie experiences conflict between her sweet tooth and her goal to maintain good health. Similarly, the recovering alcoholic may experience conflict between the urge to drink and the good sense not to. While the literature in psychology and economics offers rigorous theorizing about what individuals do to ensure goal pursuit in anticipation of self-control conflict (e.g., by placing the alarm clock away from the bed), it is rather vague about the self-control conflict itself. In particular, current psychological and economic theory is relatively silent about the joint relationship between key variables, such as payoff from attaining the goal, the strength of temptation, and willpower, which jointly ought to determine success or failure in self-control conflict. Our model specifies a relationship between the aforementioned variables and an individual's likelihood of successfully resisting a temptation.

This section outlines the model of self-control conflict, first by considering the agent's decision problem when facing temptation, second by considering the agent's resulting maximization problem, and third by deriving her reaction function. Finally, we specify the influence on the agent's decision of temptation, conceptualized as a visceral influence (e.g., Loewenstein, 1996; 2000). We thereby establish the relationship between self-control effort, on the one hand, and the payoff from the goal, the strength of temptation, and willpower on the other. This allows us to examine the joint relationship between these factors and the individual's probability of successful goal pursuit and her associated behavior.

3.1. The agent's decision problem.

Figure 1 outlines the agent's decision problem. The baseline model consists only of a conflict stage. We assume two mutually exclusive choice alternatives, g and a. Moreover, we assume that $g > a$, such that the "payoff from the goal" g is larger than the "payoff from the tempting alternative." This assumption defines the domain of the self-control problem, wherein the agent would prefer to choose the action that yields the goal payoff, but, due to visceral influences, she might instead choose the action that yields the tempting alternative. We define the visceral influence as a property of a tempting alternative $a > 0$. The visceral influence acts upon the agent's decision as a force of attraction, pulling the agent towards an inferior payoff a, thus away from the superior payoff *g*. The relationship between the payoff from the tempting alternative and the visceral influence is discussed in Section 3.3.

Figure 1: The agent's decision problem



At the conflict stage, the agent has to determine the level of self-control effort $e \geq 0$ to commit to the conflict with the visceral influence $v > 0$. We let $\frac{e}{w}$ denote the linear cost of self-control effort, where $w > 0$ denotes the agent's exogenous willpower parameter, augmenting the cost of self-control effort.[4] If the agent is successful at the conflict stage, she thus gains $g - (e/w)$. In case she is unsuccessful, she gains $a - (e/w)$.

The outcome at the conflict stage is determined stochastically by a contest-success function. wherein the success probability of the agent is equal to her effort e divided by the sum of the agent's effort and the strength of the visceral influence $(e+v)$ (for more on contest-success functions, see Skaperdas, 1996). The agent's success probability is therefore $p_{win} \equiv e/(e+v)$, and the loss probability is $1 - p_{win} \equiv v/(e+v)$. In the remaining part of this section, we first solve the decision problem, and then we discuss the effect of visceral influence on choice (i.e., the strength of temptation).

3.2. The conflict stage: The agent's maximization problem

We assume that the objective of the agent is to maximize the expected value of the utility function $u(x)$, with the properties $u'(x) > 0, u''(x) = 0$. That is, the agent is risk neutral, and she cares only about her payoff. In the present paper, we restrict attention to the linear case; we postpone consideration of risk preferences to later work, though we expect this to be interesting.

The agent's problem at the conflict stage is to maximize (1) with respect to self-control effort e, subject to the constraint in (2), where (1) is the expected payoff from conflict (denoted (A)); equation (2) states that effort cannot be negative.[5]

---

[4] Due to the non-linearity of the probability, which implies a concave benefit function, this is equivalent to a convex cost function.

[5] A first corner solution is to provide zero effort $e^* = 0$ whenever $v = 0$. That is, there is no visceral influence, so the agent immediately attains payoff g without having to exert any effort. This is simply a situation where there is no self-control problem, as $v = 0$. Due to the functional form of the contest success function, the solution is ill-

$$\max_{e} E\left[u(x)\right] = \frac{e}{e+v}\left(g - \frac{e}{w}\right) + \frac{v}{e+v}\left(a - \frac{e}{w}\right) \equiv \mathrm{A} \quad (1)$$

subject to $e \geq 0$ (2)

Proposition 1 (Optimal self-control effort) *Optimal self-control is given by the reaction function in (3).*

$$e^{*} = -v + \sqrt{wv(g-a)} \quad (3)$$

Proof in Appendix A.

Proposition 2 (Concave optimal effort) *The effort reaction function is concave in the visceral influence.*

Proof in Appendix A.

These results indicate an ambiguous effect on effort by a rise in the visceral influence. Initially, from the point of zero effort, e^{*} rises with increasing visceral influence; the force acting on the agent to choose a increases, and so the agent must exert increasing effort to ensure choice of g. Eventually, when optimal effort reaches its maximum value at $a = (gw - 4v)/w$, the visceral influence has reached such a magnitude that exerting further effort to resist it becomes too costly. Beyond this point, therefore, the agent's effort declines to zero with rising visceral influence. The effort condition (4) defines the space in which the agent exerts effort.

---

defined whenever $e = v = 0$. However, because this case is outside the domain of the self-control problem, we require that $v > 0$ and focus on interior solutions.

This concavity result is due to functional form assumptions. In particular, the cost of effort is linear, and the benefit function is concave. Notably, this is equivalent to a convex cost of effort and a linear benefit function. With a convex cost function, the same result holds whenever the benefit function is not strictly convex. That a benefit function is not strictly convex is a standard assumption in economics. Based on the results presented in an experiment by Shiv and Fedorikhin (1999), where subjects were more likely to indulge when having to memorize a 7-digit number than when not to, Fudenberg and Levine (2006)[6] argue that a convex cost of self-control adequately represents the psychological evidence that self-control indeed is a limited resource (e.g. Baumeister, 2000).[7] From these two propositions, we derive some corollary results.

Corollary 1 (Effort condition) *The agent exerts effort only if willpower is larger than the ratio of the visceral influence and the difference in payoffs from the goal and the tempting alternative.*

Proof Setting $e^* > 0$ and solving for w reveals that the agent only exerts effort whenever

$$w > v/(g - a) \tag{4}$$

Corollary 2 (Comparative statics) *Effort increases in payoff from the goal and willpower, but falls in the payoff from the tempting alternative*

Proof in Appendix A.

The interpretation of condition (4) is that an agent with a sufficiently low willpower would prefer to yield to temptation immediately and gain the payoff a, rather than exert any costly effort. In

[6] The authors augment their base line model to account for cognitive load. Our modeling approach is equivalent to their "assumption 5 (cost of self-control with cognitive load)", in section V.

[7] Such "cognitive load" makes it harder to resist temptation as cognitive resources are burdened.

other words, with a sufficiently low willpower parameter, effort would be "too costly" given the strength of the visceral influence and the expected gain from conflict.

Corollary 2 states that higher willpower w always implies increased effort. This is because the exertion of effort, in effect, becomes cheaper as w rises. Higher payoff from the goal g also increases the level of effort. This is because the expected value from conflict rises. In contrast, higher payoff from the tempting alternative a reduces effort.

As far as the agent's decision is concerned, the critical factor is the difference between the two payoffs. As the distance increases, choosing to attempt resisting the temptation becomes more worthwhile, leading to more effort. We next consider the behavioral implications of these parameters.

Proposition 3 (Behavioral implications) *Increases in the payoff from the goal and increases in willpower increase the probability of success, while increases in the payoff from the tempting alternative and the visceral influence reduce it. Choice probabilities are thus monotonic in payoffs.*

Proof in Appendix A.

Willpower and payoff from the goal , via their effects on effort, both have a positive effect on the probability of success. This is because the success probability increases in agent effort. Conversely, increasing the payoff from the tempting alternative decreases effort, thereby reducing the probability of success. This is because losing at the conflict stage becomes less costly as a approaches g. Moreover, just as agent effort increases the probability of success, the visceral influence decreases it. This is the subject of the next section.

3.3. Visceral influence on choice

Thus far we have not examined why an individual might act against her better judgment and choose to indulge in a when the payoff from the goal g is larger. This section discusses the visceral influence on choice, relating the visceral influence v to the payoff from the tempting alternative a. To capture visceral influence on choice by the temptation, we assume the simplest possible, though not far-fetched, functional form:

$$v = a \tag{5}$$

This simply means that the payoff from the temptation is identical to the degree to which it influences the agent. For example, the more alluring the cookie, the more pleasure the agent would derive from consuming it. While we believe that this assumption indeed is plausible, it certainly is debatable. It is possible that there exists a non-linear relationship between the visceral influence (e.g., a craving for drugs) and the actual value realized upon consumption of the temptation (e.g., the consumption of drugs). Exploring our model for other functional forms of $v(a)$ is indeed worthwhile, but we postpone that to later work.

Using the visceral influence function (5) in (3) yields the augmented reaction function:

$$e^* = -a + \sqrt{w\left(ga - a^2\right)} \quad . \tag{6}$$

In what follows we will rely on the effort function in (6). While slightly simplifying analysis, none of our qualitative results depend on the assumption (5).

3.4. A numerical example

To illustrate the model, we use an arbitrary numerical example. The results are general, and they may be derived for any parameter values as long as $g - a > 0$ holds. We assume that $g = 2$ and $w = 1$, and let a vary, starting at 0. The reaction function becomes:

$$e^* = -a + \sqrt{\left(2a - a^2\right)} \quad . \tag{7}$$

Figure 2: Optimal effort, expected payoff from conflict given optimal effort, payoff from the tempting alternative, and payoff from the goal vs. the visceral factor. *g* = 2, *w* = 1.

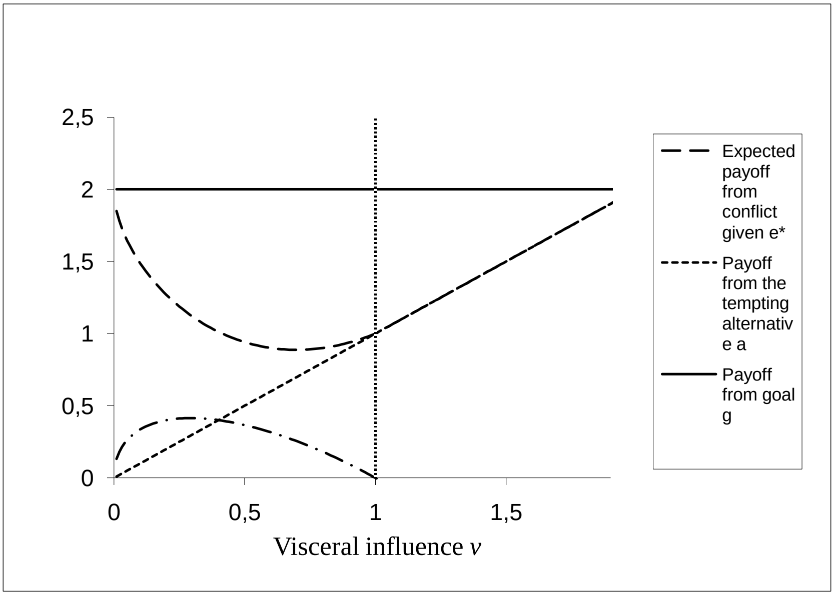


Figure 2 displays the expected payoff from conflict A given optimal effort, the payoff from the tempting alternative, and the payoff from the goal as a function of the visceral influence and the effort condition from (4). The agent chooses to enter the conflict only when the expected payoff from the conflict is higher than the payoff from the tempting alternative a. This is illustrated by the area to the left of the effort condition.

First, we note that whenever the visceral factor is zero, the agent's effort also is zero, yielding payoff g since the cost of effort at this point is zero.[8] Once the payoff from the tempting alternative starts rising, however, effort sharply increases towards the maximum. This maximum exists because the payoff from the tempting alternative rises with the visceral influence. The agent is then required to devote more effort to the struggle, and does so as the expected payoff from conflict is rising. Beyond the maximum, the agent's effort starts to fall; chances of success fall faster than the expected payoff increases, and devoting more effort then becomes too costly.

[8] Since visceral influence is zero, there is no force acting on the agent and so "standard" decision making conditions apply.

Moreover, the opportunity cost of losing becomes smaller and smaller, reducing effort (Corollary 2).

The agent's effort drops to zero at the point where the expected value of conflict is equal to the payoff from the tempting alternative. Beyond this point, she will prefer to avoid the struggle altogether and go immediately for the payoff a. That is, for visceral influences higher than that indicated by the entrance condition, the agent thus refrains from entering the conflict stage altogether; the payoff rises in a, toward and beyond the payoff from the goal g (Corollary 1).

As illustrated in Figure 2, our model predicts that the agent does not simply increase her effort at the conflict stage as the temptation grows stronger (Proposition 3). Rather, she strategically allocates her effort to the struggle according to the reaction function, which is shaped like an inverted U. Although at low levels a stronger visceral influence leads the agent to exert more effort to resist it, at higher levels she reduces her effort since it becomes too costly. That is, an individual might exert little effort to restrain herself when facing a strong temptation, not because she is weak, but because the temptation is too strong.

4. Two forms of sophistication: Pre-commitment, naïveté and overconfidence[9]

This section extends the basic decision problem to include a possibility for pre-commitment to pre-empt self-control conflict. We also allow for a varying degree of agent sophistication, in the spirit of O'Donaghue & Rabin (1999). In their model, the agent's sophistication is defined as the agent's ability to correctly anticipate her self-control problem, i.e. the true value of β in their model. Exploring the self-control domain of procrastination, they assume a framework where the agent at some point in time has to perform a specific task, but that she will always prefer to delay this task. Sophisticated agents realize that if they wait until the next period, they will wait for yet another period and so on, thereby suffering a welfare loss. Anticipating such procrastination, they perform the task too soon. Naive agents, on the other

[9] Mention the possibility that people overestimate future visceral influences (probably true for some). Is there any evidence for your assumptions?

hand, do not realize the extent of their self-control problem, and wait until the last minute to complete the task, but also suffer a welfare loss.

O'Donaghue & Rabin (1999) make no distinction between the tendency to overestimate one's ability to control visceral influences, reflecting overconfidence about ones willpower, on the one hand, and on the other, the tendency to underestimate the impact of visceral influences on behavior. We introduce the distinction into our framework, and we explore which is more detrimental for welfare losses. That is, we define two forms of sophistication: the ability to correctly anticipate the impact of visceral influences on one's behavior, which we refer to as "degree of sophistication", and "overconfidence" in one's willpower.

Accordingly, we augment the agent's decision problem from Section III in three ways, as outlined in Figure 3. The first extension concerns the pre-commitment stage, where the agent has the possibility to choose a technology that guarantees successful goal pursuit, albeit at a cost $c \geq 0$. This cost of pre-commitment is sometimes trivial, e.g., not bringing high calorie foods into one's house. Other times it is quite significant, e.g., buying smaller packs of cigarettes for fear of overconsumption, using non-interest bearing savings accounts due to fear of overspending, or attending dieting clinics that promise not to feed their clients (see Thaler and Shefrin, 1981). [10]

The second extension concerns the perception of the conflict stage at the pre-commitment stage. We allow for the agent's degree of sophistication $s \in [0,1]$ in correctly anticipating the strength of the visceral influence experienced at the conflict stage. The agent's expected visceral influence is thus $E[v] = sa$, rather than the true visceral influence $v = a$. When the degree of sophistication equals one, the agent has full appreciation of the strength of the visceral influence and may take appropriate measures of pre-commitment. As the degree of sophistication approaches zero, the agent's naiveté with respect to the visceral influence is complete, and the perception of conflict is more favorable to entry as $E[v] = 0$.[11] Thus, the expected visceral influence can be stated as $E[v] \in [0, a]$.

---

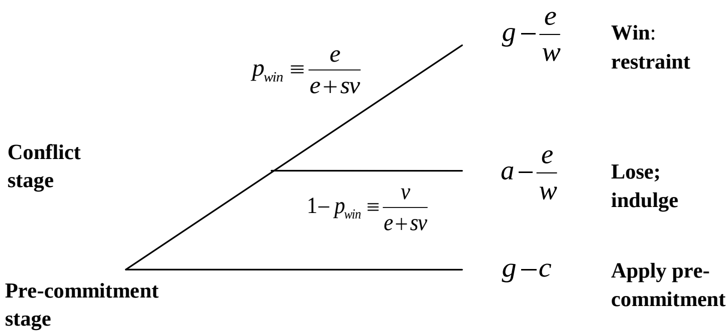
[10] Cost is sometimes significant as in the case of dieters purposefully consuming worm eggs, the spawns of which are intended to consume part of of the host's ingested food..

[11] In our analysis we do not allow for the individual to be completely naïve since the solution to the problem is ill-defined for $s = 0$. It is, however, quite intuitive that if the agent is completely naïve in this sense, she does not

Third, we introduce a parameter $\beta \geq 1$ to measure the agent's overconfidence in willpower. The agent's perceived willpower is thus βw. When the agent's has no overconfidence, the expected willpower is equal to the actual willpower. As overconfidence approaches infinity, the expected willpower also approaches infinity. We summarize this as $E[w] \in [w, \infty)$.

In formal terms, the naïve expected payoff from conflict, denoted N, is an element of the interval from the actual expected payoff from conflict, denoted A, and the payoff from the goal g. We state this as $\mathrm{N} \in \left[\mathrm{A}, g\right]$. The lower the degree of sophistication s, or the higher is overconfidence, the higher is N. Moreover, when $s = \beta = 1$, the agent's sophistication is complete, and hence $\mathrm{N} = \mathrm{A}$.

Figure 3: The agent's decision problem with pre-commitment



The decision problem in Figure 3 includes all three augmentations. At this point, it is necessary to distinguish between what the agent at the pre-commitment stage perceives about the conflict stage and how this perception changes when she finds herself at the entrance stage.

---

perceive of any conflict at all. From this point of view, the problem is simply one of choosing between g and a, without any apparent problems.

At the pre-commitment stage, the agent's sophistication affects her perception of her optimal effort at the conflict stage. Since her degree of sophistication influences her perception of the visceral influence and her overconfidence influences her perceived stock of willpower, her expected optimal effort is

$$e_s^* = -sa + \sqrt{\beta ws\left(ga - a^2\right)} \tag{8}$$

We denote the perceived success probability $p_{win}^s \equiv e_s^* / (e_s^* + sv)$. When s is close to zero, the agent is close to completely naïve. She believes that she will succeed at the conflict stage with near probability one since p_{win}^s approaches one as s approaches zero.[12] In contrast, when $s = 1$, the agent has perfect sophistication. Her expected probability of success is therefore $p_{win}^s \equiv e/(e+v) = p_{win}$, as in Figure 2. The perceived success probability is thus always larger than or equal to the true success probability, and it has a maximum value of one, such that $p_{win}^s \in \left[p_{win}, 1\right]$.

We make the additional assumption that the agent learns the true strength of the visceral influence once she has foregone pre-commitment. That is, the true visceral influence is revealed as $v = a$ when she faces the tempting alternative a. Similarly, the agent becomes aware of her true willpower parameter when she engages in self-control conflict. Because the decision problem from Figure 2 is then contained within Figure 3, we can retain all the results from the previous analysis. What we lack is the naïve expected payoff from conflict N as perceived at the pre-commitment stage.

The agent's maximization problem extends in a straightforward way to include the degree of sophistication by substituting in the perceived success probability p_{win}^s. The perceived maximization problem if entering thus becomes:

[12] s cannot be equal to zero as the solution to the maximization problem at the conflict stage is ill-defined when e+v=0, or as in this case (0)a=0 (as effort would be zero).

$$\max_e u(x) = \frac{e}{e+sv}\left(g - \frac{e}{\beta w}\right) + \frac{sv}{e+sv}\left(a - \frac{e}{\beta w}\right) \equiv \mathrm{N} \ , \tag{9}$$

$$\text{subject to } e \geq 0 \ . \tag{10}$$

Maximization of the payoff from conflict with respect to agent effort e, using the visceral influence function, yields the slightly modified reaction function (11):

$$e_s^* = -sa + \sqrt{s\beta w\left(ga - a^2\right)} \quad , \tag{11}$$

where e_s^* now is the perceived optimal effort depending on the level of sophistication, which takes values from zero, and $\bar{e}_s^*$ such that $e_s^* \in \left[0, \bar{e}_s^*\right]$ where $\bar{e}_s^*$ is the maximum perceived effort.

For completeness, we consider the effect on perceived optimal effort of a change in sophistication s and overconfidence β.

Corollary 3 (Self-control effort and naiveté) *Changing the degree of naiveté has an ambiguous effect on perceived optimal effort.*

Proof The derivate of effort with respect to degree of sophistication is

$$\partial e_s^* / \partial s = -a + \frac{1}{2s} > 0 \text{ if } \frac{1}{2a} > s \tag{12}$$

Corollary 4 (Self-control effort and overconfidence) *Increasing overconfidence has a positive effect on perceived optimal effort.*

Proof in Appendix A.

Corrollary 3 follows from the fact that optimal effort is concave in the visceral influence (proposition 2). Varying the perception of the payoff of the tempting alternative causes the same response as varying visceral influence itself. Initially, when sophistication increases, so does the expected payoff from conflict. Eventually, however, the cost of fighting looms too large, and hence the optimal effort devoted to conflict declines to zero. Corollary 4 is driven by the same mechanism as increasing willpower itself. Increasing overconfidence leads to a lower perceived self-control cost, and so the agent expects to exert more effort.

Given that self-control effort is costly, facing (and not pre-empting) the self-control problem is costly too. This gives rise to the possibility that the agent would be willing to pay for a pre-commitment mechanism depending on its relative cost. Because both lack of sophistication and oinfluences perception of the visceral influence and overconfidence influences perceived stock of willpower, it is possible that the agent's willingness to pay for pre-commitment might be inordinately low. Consequently, the agent might choose to forego pre-commitment in cases where she would have benefited from it. We explore willingness to pay for pre-commitment and the resulting welfare effects in the next section.

5. Welfare and willingness to pay for pre-commitment[13]

Our conceptualization of the self-control problem as a maximization problem under a visceral influence proves particularly advantageous for examining welfare effects; we assume that the payoff accrues to one agent rather than by distribution to multiple selves. A self-control conflict always is costly for the individual since she can never guarantee successful goal pursuit; there is always a chance that the influence of the temptation may compel a suboptimal choice. In many cases, the agent has opportunity to employ a pre-commitment technology to ameliorate welfare loss. The application of such a technology, however, is not unproblematic. In this section, we make an attempt to delineate which welfare effects arise due to lack of sophistication and which arise due to overconfidence.

[13] Note that welfare implications depend crucially on non-testable details of your model (c.f. Bernheim, 2009 JEEA)

The agent only is willing to pay for pre-commitment if it will lead to a higher payoff than expected from engaging in conflict. Because this implies that the cost of pre-commitment cannot be too large, we may write:

$$g - c > \mathrm{N}. \tag{13}$$

As long as (13) holds, the agent prefers pre-commitment. Using (13) we can conclude that the agent's willingness to pay, wtp, may be written as:

$$wtp = g - \mathrm{N} \tag{14}$$

In cases where the naïve expected payoff from conflict is larger than the payoff from pre-commitment (15), the agent will prefer to enter and consequently suffer a welfare loss.

$$\mathrm{N} > g - c \tag{15}$$

This welfare loss is then the difference between the payoff from pre-commitment and the payoff from actual conflict.[14] We may thus write:

$$\omega = (g - c) - \mathrm{A}, \tag{16}$$

where ω is the welfare loss. The severity of the welfare loss will depend on the difference between the two expected payoffs, $\mathrm{N} - \mathrm{A}$. This confirms the intuition that welfare losses resulting from naïveté are less severe for individuals with stronger willpower. At this stage, we explore when the agent will and will not forego pre-commitment. Variation in all parameters, however, can lead to welfare losses under naïveté, which we summarize in the following proposition.

[14] Note that this welfare loss is distinct from the welfare loss suffered due to the presence of the self-control problem; the latter simply is $g - \mathrm{A}$.

Proposition 2 (welfare effects: under estimating visceral influences) *For* $\beta = 1$, *increases in all parameters,* w, g, a *and* c *can lead to welfare loss when* $s < 1$

Proof We prove this result by constructing examples. Each example contains the actual and naïve payoffs from conflict, i.e., what the agent thinks she will get from foregoing pre-commitment (N) and what she actually gets (A). In addition, the figures display the payoff from pre-commitment $g - c$ and welfare loss ω, in case the agent chooses to play no pre-commitment. Of course, this welfare loss is only realized whenever the naïve expected payoff is larger than the payoff from pre-commitment, such that $\mathrm{N} > g - c$ holds. We will next look at the welfare effects from varying the agent's willpower w; the cost of pre-commitment c; the payoff from the goal g; and the payoff from the tempting alternative a .

We start by assuming that the agent has no overconfidence, i.e. $\beta = 1$. The condition for foregoing pre-commitment is stated in (17) below.

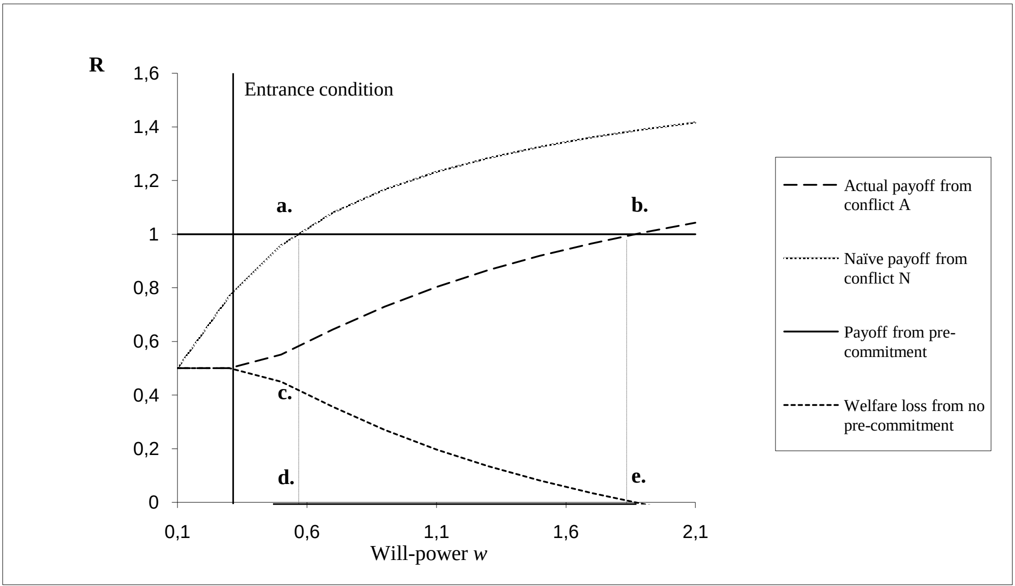
$$\mathrm{N} \geq g - c > \mathrm{A} \tag{17}$$

Using the expression for the naïve expected payoff from conflict and optimal effort, the condition for foregoing pre-commitment is stated in (18). We use this expression for calculations in the following examples.

$$\mathrm{N} \equiv \frac{e_s^*}{e_s^* + sa}(g - a) + a - \frac{e_s^*}{w} > g - c \tag{18}$$
$$e_s^* = -sa + \sqrt{ws\left(ga - a^2\right)}$$

Example 1 (Lack of sophistication and willpower) Assume that $g = 4$, $a = 2$, $c = 1$ and $s = .5$. Then the agent suffers a welfare loss when willpower is in the range $.55 \leq w \leq 1.9$ Hence, the agent would in some cases be better off with lower willpower (Figure 4).

Figure 4: Payoff from pre-commitment, perceived and actual conflict and welfare loss from no pre-commitment vs. willpower. *g* = 2, *v* = .5, *s* = .3, *c* = 1

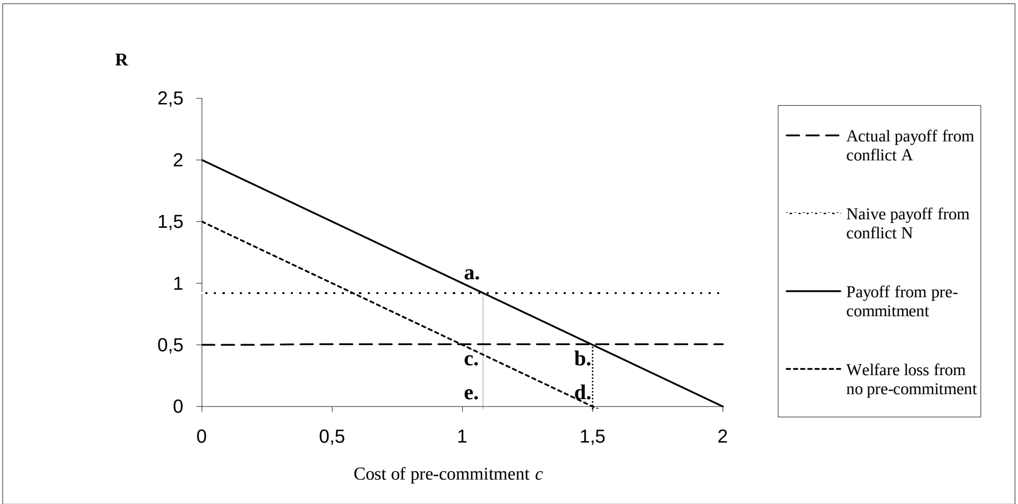


At first glance, it may appear counterintuitive that increasing the agent's willpower could lead to welfare losses. This happens when the agent foregoes the payoff from pre-commitment $g-c=1$ in favor of the naïve payoff from conflict $\mathrm{N}>1$, which holds at point **a** in the figure. The welfare loss is then given by ω in (31) up to the point where $\mathrm{A}\geq g-c=1$ (**b** in the figure); beyond this point actual payoff from conflict is greater than or equal to the payoff from pre-commitment. Welfare loss is thus given by the area in the figure enclosed by **c**, **d**, and **e**.

The same analysis holds for any $s<1$ and $c>0$. The reason is that lack of sophistication drives a wedge between the naïve and actual payoff from conflict, and increasing willpower drives the naïve payoff from conflict above the payoff from pre-commitment. The condition $s<1$ ensures that the agent is not fully sophisticated and $c>0$ that pre-commitment is not costless. Were pre-commitment costless, pre-commitment always would be preferred.

Example 2 (Lack of sophistication and cost of pre-commitment) Assume $w = .3$, $a = .5$, $g = 2$ and $s = .3$. Then the agent suffers a welfare loss when the cost of pre-commitment is in the range $1.2 \leq c \leq 1.5$. Hence, the agent in some circumstances will be better off with a higher cost of pre-commitment (Figure 5).

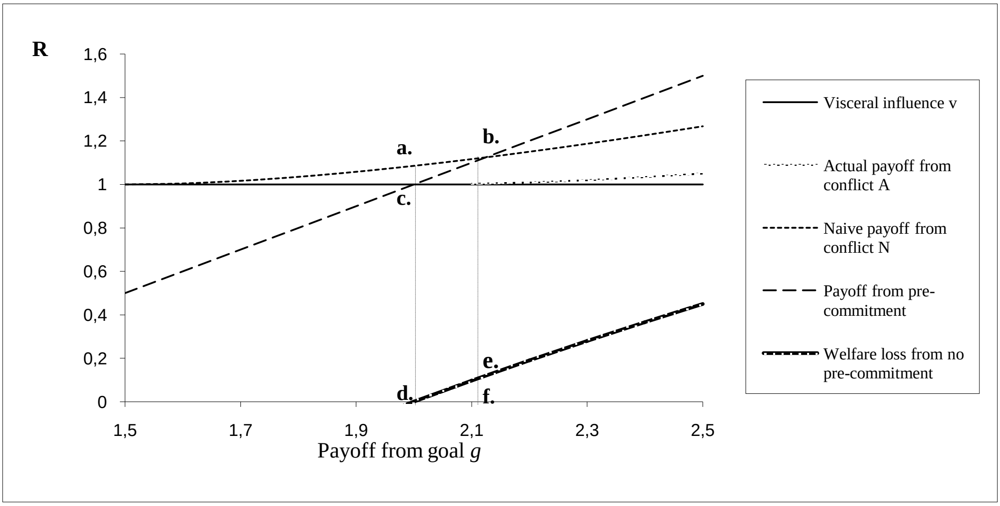
Figure 5: Payoff from pre-commitment, perceived, and actual payoff; and welfare loss from no pre-commitment vs. cost of pre-commitment. *g* = 2, *v* = .5, *w* = .3, *s* = .3



Again, although the simulation values are arbitrary, the results require very few restrictions to hold, specifically that $g - c < \mathrm{N}$. The agent then suffers a welfare loss by attaining the payoff $\mathrm{A} = .5 \leq g - c$ up to the point where the actual payoff from conflict is larger than that from pre-commitment $\mathrm{A} \geq g - c$. Welfare losses are indicated in Figure 5 by the area enclosed by **c**, **d**, and **e**.

Example 3 (lack of sophistication and payoff from the goal) Assume that $w=1$, $a=2$, $c=1$ and $s=.5$. Calculations show that an agent with lack of sophistication mistakenly will forego pre-commitment and suffer welfare loss when the payoff from the goal is in the approximate range of 2 and 2.1.

Figure 6: Perceived and actual payoff from conflict and welfare loss from no pre-commitment vs. payoff from goal. v = 1, c = 1, s =.5, w = 1.



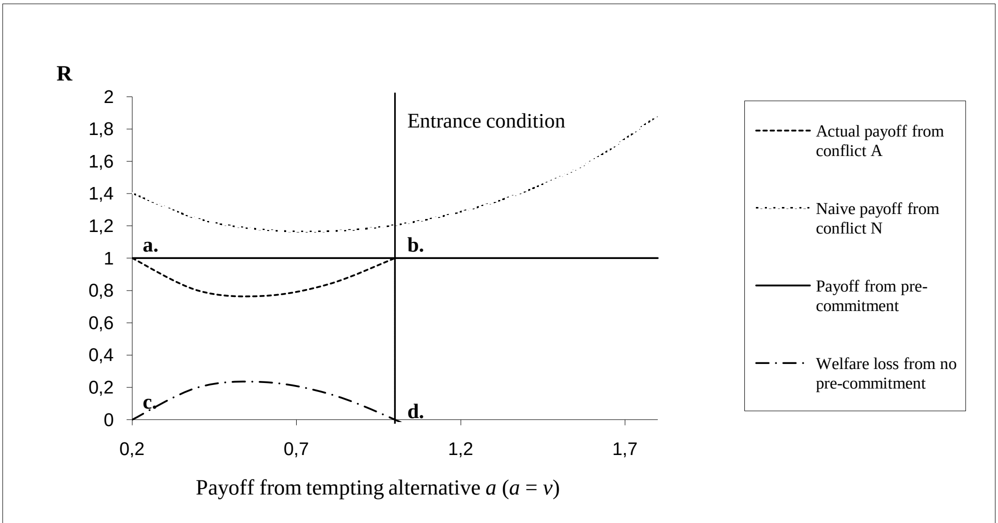
That increasing the payoff from the goal could prove harmful from a welfare point of view is quite interesting. The mechanism, however, turns out to be similar to that of increasing willpower. The results hold whenever $\mathrm{N} > g-c > \mathrm{A}$. This condition shows that welfare losses occur when the agent foregoes pre-commitment in favor of the naïve payoff from conflict. This happens when the payoff from pre-commitment $g-c \leq \mathrm{N}$, which is approximately when $g=2$. The welfare loss is then given by ω in (16) up to the point where $\mathrm{A} \geq g-c$, which is approximately when $g=2.1$; from there on pre-commitment is preferred to the naïve payoff from conflict $g-c > \mathrm{N}$. The welfare loss, then, is given by the area enclosed by **d**, **e**, and **f.** (Figure 6).

Lack of sophistication drives a wedge between the naïve and actual payoff from conflict, and increasing payoff from the goal drives the naïve payoff from conflict above the payoff from pre-commitment.

Example 4 (Lack of sophistication and payoff from the tempting alternative) Assume $w=1$, $a=.5$, $g=2$ and $s=.3$.

With naïveté, increasing the payoff from the tempting alternative increases the naïve payoff from conflict more than it increases the value of the tempting alternative itself; the agent does not fully appreciate the full extent of the visceral influence at the conflict stage. To examine these effects, we assume parameter values g = 2, c = 1, s =.3, and w = 1. The effects from this analysis hold as long as $g-c>\mathrm{A}$ and $\mathrm{N}>g-c$. The latter holds if $s<1$ and $c>0$. The welfare loss from lack of sophistication is therefore given by the area between **c** and **d**, as seen in Figure 7.

Figure 7: Perceived and actual payoff from conflict and welfare loss from no pre-commitment vs. payoff from tempting alternative. *g* = 2, *c* = 1, *s* =.3, *w* = 1.



When the visceral influence (the payoff from the tempting alterative) is in the range of $0.2 \le v \le 1$, the agent will prefer to forego pre-commitment since $\mathrm{N} > g - c$. In fact, this is always the case when $v > 1$, as well. However, it is not associated with a welfare loss, since the actual payoff from conflict is also larger than the pre-commitment payoff. Specifically, $\mathrm{A} > g - c$ implies that choosing to face conflict is not an incorrect decision, although the agent will receive a smaller payoff than expected.

Proposition 3 (welfare effects: overconfidence) *Given $s = 1$, the agent will prefer the naïve expected payoff from conflict rather than the strictly larger payoff from pre-commitment whenever she is "too" overconfident*

Proof Consider the following condition $g - c = \mathrm{N} > \mathrm{A}$, $s = 1$, but where β is such that the agent is just indifferent between purchasing pre-commitment and not. Because $g - c > \mathrm{A}$, foregoing pre-commitment will yield welfare loss. Using $g - c = \mathrm{N} > \mathrm{A}$, we derive a threshold level $\overline{\beta}$ of overconfidence beyond which the agent is "too" overconfident in the sense that she will prefer conflict instead of pre-commitment when she ought not to. This expression is displayed in (19) below.

$$\beta \ge \overline{\beta} \equiv \frac{a^2}{w\left[(g-a)^2 + a^2 + \frac{a^2}{w^2} - (g-c)^2\right]}. \tag{19}$$

From a welfare point of view, overconfidence is not a problem unless the agent is sufficiently overconfident to forego pre-commitment when she ought not to. Moreover, the level of overconfidence that becomes detrimental is jointly determined by the parameters in the model. Having derived the threshold, we show that increasing any of the aforementioned parameters allows for $\mathrm{N} > g - c > \mathrm{A}$ to hold.

Corollary 5 (Overconfidence and willpower) *Increasing willpower leads to an increase in the threshold* $\overline{\beta}$ *if willpower is sufficiently low, such that* $a > w$ *. Otherwise it leads to a decrease.*

Proof in Appendix A.

Corollary 5 reveals that the welfare loss resulting from overconfidence is more severe when willpower is high. Only for a small range of parameter values, i.e. $a > w$, is the effect on the threshold positive. Whenever willpower exceeds this level, the agent mistakenly will forego pre-commitment. Notably, however, if willpower is sufficiently high, the likelihood of being successful at the conflict stage approaches one; beyond a certain level of willpower, overconfidence does not matter for payoffs. The next result reveals a similar effect from the payoff from the goal

Corollary 6 (Overconfidence and payoff from the goal) *Increasing payoff from the goal leads to an increase in the threshold whenever* $a > c$ *; otherwise it leads to a decrease.*

Proof in Appendix A.

This result reveals that increasing the payoff from the goal has an ambiguous effect on the overconfidence threshold. This is because increases in g increase both the expected payoff from conflict as well as the payoff from pre-commitment. While increasing the expected payoff from conflict decreases the threshold as conflict becomes more favorable, increasing the payoff from pre-commitment increases the threshold. The overall effect on the overconfidence threshold thus depends on the relative strength of these two effects. The point where the payoff from pre-commitment dominates this effect is where $a > c$.

Corollary 7 (Overconfidence and payoff from the tempting alternative) *Increasing payoff from the tempting alternative always increases the threshold.*

Proof in Appendix A.

Corollary 8 (Overconfidence and cost of pre-commitment) *Increasing cost of pre-commitment always lowers the threshold level.*

Proof in Appendix A.

Corollary 7 reveals that increasing the payoff from the tempting alternative has an unambiguous effect on the threshold. While this may seem intuitive, it may also not be as clear-cut as it appears at first glance. Increasing the payoff from the tempting alternative influences the threshold by two distinct mechanisms. On one hand, it increases the naïve expected payoff from conflict N, which reduces the threshold. On the other hand, it increases the strength of temptation, making conflict less favorable, thus increasing the threshold. Our result confirms that the effect of increasing temptation dominates that from increasing expected payoff. Hence, the net effect on the threshold is positive.

In contrast, the result in Corollary 8is intuitive; increasing the cost of pre-commitment reduces the payoff from pre-commitment relative to that from conflict. Hence, the effect on the threshold should be negative.

Proposition 4 (Overconfidence and lack of sophistication: optimal effort) *Changes in optimal effort always increases faster with increases in overconfidence than with decreases in lack of sophistication, according to the condition*

$$\beta\left(a^2+\left[w\left(ga-a^2\right)\right]^2\right)\Big/\left[w\left(ga-a^2\right)\right]^2=\underline{s}>s$$

Proof in Appendix A.

This result reveals an important insight about the different between overconfidence in willpower and underestimation of visceral influences. From the above condition, the threshold $\underline{s}$ depends on all parameters, and the comparative statics are both straightforward and instructive. Increases in all parameters raise the threshold since the numerator always is larger than the denominator given that $\beta \geq 1$. Furthermore, since the threshold $\underline{s}$ never can be less than one, and one by definition is the maximum value of s, this condition always holds. Consequently, overconfidence always has a larger effect on anticipated optimal effort than does lack of sophistication. We show with the next proposition that this result has important implications for welfare loss following from foregoing pre-commitment.

Proposition 5 (Overconfidence and lack of sophistication: willpower) *Naïve expected payoff from conflict increases faster with overconfidence than with lack of sophistication as long as willpower is large enough such that*

$$w \geq \frac{a^2 (1-s)^2}{(\beta - s)(ga - a^2)} \tag{20}$$

Proof in Appendix A.

This result illustrates an interesting point about overconfidence in general. Overconfidence is more detrimental to optimal decision-making is than lack of sophistication when the agent has a high stock of willpower. With a high stock of willpower, even minimal overconfidence leads to inferior choices. However, the opposite is true for individuals with low stock of willpower. For knowingly "weak" individuals, overconfidence must be substantial before it leads to inferior choices.

Corollary 9 (Overconfidence and sophistication: willpower threshold) *The willpower threshold falls when the difference between the overconfidence and sophistication parameters increases.* $w \geq \frac{a^2(1-s)^2}{(\beta-s)(ga-a^2)}$

Proof As the denominator of (19) increases the expression on the right hand side decreases.[15]

This result is interesting since it illustrates the substitutability of overconfidence and lack of sophistication. As long as the difference $\beta - s$ remains constant, so does the threshold. In this way, being slightly overconfident can be compensated for by an increase in sophistication. In our model, this result is only possible to maintain over a certain range since β is not bounded above, while s is.

6. General Discussion

We have proposed a new model of self-control conflict that is grounded in psychological theory. The model captures the interrelationship between key psychological variables involved in self-control conflict: the strength of temptation, the payoff from the conflicting goal, and the stock of willpower. We conceptualize self-control conflict as a struggle of an agent, who prefers goal pursuit, against the visceral influence of temptation, which impels behavior that is inconsistent with goal pursuit.

Our model contributes to the self-control literature in two respects. First, we have attempted to construct the model bottom-up, using insights into the psychological mechanisms of the self-control conflict. Our model attempts not only to account for behavioral outcomes observed in the field and in the lab; it also tries to do so by integrating previously disparate strains of psychological theory into a more comprehensive framework, reflecting our joint understanding of the psychological mechanics of self-control conflict. Second, our model

[15] As before, we cannot allow $s = 0$.

provides novel predictions about behavior not only in the midst of self-control conflict, but also in anticipation thereof. For example, the model predicts that an agent with limited cognitive resources, the expenditure of which is costly, will not necessarily monotonically increase effort to resist temptation in response to stronger temptation. Rather, her reaction function will slope upwards at first, but have a negative second derivative, implying that she eventually will reach a point of maximum effort in response to stronger temptation, beyond which effort will be reduced to the point where the temptation is sufficiently strong to offset any expected benefits from exerting effort, which implies that the agent will not resist at all. While our model emphasizes the conflict between an agent and a conflicting visceral influence, earlier models emphasized a game between multiple selves to explain and explore the role of pre-commitment in decision making.

Schelling (1978, 1984) discusses the dichotomy between what one wants ex-ante and what one wants ex-post. This is reminiscent of the hyperbolic discounting literature where the decision maker may prefer $50 today to $100 in a year while simultaneously preferring $100 in six years to $50 in five, even though these decisions are normatively identical. Schelling discusses the different strategies one might undertake to "game" oneself ex-ante. For example, a woman about to give birth might request that anesthesia be made unavailable during delivery if she knows she will use it if it is available. Our approach does not speak to the explicit nature of these strategies, but it allows for pre-commitment. Moreover, it can account for the apparent switch in preferences that have been documented since Strotz (1956) and that now fall under the rubric of hyperbolic discounting. Laibson (1997) studied the class of discount functions that lead to time-inconsistent preferences, in particular the now common β, δ functional form, which is steeper than the conventional exponential discount function that is standard in economics. Only minor adjustments are required for our framework to account for behavior consistent with hyperbolic discounting. Specifically, we augment the visceral influence function by adding time as an argument. The function then becomes $v = v(a,t)$. Adding the property that $\delta v / \delta t < 0$, the visceral influence decreases in time. This means that the visceral influence is stronger for decisions closer in time and is therefore associated with more indulgence. This assumption is well-grounded in the psychological literature (e.g., see Loewenstein, 1996).

Although our framework can account for time-inconsistent behavior for a single agent, the papers closest to our work view the decision maker as consisting of multiple selves with competing preferences. Thaler and Shefrin (1981) pioneer such modeling efforts in the economics literature by assuming a farsighted "planner" and a sequence of shortsighted "doers." They assume that the planner attains utility both from the present and future periods. The doer, however, lives only for one period and thus only gets utility from the present. While the planner would maximize her utility by exercising restraint in the present in favor of the future, the doer would rather consume all in the present. This gives rise to an intra-personal conflict between the two selves. The planner then attempts to manipulate the doer in a variety of ways. One way is to use pre-commitment strategies, as exemplified by the dieter not bringing cookies into the home. Another is to change the doer's preferences, for example by consuming the Antabuse drug to avoid alcohol consumption (Antabuse and alcohol in combination cause sickness). Thaler and Shefrin (1981) aim to rationalize pre-commitment, Bénabou and Tirole (2006) develop a theory of "internal commitment" or "personal rules," providing a mechanism for exercising restraint when facing temptation based on imperfect recall of past motives and feelings. While we do not explicitly account for such a cognitive mechanism, it could be incorporated into our framework by modifying the evolution of cost of self-control in a dynamic framework.

The model closest to ours is that of Fudenberg and Levine (2006). They present a generalized model of Thaler and Shefrin (1981). Their model assumes a planner who gains utility from the discounted sum of all doers' utility. Although our conceptual starting point is slightly different from theirs, the two models are closely connected. Their specification of self-control cost is "opportunity based" In a state where the short-run self has a feasible action that is more valuable than some other feasible action in another state, cost of self-control is higher. While our cost of self-control does not directly depend on preferences by the short-run self, the two approaches are similar; a higher visceral influence leads to a higher cost of self control, which can be interpreted as a short-run utility. Their preferred specification of self-control cost is convex, accounting for cognitive load. In our model, costs are linear, but the benefit function is concave due to the contest-success function in the maximization problem. Because this is equivalent to a convex cost, the two models share this feature as well as the resulting unique equilibrium. For this reason, neither model satisfies the axioms of the Gul and Pesendorfer (2001, 2004) framework.[16]

A reduced form version of the Fudenberg and Levine (2006) model becomes a simple maximization problem just as in our model. Their model is slightly more general in this respect since they account for both deterministic and stochastic decisions, while ours is stochastic as long as effort is positive, after which it takes on a deterministic nature by always resulting in indulgence. Furthermore, while they allow for a larger action set, we only account for two actions. Our model can, however, be modified to include more actions by assuming that the goal payoff is chosen from a larger action set. In such a version, our framework would imply the refinement introduced by Fudenberg and Levine (2006), where the short run self always best responds. In our framework, the payoff from the tempting alternative is fixed and the conflict condition $g-a>0$ is positive, which is similar to a subsystem best responding. In their model, increasing the payoff from the most preferred alternative has no consequence; it does not affect the marginal cost of self-control. [17] In contrast, an increase in g in our model implies a reduction in the cost of self-control. Our model's short-run self would not be indifferent between options only affecting future states; the preferences of our short-run self would instead depend on the temporal distance of the realized payoffs.

Fudenberg and Levine (2006) focus their applications on "sophisticated" agents, who correctly estimate the self-control cost parameter. Our focus is on agents who either underestimate visceral influences or overestimate their willpower parameter. Moreover, our model yields predictions regarding naïve agents that cannot be derived using either the framework of Fudenberg and Levine (2006) or that of Gul and Pesendorfer (2001, 2004). Fudenberg and Levine (2006) assume that the planner is farsighted and has rational expectations, and our framework is easily modified to account for accuracy of such expectations, as in the spirit of O'Donaghue and Rabin (1999, 2001).

[16] Fudenberg and Levine (2006) show how convex costs imply violation of the set-betweeness axiom. See their section VI.

[17] Fudenberg and Levine (2006) show that this violates the independence axiom. Like the authors, we feel that the independence axiom should be relaxed.

Moreover, our model also extends Thaler & Shefrin's (1981) planner-doer framework by allowing a more detailed analysis of an agent (in their terminology, the "planner"), who is anticipating a future encounter with temptation. The agent knows that she can resist some temptations, but not all, and certainly not all temptations all the time. Ulysses, who anticipated the song of the Sirens, had to evaluate the costs and benefits of trying to resist the Sirens and of giving in to them. Then he had to compare the best option, conditional on facing the Sirens, with the costs and benefits of pre-committing to being tied to the mast. Considering that seduction by the Sirens entailed death, this analysis was pretty straightforward, and most of us can appreciate why he tied himself to the mast. However, succumbing to many of our everyday temptations does not entail instant death. When anticipating more common and less dramatic temptation, the decision of whether or not to pre-commit is not obvious. Rather, the agent's optimal choice depends on the strength of the temptation, the importance of the conflicting goal, and willpower, each of which is featured in our model of self-control conflict.

Gul and Pesendorfer (2001, 2004) consider a single player who has preferences regarding choice sets that include the desire to limit the available alternatives. That is, the agent may prefer a subset of alternatives to the set itself. These different choice sets are motivated by the presence of visceral influences such as hunger, just as is in our framework. Under various axioms over choices over menus of lotteries, including the "set betweenness axiom," they show that the decision process can be represented by a utility function with a cost of self-control or a disutility from the presence of a tempting alternative. The result is a set of preferences that explains behavior when facing temptation as the outcome of a rational decision process even though the agent benefits from pre-commitment technologies to reduce future temptations. Benabou and Pycia (2002) use a costly influence game between the planner and the doer to show that the implied outcome probabilities are equal to those implied by Gul and Pesendorfer's representation (2001, 2004). In contrast, our approach treats the doer not as a strategic player, but as a visceral influence that pulls the agent toward an inferior choice. We preserve the result from Gul and Pesendorfer (2001, 2004) that the individual can benefit from pre-commitment, whenever it is not too costly. Moreover, we depart from the interpretation that two different preferences over sets both generate and resolve the conflict. Thus, our outcome probabilities differ from those derived by Benabou and Pycia (2002), as well. Our conflict interpretation allows us to analyze the interplay among the payoff from the goal, the strength of the visceral influence, and willpower in a simple framework. We emphasize that the visceral influence acts on the agent, thereby reducing the single period equilibrium to the result of a simple optimization problem.

With respect to the cost of self-control, however, we find that the effect of an increase in the visceral influence is ambiguous with respect to the effort exerted by the agent, the resulting self-control cost, and the implied welfare loss due to absence of pre-commitment. While this result can be derived using the models of Fudenberg and Levine (2006) and Gul and Pesendorfer (2001, 2004), we emphasize this here since it has important consequences for pre-commitment for which our model yields novel insights. Our model yields novel predictions about the agent's decision of whether or not to pre-commit, especially if the agent underestimates the influence of anticipated temptations on behavior or is overconfident about her own willpower. In such cases, the agent will exaggerate the expected value of trying to resist temptation, and hence mistakenly choose to forego pre-commitment when she would have benefited from it. Finally, a welfare analysis of the model yields the surprising prediction that stronger willpower reduces welfare under some circumstances. Equally surprisingly is the finding that the same is true for a higher payoff from the goal.

While we here only treat the static case of self-control conflict, the present model could prove fruitful as a building block for a dynamic analysis. Since the agent has limited cognitive resources to fight temptation (see also Baumeister et al., 1999), fighting a present temptation implies reduced resources to fight future temptation. For example, the dutiful student on a diet, who faces a tempting piece of chocolate during her final examinations, might strategically choose to indulge in the chocolate to ensure that her energy reserves are sufficient to subsequently fight the urge to stray from painful last-minute cramming. Such an analysis, of course, requires careful assumptions about the reservoir of cognitive resources available for self-control. These assumptions could be informed by experimental psychology, which has utilized the metaphor of the muscle to describe willpower, noting that it gets temporarily depleted due to exertion, that it eventually gets replenished, and that it might grow stronger in the long run from repeated use (e.g., Baumeister, Heatherton, & Tice, 1994; Baumeister et al., 1998; Muraven & Baumeister, 2000).

References

Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology 74*, 1252 – 1265.

Baumeister, R.F., Heatherton, T.F., and Tice, D.M. (1994). Losing control: How and why people fail at self-regulation. San Diego, CA: Academic Press.

Benabou, R., and Pycia, M. (2002). Dynamic Inconsistency and Self-Control: A Planner–Doer Interpretation," *Economic Letters 77,* 419–424

Benabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy 112,* 848-886.

Elster, J. (1977). Ulysses and the Sirens: a theory of imperfect rationality. *Social science information XVI, 5*, 469-526.

Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature 36,* 47-74.

Fishbach, A., and Shah, J. Y. (2006). Self Control in Action: Implicit Dispositions toward Goals and Away from Temptations. *Journal of Personality and Social Psycholoy 90*, 820-832.

Fudenberg, D. and Levine, D.K. (2006). A Dual Self Model of Impulse Control. *The American Economic Review 96,* 1449–1476

Gul, F. and Pesendorfer, W. (2001). Temptation and Self-Control, *Econometrica 69*,1403—1436.

Gul, F. and Pesendorfer, W. (2004). Self-Control and the Theory of Consumption." *Econometrica 72*, 119-158.

Hirshleifer, J. (1989). Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Publ. Choice 63,* 101-112

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist 58,* 697–720

Laibson, D.I. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics 112*, 443-477

Loewenstein, G. (2000). Preferences, behavior and welfare: emotions in economic theory and economic behavior. *The American Economic Review 90*, 426-432

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes 65*, 272–292.

Metcalfe, J., and Mischel,W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review 106*, 3–19.

Muraven, M., and Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin 126*, 257 – 259.

Myrseth, K.O.R. and Fishbach, A. (forthcoming). Self-control: Knowing when and how to exercise restraint, *Current Directions in Psychological Science.*

Myrseth, K.O.R. and Fishbach, A. (2010). Seeing self-control conflict: Isolated vs. interrelated temptations. Working paper, the University of Chicago.

O'Donaghue, T. and Rabin, M. (1999). Doing it now or later? *The American Economic Review 89*, 103-124.

Schelling, T.C. (1984). Self-Command in Practice, in Policy, and in a Theory of Choice, *The American Economic Review 74*, 1-11

Skaperdas, S. (1996). Contest success functions. *Economic Theory 7*, 283-290.

Sloman S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin 119*, 3-22.

Strotz, R. H. (1955-56). Myopia and Inconsistency in Dynamic Utility Maximization, *Review of Economic Studies 23*, 165—180.

Thaler, R. H. & Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy 89*, 392-406.

Trope, Y., & Fishbach, A. (2000). Counteractive self control in overcoming temptation. *Journal of Personality and Social Psychology 79*, 493-506.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient visceral feelings. *Personality and Social Psychology Bulletin 29*, 1159–1168.

Appendix A. Proofs

Proposition 1.

The Lagrangean for this problem becomes:

$$\Gamma = \frac{e}{e+v}\left(g - \frac{e}{w}\right) + \frac{v}{e+v}\left(a - \frac{e}{w}\right) - \lambda_1(-e) \quad . \tag{21}$$

First order conditions:

$$\frac{\partial \Gamma}{\partial e} = \left(\frac{e}{e+v}\right)\left(-\frac{1}{w}\right) + \left(g - \frac{e}{w}\right)\left(\frac{v}{(e+v)^2}\right) + \left(\frac{v}{e+v}\right)\left(-\frac{1}{w}\right) + \left(a - \frac{e}{w}\right)\left(\frac{-v}{(e+v)^2}\right) - \lambda_1 = 0 \quad , \tag{22}$$

$$\frac{\partial \Gamma}{\partial \lambda_1} = -e \le 0;\ \lambda_1 \ge 0;\ \lambda_1(-e) = 0 \quad . \tag{23}$$

The interior solution is derived by solving the first order conditions in (22) for e yielding the reaction function e^* in (24).

$$e^* = -v + \sqrt{wv(g-a)} \tag{24}$$

Proposition 2.

Recall the definition of a concave function. A function $f : \mathbb{R}^n \supseteq \mathrm{X} \to \mathbb{R}$, where X is a convex set, is concave if given any two points x' and x'' in X we have that

$$(1-\lambda) f(x') + \lambda f(x'') \le f\left[(1-\lambda)x' + \lambda x''\right] \equiv f(x^\lambda)\ \forall \lambda \in (0,1). \tag{25}$$

Let $v^{'}$ and $v^{''}$ be two visceral influences in the domain of e^{*}. Then, using (24) in (25) we have that

$$\begin{aligned}&(1-\lambda)\left(-v^{'}+\sqrt{wv^{'}(g-a)}\right)+\lambda\left(-v^{''}+\sqrt{wv^{''}(g-a)}\right)\leq \\ &-\left((1-\lambda)v^{'}+\lambda v^{''}\right)+\sqrt{w\left((1-\lambda)v^{'}+\lambda v^{''}\right)(g-a)}\end{aligned}. \qquad (26)$$

Solving for λ yields the condition

$$\lambda \leq \frac{\left(v^{''}-v^{'}\right)(g-a)}{\left(v^{''}-v^{'}\right)(g-a)} \equiv 1, \qquad (27)$$

Which holds $\forall \lambda \in (0,1)$.

Proof Corollary 2.

The derivates of effort with respect to g, w and a are positive, positive and negative, respectively.

$$\partial e^{*}/\partial w = \frac{gv-va}{2\sqrt{w(gv-va)}} > 0 \quad , \qquad (28)$$

$$\partial e^{*}/\partial g = \frac{wv}{2\sqrt{w(gv-va)}} > 0 \quad , \qquad (29)$$

$$\partial e^{*}/\partial a = -v - \frac{1}{2}\left[wv(g-v)\right]wv < 0 \qquad (30)$$

Proof Proposition 3.

After the choice of optimal effort, the probability of success at the conflict stage is given by $p_{win} = e^{*}/\left(e^{*}+v\right)$. Since the success probability p_{win} increases in effort and falls in visceral influence v, it is straightforward from (28) and (29) that g and w must increase p_{win} since these increase effort. Similarly, a and v decrease effort leading to a fall in p_{win}. Furthermore, v

operates directly to reduce p_{win}. Hence, increases in g and w increase the likelihood of success while a and v decrease it.

Proof Corollary 4 The derivate is effort with respect to overconfidence is

$$\partial e_s^* / \partial \beta = \frac{1}{2}\left[s\beta w\left(ga - a^2\right)\right] sw\left(ga - a^2\right) > 0 \tag{31}$$

Proof Corollary 5

The derivate of $\overline{\beta}$ is positive only if $a > w$.

$$\frac{\partial \overline{\beta}}{\partial w} = \frac{a^4}{w^3\left[a^2 + (g-a)^2 - (g-c)^2 + \frac{a^2}{w}\right]} - \frac{a^3}{w^2\left[a^2 + (g-a)^2 - (g-c)^2 + \frac{a^2}{w}\right]} > if\ a > w \tag{32}$$

The derivate of $\overline{\beta}$ is negative as long as the payoff from the tempting goal is positive, which it is by definition.

Proof of Corollary 6

The derivate of $\overline{\beta}$ is negative as long as the payoff from the tempting alternative is larger than the cost of pre-commitment.

$$\frac{\partial \overline{\beta}}{\partial g} = \frac{2a^2\left[(g-a)^2 - (g-c)^2\right]}{w\left[a^2 + (g-a)^2 - (g-c)^2 + \frac{a^2}{w}\right]^2} > 0\ if\ a > c \tag{33}$$

Proof of Corollary 7

$$\frac{\partial\overline{\beta}}{\partial a}=-\frac{2a^2\left[(g-a)^2-(g-c)^2+\frac{a}{w}\right]}{w\left[a^2+(g-a)^2-(g-c)^2+\frac{a^2}{w}\right]^2}+\frac{2a}{w\left[a^2+(g-a)^2-(g-c)^2+\frac{a^2}{w}\right]}>0 \quad (34)$$

$$if\ \ g>-a/\sqrt{1-a}<0$$

Since $g-a>0$, $g>0$ the condition holds.

Proof of Corollary 8

The derivate of $\overline{\beta}$ with respect to c is always negative.

$$\frac{\partial\overline{\beta}}{\partial c}=-\frac{2a^2(g-c)}{w\left[a^2+(g-a)^2-(g-c)^2+\frac{a^2}{w}\right]^2}<0 \quad (35)$$

Proof Proposition 4

Consider the derivatives of optimal effort with respect to the two variables when the other is equal to one. These are

$$\frac{\partial e^*_{s,\beta=1}}{\partial s}=-a+\frac{w\left(ga-a^2\right)}{2\sqrt{ws\left(ga-a^2\right)}},\frac{\partial e^*_{s,s=1}}{\partial\beta}=\frac{w\left(ga-a^2\right)}{2\sqrt{w\beta\left(ga-a^2\right)}} \quad (36)$$

Since the derivative $\frac{\partial e^*_{s,\beta=1}}{\partial s}$ refers to an increase in sophistication and $\frac{\partial e^*_{s,\beta=1}}{\partial s}$ refers to an increase in overconfidence, we must negate $\frac{\partial e^*_{s,\beta=1}}{\partial s}$ so that we compare a reduction in sophistication with an increase in overconfidence.

$$\frac{\partial e^*_{s,s=1}}{\partial \beta} > -\frac{\partial e^*_{s,\beta=1}}{\partial s} \; if \; \frac{\beta\left(a^2 + \left[w\left(ga - a^2\right)\right]^2\right)}{\left[w\left(ga - a^2\right)\right]^2} = \underline{s} > s \tag{37}$$

The increase in effort due to lack of sophistication is thus larger than due to an increase in sophistication only when s is small enough. Beyond $\underline{s}$, the increase in overconfidence has a larger effect on effort.

Proof Proposition 5.

Consider the derivatives of the naïve expected payoff from conflict with respect to willpower when either s or β is equal to one. These derivatives are

$$\frac{\partial \mathrm{N}_{|s=1}}{\partial w} = \frac{-a + \sqrt{\left(ga - a^2\right) w\beta}}{w^2}, \frac{\partial \mathrm{N}_{|s=\beta}}{\partial w} = \frac{-sa + \sqrt{\left(ga - a^2\right) ws}}{w^2} \tag{38}$$

$$\frac{\partial \mathrm{N}_{|s=1}}{\partial \beta} > \frac{\partial \mathrm{N}_{|s=\beta}}{\partial s} \; if \; w \geq \frac{a^2\left(1-s\right)^2}{\left(\beta - s\right)\left(ga - a^2\right)} \tag{39}$$

The expression on the right hand side of (39) is positive as long as $s<1$ since $g-a>0$ and $\beta \geq 1$.

Chapter 2

Self-Control in Games

Conny Wollbrant
University of Gothenburg, Sweden

April 14, 2010

Abstract

People are often tempted to deviate from their optimal strategies. A situation reflecting such interference by temptation is defined as a self-control game where each player consists of two cognition types. One type generates biases in decision making by producing visceral influences. In contrast, another cognition type can ameliorate visceral influences by exercising self-control. The set of outcomes reflecting perfect self-control are called "self-control equilibria" and is equal to the set of subgame perfect Nash equilibria. In contrast, the set of "temptation equilibria" reflects imperfect self-control and is a superset if will-power is "high enough." We explore implications for several instances of social interaction when players are altruists tempted to be greedy.

Keywords: Self-Control, Game Theory, Temptation, Prosocial behavior
JEL Classification: C79, D01, D03

Acknowledgements: I am grateful to Martin Dufwenberg, Olof Johansson-Stenman, Kristian Ove Myrseth, Katarina Nordblom, Amrish Patel, Robert Östling, and seminar participants at the University of Gothenburg for helpful comments. Remaining errors are my own.

1 Introduction

The role of self-control in strategic interaction is potentially very powerful. Individuals prove susceptible not only to powerful emotional influences, also known as visceral influences, from sources exogenous to the game (such as a rising room temperature causing more aggressive posturing in negotiation), but also to impulses endogenous to the game (such as escalating aggression in response to rude behavior). Moreover, emotions may influence behavior by quite distinct routes, either by altering the perceived payoffs (such as the relief of ending negotiation due to an unpleasant room temperature) or by acting directly on the individual as a visceral influence or “force” (such as a surging aggressive impulse to punish the counterparty for her annoyances). Whenever emotional forces and optimal strategies provide conflicting prescriptions for behavior, the individual will have to leverage her willpower in order to stay on course. Moreover, awareness of emotional forces acting on others suggests that these should form part of a successful strategy. In this way, players should take into account their own willpower, or lack thereof, as well as that of others. The objective of this paper is to provide a game theoretic framework to analyze the strategic aspects of self-control and emphasize the importance of emotions, in particular visceral influences, for social outcomes. To this end, we apply a self-control model that reduces the self-control problem to a simple optimization problem and allows us to characterize visceral influences as functions of expected play.

The venture to incorporate the role of emotion in models of strategic games may shed new light on the strategic structure of social interaction, and, thereby, yield entirely new predicted equilibria, due to, e.g., guilt or anger leading to negative reciprocity (e.g., Battigalli and Dufwenberg, 2006, 2009; Rabin, 1993).[1] Starting with Geanakoplos et al. (1989), the literature on psychological games aims to model belief-dependent motivations in games as many aspects of a variety of emotions depend on individuals' expectations or beliefs. As an example, suppose a man walking on the sidewalk is pushed into the road by another. Is the pushing man being kind or unkind? The answer to this question depends on what the man being pushed believes about the beliefs of the pushing man. Is he pushing because he believes that something dangerous is about to occur on the sidewalk? If so, he is being kind. Or, does he want to harm the man by pushing him into traffic? If so, he is being unkind. In this way, the appropriate response for the man being pushed depends on his evaluation of the intentions of the man doing the pushing. Specifically, it depends on his beliefs about the other individual's beliefs. Geanakoplos et al. (1989) develop a game theoretic framework that accommodates utility functions that incorporate such a hierarchical belief structure, and thus manages to add belief-dependent utility to the game tree.

The first application of this framework is developed by Rabin (1993) for the specific case of fairness. It is assumed that individuals have preferences involving assessments of kindness. In particular, individuals want to be kind to those who are kind to them, and unkind (e.g., due to anger) to those who are unkind, leading to the solution concept of an "fairness equilibrium." The analysis leads to several interesting insights. For instance, a fairness equilibrium arises when each player maximizes the others' payoff. Hence, in a prisoners dilemma, cooperate-cooperate is a fairness equilibrium since each player believes that the other player is trying to be kind and hence wishes to reciprocate by being kind in return. However, when a player is forced to cooperate, the other player does not consider this an act of kindness and will hence defect rather than cooperate. While Rabin (1993) is limited to two-by-two normal form games, Dufwenberg and Kirschsteiger (2004) extend this framework to develop a general model of reciprocity for sequential games with n players.

[1] The venture to incorporate the role of emotions in economic theory is not new. See for example Elster (1998).

Notably, most of the work to date has focused on the influence of emotions on an individual's payoffs (e.g., Rabin, 1993). While incorporating this route of emotional influence, we here predominantly explore the second route, whereby emotion acts as a direct visceral influence on behavior (e.g., Loewenstein, 1996, 2000; Van Boven and Loewenstein, 2003; Loewenstein & O'Donoghue, 2007). The role of visceral influences is important because they often are not aligned with the individual's best interest, thereby giving rise to conflict with better judgment. Because individuals' capacity to exercise restraint in the face of temptation is limited, as the classical writers well knew (Plato, 1986/380 B.C.) and as social psychologists and philosophers recently reaffirmed (e.g., see Baumeister et al., 1994; 1998, Elster, 1977), self-control conflict presents the possibility of acting against one's better judgment.

Recently, it has been suggested that self-control problems extend to social interaction (e.g., Loewenstein, 1996, 2000; O'Donoghue and Loewenstein, 2007) and are not limited to issues of, for example, consumption such as dieting, procrastination and saving (see, e.g., Strotz, 1955-56; O'Donoghue and Rabin, 1999; see also Fredericks et al. 2002, for a review). On the contrary, the scope of potential emotional influences on behavior in strategic settings seems larger than in individual decision-making situations as many emotions are social and strategic interaction, by definition, involves more than one agent. For example, in envy, one is envious of something that another possesses, and at greater intensities, this might lead to spite, causing the individual to act in a mutually destructive manner, thus leading to socially inferior outcomes. Similarly, while one might occasionally be angry at oneself, it is more common to be angry with someone else, perhaps leading to retaliation that would be costly to both parties.[2] As many emotions depend on another, or even a third individual being present (e.g. jealousy), self-control conflict is more likely to occur in instances of social interaction than in instances of individual decision making. Moreover, since many visceral influences are accompanied by aversive sensations such as hunger or pain, they also undermine altruism in general because the aversiveness makes the individual focus inwards, e.g., on the hunger or pain (Damasio, 1994; Loewenstein, 1996).

Some recent work has focused on the conflict between pro-social preferences in terms of altruism and conditional cooperation and the visceral influence produced by greed (Martinsson et al. 2010a; 2010b). They find that a measure of individual trait self-control positively predicts pro-social behavior. In the present paper, we outline strategic implications that follow from games featuring agents with a propensity to encounter self-control conflict due to a visceral influence induced by greed, and thus elaborate on the idea that pro-social vs. selfish behavior represents a self-control conflict.

[2] Of course, the list of social emotions is very long.

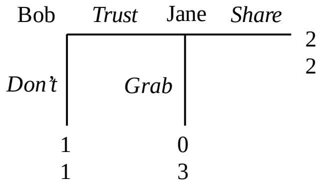


Figure 1. A trust game.

To further illustrate the relevance of self-control to strategy and pro-social behavior, consider the trust game in Figure 1 with material payoffs. When Bob is about to make his decision, he is aware that with preferences specified over material payoffs, Jane will surely choose $Grab$ rather than $Share$. Hence, Bob's best response is to choose $Don't$, leading to a suboptimal outcome for both players. Assume instead that players are motivated by reciprocity, in the spirit of Rabin (1993) and Dufwenberg and Kirschsteiger (2004) (henceforth, RDK). In such a case, players want to be kind to those who are kind and unkind to those who are unkind. As a result, when Bob chooses $Trust$, Jane will consider Bob kind and will want to be kind in return and so chooses to $Share$, leading to an improvement for both players. But what if Jane is *tempted* by greed to choose $Grab$ anyway? After all, once Bob has chosen to $Trust$, the material payoff of $Grab$ looms larger and perhaps this will dissuade Jane from choosing $Share$. Jane's decision will then come down to her willpower and in turn, it will determine whether Bob will choose to $Trust$ or play $Don't$. Thus, even with preferences for reciprocity, willpower is central to the outcome and we need a framework to determine when the reciprocal profile $(Trust, Share)$ can be expected to occur and when play will result in $(Don't, Grab)$. In a later section, we will augment the game in Figure 1 and solve the resulting self-control game.

It is not always the case that self-control considerations will strike a balance between a preference-motivated outcome and a temptation-motivated outcome and lead to extremities. Instead, self-control problems can lead to strategies that result in intermediate outcomes, as the following illustrates (Figure 2).

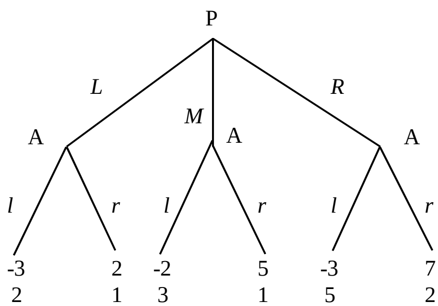


Figure 2. The Social Planner's Problem.

In this example, which we term "The Social Planner's Problem," we have in mind a social planner P and an agent A. The social planner has as her objective to maximize social welfare by maximizing the sum of material payoffs. The payoff at the top of each branch (the social planner) may be thought of as the material payoff to everyone else in society, except the agent, whose material payoff is denoted at the bottom of each branch. We will assume that the agent, just like the social planner, has preferences for "simple altruism," such that she wishes to maximize the sum of payoffs to society, although she might be tempted by greed to behave selfishly.

The planner has three possible actions available to her, Left (L), Middle (M) and Right (R), while the agent has only Left (l) and Right (r) in all subgames. Both the planner and the agent would prefer the outcome (R, r) to all other outcomes, but in order to get there, the agent must overcome a significant self-control problem. This is because the material gain from playing l is larger than that from playing r in the subgame starting with the social planner playing R. It seems plausible that this leads to a temptation strength proportional to the difference between these payoffs $(5 - 2)$. If the agent does not have sufficient willpower, society risks ending up with the less attractive outcome of (R, l). To avoid this, the planner might consider a less ambitious policy by choosing M since this might lead to the second best outcome (M, r). The agent would prefer this outcome to (M, l), but is tempted by the larger material payoff (M, l) since the difference in material payoffs is $(3 - 1)$. If the agent is not equipped with sufficient willpower, she might succumb and the inferior outcome (M, l) will result. Again, this can be avoided by the social planner choosing L, potentially leading to the outcome (L, r). While the agent will be tempted in this subgame too, the strength of temptation is relatively weak $(2 - 1)$. Finally, should the agent not be able to resist this urge either, it might be better for the social planner to pick a more ambitious policy, e.g., R, and have the agent fail, since the payoff from (R, l) is larger than that from (L, l). Thus, the optimal policy will depend on the agent's ability to control the urge to be greedy. These examples illustrate that self-control problems are important strategic considerations. In addition, one can analyze interactions between temptation and different "underlying" preferences. In our examples, we first made use of a simple trust game with a slightly more complicated preference for fairness, and second, we used a slightly more complicated game, but with preferences for simple altruism, assuming the agent wishes to maximize the benefits to society. It is to this end that we develop our framework for self-control in games.

While problems of self-control to date have not been incorporated into game theoretic models of strategic interaction between individuals, game theoretic models of strategic interaction have been applied to understand self-control conflict as a strategic interaction within individuals, between "multiple selves" (Thaler and Shefrin, 1981; Laibson, 1997; Fudenberg and Levine, 2006). We choose, however, not to model intrapersonal self-control conflict as a game between multiple temporally separated selves. While many strategic considerations are likely to be independent of the underlying self-control model, we choose to adopt the framework proposed by Myrseth and Wollbrant (2010), which conceptualizes self-control conflict as a struggle between a rational agent of limited willpower and a visceral influence, which acts like a force on the agent. This model does not only confer the advantage of reflecting key psychological processes thought to underlie self-control conflict, such as visceral influences (e.g., see Loewenstein, 1996; Loewenstein and O'Donoghue, 2007) and limited will-power (e.g., see Baumeister et al., 1994, 1998; Muraven and Baumeister, 2000), but also the advantage of reducing the question of self-control to a simple constrained optimization problem, allowing for modeling visceral influences directly and as part of the solution.[3] Furthermore, unlike alternative models of self-control that exclusively consider inter-temporal intrapersonal conflict, such as the planner at t_0 strategizing against the myopic doer at t_1 (Thaler and Shefrin, 1981; Schelling, 1984) or the struggle between competing discount functions at different points in time (e.g., Ainslie, 1975; 2001), the model at hand also considers intra-temporal intra-personal conflict.[4] That is, the model may describe an ongoing struggle between an agent and a conflicting visceral influence, thereby allowing modelers to analyze self-control conflicts that lack inter-temporal components and that thus prove intractable with aforementioned frameworks. Visceral influences may in principle be at work in any strategic interaction, and they may take many forms. We have already used the example of aggression, but one could easily imagine others, such as fear, greed, and sympathy, to mention just a few.

2 A Model of Self-control Games

The general model of self-control in games captures the notion of a conflict between "System 1" and "System 2" cognition in strategic interaction. The System 1 and System 2 terminology is adopted from psychological dual-process models that classify cognition into two distinct processing modes (e.g., Kahneman, 2003; Sloman, 1996; Metcalfe and Mischel, 1999):

- System 1: effortless, parallel processing; associative reasoning; "hot," emotional influences; uncontrolled
- System 2: effortful, serial processing; rule-based reasoning; "cool" thinking; controlled.

The terminology also reflects the evolutionary sequencing of these two cognition modes. System 1 may be thought of as corresponding to the brain's limbic system, the part of the brain that we share with animals, such as lizards, and was developed first. System 2, by contrast, is a much later development. It is responsible for more abstract thought processes, corresponding to the brain's pre-frontal cortex. Or, as in the terminology of Thaler and Shefrin (1981), the

[3] This is a feature of Fudenberg and Levine (2006) and Gul and Pesendorfer (2001; 2004) too. Under certain assumptions, our model is closely connected to the reduced form of Fudenberg and Levine (2006).

[4] Bénabou and Tirole (2004) presents a model that emphasizes the role of personal rules for self-control behavior.

"Planner" (System 2) and the "Doer" (System 1).[5] Similar to Loewenstein (1996), we conceptualize temptation as a System 1 "visceral influence" on the agent. This visceral influence may be thought of as a drive-state, and acts like a force on the agent, impelling the agent to act in a specific manner. Typical examples of such drive-states include hunger and pain relief, but also more complex emotions such as fear, anger, and greed.

While each system possesses individual preference relations, it is assumed that the preference of System 2 is the relevant preference for evaluation of outcomes while System 1 preferences provide the foundation for visceral stimuli. To outline the modelling framework, we start by specifying an extensive form "base game" after which we add the additional components required for the definition of the "self-control game". Most of the additional components required for the self-control game are related to the two different types of cognition of each player and thus, we slightly delay the introduction of preferences to the definition of the self-control game

2.1 A general model

Definition 1 (Base game) *An extensive game form, excluding preferences, with perfect information contains the following components (Osborne and Rubinstein, 1994).*

- A finite set N (the set of players)·
- A non-empty set of actions A_i for each player $i \in N$
- A set H of finite sequences that satisfies the following properties
  - The empty sequence $\varnothing$ is an element of H.
  - If $\left(c^k\right)_{k=1,\ldots,K} \in H$ and $L < K$ then $\left(c^k\right)_{k=1,\ldots,L} \in H$
- A function f_c that associates with every history h for which $P(h) = c$, a probability measure $f_c(\cdot|h)$ on $A(h)$ (where c denotes *chance* or the state of *nature*; exogenous uncertainty).
- A player function P that associates each non terminal history $h \in H \backslash Z$ with an element in $N \cup \{c\}$,where Z is the set of terminal histories.

The base game is thus the initial form, excluding preferences, to which we add the modifications implied by the self-control game. While the primitives of the base game in definition 1 are standard, what follows in definition 2, with the

[5] The fact that system 2 is referred to as the "planner" is itself quite revealing as planning is inherently an abstract thought process.See Bénabou and Pycia (2002) for a Planner-Doer interpretation of Gul and Pesendorfer (2001).

exception of preferences, is not; the components of definition 2 enter our model of a self-control game.

Definition 2 (Self-control game) *A Self-control game in extensive form with perfect information contains the following additional components*

- A set of efforts $E_i = [0, \infty)$ for every player $i \in N$.
- The set of cognition types $M = \{1, 2\}$ corresponding to "System 1" and "System 2" cognition.
- A visceral influence function v_i (associates a subset of $A_i(h)$ with a real positive number interpreted as the strength of the visceral influence).
- A resolution mechanism R.

We will now define the components of the game. Since we need to distinguish between actions in the conventional sense, i.e., observed behavior, and unobservable efforts, we introduce the qualifiers "pure" and "combined" to mean an element from the conventional (pure) action space A_i , and a combination of a pure action and an effort to mean an element from C_i, respectively.

Definition 3a (pure actions) *A "pure action" is an action in the conventional sense where $a_i \in A_i$ is the set of actions for player i.*

Definition 3b (combined actions) *A combined action is a pair consisting of a pure action $a_i \in A_i$ and an effort $e_i \in E_i$. The combined action space is the product $C_i = E_i \times A_i$ for every player $i \in N$.*

With "pure" actions, we refer to physically observed behavior, e.g. "left," "middle," "right," and so on. A pure action is a one-dimensional choice while a combined action is two-dimensional. We assume that System 1 makes the effortless impulsive choice of pure action, while System 2, taking System 1 pure action as given, needs to decide on a pure action as well as an effort level to combat System 1 impulses.

The strength of temptation thus depends on System 1 "preferences." Each action has associated with it, a degree of "relief" from visceral influences arising from any, or a number of, emotions pulling the agent in different directions. These influences differ strength and so, the relief provided by a certain action can be represented by a number, and actions can thus be ordered accordingly, which allows us to represent visceral influences as a set of preferences over endnodes.

Among the possible actions that would provide relief, it is assumed that System 1 chooses the pure action with the highest relief value. As a result, whenever System 2 cognition chooses a pure action with a lower relief value, more effort is needed to combat System 1. This is so since the discrepancy between the relief value of System 1's best response and the relief value implied by System 2's choice of pure action, leads to greater frustration of System 1's preferences and hence larger resistance, which we interpret as the conflicting visceral influence.[6]

To illustrate, consider the decision problem in the example below. The payoff accruing to System 1 cognition is displayed to the left and, that accruing to the System 2 to the right in the payoff vectors. System 2 cognition prefers L to M and M to R while System 1 cognition ranks the alternatives in reverse order. If System 2 cognition were to insist on L, she could potentially reach a high payoff of 100 but would face a very strong temptation. The reason is that System 1 cognition stands to gain a relief value of 90 from R, and the discrepancy between relief from R and L provides a powerful stimulus. To resist, System 2 cognition would need to expend a large amount of "costly effort." If, however, System 2 instead were to pick pure action M, the payoff of 99 is almost as good as that of 100 from L, with the added benefit that the temptation would be much lower than before. The difference in payoff for System 1 is just $90-80=10$ compared to the previous 90, implying an intermediate level of costly effort for System 2. Pursuing pure action M then seems like a good compromise for System 2 cognition.

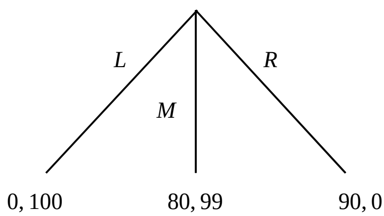


Figure 3. A three action decision problem.

Since effort is costly for System 2, these preferences must be specified over combined actions, i.e., pure actions and efforts. System 1, however, does not have to exert any costly effort, and hence these preferences are specified only over pure actions. This leads us to the following definition of the utility functions of each player's cognition

Definition 4 (utility functions) *The utility functions* u_{im} *denote the preference relations* $\succeq_{im}$ *such that* $u_{i1}(a) \geq u_{i1}(a')$ *whenever* $a \succeq_{i1} a'$ *and* $u_{i2}(c) \geq u_{i2}(c')$ *whenever* $c \succeq_{i2} c'$.

[6] This is similar to the "opportunity based cost of self-control" in Fudenberg and Levine (2006). Their idea is that controlling the short-run self will be more costly in a state where the short-run self would attain a higher utility than in a state with a lower utility. While our formulation differs, it is closely connected, since a higher visceral influence, i.e., strenght of temptation, will require more costly effort.

System 2's choice of effort will depend on the visceral influence arising due to the discrepancy in relief value for System 1, which we define here as the *visceral stimulus*. The visceral stimulus, *visceral influence* and *best responses* are determined simultaneously. We begin by introducing the visceral stimulus, after which the other two definitions will follow.

Definition 5 (visceral stimulus) *The visceral stimulus s is the difference $u_{i1}(B_{i1}) - u_{i1}(B_{i2})$. Whenever $s = u_{i1}(B_{i1}) - u_{i1}(B_{i2}) > 0$, we have that $s \in (0, u_{i1}(B_{i1})]$.*

Where B_{im} is the best response function for cognition type m of individual i (to be defined shortly). This definition of visceral stimulus allows for modeling a variety of immediate emotions. In our applications, we assume that the visceral influence is greed, so the only discrepancy we need to consider is the difference in material payoffs. In principle, however, other more complicated System 1 emotions could be modeled, such as anger or annoyance with another player's behavior.

It is assumed that visceral stimulus arises whenever System 2 cognition deliberates over her best course of action. B_{i2} will in some cases lead to a less than optimal payoff for System 1 cognition and hence $s > 0$. Consider Figure 3 where $B_{i1} = R$, which leads to the highest value for System 1, i.e., 90. Should System 2 choose L the consequence for System 1 would be a difference between 90 and the value attained from L, which is 0.This would then imply visceral stimulus of $90 = 90 - 0$. Should System 2 instead choose M,then System 1 would attain 80, implying a stimulus of $10 = 90 - 80$. It is this basic tension between the two best response functions that leads to a self-control problem via the stimulus generation leading to a visceral influence. Next, we define the visceral influence function.

Definition 6 (visceral influence function) *The visceral influence function V of player i associates every visceral stimulus $s \in (0, u_{i1}(B_{i1})]$ with a real number, such that $V_i : [0, u_{i1}(B_{i1})] \rightarrow R$, for every non-terminal history $h \in H \backslash Z$ ($V_i(h, B_{i1})$ is the visceral influence of $B_{i1} \in A_i(h)$ whenever $P(h) = i$).*

Embedded in the definition of visceral influence is the notion that the visceral influence arises whenever System 1 and System 2 cognition disagree. This is also clear from the definition of visceral stimulus, since visceral stimulus is weakly positive for all pure action elements other than the one providing maximum relief value to System 1. In line with the example above, this leads to a set of pure actions for which System 1 and System 2 cognition would disagree which we term *contested actions*. In Figure 3, the set of contested actions consist of L and M. In this way, the only action that would not lead to a contest is if System 2 decided to attempt a conflict involving R, which is indeed the most preferred pure action element of System 1 in this problem. Given the utility functions and the visceral influence function, we can also state the best response functions of System 1 and System 2.

Definition 7 (best response functions) *The best response functions of player i, one for each cognition type $m \in M$, are defined as the correspondences B_{im}*

$$B_{i2}(a_{-i}, v_i) = \{c_i \in C_i : (a_i, c_i|v_i) \succsim_{i2} (a_i, c_i'|v_i) \forall c_i' \in C_i\}, \tag{1}$$

$$B_{i1}(a_{-i}) = \{a_i \in A_i : (a_{-i}, a_i) \succsim_{i1} (a_{-i}, a_i') \forall a_i \in A_i\}. \tag{2}$$

Technically, the definition of B_{i1} of System 1 cognition is straightforward as it corresponds to the standard best response function where a player chooses an action in order to maximize her utility, given the actions of all other players. The interpretation, however, is slightly different. Implicit in the concept of System 1 best responding is the aforementioned idea that System 1 chooses the pure action with the highest relief value. That is to say, and individual would never choose a tempting action that is less tempting than some other action. Implicitly, this is a refinement that enables the agent's System 2 cognition to treat temptations as parameters for choice.[7]

The interpretation of the best response function of System 2 cognition B_{i2}, is also slightly different. System 2 cognition has to best respond, not only to actions by others, but also to her own urges v_i by picking a combined action c_i^* consisting of pure action a_{i2}^* and an effort level e_{i2}^* for a given visceral influence. The mechanism by which System 2 cognition best responds is assumed to be the self-control conflict mechanism of Myrseth and Wollbrant (2010) which we will elaborate on below.[8]

Due to the two cognition types and the difference between combined and pure actions, we make a distinction between a strategy for System 2 cognition and System 1 cognition.

Definition 8 (strategy) *A strategy s_{im} 1(2) of player i is a function that assigns a pure (combined) action in $A_i(h)$ $(C_i(h))$ to every non-terminal history $h \in H \backslash Z$ for which $P(h) = i$.*

The interpretation of a strategy s_{i1} is that System 1 cognition at each decision node strives toward the action leading to the highest relief value whenever

[7]System 1 best-responding might be regarded as the refinement implicit in Fudenberg and Levine (2006) wherein their "Short Run" self always best responds. This is congruent with the idea that system 1 acts on impulse and is one of the reasons why Fudenberg and Levine's theory, as well as my own, avoids multiplicity of equilibria, in contrast to models of hyperbolic or quasi-hyperbolic discounting (e.g., Laibson, 1997).

[8]Though this model is an elaboration of Myrseth and Wollbrant (2010), it can easily be augmented to fit other frameworks too, such as those of Gul and Pesendorfer (2001, 2004) or Fudenberg and Levine (2006). The hyperbolic discounting model is also an alternative, though it would yield multiple equilibria. Furthermore, it is unclear to us how discount rates would be allocated over choice sets.

it is player $i's$ turn to move. The strategy of System 2 s_{i2} is necessarily slightly more complicated since she needs to take System 1's insistence on a specific pure action as given. System 2's choice of effort will depend on the visceral influence arising due to the discrepancy in relief value for System 1, which we define here as the *visceral stimulus.*

Definition 9 (contested actions) *Contested actions is a set*

$$D(h) = \{(B_{i1}(h), a_i'(h)) : B_{i1} \neq a_i'\}.$$

A point $d \in D(h)$ contains two pure action elements, $B_{i1}(h)$ being the maximizer of the cognition System 1 preference and $a_i'(h)$ being the pure action element in $C_i(h)$ for which $B_{i1} \neq a_i'$ for all $a_i' \in C_i(h)$.

When $D(h) = \varnothing$, either System 1 cognition is indifferent between all alternatives, or the best response functions of both types prescribe the same pure action for some history $h \in H$. Since self-control games have to involve visceral influences, and hence a non empty set $D(h)$, self-control games can only involve mixed strategies at nodes where $D(h) = \varnothing$. Mixed strategies require that the decision maker, in this case System 2, is indifferent, and such indifference only occurs when $D(h) = \varnothing$. If at some node System 2 is indifferent and System 1 is not, System 2 would prefer to agree with System 1 rather than engage in costly conflict for an alternative that would yield the same payoff.

Since System 2 cannot pick more than one pure action, she only engages with a single visceral influence. This leads us to the closely related concept of cognition conflict, arising whenever the best response of System 1 (a pure action) is not contained within the best response of System 2 (a combined action)

Definition 10 (cognition conflict) *Cognition conflict arises whenever $B_{i1}(h) \nsubseteq B_{i2}(h)$ (cognition preferences prescribe conflicting courses of action) or, alternatively, if it is the case that $u_{i1}(B_{i1}) - u_{i1}(B_{i2}) > 0$.*

Given that cognition conflict arises, the conflict is resolved by means of a resolution mechanism. The resolution mechanism employed here is that of Myrseth and Wollbrant (2010). They use a contest success function to specify a probability distribution over pure actions in the face of cognition conflict. System 2 effort and visceral influence are analogous to efforts entering the contest success function, but the prize from being successful is to attain the most preferred alternative. The probability of being successful rises in System 2 (costly) effort, and falls in the visceral influence[9].

Definition 11 (resolution mechanism) *A resolution mechanism $R(h)$ of player $i \in N$ assigns a probability distribution $r(\cdot|B_{i1}(h), a_i'(h), v_i(h))$ over the set of contested actions $D(h)$ for every non-terminal history $h \in H\backslash Z$.*

[9]For more on contest success functions, see Skaperdas (1996) and Hirshleifer (1989).

The outcome probability r denotes the probability that System 2 "wins" the conflict. Thus, at a given non -terminal history, r is interpreted as the likelihood of $a_i'(h)$ occurring and $(1-r)$ the likelihood of $B_{i1}(h)$ occurring whenever $B_{i1} \neq a_i'$. It is implicit that whenever System 2 cognition has chosen to pursue a particular pure action, only $a_i'(h)$ and $B_{i1}(h)$ can occur with positive probability. Given these components, a self-control game can be summarized as follows.

Definition 12 (summary self-control game) *A self-control game Γ^s in extensive form with perfect information can thus be summarized as a 9-tuple:*

$$\Gamma^s = (N, M, A, E, H, P, u_{im}, v_i, R) .$$

An alternative modeling approach would have been to model System 1's influence on System 2 as a "procedural preference" over actions at each node rather than as a preference over outcomes at endnodes. Such an approach would be more conceptually aligned with the idea of "action tendencies," according to which emotional influences result in a particular behavior such as running away when faced with a fearful situation. While our approach relies on this dual preference structure, we believe this to be an advantage in deriving the visceral influences of the game. In particular, we view this underlying preference as providing a foundation for why a particular "action tendency" occurs at any given node, since anticipated play will change action tendencies.

2.2 Equilibrium

The equilibrium concept applied here is that of Subgame Perfection. Allowing for long histories and solving the game with backwards induction, this contrast with models where the action of the short-run self is either the result of optimization at every node of the game (e.g., Fudenberg and Levine, 2006) or an alternative solution concept where the agent makes mistakes, such as the quantal response equilibrium (McKelvey and Palfrey, 1995; 1998). While it may appear as if this requires a great deal from our conceptualization of System 1, we believe the solution concept to be appropriate.

Stimuli for visceral influences are both internal and external. A rising room temperature that makes a player more aggressive is an example of an external stimulus and anticipated opportunities for material gain, stimulating greed in the player, is an example of internal stimulus. In the latter case, the material gains need not be immediate, but can be the product of a very elaborate strategy. In our interpretation, it is thus the deliberation of System 2 over choices that provides stimuli. Thus, visceral influences are in some cases endogenous to deliberation and the solution concept, and we favor this interpretation for the case of strategy. Since emotions in principle become functions of expected play, an emotion-specific visceral influence function provides a convenient representation in the game tree. Using subgame perfection as a solution concept is thus both convenient and compelling.

To apply it, however, we first augment the player's decision process by introducing the self-control model. We assume that in the face of cognition conflict, every player's System 2 cognition has to maximize expected utility via the resolution mechanism R for a given visceral influence $v_i(h)$, by picking a combined action $c_i \in C_i(h)$. We assume that effort is costly and capture this with the cost function $c(e)$ with the following properties: $c'(e) \geq 0, c^{''}(e) \geq 0, c(0) = 0$. The exogenous parameter $w \in [0, \infty)$ is the cost of effort augmenting willpower.[10] The outcome of the resolution mechanism is thus due to the solution of the following maximization problem:

$$\max_{c_i \in C_i(h)} E(u_{i2}) = r(e,v)u_{i2}(a_i, a_{-i}) + (1 - r(e,v))\, u_{i2}(B_{i1}, a_{-i}) - \frac{c(e)}{w} \tag{3}$$

$$v = v(a_i), v'(a_i) \geq 0, v''(a_i) \geq 0$$

Differentiating with respect to e_i yields the first order condition

$$r_e(e, v(a_i))\,(u_{i2}(a_i, a_{-i}) - u_{i2}(B_{i1}, a_{-i})) \leq \frac{c'(e)}{w}, \tag{4}$$

and differentiation with respect to a_i yields

$$r_{a_i}(e, v(a_i))\, v^{'}(a_i)\,[u_{i2}(a_i, a_{-i}) - u_{i2}(B_{i1}, a_{-i})] + u_{i2}^{'}(a_i, a_{-i}) r(e, v(a_i)) = 0. \tag{5}$$

These two equations jointly determine the optimal combined action; behavior will be determined by the probability $r_e(e^*, v(a_i^*))$ inducing a probability distribution over B_{i1} and a_i^*. Effort affects the maximum of the problem by first increasing the probability r of resisting temptation and second by increasing the cost of effort $\frac{c(e)}{w}$. The condition in (4) states that the expected benefit from conflict must be less than or equal to the willpower-augmented marginal cost of effort. When $r_v > 0$ and $r_{vv} < 0$, the benefit function, i.e., the positive components in (3), is concave. This functional form assumption reflects how we typically think about visceral stimulus. At low levels of stimulus, the visceral influence is also low; increasing the stimulus might lead to very high levels of visceral stimulus. Since the cost function is linear, it is well-known that this problem has a solution.

[10] Note that when the willpower parameter is infinitely larger, effort will be costless and hence the resolution probability is equal to 1, for any pure action and visceral influence.

The term in brackets in (5) is the difference between the System 2 utilities, where $u_{i2}(a_i, a_{-i})$ is the utility System 2 attains when resisting temptation and $u_{i2}(B_{i1}, a_{-i})$ the the utility it attains from succumbing. Since by definition, resisting is preferred to succumbing, this payoff difference must be positive. Here we can state a useful result:

Lemma 1 (conflict payoff positivity) *The difference between the payoff from the goal and the payoff from the tempting alternative must be positive* $u_{i2}(a_i, a_{-i}) - u_{i2}(B_{i1}, a_{-i}) > 0$ *for a self-control problem to occur.*

Increasing the difference also increases optimal effort, since the expected payoff from resisting increases. The solution of the first order conditions yields a reaction function in terms of System 2 payoffs from resisting and succumbing, willpower and visceral influence. For a subgame perfect equilibrium to exist in the self-control game, we require that System 2 cognition always has a best response at every history. Since the solution to this problem is quasi-convex in the visceral influence, leading to a quasi-convex cost of self-control $\frac{c(e)}{w}$, it is well known that the maximum of this problem is obtained as long as the benefit function is quasi-concave; hence, a minimum requirement is that either the benefit function is linear and the cost is convex, or the cost is linear and the benefit function is concave.[11]

We now proceed by defining the subgame and the subgame perfect equilibrium of a self-control game.

Definition 13 (subgame) *A subgame of the extensive self-control game with perfect information* $\Gamma^s = (N, M, A, E, H, P, u_{im}, V, R)$ *is the extensive self-control game*

$$\Gamma^s(h) = (N|_h, M, A|_h, E, H|_h, P|_h, u_{im}, V, R) .$$

Definition 14 (subgame perfect equilibrium of the self-control game) *A subgame perfect equilibrium of the extensive self-control game with perfect information* Γ^s *is a strategy profile* s^*_{i2} *such that for every player* $i \in N$ *and every non-terminal history* $h \in H \backslash Z$ *for which* $P(h) = i$, *we have that* $O_h(s_{-i2}|_h, s^*_{i2}|_h) \succeq_{i2} O_h(s_{-i2}|_h, s_{i2}|_h)$ *for every strategy* s_{i2} *of player* i *in the subgame* $\Gamma^s(h)$.

Thus, the subgame perfect equilibrium concept applied to the self-control game requires that for each player i, for $P(h) = i$, System 2 cognition best responds by choosing a combined action $c^*_i \in C_i(h)$ consisting of a pure action $a_i \in A_i(h)$ and an effort $e_i \in E_i$ in the subgame $\Gamma^s(h)$ for every non-terminal history $h \in H \backslash Z$. This definition is contained within the normal definition

[11] In our applications, and in Myrseth and Wollbrant (2010), we make use of a contest success function where the probabilty in the expected value is at least quasiconcave in the probability and hence the solution to the problem is guaranteed.

of a subgame perfect equilibrium with two-dimensional choice sets, except for the fact that here the equilibrium is defined by System 2 preferences and the resolution mechanism.

Since most outcomes do not occur with certainty, we can state the outcome as a probability distribution over pure action profiles of the self-control game, which we refer to as the "subgame perfect outcome," which we will make frequent use of in our applications.

Lemma 2 (subgame perfect outcome) *The subgame perfect equilibrium outcome of the self-control game with perfect information Γ^s is the probability distribution $r(\cdot|s^*)$ over end nodes Z that results whenever all players' System 2 cognition adhere to their optimal strategies where r is a probability distribution*

$$r(\cdot|s^*) = \prod_{t=0}^{T} r(B_{i1}|_h, B_{i2}|_h). \qquad (6)$$

To illustrate, consider a two-period game with self-control conflicts at every node. In the final period, for a sufficiently high willpower, System 2 of the second mover will decide on a positive effort level. As a result, the pure action in this period is stochastic, i.e., a there is a probability weight on each of the two conflicted pure actions summing to one. In the first period, system 2 of the first mover decided (for a sufficiently high willpower) on a positive effort level leading to a probability weight. Thus, since both players decisions generated two probability weights, the probability of any outcome profile is determined by the product of these weights.[12]

Given Lemma 1, a self-control game can lead to two classes of equilibria, which we elaborate upon below.

Definition 15 (self-control equilibrium) *A self-control equilibrium is a pure action profile with the property that the resolution probability of each player i assigns probability one to the most preferred pure action element of System 2 cognition at each decision node, whenever it is 'i's turn to move.*

Intuitively, the pure action profile in this equilibrium occurs when each player is either strong enough or there is no visceral influence, so that every player behaves in accordance with her most preferred System 2 pure action. In other words, this occurs if and only if both players are perfectly controlled or there is no self-control conflict in the first place.

Theorem 1 *A self-control equilibrium is equal to the subgame perfect Nash equilibrium of the base game with System 2 preferences.*

Proof Let Θ^S denote the set of self-control equilibrium and Θ^N the set of subgame perfect Nash equilibria of the base game with System 2 preferences.

[12] This includes the degenerate case in which no effort is exerted and a single action at some node is played with certainty.

Whenever it is the case that $B_{i1}(h) \nsubseteq B_{i2}(h)$ for any history $h \in H, B_{i2}(h)$ specifies a combined action consisting of an effort $e^*(h)$ and a pure action $a^*(h)$. Let $a^*(h)$ be the most preferred System 2 pure action of a contested action pair. Given that $V_i(h, B_{i1}) > 0$, the resolution mechanism assigns probability one to $a^*(h)$ if and only if $\frac{c(e)}{w} = 0$, such that $e^*(h)$ is infinity. As $e^*(h)$ approaches infinity, the resolution probability $r(h)$ approaches one; as $\frac{c(e)}{w} = 0$ when $w \to \infty$, the player has perfect self-control. Hence, all elements of Θ^S have the property that $r(h)a^*(h) + (1 - r(h))a_i(h) = a^*(h)$.Then, player i will always play her most preferred pure action and the player always best responds in each subgame of the base game. Hence, action profiles in the self-control game with infinite willpower (perfect self-control) are subgame perfect Nash equilibria of the base game, and the set of self-control equilibria Θ^S is equal to the set of subgame perfect Nash equilibria Θ^N. □

Since we are interested in self-control problems, the class of equilibria that is most interesting is that where players neither are perfectly controlled (willpower is infinite) nor when there are no visceral influences. In addition, these "temptation equilibria" where players instead pursue suboptimal pure actions are the most important for applications. In fact, in all temptation equilibria, players put positive probability weights on some suboptimal choice leading to equilibria that differ from the standard subgame perfect predictions. We thus arrive at our main result.

Theorem 2 (temptation equilibrium) *A temptation equilibrium puts positive probability weights on non subgame perfect Nash equilibrium profiles of the base game.*

Proof Let Θ^T denote the set of temptation equilibria such that for all elements, if it is the case that $r(h)a^*(h) + (1 - r(h))a_i(h) \neq a^*(h)$. It follows from theorem 1 that for any $V_i(h, B_{i1}) > 0$, the resolution mechanism will not assign probability one to the pure action element of $B_{i2}(h)$, for any $d \in D(h)$, unless $\frac{c(e)}{w} = 0$, which for $c(e) > 0$ occurs only when $w \to \infty$. Hence, $B_{i2}(h)$ will lead to the resolution mechanism assigning positive probability to some pure action $a_i(h) \not\supseteq B_{i2}(h)$ for some history $h \in H$. □

Thus, whenever a temptation equilibrium is played, it implies that at least one player at some node has, in the face of some positive visceral influence, decided on an effort level such that a (System 2) utility-inferior pure action is played with some positive probability. Since this is the case, play of the game may result in behavior that departs from the optimal path of the base game at this node. With this possibility, behavior may depart from the subgame perfect path. Due to this result, it follows that temptation equilibria cannot be subgame perfect equilibria of the base game.

Corollary 1 *No temptation equilibrium is a subgame perfect equilibrium of the base game.*

Proof From Corollary 1, all elements of Θ^T contain at least one pure action $a^*(h)$, such that $r(h)a^*(h)+(1-r(h))a_i(h) \neq a^*(h)$. Consequently, player i does not play her best response with probability one at some node in the base game. Hence, no element of Θ^T is contained in Θ^N, such that $\Theta^T \cap \Theta^N = \varnothing$ □

Note that the pure action profiles of Θ^S may still result even if an element from Θ^T is being played (albeit not with probability one). It then appears as if the Θ^T is larger than Θ^S. In general, however, the set Θ^T is not always a super-set of Θ^S.

Corollary 2 *$\Theta^S \cap \Theta^T \neq \varnothing$ if the subgame perfect outcome does not assign zero probability to action profiles in Θ^S.*

Proof The set Θ^S contains only the most preferred pure action elements at each history, $a_i^*(h) \in A_i(h)$. For $\Theta^T \supset \Theta^S$ to hold, it must be the case that the temptation equilibrium profile assigns positive probability weight to the pure action element in Θ^S.Given the contested action pair $(B_{i1}(h), a_i^*(h))$ where $a_i^*(h)$ is the pure action element of System 2 cognition (otherwise $\Theta^T \supsetneq \Theta^S$), the resolution probability $r(h)$ must be positive for all $h \in H$, which occurs only when System 2 assigns positive a effort to the conflict. In turn, this occurs only if the cost of effort is sufficiently low, implying that willpower is sufficiently high as to equate marginal cost and benefit of effort □

Representation of a self-control game requires only little modification of the standard model. While only System 2 preferences would be represented in the standard framework, we now have two preference relations for each individual. Hence, it is necessary to specify both payoffs as part of the game. While the utility of System 2 would include the cognitive cost of self-control effort, we do not display these costs in the game tree prior to applying a solution concept, as we first need to solve the game in order to specify the self-control cost.

Definition 16 (Dual payoff representation) *The dual payoffs representation of a base game displays an $i \times m$ dimensional matrix of payoffs u_{im} at each end node, where each entry is a pair of payoffs π_{im}, each row $i = \{1, ..., N\}$ is a player and each column $m = \{1, 2\}$ is a cognition type; payoff for cognition System 1 is displayed to the left and payoff for System 2 to the right.*

$$u_{im} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ . & . \\ . & . \\ \pi_{N1} & \pi_{N2} \end{pmatrix}$$

Van Damme (1996) is conceptually close to our framework. He outlines a model of control costs where the agent has to suffer a cost in order to implement her best response. In our model, this is true for the set of contested actions only, since our model in principle allows for some costless best responses. Having outlined the general model of self-control games, we next apply it to instances of social interaction.

3 Applications: Modeling Social Interaction

In this section, we will apply the general framework to a couple of different games. Our goal here is two-fold. First, we will illustrate how the general framework operates, and, second, we will seek to show how the framework yields novel insights in games that are isomorphic to relevant real-world settings of strategic interaction. We consider three games, namely the classic sequential prisoner's dilemma, a trust game and a dictator game.

All three games involve players with pro-social preferences since these are more descriptive of most real people than are purely selfish preferences (see e.g., Fehr and Schmidt, 2006, or Elster, 2006, for surveys, and Kahneman et al., 1986a, and Kahneman et al., 1986b, for pioneering work in this area). Furthermore, and in line with Martinsson et al. (2010a, 2010b), we assume that players may experience a visceral influence, or temptation, to be purely selfish, which may act against their better judgment and thereby cause self-control conflict.[13] Damasio (1994) and Loewenstein (1996) argue that since most visceral influences cause aversive sensations in the individual, they direct attention inwards and undermine altruism in general (see also O'Donoghue and Loewenstein, 2007). Hauge et al. (2009) hypothesize that a higher cognitive load makes temptations harder to resist. Using a dictator game under different conditions of cognitive load in the spirit of Shiv and Fedorikhin (1999), they find no effect on giving. Curry et al. (2008), however, find that individuals with higher discount rates contribute less to a public good. That is to say, the more patient an individual is, the more she contributes, which is consistent with the hypothesis. Accounting for "conflict identification" in their design, Martinsson et al. (2010a; 2010b) provide direct evidence by showing that individuals with high self-control donate more in a dictator game and contribute more, both conditionally and unconditionally, in a public goods game once conflict has been identified, while pro-social behavior was no higher in conditions where individuals were less likely to identify conflict.[14]

While deriving implications for pro-social behavior, our applications mainly serve as a starting point to illustrate the workings of our model and does not directly rely on the assumptions made on underlying preferences. One can equally

[13] Consistent with this idea, Pronin et al. (2008) find that decisions about future, more abstract, temporally more distant "selves," resemble decisions about others. Both classes contrast with decisions about more present, "less abstract selves." See also Elster (1985).

[14] This suggests that the effect of conflict identification is not trivial and helps explain why evidence for the link between pro-social behavior and self-control has been hard to provide. Conflict identification and its effect in other domains are explored in Myrseth and Fishbach (2009a; 2009b).

imagine selfish individuals tempted by empathy to be kind as in the case of a man determined not to give money to beggars. When faced with a beggar's condition, however, he might act against his better judgment and donate. We begin by stating assumptions and results that will be used throughout. The first assumption is related to the specific functional form of the contest success function embedded in the resolution mechanism. Following Myrseth and Wollbrant (2010), we assume the ratio-form.[15]

Assumption (the probability of success) *The probability of success can be characterized by the functional form*

$$r_{it} = \frac{e^*_{it}}{e^*_{it} + v_{it}}. \tag{7}$$

Using the definition of the visceral stimulus $s = u_{i1}(B_{i1}) - u_{i1}(B_{i2}) \geq 0$, we further characterize the visceral influence as[16]

$$v_{it} = \theta_i \left(B_{i1}(h) - a_{i2}(h) \right), \text{ for } a_{i2}(h) \neq B_{i1}(h), \tag{8}$$

where $\theta_i \geq 0$ is a sensitivity parameter measuring how much individual i responds to a given visceral stimulus. In general, since the effect of θ_i could be modeled by simply changing System 1 preferences, this parameter would not be needed. In our case, however, we assume that System 1 has preferences over material payoffs. Thus, for intuitive appeal, we use θ_i to rescale System 1 utilities rather than denoting them explicitly.

Assuming (7) and solving the maximization problem in (3) for optimal combined action yields a reaction function in terms of System 2 payoffs from resisting and succumbing, willpower, and visceral influence. Using the ratio functional form for the resolution probability and linear self-control cost, we can write[17]

Lemma 3 (optimal effort) *Given a visceral influence for individual i in period t, v_{it}, optimal effort can be characterized generically in the case of two pure actions as*

$$e^* = -v^* + \sqrt{wv^* \left(u^*_{i2} - u_{i2} \right)} \tag{9}$$

$$r = \frac{e^*}{e^* + v^*} \tag{10}$$

[15] This implies that success probabilites are determined by the ratio of efforts rather than, as in the other common "logit functional form," determined by the difference in efforts.

[16] This is closely connected to the "opportunity based self-control" approach used by Fudenberg and Levine (2006).

[17] We omit the solution to the maximization problem; details are available upon request.

Where v^* denotes the visceral influence when optimal combined action is played. Corner solutions may be obtained when, for one of two reasons, it is optimal for System 2 to exert zero effort. First, zero effort is exerted for an insurmountable level of visceral influence. That is, the strength of temptation is so large that there is no gain to be made from attempting any conflict. Second, when the visceral influence is zero. That is to say, as visceral influences become smaller and smaller, exerted effort falls too. Eventually, visceral influence hits zero at which point zero effort is exerted too, since no effort will be needed to attain the most preferred pure action of System 2.

In some of our applications below pure action spaces consist of only two actions and so it is useful to state a generic expression for the resolution probability.

Lemma 4 (resolution probability) *Given that conflict payoff positivity holds and given optimal effort, we can write the resolution probability as*

$$r_{it} = 1 - \left[\frac{v_{it}}{[w_i (u_{i2}^* - u_{i2})]} \right]^{1/2}. \tag{11}$$

This greatly simplifies the analysis since, for each conflict we only need to characterize the difference $u_{i2}^* - u_{i2}$ together with the associated value of v_{it} to be able to state the resolution probability r_{it}. Given these results, we now proceed to illustrate the application of the model to a familiar setting.

3.1 A Sequential Prisoners Dilemma

Consider a standard sequential prisoners dilemma where $i = \{Bob, Jane\}$. We assume that Bob moves first and Jane second, and that the pure action spaces are $A_B = A_J = \{C, D\}$, "cooperate" or "defect." With material payoff-maximizing preferences this leads to the familiar outcome where Jane in each subgame always chooses to defect and hence, Bob at the beginning of the game chooses to defect too, since this secures a payoff of $2x$ rather than x.

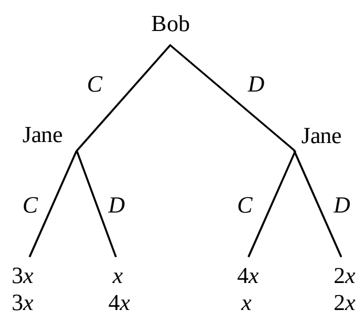


Figure 4. A Sequential Prisoner's Dilemma.

To extend this game to a self-control game, we first specify the preference relations of each cognition type $m = \{1, 2\}$ for both Bob and Jane. Let (selfish) System 1 cognition of each individual maximize individual material payoff represented by the utility function

$$u_{i1} = \pi_i, \tag{12}$$

so that u_{i1} is linear and equal to π_i. We will refer to the preferences of System 1 as *egoist*. To set the scene for cognition conflict, we assume that System 2 of each player is pro-social. Specifically, we assume that System 2 cognitions aims to maximize the sum of individual payoffs such that

$$u_{i2} = \pi_i + \alpha\pi_{-i}, \ \alpha = 1 \tag{13}$$

In line with the literature, we refer to these preferences as *altruist*. Although it is common to maintain the parameter α in analysis, we have restrained it to be equal to one. While it is not realistic that a player should care as much about the other player's payoff as she cares about her own, this is qualitatively unimportant for our results as long as $\alpha > 1/2$. Moreover, since we consider greed as a temptation, it is rather important that a player's System 2 cognition is "altruistic enough" since otherwise, greed does not lead to any problems. Given these assumptions, the dual payoff representation of this game is provided by Figure 5.

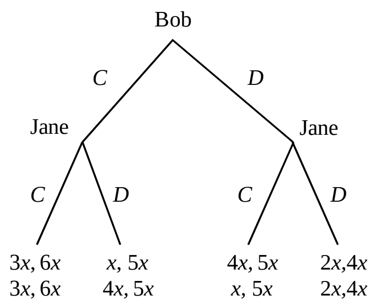


Figure 5. A Sequential Prisoners Dilemma in Dual Payoff form

The prisoner's dilemma game in Figure 5 illustrates the perfect information extensive game form with the dual payoffs denoted by the numbers at the end of branches. We follow the convention of the top numbers denoting first mover utility (Bob) and the bottom numbers the utility of the second mover (Jane). Recall that we display System 2 cognition (altruist) payoffs to the left in each bracket, while the materialist payoff is positioned to the right.

From Figure 5 we can deduce the preferences of players' cognition types of pure action profiles. For Bob's System 1 cognition we have the ranking $(D,C) \succeq_{B1} (C,C) \succeq_{B1} (D,D) \succeq_{B1} (C,D)$, while his System 2 cognition has $(C,C) \succeq_{B2} (D,C) \sim_{B2} (C,D) \succeq_{B2} (D,D)$. Similarly for Jane, we have that her System 1 cognition ranks pure action profiles $(C,D) \succeq_{J1} (C,C) \succeq_{J1} (D,D) \succeq_{2J1} (D,C)$, while her System 2 cognition ranks the pure action profiles $(C,C) \succeq_{J2} (D,C) \sim_{J2} (C,D) \succeq_{J2} (D,D)$. While the selfish preferences of System 1 are appropriate for capturing selfishness, the altruist preferences may represent altruism.

Once we have stated the preference relations over pure action profiles, however, it is clear both from rankings and Figure 5 that System 1 and System 2 conflict at all decision nodes, for both Bob and Jane. For example, after Bob plays D, Jane's System 2 (altruist) would like to play C instead of D, since that yields the altruist payoff of $6x$ rather than $4x$. Jane's System 1 (selfish) cognition, however, would prefer to play D instead of C since this yields material payoff of $2x$ rather than x. So, Jane's set of contested actions are $D(\varnothing, C) = D(\varnothing, D) = \{C, D\}$.

Let subscript it denote player i in period t throughout and let superscript denote exponents. Applying Lemma 1 to the game, we can state our first result as

Proposition 1 *Let $\Theta(a_B, a_J)$ denote a pure action profile of the game. The outcome probabilities of the sequential prisoner's dilemma are*

$$
\begin{aligned}
\Pr[\Theta(C,C)] &= \left[1-(\theta_B/2w_B)^{1/2}\right] \times \left[1-(\theta_J/2w_J)^{1/2}\right], \\
\Pr[\Theta(C,D)] &= (\theta_B/2w_B)^{1/2} \times \left[1-(\theta_J/2w_J)^{1/2}\right], \\
\Pr[\Theta(D,C)] &= (\theta_J/2w_J)^{1/2} \times \left[1-(\theta_B/2w_B)^{1/2}\right], \\
\Pr[\Theta(D,D)] &= (\theta_B/2w_B)^{1/2} \times (\theta_J/2w_J)^{1/2},
\end{aligned}
$$

where w_i denotes willpower and θ_i is the visceral stimulus parameter.

Proof We solve the game using backward induction. We start by solving simultaneously for the two subgames following Bob either playing C or D. By symmetry, we need only solve for the resolution probability once. First, note that $v_{J2} = \theta_J x$, for the histories where Bob played either C or D.Using Lemma 2 and Lemma 3, we obtain

$$r_{J2} = 1 - (\theta_J/2w_J)^{1/2}, \tag{14}$$

where r_{J2} is the probability that Jane plays cooperate. Thus, in both subgames, Jane will resist temptation and play C with probabilityr_{J2} and succumb and play D with probability

$$(1 - r_{J2}) = (\theta_J / 2w_J)^{1/2}.$$

Now consider Bob. Knowing r_{J2}, he needs to resolve his choice in period 1. For Bob's System 2 cognition, playing C is always at least as good as playing D. In contrast, for his System 1 cognition, playing D is always better than playing C. Hence, Bob's set of contested actions is $D(\varnothing) = \{C, D\}$. Proceeding in the same manner as for Jane, we first note that Bob's visceral influence can be stated as

$$v_{B1} = \theta_B x.$$

Again using Lemma 2 and Lemma 3, we obtain

$$r_{B1} = 1 - (\theta_B / 2w_B)^{1/2}.$$

Using Lemma 1, we obtain the whole outcome probability distribution. □

Corollary 3 *The outcome $\Theta(C, C)$ is played with probability one, if and only if both players possess infinite willpower.*

Proof Consider the probability that pure action profile $\Theta(C, C)$ results: $\Pr[\Theta(C, C)] = \left[1 - (\theta_B / 2w_B)^{1/2}\right] \times \left[1 - (\theta_J / 2w_J)^{1/2}\right]$. The limit with respect to w_B is

$$\lim_{w_B \to \infty} \left[1 - (\theta_B / 2w_B)^{1/2}\right] \times \left[1 - (\theta_J / 2w_J)^{1/2}\right] = \left[1 - (\theta_J / 2w_J)^{1/2}\right]. \quad (15)$$

In turn, the limit of (15) with respect to w_J is

$$\lim_{w_J \to \infty} \left[1 - (\theta_J / 2w_J)^{1/2}\right] = 1, \quad (16)$$

confirming the intuition that when both altruistic players have infinite willpower, they should play the outcome (C, C) with certainty. From Theorem 1, this is a self-control equilibrium and a subgame perfect Nash equilibrium of the base game with altruistic preferences. □

Corollary 4 *The outcome $\Theta(C, C)$ is played with probability zero if either player is infinitely sensitive to visceral stimulus.*

Proof Consider again the probability that pure action profile $\Theta(C, C)$ results: $\Pr[\Theta(C, C)] = \left[1 - (\theta_B / 2w_B)^{1/2}\right] \times \left[1 - (\theta_J / 2w_J)^{1/2}\right]$.The limit with respect to θ_B is

$$\lim_{\theta_B \to \infty} \left[1 - (\theta_B/2w_B)^{1/2}\right] \times \left[1 - (\theta_J/2w_J)^{1/2}\right] = 0, \tag{17}$$

confirming the intuition that if either of the altruistic players is infinitely sensitive to visceral stimulus, the cooperative outcome (C, C) is impossible to obtain. □

A subgame perfect equilibrium analysis applied to the base game with altruistic preferences would lead to both players choosing to cooperate. Here we see that this outcome occurs only as a limiting case when both players either posses infinite willpower, or are completely insensitive to visceral stimulus. In all other cases, there is room for players to act inconsistently in the sense that they both prefer to cooperate but might act against their better judgement and defect. Due to the particular payoff structure above, both players are tempted at all nodes irrespective of what the other player does, ruling out certain results. Next we consider a game in which this is not always the case.

3.2 A Trust Game

Again consider Bob and Jane, this time facing a trust game like the one discussed in the introduction. For convenience, below we reproduce the game in Figure 1 in dual payoff form. As first mover, Bob has to decide whether to $Trust$ Jane or play $Don't$. In case Bob plays $Don't$, both get a material payoff of 1. In case Bob decides to trust Jane, she has to decide whether to $Share$ the payoff with Bob or $Grab$. In case she shares, both get a material payoff of 2. In case she grabs, she gets a payoff of 3 while Bob gets nothing. With altruistic preferences, as expressed in (13), the solution to this game is straightforward. At the second stage, Jane will prefer to $Share$ and hence Bob will $Trust$. Having established the standard subgame perfection outcome (and thus also the self-control equilibrium) we now solve the self-control version. Since there are only two periods and each player only has one decision, we suppress the time notation.

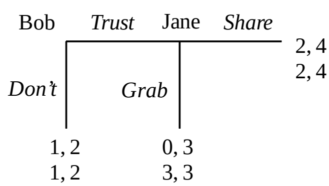


Figure 6. A Trust Game in Dual Payoff form

Proposition 2 *Let $\Theta(a_B, a_J)$ denote a pure action profile of the game. If the probability of Jane playing share is less than a half, $r_J < 1/2$, the outcome probabilities of the trust game are*

$$
\begin{aligned}
\Pr[\Theta(Trust, Share)] &= \left[1-\left(\frac{\theta_B\,(1-2r_J)}{w_B\,(r_J+1)}\right)^{1/2}\right] \times \left[1-\left(\frac{\theta_J}{w_J}\right)^{1/2}\right], \\
\Pr[\Theta(Trust, Grab)] &= \left[1-\left(\frac{\theta_B\,(1-2r_J)}{w_B\,(r_J+1)}\right)^{1/2}\right] \times \left(\frac{\theta_J}{w_J}\right)^{1/2}, \\
\Pr[\Theta(Don't, Share)] &= \left(\frac{\theta_B\,(1-2r_J)}{w_B\,(r_J+1)}\right)^{1/2} \times \left[1-\left(\frac{\theta_J}{w_J}\right)^{1/2}\right], \\
\Pr[\Theta(Don't, Grab)] &= \left(\frac{\theta_B\,(1-2r_J)}{w_B\,(r_J+1)}\right)^{1/2} \times \left(\frac{\theta_J}{w_J}\right)^{1/2}, \\
where\ r_J &= 1-\left(\frac{\theta_J}{w_J}\right)^{1/2}.
\end{aligned}
$$

Proof Jane will always experience a self-control conflict since payoff positivity holds ($4 > 3$) at the same time as Jane's System 1 prefers to grab. Her visceral influence is $v_J = \theta_J\,(3-2)$. For Bob, payoff positivity always holds ($4r_J + (1-r_J)3 > 2$), but System 1 will prefer $Don't$ whenever $r_J < 1/2$ (since then $1 > 2r_J$), giving rise to self-control conflict as $v_B = \theta_B\,(1-2r_J)$. From Lemma 3 and 4, we derive the probabilities.□

Since Jane always suffers a self-control problem, the pivotal issue is Bob's self-control problem. From the result we see, however, that the extent of Bob's self-control problem depends on Jane's resolution probability. In particular, the severity of Bob's problem diminishes as r_J increases (indeed it is equal to one when $r_J \geq 1/2$) and vice versa. In fact, when $r_J = 0, r_B = 1 - \sqrt{\theta_B/w_B}$, similar to r_J in that it only contains the willpower and sensitivity parameters. In this way, Jane's willpower has a dual effect on the likelihood of the profile $(Trust, Share)$ occurring. Though it initially looks as if Bob will be the pivotal player in determining the outcome, it is in fact Jane who has the opportunity to solve everyone's problem while Bob is the victim of both his own and Jane's lack of willpower.

As Jane's willpower is insufficient to solve the problem, a way to solve the problem is for Bob to offer Jane a transfer such that she will manage to choose *Share* rather than *Grab*.

Corollary 5 Bob can make a transfer T to Jane such that Jane's self-control problem is resolved, resulting in Bob playing $Trust$.

Proof It is straightforward to see that such a transfer is possible. Both Bob and Jane get a material payoff of 2 if Jane decides to *Share*. However, if Jane *Grabs*, she will get a payoff of 3. Bob can then transfer 1 to Jane such that after sharing, she gets 3 and Bob gets 1. Bob's material payoff from $Don't$

is also 1, so that from a a selfish point of view he is indifferent and will feel no greed. However, from a social point of view, $Trust$ is better than $Don't$ since now $4 > 2$, hence, Bob will trust.□

Of course, while $T = 1$ solves Jane's self-control problem, it might not be the optimal choice for Bob since it turns out to be quite costly. From a material point of view, it might be better for him to transfer something slightly less such that, given Jane's self-control problem, Bob gets a higher material payoff in expectation than when $T = 1$. Since Bob is maximizing the sum of payoffs, however, this mechanism is irrelevant in the present case since the transfer only amounts to a redistribution of payoffs. Hence, Bob's payoff is 4 both before and after the transfer.

To test the theory, one can take a cue from Proposition 2. One approach to testing this result is to randomly match participants into either the role of Bob or Jane, such that one has to make the decision of trusting, and the other of sharing. Since r_B is an increasing function of r_J and w_B, it appears that a relevant approach would be to collect "Bob's" beliefs about "Jane's" behavior in order to proxy r_J. In addition, one could collect measures of self-control such as the Rosenbaum Self-Control Schedule (henceforth RSS) (Rosenbaum, 1980) in order to proxy w_B. An increase in the interaction $r_J\ w_B$ should lead to an increase in r_B, in turn resulting in more trusting behavior for individuals with higher self-control. The hypothesis is then that an interaction of beliefs and the self-control measure should positively predict transfers in the trust game.

3.3 A Dictator Game

So far we have limited the analysis to games that include only discrete choice sets. To show that the framework also accommodates continuous choice sets, we make use of a simple version of the standard dictator game. In this game, donating is inherently efficient such that whenever a player transfers a fraction α of the endowment E, the recipient receives $(1+\epsilon)\,\alpha E$, where the interpretation of $\epsilon > 0$, is that of a very small number, as is conventional. We show that when an altruistic player is tempted to be selfish, she will face, not only a self-control conflict leading to a stochastic choice between transferring a zero amount, thus being completely selfish, and some other positive transfer, but also that this positive transfer itself is lower than what is optimal in the absence of a self-control problem.

Consider a situation in which Bob has an endowment E from which he can transfer a fraction α to Jane. The material payoffs become

$$\begin{aligned} \pi_B &= (1-\alpha)E, \\ \pi_J &= (1+\epsilon)\,\alpha E. \end{aligned}$$

Assume further, in line with (13), that Bob has an altruistic utility function, while he is tempted to be completely greedy and maximize material payoffs. Then we can write

$$u_{B2} = (1-\alpha)\,E + \alpha\,(1+\epsilon)\,E.$$

In the absence of any self-control problem, Bob would choose to transfer his whole endowment so that $\alpha = 1$. The visceral stimulus arises from the difference between the impulse to give nothing, following the π_B maximizing decision of setting $\alpha = 0$, and the altruistic decision of transferring some positive amount, evaluated with u_{B2}. The visceral stimulus thus becomes

$$E - ((1-\alpha)E) = \alpha E.$$

Using a quadratic[18] visceral stimulus function, we attain[19]

$$v_B = (\alpha E)^2 .$$

Note that should Bob succumb to temptation and give nothing, his utility will be equal to zero. The maximization problem for Bob thus becomes

$$\max_{\alpha,e} E[u(\alpha,e)] = \frac{e\epsilon\alpha E}{e+(\alpha E)^2} + E - \frac{e}{w}$$

where choosing α and e is tantamount to choosing the optimal combined action.

Proposition 3 *The optimal transfer is given by* $\alpha^* = \frac{\epsilon w_B}{4E}$. *This is equal to one when* $\frac{4E}{\epsilon} \leq w_B$, *i.e., when willpower is sufficiently large.*

It is interesting to note that the optimal transfer depends on the size of the endowment. The intuition for this is that when the endowment is larger, the material payoff to withhold becomes larger, leading to a stronger temptation. Technically, this is because the distance between the maximum material payoff to be obtained and a given fraction transferred grows, and so the visceral stimulus increases.

Proof $\frac{\partial E[u(\alpha,e)]}{\partial \alpha} = \frac{2\alpha^2 e\epsilon E^3 - e^2\epsilon E - \alpha^2 e\epsilon E^3}{(e+\alpha^2E^2)^2} = 0$ implies that $e = \alpha^2E^2$. Using Lemma 3, we have $e^* = -\alpha^2E^2 + \sqrt{w_B\alpha^2E^2[aE]}$. Setting $e = e^*$ yields the result.□

This result predicts that with imperfect self-control, players should not only experience lower chances of successfully leveraging their willpower, but should also be expected to lower their intended transfers (i.e. α^*). This is thus an instance where the size of the endowment and relative income should matter for pro-social behavior as the next result illustrates.

[18] In order not to attain solutions of $\alpha \in (0,1)$, it can be shown that the visceral influence function must be convex enough to attain $\alpha > 1$, but not "too convex," or else $\alpha = 0$, since cost of self-control in essence becomes too high.

[19] For simplicity, we set $\theta_B = 1$.

Corollary 6 *Increasing willpower has a positive effect on intended transfer. This effect is stronger when the endowment is relatively small.*

Proof The derivative $\frac{\partial \alpha^*}{\partial w} = \frac{\epsilon}{4E} > 0$ while $\frac{\partial \alpha^*}{\partial w \partial E} = \frac{-4\epsilon}{E^2} < 0.\Box$

The intuition for this result is that while self-control indeed has a positive effect, this effect is smaller when the visceral influence i.e., temptation, is stronger due to a larger endowment. This can reconcile conflicting evidence in experimental economics in which stake size sometimes matters and sometimes does not (e.g. Carpenter et al., 2005; Forsythe et al., 2004; Johansson-Stenman et al., 2005). A suggestion for how to test this might be to use experimental treatments varying the stake size, but at the same time collect measures of self-control such as the Rosenbaum Self-Control Schedule (Rosenbaum, 1980). The hypothesis is that self-control measures should be more significant for treatments in which endowments are small. A due caveat is that the visceral influence as a function of the stimulus provided by the availability of money has to be "convex enough" for the relevant increment in endowment in order to increase self-control costs.

4 Discussion and concluding remarks

Emotions have a powerful influence on behavior, not only by augmenting payoffs, but also through their action tendencies. These emotions, also termed visceral influences, are structured and possible to incorporate into models of strategic interaction. Furthermore, visceral influences are often not aligned with an individual's preferences, resulting in a self-control conflict between the individual's two systems of cognition. This is important for strategy since players not only have to know the preferences of others, but also consider the emotions that might be generated by their own and others' behavior, as well as the abilities of all players to implement their strategies. We developed a framework for analyzing self-control in extensive games with perfect information which in principle allows for any kind of preference relation conflicting with any kind of visceral influence.

We showed that self-control equilibria are equal to Nash equilibria of the base game. This accords well with the notion that as an agent's cognitive powers approach those of homo economicus, behavioral predictions should coincide. We also showed that temptation equilibria can be a super set of self-control equilibria, when players' self-control was not "high enough" so that the intersection between the two sets was non empty. Regardless of the type of equilibrium resulting, however, an outcome probability matrix can be derived, stating the probability of each possible action profile resulting.

We demonstrated the usefulness of this approach by applying it to three games where players are assumed to hold pro-social preferences, but are tempted by greed to act in a selfish manner. Exploring pro-sociality vs. selfishness makes application particularly simple, since both selfish and pro-social behavior have been extensively studied. In reality, however, people are probably pro-social to some degree, and in some cases, behavioral outcomes reflect the resolution of self-control conflict. We thus showed in the prisoners dilemma game that mutual cooperation only occurs with certainty when both players have infinite willpower. Similarly, mutual cooperation can never occur if either player is infinitely sensitive to visceral stimulus. These two results confirm intuitions about when pro-social individuals might be expected to cooperate and when they might be expected to defect. Other applications to social interaction, where emotions such as shame, fear, and anger are likely to play a prominent role, is likely to prove very fruitful

In the trust game, we demonstrated how self-control problems can arise endogenously and derived novel predictions relevant both for self-control problems and for issues of trust. First, when considering the behavior of other players, temptations may arise and lead to self-control problems, revealing the complexity of self-control strategies. For example, the severity of one player's self-control problem may depend on the problem of another player. A players willpower may have a dual role in trust games by affecting both her own self-control problem and that of another player. Second, if the probability of a partner acting prosocially is sufficiently low, payoff positivity might not be satisfied, and hence a player anticipating this may prefer not to trust. In other instances, such as in the dictator game, it may lead to the player choosing to aim for less than optimal transfers simply because the implied self-control conflict would be too costly to engage in. In this way, even if the agent is successful at the self-control conflict, she would transfer less than implied in the absence of a self-control conflict. The dictator game also yielded an important prediction for self-control in general, i.e., that the optimal transfer depended negatively on the size of the endowment. This accords well with the intuition that the extent of opportinuty for selfish action should matter and illustrates how greed can be generated both in instances where individuals have to give, and when they are presented with chances to keep.

References

[1] Ainslie, G. (1975). "Specious reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*, 82, 463-496.

[2] Ainslie, G. (2001). "Precis of Breakdown of Will." *Behavioral and Brain Sciences*, 8:5, 635-650.

[3] Battigalli, P. and Dufwenberg, M. (2007). "Guilt in Games." *The American Economic Review*, 97:2, 170-176.

[4] Battigalli, P. and Dufwenberg, M. (2009). "Dynamic Psychological Games." *Journal of Economic Theory*, 144:1,1-35.

[5] Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). "Ego depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology*, 74,1252 – 1265.

[6] Baumeister, R.F., Heatherton, T.F., and Tice, D.M. (1994). "Losing Control: How and Why People Fail at Self-Regulation." San Diego, CA: Academic Press.

[7] Bénabou, R., and Pycia, M. (2002). "Dynamic Inconsistency and Self-Control: A Planner–Doer Interpretation." *Economic Letters*, 77, 419–424.

[8] Bénabou, R. and Tirole, J. (2004). "Willpower and Personal Rules." *Journal of Political Economy*, 112, 848-886.

[9] Carpenter J.,. Verhoogen, E. and Burks, S. (2005). "The Effect of Stakes in Distribution Experiments." *Economics Letters*, 86, 393–398.

[10] Curry, O., Price, M.E., and Price, J. G. (2008). “Patience is a Virtue: Cooperative People Have Lower Discount Rates.” *Personality and Individual Differences*, 44(3): 780-85.

[11] Damasio, A. R. (1994). "Descartes’ Error: Emotion, Reason, and the Human Brain." New York: Putnam.

[12] Dufwenberg, M. and Kirschsteiger, G. (2004). "A Sequential Theory of Reciprocity." *Games and Economic Behavior*, 47:2, 268-298.

[13] Elster, J. (1977). "Ulysses and the Sirens: A Theory of Imperfect Rationality." *Social Science Information*, 16:5, 469-526.

[14] Elster, J. (1985). "Weakness of Will and the Freerider Problem." *Economics and Philosophy*, 1, 231-265.

[15] Elster, J. (1998). "Emotions and Economic Theory." *Journal of Economic Literature*, 36, 47-74.

[16] Elster, J. (2006). "Altruistic Behavior and Altruistic Motivations." ch. 3 in Kolm, S. and Mercier Ythier, J. Eds. (2006). *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1. Handbooks in Economics 23, North-Holland: Amsterdam.

[17] Fehr, E. and Schmidt K. M. (2006). "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories." ch 8 in Kolm, S. and Mercier Ythier, J. Eds. (2006). *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1. Handbooks in Economics 23, North-Holland: Amsterdam.

[18] Forsythe, R., Horowitz, J. Savin, N., and Sefton, M. (1994). "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior* , 6, 347-369.

[19] Fredericks, S., Loewenstein, G. and O'Donoghue, T. (2002). "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 15, 397-404.

[20] Fudenberg, D. and Levine, D.K. (2006). "A Dual Self Model of Impulse Control." *The American Economic Review*, 96:5, 1449-1476.

[21] Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989). "Psychological Games and Sequential Rationality." *Games and Economic Behavior*, 1, 60-79.

[22] Gul, F. and Pesendorfer, W. (2001). "Temptation and Self-Control." *Econometrica*, 69,1403-1436.

[23] Gul, F. and Pesendorfer, W. (2004). "Self-Control and the Theory of Consumption." *Econometrica*, 72, 119-158.

[24] Hauge, K., Brekke, K., Johansson, L., Johansson-Stenman, O., and Svedsäter, H. (2009). "Are Social Preferences Skin Deep? Dictators under Cognitive Load." *Working Papers in Economics*, 371, The University of Gothenburg.

[25] Hirshleifer, J. (1989). "Conflict and Rent-Seeking Success Functions: Ratio vs. Difference Models of Relative Success." *Publ. Choice*, 63, 101-112.

[26] Johansson-Stenman, O., Mahmud, M. and Martinsson, P. (2005). "Does Stake Size Matter in Trust Games?" *Economics Letters*, 88:3, 365-369.

[27] Kahneman, D. Knetsch, J. L. and Thaler, R. H. (1986a). "Fairness as a Constraint on Profit-Seeking: Entitlements in the Market." *The American Economic Review*, 90:2, 272-292.

[28] Kahneman, D. Knetsch, J. L. and Thaler, R. H. (1986b). "Fairness and the Assumptions of Economics." *Journal of Business*, 59:4, 285-300.

[29] Kahneman, D. (2003). "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist*, 58, 697-720.

[30] Laibson, D.I. (1997). "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112, 443-477.

[31] Loewenstein, G. (1996). "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65, 272-292.

[32] Loewenstein, G. (2000). "Preferences, Behavior and Welfare: Emotions in Economic Theory and Economic Behavior." *The American Economic Review*, 90:2, 426-432.

[33] Loewenstein, G. and O'Donoghue, T. (2007). "The Heat of the Moment: Modeling Interactions Between Affect and Deliberation." http//:www.arts.cornell.edu/econ/edo1/heat.pdf.

[34] Martinsson, P., Myrseth, K. O. R., and Wollbrant, C. (2010a). "Reconciling Pro-Social vs. Selfish Behavior: Evidence for the Role of Self-Control." Working Paper, the University of Gothenburg.

[35] Martinsson, P., Myrseth, K. O. R., and Wollbrant, C. (2010b). "Conditional Cooperation and Self-Control." Working paper, the University of Gothenburg.

[36] Metcalfe, J., and Mischel,W. (1999). "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower." *Psychological Review*, 106, 3–19.

[37] Muraven, M., and Baumeister, R. F. (2000). "Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?" *Psychological Bulletin*, 126, 257 – 259.

[38] Myrseth, K. O. R. and Fishbach, A. (2009a). "Self-Control: A Function of Knowing When and How to Exercise Restraint." *Current Directions in Psychological Science*, 18, 247 - 252.

[39] Myrseth, K. O. R., and Fishbach, A. (2009b). "Activating Self-Control: Isolated vs. Interrelated Temptations." Working paper, the University of Chicago.

[40] Myrseth, K. O. R., and Wollbrant, C. (2010). "A Theory of Self-Control Conflict: The Pyrrhic Motions of Reason and Passion." Working paper, the University of Gothenburg.

[41] McKelvey, R., and Palfrey, T. (1995). "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior*, 10, 6–38.

[42] McKelvey, R., and Palfrey, T. (1998). "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1, 9–41.

[43] O'Donoghue, T. and Rabin, M. (1999). "Doing it Now or Later?" *The American Economic Review*, 89:1, 103-124.

[44] Osborne, M. J. and Rubinstein, A. (1994). "A Course in Game Theory." Cambridge Massachusetts and London England: The MIT Press.

[45] Plato. (1986). "Protagoras." In *The Dialogues of Plato*, Trans. Anthony F. Hubbard and Ellen S. Karnofsky, New York: Bantam Books. (Orig. pub. 380 B. C.).

[46] Pronin, E., Olivola, C. Y. and Kennedy, K. A. (2008). "Doing Unto Future Selves as You Would Do Unto Others: Psychological Distance and Decision Making." *Personality and Social Psychology Bulletin*, 34:2, 224-236.

[47] Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *The American Economic Review*, 83:5, 1281-1302.

[48] Rosenbaum, M. (1980). "A Schedule for Assessing Self-Control Behaviors: Preliminary Findings." *Behavior Therapy*, 11, 109-121.

[49] Schelling, T.C. (1984). "Self-Command in Practice, in Policy, and in a Theory of Choice." *The American Economic Review*, 74:2, 1-11.

[50] Shiv, B., and Fedorikhin, A. (1999). "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making." *The Journal of Consumer Research*, 26:3, 278-292.

[51] Skaperdas, S. (1996). "Contest Success Functions." *Economic Theory*, 7:2, 283-290.

[52] Sloman S.A. (1996). "The Empirical Case for Two Systems of Reasoning." *Psychological Bulletin*, 119, 3-22.

[53] Strotz, R. H. (1955-56). "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies*, 23, 165-180.

[54] Thaler, R. H. and Shefrin, H. M. (1981). "An Economic Theory of Self-Control." *Journal of Political Economy*, 89:2, 392-406.

[55] Trope, Y., and Fishbach, A. (2000). "Counteractive Self-Control in Overcoming Temptation." *Journal of Personality and Social Psychology*, 79, 493-506.

[56] Van Boven, L., and Loewenstein, G. (2003). "Social Projection of Transient Visceral Feelings." *Personality and Social Psychology Bulletin*, 29, 1159–1168.

[57] Van Damme, E. (1996). "Stability and Perfection of Nash Equilibria." 2nd Ed. Berlin, Germany: Springer-Verlag.

Chapter 3

Reconciling Pro-Social vs. Selfish Behavior
- Evidence for the Role of Self-Control

Peter Martinsson, University of Gothenburg[A]
Kristian Ove R. Myrseth, ESMT European School of Management and Technology[B]
Conny Wollbrant, University of Gothenburg[C]

Abstract

We test the proposition that individuals may experience a self-control conflict between short-term temptation to be selfish and better judgment to act pro-socially. Using a public goods game and a dictator game, we manipulated the likelihood that individuals identified self-control conflict, and we measured their trait ability to implement self-control strategies. Consistent with our hypothesis, we find that trait self-control exhibits a positive and significant correlation with pro-social behavior in the treatment that raises likelihood of conflict identification, but not in the treatment that reduces likelihood of conflict identification.

JEL Classification: D01, D03, D64, D70.
Keywords: self-control, pro-social behavior, altruism

Acknowledgments: Financial support from the Swedish Research Council (*Vetenskapsrådet*) and the Swedish International Development Cooperation Agency (Sida) to the Environmental Economics Unit at the University of Gothenburg is gratefully acknowledged. We are grateful to Katarina Nordblom and Robert Östling for helpful comments. Horacio Antonio Villegas Rivera provided excellent logistical support and Angela Maria Gonzales provided excellent research assistance.
[A] Department of Economics, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 52 55; Fax +46 31 786 10 43; E-mail: peter.martinsson@economics.gu.se
[B] ESMT European School of Management and Technology, Schlossplatz 1, 10178 Berlin, Germany; Ph +49 (0) 30 212 31 1529; Fax +49 (0) 30 212 31 1281 E-mail: myrseth@esmt.org
[C] Department of Economics, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 26 15; Fax +46 31 786 10 43; E-mail: conny.wollbrant@economics.gu.se

1. Introduction

Lured by temptation, individuals may find themselves acting against their better judgment. Self-control failure, famously termed "akrasia" in Plato's Protagoras (Plato, 1986/B.C. 380) represents a central issue of both philosophy and modern-day social sciences. Problems of self-control persist throughout domains of our daily life.[1] For example, the dieter faced with the opportunity to indulge in a delicious creamy cake may perceive a conflict between indulging and maintaining a good figure. The student may feel conflicted between the desire to go to the cinema and her better judgment to stay home and study (e.g. O'Donoghue and Rabin, 1999; Burger *et al.*, 2009). And, similarly, the fashionista might feel conflicted between the temptation to purchase new boots and her better judgment to maintain a responsible budget.

Perhaps less intuitively, but no less importantly, the question of pro-social versus selfish behavior may be understood in similar terms. This conceptualization may help reconcile conflicting notions in economics of the selfish Homo Economicus and the pro-social Homo Behavorialis. That individuals should care much about their own self-interest seems almost tautological and requires little further exposition, but that individuals also should care about the interests of others – even at the expense of those of their own – has attracted significant interest (for overview on social preference see e.g. Fehr and Schmidt, 2006).[2] For example, many individuals voluntarily contribute to charity or to public goods (e.g. recycling), and they pay their taxes even though economic theory expects them not to, given low likelihood of punishment.[3] Nonetheless, one could imagine that even individuals of pro-social inclination on occasion may feel tempted to act selfishly and hence underreport income to the authorities. That is, pro-social preferences potentially fly in the face of basic urges for personal gain – or greed – and the individual may thus experience a self-control conflict between better judgment to act pro-socially and temptation to act selfishly.

---

[1] For work on self-control and time inconsistency, see e.g. hyperbolic and quasi-hyperbolic discounting models by Strotz (1955) and Laibson (1997), the "planner-doer" model by Thaler and Shefrin (1981), and the dual-self model by Fudenberg and Levine (2006).

[2] For pioneering work in this area see e.g., Kahneman *et al.*, (1986a; 1986b).

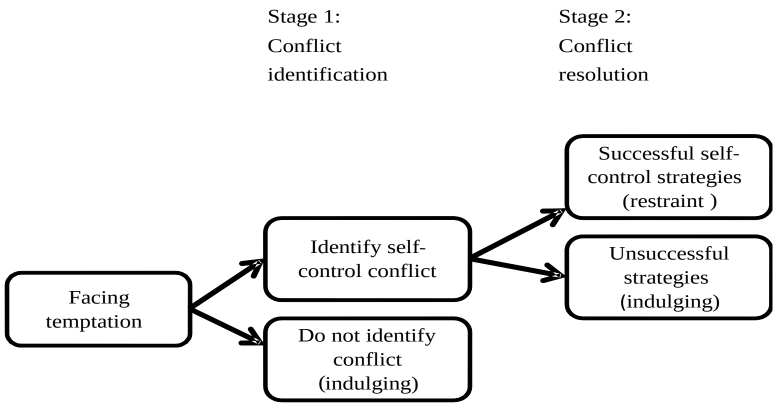
[3] There exists an extensive literature on the motivation behind pro-social behavior. For example, Bénabou and Tirole (2006) classify the motivations into three broad categories: intrinsic, extrinsic and image motivation, and a similar classification is found in Ariely *et al.* (2009)

Only recently has the psychological literature started to explore how the question of pro-social versus selfish behavior relates to that of self-control. Loewenstein (1996; 2000) suggests that selfish behavior may be motivated by visceral urges or drive-states, resembling cravings for relief of hunger, pain, and sexual deprivation. O'Donaghue and Loewenstein (2007) argue that such selfish urges often may conflict with the "colder," more abstract preferences for altruism, as visceral urges for sweets could conflict with more abstract preferences for a fine figure or good health. At present, there is but preliminary evidence for this idea. Most notably, Pronin *et al.* (2008) show that decisions about others resemble decisions about "future selves," both classes of which contrast to decisions about less abstract "present selves." Moreover, Curry *et al.* (2008) find in a standard public goods game that individuals' discount rates are negatively associated with their contribution to the public good. That is, more "impatient" individuals contributed less to the public good than did "patient" ones. While indeed supportive of the hypothesis that the question of pro-social versus selfish behavior may represent a problem of self-control, existing empirical evidence is not conclusive.

In this paper we attempt a direct test of the hypothesis that pro-social versus selfish behavior may represent a self-control problem. In so doing, we rely on two conditions necessary for successfully exercising restraint in the face of temptation; Myrseth and Fishbach (2009) propose a two-stage model of self-control, which postulates that an individual in the face of temptation first identifies conflict or not between indulging and pursuing a higher-order goal and, second, that the individual next employs self-control strategies only if conflict was identified at the first stage (see Figure 1). Such self-control strategies may take a variety of forms, and common examples include willpower (e.g., Baumeister *et al.*, 1994), and pre-commitment (Thaler and Shefrin, 1981; Schelling, 1984).

Critically, self-control strategies are relevant to the decision to indulge only when the individual has identified self-control conflict. Therefore, one strategy for investigating whether the problem of pro-social versus selfish behavior resembles one of self-control is to test whether self-control strategies are positively associated with pro-social behavior when individuals have identified self-control conflict, but less so or not at all when individuals have not.

Figure 1. The two-stage model of self-control.



Source: Myrseth and Fishbach (2009).

Determinants of conflict identification in the face of temptation have been explored only recently. In some contexts, the question is almost trivial and identification of conflict virtually obvious. For example, the diabetic dieter probably knows that having even a single, tempting chocolate may incur major costs. However, the question of self-control conflict is more ambiguous for the non-diabetic dieter who faces the same chocolate. Having this one chocolate alone will not incur major costs, but doing so regularly might. Similarly, the good citizen may find that not reporting his annual income would represent a major threat to his self-image (and possibly also to his criminal record), but failing to report but a few small windfalls is a more ambiguous matter. Myrseth and Fishbach (2010) use the term epsilon cost temptation to denote tempting opportunities that incur nothing but trivial costs when consumed in small amounts but potentially serious costs when consumed extensively. They argue that individuals identify self-control conflict in the face of epsilon cost temptation if and only if two conditions are met: (a) the focal consumption opportunity must be viewed in relation to multiple additional opportunities, and (b) the decision maker must assume that similar choices are made for each opportunity. That is, considering the question of whether or not to have a delicious creamy cake will evoke self-control conflict in the dieter if the dining opportunity is viewed in relation to future opportunities for dessert consumption, but not if the dining opportunity is viewed in isolation, as a singular episode. Similarly, the question of whether or not to withhold from the tax authorities a few small amounts may elicit self-control conflict in the good citizen if the income reporting is viewed in relation to future reports, but not if the reporting is viewed in isolation.

Myrseth and Fishbach (2010) show that subtle framing manipulations are sufficient to influence identification of self-control conflict in the face of epsilon cost temptation. They find that presenting a calendar displaying the current month with a grid separating the dates increased participants' subsequent consumption of potato chips compared to participants whom were presented a non-gridded calendar of the current month. The reason for this, they argue, is that the gridded calendar activated an isolated (versus interrelated) frame of the choice opportunity; it made participants more likely to isolate the date in question and thus less likely to see the decision task in relation to similar future opportunities. Consequently, the grid reduced the likelihood that participants would identify a conflict between the temptation to have chips and long-term health or dieting goals. Indeed, participants who were viewing the gridded calendar reported that they experienced less conflict during their decision to have chips or not than did those who were viewing the non-gridded calendar.[4] Furthermore, participants' trait ability to implement self-control strategies, measured by Rosenbaum's (1980) psychometric scale[5], positively predicted chips consumption for those who were viewing the calendar without the grid (and who identified conflict) but not for others who were viewing the calendar with (and who were less likely to identify conflict). That is, participants who viewed the calendar without the grid, more likely than those who viewed the calendar with, identified self-control conflict and therefore leveraged their self-control strategies to resist the tempting chips.

To explore our hypothesis that the problem of pro-social versus selfish behavior may represent one of self-control, we have applied the empirical strategy from Myrseth and Fishbach

[4] Experienced conflict was assessed by averaging participants' answers to two questions: (1) to what extent they felt mixed feelings when deciding whether or not to have more potato chips, and (2) to what extent they felt conflicted when deciding whether or not to have more chips. The questions were posed immediately after participants finished consuming potato chips.

[5] The Rosenbaum scale is further discussed in the Experimental Design section. Also see Appendix B.

(2010) in two classic experimental games: the dictator game and the public goods game. These games pit pro-social behavior against self-interest. If pro-social versus selfish behavior could represent a self-control conflict, we would expect participants' trait self-control, as measured by Rosenbaum's (1980) scale, to positively predict pro-social behavior for participants who had just previously viewed a calendar without a grid, but less so or not at all for participants who had viewed a calendar with.

2. Experimental Design

2.1. Experimental Treatments

In both games, we employed three between-subjects treatments – the isolated frame, the standard frame, and the interrelated frame. The isolated and interrelated frames were manipulated with the procedure from Myrseth and Fishbach (2010). Participants viewed a calendar showing the present month, and the calendar contained either a grid that separated the dates or no such grid (see Appendix A). Moreover, the date of the experiment was highlighted in grey in the gridded calendar, but not in the non-gridded calendar. Because we expected participants who viewed the gridded calendar to adopt a more isolated view of their subsequent choice opportunities, we refer to this treatment as the isolated treatment. Conversely, because we expected participants who viewed the calendar with no grid to adopt a less isolated frame, whereby the choice opportunities would be viewed relatively more related to each other, we refer to this as the interrelated treatment. We denote the third treatment, without a calendar, as the standard treatment.

To capture individuals' self-control, we used the Rosenbaum Self-Control Schedule (Rosenbaum, 1980). The psychometric scale measures individuals' cognitive skills for exercising self-control in the face of temptation. Each subject is asked to respond to 36 statements using a 6-point Likert-scale.[6] Cognitive skills, such as willpower, have been found to be relatively stable within individuals across time, and thus may be said to represent a personality trait, which we refer to as trait self-control. The Rosenbaum Self-Control Schedule has been externally validated against several criteria, such as coping with seasickness (Rosenbaum and Rolnick, 1983) and

[6] Each statement is graded from -3 to +3. Thus, "perfect" self-control corresponds to +108 and no self-control to -108. See appendix B.

saving versus spending (Romal and Kaplan, 1995). Henceforth, we refer the outcome of the Rosenbaum Self-Control Schedule only as the Rosenbaum score.

We expect pro-social behavior to depend on the interaction between identification of self-control conflict (induced by the treatments) and success at the conflict stage (see Figure 1). The isolated treatment yields a lower probability of conflict identification relative to that of other treatments. Hence, trait self-control as measured by the Rosenbaum score is expected to exhibit a weaker correlation with pro-social behavior. In contrast, the interrelated treatment yields a higher probability of conflict identification. Hence, trait self-control is expected to exhibit a stronger positive correlation with pro-social behavior.

2.2. Games

We recruited subjects from various undergraduate and graduate classes at three universities in Medellín, Colombia, 2008. For the dictator game, we held six sessions with 18-31 participants per session. Individuals were randomly assigned within a session to one of the three treatments. In the public good game, we held six sessions, two for each of the experimental treatments, with 24-28 participants per session. Individuals were randomly assigned to sessions and thus to treatments. Nobody participated in more than one experimental session, and none were students of mathematics, psychology, or economics.

Dictator Game

We employed a standard dictator game, designating the Colombian Red Cross as recipient (e.g., similar to Eckel and Grossman, 1996). Average session earnings were 13,000 Colombian Pesos (including a 5,000 Peso show-up fee).[7] A receipt of the donations was posted on a bulletin board within five days of the completed session in the building adjacent to that in which the experiment was conducted.[8] A session lasted on average about one hour.

[7] The exchange rate at the time of the experiment was approximately 1 USD=1,762.00 Colombian Pesos.
[8] To ensure credibility, invitations to the experiment were done jointly by the experimenters and the head administrator who later posted experimental id numbers, and their respective donations, as well as the total amount donated to the Red Cross Colombia. This procedure was outlined as part of recruitment as well as in the instructions.

Public Goods Game

We employed a standard linear public goods experiment. Each group consisted of four members. Each member was endowed with 20 tokens, to be divided between a public and a private good. The payoff for member *i*, measured in units of tokens, was calculated according to the payoff function

$$\pi_i = 20 - c_i + 0.4\sum_{i=1}^{4} c_i \,, \tag{1}$$

where c_i is member i's contribution to the public good. The contribution to the public good yielded a marginal return to each member of 0.4 tokens. The choice of parameter values reflects the features of a public good; full contribution to the public good is Pareto optimal, while the dominant strategy is zero contribution. In other words, the dominant strategy is to free-ride.

Our game followed the experimental design of Fischbacher *et al.* (2001), also used in numerous follow-up studies (e.g., Fischbacher and Gächter, forthcoming). Participants were asked to make both an unconditional and a conditional contribution to the public good. In the case of unconditional contributions, they were simply asked how much they would like to contribute to the public good (as in a standard public goods game). In the case of conditional contributions, participants were asked how much they would like to contribute conditional on the average contribution of other group members, the contribution of which ranged from 0 to 20, rounded to the nearest integer. Participants were randomly assigned to groups of four (from the same session).[9] To make each decision incentive compatible, the unconditional contribution was the payoff-relevant decision for three randomly selected members. Using their average unconditional contribution, the contribution of the fourth member was given by her conditional contribution table. Then, each member's monetary payoff is computed by equation (1). Each token in the experiment was exchanged for 750 Colombian Pesos. The average earnings per participant were 25,000 Colombian pesos (including a 5,000 show-up fee). A session lasted about 1.5 hours.

---

[9] The selection was anonymous. Hence no participant knew to which group he/she belonged.

3. Results

3.1. Dictator Game

In table 1, we summarize the descriptive results from the dictator game. We cannot based on a Kruskal-Wallis test reject the null hypothesis of no difference in donations across treatments. Similarly, we cannot reject the null hypothesis of no difference in Rosenbaum score across treatments. This implies that participants in the three treatments had the same level of trait self-control.

Table 1. Descriptive statistics - the dictator game

	Isolated treatment		**Standard treatment**		**Interrelated treatment**		**HO: No difference between treatments (Kruskal-Wallis p)**
Variable	Obs	Mean	Obs	Mean	Obs	Mean	
Donation	51	7892.16 (4158.50)	49	8321.43 (4608.642)	46	8691.30 (4959.91)	0.646
Rosenbaum score	47	35.85 (25.73)	48	34.04 (24.54)	45	33.13 (23.12)	0.777

Note. Standard deviations in parentheses

We hypothesized that participants' trait self-control, as measured by the Rosenbaum score, would exhibit a stronger positive correlation with charitable giving in the interrelated treatment; participants in the interrelated treatment more likely would identify self-control conflict than would participants in the other two treatments. We tested this hypothesis with an OLS regression, and we report the results in table 2. We included an interaction between the Rosenbaum and the dummy variable identifying the interrelated treatment, but we did include not the dummy variable alone; the interrelated treatment represented the baseline treatment. We also included two dummy variables to identify the treatments (the isolated treatment and the standard treatment) and interaction variables between the Rosenbaum score and dummy variables for each of the other two treatments.

Table 2. Estimation results - the dictator game.

Dep. var: Donation in 1,000	OLS
	Coef.
Isolated treatment	2.28
	(1.49)
Standard treatment	1.22
	(0.78)
Rosenbaum score x Isolated treatment	-0.01
	(-0.42)
Rosenbaum score x Standard treatment	0.03
	(1.13)
Rosenbaum score x Interrelated treatment	0.08***
	(2.98)
Constant	6.02***
	(6.61)
Number of observations	140
R-squared	0.07

Note: *** denotes significance at the 1% level, ** at the 5% significance level, * at the 10% significance level.
Note: the regression controls for the university where sessions were run but the result isomitted; t-statistics in parenthesis; robust standard errors.

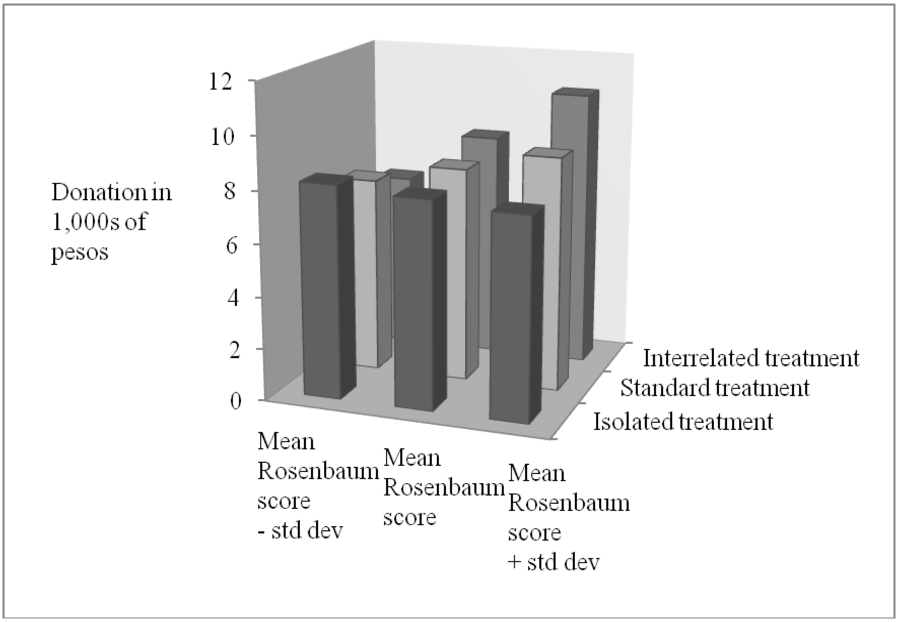
We expected the sign to be positive on the interaction between the Rosenbaum and the dummy for the interrelated treatment. Furthermore, we predicted that coefficients on the interactions between the Rosenbaum and the other two treatment dummies would be smaller than that on the interaction between the Rosenbaum and the interrelated treatment dummy.

The estimated parameter for the interaction between the interrelated treatment and the Rosenbaum score is positive and significant at the 1% level.[10] This means that higher self-control in the interrelated frame is correlated with larger donations. The regression shows that the estimated parameter for the interaction between the isolated treatment and the Rosenbaum is negative and not significant, suggesting a weaker correlation between trait self-control and giving in the isolated than in the interrelated treatment (also see Figure 2). Moreover, the regression

[10] Results are robust also when using Tobit rather than OLS with robust standard errors.

shows that the estimated parameter for the interaction between the standard treatment and the Rosenbaum score is positive and not significant. Taken together, the regression provides evidence for our hypothesis that trait self-control exhibits a stronger positive correlation with donating in the interrelated treatment (calendar without a grid) than in the isolated treatment (calendar with a grid).

Figure 2. Donation in 1,000s by treatment and Rosenbaum score.



Plot of predicted values of donation in 1,000's using the regression model in table 2. Within each treatment, predicted values are calculated for three values of the Rosenbaum score: the mean, the mean less one standard deviation, and the mean plus one standard deviation.

This effect is of economic significance. The marginal effect of the Rosenbaum score in the interrelated treatment is 0.08. In the interrelated treatment, a one standard deviation increase in the Rosenbaum score (approximately 33 units in the test score) increases donations by about

1,862 Colombian Pesos. Compared to the predicted mean contribution of 8,688, this corresponds to a 21% increase in donations relative to the aforementioned baseline.

3.2. Public Goods Game

In table 3, we present the descriptive results from the public goods game. We cannot based on a Kruskal-Wallis test reject the null hypothesis of no difference in unconditional contributions across treatments. Similarly, we cannot reject the null hypothesis of no difference in Rosenbaum score across treatments.

Table 3. Descriptive statistics - the public goods game

	Isolated treatment		**Standard treatment**		**Interrelated treatment**		**HO: No difference between tratments (Kruskal-Wallis p)**
Variable	Obs	Mean	Obs	Mean	Obs	Mean	
Unconditional contribution	56	8.05 (6.28)	47	7.68 (5.96)	53	8.72 (6.61)	0.73
Rosenbaum Score	55	31.51 (20.32)	48	27.98 (19.70)	51	29.41 (19.86)	0.7

Note. Standard deviations in parentheses

We hypothesized that participants' trait self-control, as measured by the Rosenbaum score, would exhibit a stronger positive correlation with unconditional contribution in the interrelated treatment; participants in the interrelated treatment more likely would identify self-control conflict than would participants in the other two treatments. We test this hypothesis with an OLS regression, results reported in table 4. We included an interaction between the Rosenbaum and the dummy variable identifying the interrelated treatment, but we did include not the dummy variable alone; the interrelated treatment represented the baseline treatment. We also included two dummy variables to identify the treatments (the isolated treatment and the standard treatment) and interaction variables between the Rosenbaum score and dummy variables for each of the treatments.

Table 4. Estimation results - the public good game.

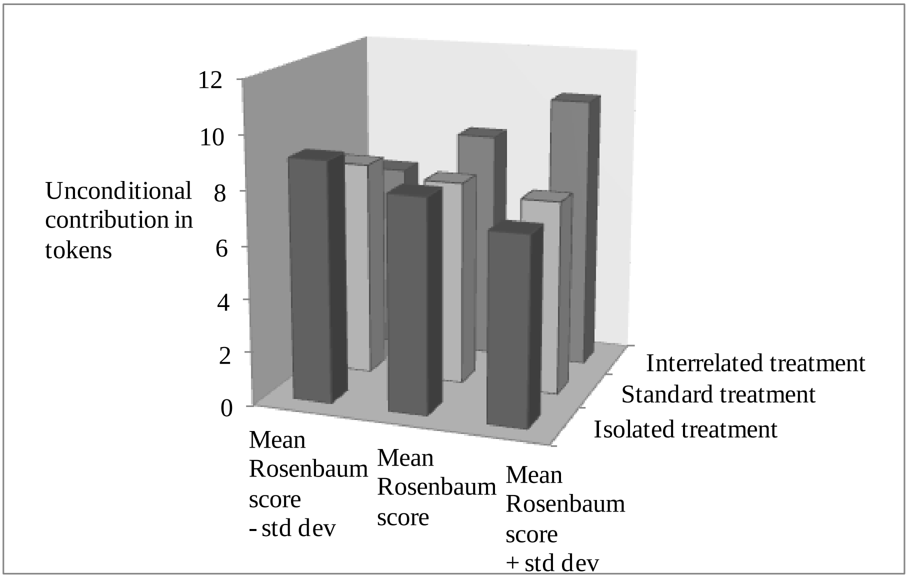
Dep. var: Unconditional contribution in tokens	OLS Coef.
Isolated treatment	3.23
	(1.57)
Standard treatment	1.99
	(1.01)
Rosenbaum score x Isolated treatment	-0.05
	(-1.03)
Rosenbaum score x Standard treatment	-0.02
	(-0.57)
Rosenbaum score x Interrelated treatment	0.08**
	(1.98)
Constant	6.34
	(4.52)
Number of observations	153
R-squared	0.04

Note: *** denotes significance at the 1% level, ** at the 5% significance level, * at the 10% significance level.
Note: the regression controls for the university where sessions were run but the result isomitted; t-statistics in parenthesis

We expected the sign to be positive on the interaction between the Rosenbaum and the dummy for the interrelated treatment. Furthermore, we predicted that coefficients on the interactions between the Rosenbaum and the other two treatment dummies would be smaller than that on the interaction between the Rosenbaum and the interrelated treatment dummy.

As in the dictator game, the coefficient for the interaction between the interrelated treatment and the Rosenbaum is both positive and significant at the 5%-level, indicating that higher self-control in the interrelated treatment is positively correlated with contribution to the public good. The estimated coefficients for the interaction terms between the isolated and standard treatment are both negative and not significant, suggesting a weaker correlation between trait self-control and contribution in the isolated and standard treatments than in the interrelated treatment (also see Figure 3). Taken together, the regressions provide evidence for our hypothesis that trait self-control exhibits a stronger positive correlation with contribution to the public good in the interrelated than in the isolated treatment.

Figure 3. Unconditional contribution in tokens by treatment and Rosenbaum



Plot of predicted values of unconditional contribution in tokens using the regression model in table 4. Within each treatment, predicted values are calculated for three values of the Rosenbaum score: the mean, the mean less one standard deviation, and the mean plus one standard deviation.

As with the dictator game, this effect is of economic significance. The marginal effect of the Rosenbaum score in the interrelated treatment is approximately 0.08. In the interrelated treatment, a one standard deviation increase in the Rosenbaum score (approximately 20 units in the test score) increases contributions by about 1,6 tokens. Compared to the predicted mean contribution of 8.6, this corresponds to an 18% increase in donations relative to the aforementioned baseline

4. Discussion

Our objective was to test the hypothesis that individuals may experience a self-control conflict between acting in the interest of self or in that of others. In so doing, we have explored a hypothesis that would help reconcile conflicting ideas in economics about the selfish Homo Economicus and the pro-social Homo Behavioralis. While the literature to date has documented the existence of both selfish and pro-social preferences (e.g., Fehr and Schmidt, 2006), we have here explored the possibility that the same individuals may possess both. Indeed, we found evidence that individuals may experience a conflict between their better judgment to act in the interest of others and a temptation to act in that of their own. These findings shore up past evidence from psychology and improve our understanding of when and why individuals behave in the interests of others as opposed to those of themselves.

Furthermore, our findings reveal that subtle cues in the environment may prove sufficient to alter an individual's perception of an allocation opportunity between oneself and others. The cues may thereby determine the extent to which the individual uses his or her own cognitive resources to promote pro-social behavior. We demonstrated this both in the context of charitable giving using a dictator game, and in the context of a social dilemma using a public goods game. We further show that the results are of economic significance. We therefore conclude that relatively costless measures may influence individuals to use their cognitive resources to promote pro-social behavior.

These results suggest implications for policy to promote pro-social behavior. Cheap and subtle framing techniques may hold the potential to increase charitable donations, reduce free riding, and improve provision of the public good. One could imagine tax authorities, in an information letter to tax payers, highlighting how regularly underreporting just small amounts could aggregate to substantial costs for society. Similarly, authorities could remind individuals and companies that even small environmental "sins" may aggregate to major environmental problems. An alternative tactic for prompting interrelated frames might be to relate choice opportunities to individuals' self-image. For example, instead of asking people not to cheat on their taxes, authorities could urge them not to be "tax cheats." Presumably, the question of self-image more likely will prompt individuals to consider their general pattern of behavior as opposed to that of the present instance. Future research in these directions holds promise.

While we have provided evidence for the conceptualization that temptation to act in the interest of oneself may conflict with better judgment to act in the interest of others, we do not claim that this conceptualization applies universally. Rather, it may apply in situations where feelings of greed dominate those (if any) to act pro-socially. Of course, as O'Donaghue and Loewenstein (2007) suggest, there is good reason to think that the pattern may reverse under other circumstances. Specifically, when empathetic emotion is particularly strong, individuals may feel tempted to be pro-social and very generous even while knowing that they ought not to. For example, one could imagine a face-to-face interaction with a beggar whom one suspects is a "con" and who is seeking "easy" money. One knows better, but one cannot help oneself to yield to the sorry gestures. Because our present space of inquiry concerned the relationship between self-control and feelings of greed, we deliberately designed our studies to minimize feelings of empathy by keeping the recipient of pro-social behavior highly or moderately abstract (an anonymous group in the public goods game or the Red Cross, respectively). Future research may explore the opposite case, keeping feelings of greed at a minimum while emphasizing feelings of empathy. Pursuing the question of empathy in self-control and pro-social behavior may prove fruitful.

References

Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *The American Economic Review,* 99(1): 544-55.

Baumeister, Roy F., Heatherton, Todd F., and Dianne M. Tice. 1994. *Losing Control: How and Why People Fail at Self-Regulation.* San Diego, CA: Academic Press.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *The American Economic Review,* 96(5): 1652-78.

Burger, Nicholas, Gary Charness, and John Lynham. 2009. "Three Field Experiments on Procrastination and Willpower." http//:www.econ.ucsb.edu/~charness/papers/study_010809.pdf.

Curry, Oliver S., Michael E. Price, and Jade G. Price. 2008. "Patience is a Virtue: Cooperative People Have Lower Discount Rates." *Personality and individual differences,* 44(3): 780-85.

Eckel, Catherine C., and Philip J. Grossman. 1996. "Altruism in Anonymous Dictator Games." *Games and Economic Behavior,* 16(2): 181–91.

Fehr, Ernst, and Klaus M. Schmidt. 2006. "The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories." In *Handbook on the Economics of Giving, Reciprocity and Altruism,* eds. Kolm, Serge-Christophe, and Jean Mercier Ythier, 615-91. Amsterdam: Elsevier.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters,* 71, 397-404.

Fischbacher, Urs, and Simon Gächter. Forthcoming. "Social Preferences, Beliefs and the Dynamics of Freeriding in Public Good Experiments." *The American Economic Review.*

Fudenberg, Drew, and David Levine. 2006. "A Dual-Self Model of Impulse Control." *The American Economic Review,* 96(5): 1449-76.

Laibson, David I. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics,* 112(2): 443–77.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness and The Assumptions of Economics." *Journal of Business*, 59(4): 285-300.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness as a Constraint on Profit-Seeking: Entitlements in the Market." *The American Economic Review*, 76(4): 728-741.

Loewenstein, George. 1996. "Out of control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65(3), 272–92.

Loewenstein, George. 2000. "Preferences, behavior and welfare: Emotions in Economic Theory and Economic Behavior." *The American Economic Review*, 90(2): 426-32.

Myrseth, Kristian O. R., and Ayelet Fishbach. 2009. "Self-Control: A Function of Knowing When and How to Exercise Restraint." *Current Directions in Psychological Science*. 18(4): 247-52.

Myrseth, Kristian O. R., and Ayelet Fishbach. 2010. "Seeing Self-Control Conflict: The Problem of Isolated versus Interrelated Temptations." Unpublished.

O'Donoghue, Ted, and George Loewenstein. 2007. The Heat of the Moment: Modeling Interactions Between Affect and Deliberation." http//:www.arts.cornell.edu/econ/edo1/heat.pdf

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing it Now or Later?" *The American Economic Review*, 89(1): 103-24.

Plato. 1986. "*Protagoras."* In *The Dialogues of Plato,* Trans. Anthony F. Hubbard and Ellen S. Karnofsky, New York: Bantam Books. (Orig. pub. 380 B. C.).

Pronin, Emily, Cristopher Y. Olivola, and Kathleen A. Kennedy. 2008. "Doing Unto Future Selves as You Would Do Unto Others: Psychological Distance and Decision Making." *Personality and Social Psychology Bulletin,* 34(2): 224 – 36.

Romal, Jane B., and Brian J. Kaplan. 1995. "Differences in Self-Control among Spenders and Savers." *Psychology: A Journal of Human Behavior,* 32(2): 8 -17.

Rosenbaum, Michael. 1980. "A Schedule for Assessing Self-Control Behaviors: Preliminary Findings." *Behavior Therapy* , 11(1)*:* 109–21.

Rosenbaum, Michael, and Arnon Rolnick. 1983. "Self-Control Behaviors and Coping with Seasickness." *Cognitive Therapy and Research,* 7(1): 93 – 97.

Schelling, Thomas. 1984. "Self-Command in Practice, in Policy, and in a Theory of Choice." *The American Economic Review,* 74(2): 1-11.

Thaler, Richard H. and Hersh M. Shefrin. 1981. "An Economic Theory of Self-Control." *The Journal of Political Economy,* 89(2): 392-406.

Appendix A: Calendars

A.1 The isolated frame (the highlighted date is the same as today's date).

Before we continue with the experiment, please take a moment to consider this month's calendar:

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

*What is today's date?*__________

A.2 The interrelated frame.

Before we continue with the experiment, please take a moment to consider this month's calendar:

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

*What is today's date?*_________

Not for Publication

Appendix B: The Rosenbaum Self-Control Schedule

Note: * = item is reverse scored.

Directions - Indicate how characteristic or descriptive each of the following statements is of you by using the code given below

+3 very characteristic of me, extremely descriptive
+2 rather characteristic of me, quite descriptive
+1 somewhat characteristic of me, slightly descriptive
-1 somewhat uncharacteristic of me, slightly undescriptive
-2 rather uncharacteristic of me, quite undescriptive
-3 very uncharacteristic of me, extremely nondescriptive

l. When I do a boring job, I think about the less boring parts of the job and the reward that I will receive once I am finished.

-3	-2	-1	1	2	3

2. When I have to do something that is anxiety arousing for me, I try to visualize how I will overcome my anxieties while doing it.

-3	-2	-1	1	2	3

3. Often by changing my way of thinking I am able to change my feelings about almost everything.

-3	-2	-1	1	2	3

4. I often find it difficult to overcome my feelings of nervousness and tension without any outside help.*

-3	-2	-1	1	2	3

5. When I am feeling depressed I try to think about pleasant events.

-3	-2	-1	1	2	3

6. I cannot avoid thinking about mistakes I have made in the past.*

-3	-2	-1	1	2	3

7. When I am faced with a difficult problem, I try to approach its solution in a systematic way.

-3	-2	-1	1	2	3

8. I usually do my duties quicker when somebody is pressuring me.*

-3	-2	-1	1	2	3

9. When I am faced with a difficult decision, I prefer to postpone making a decision even if all the facts are at my disposal.*

-3	-2	-1	1	2	3

10. When I find that I have difficulties in concentrating on my reading, I look for ways to increase my concentration.

-3	-2	-1	1	2	3

11. When I plan to work, I remove all the things that are not relevant to my work.

-3	-2	-1	1	2	3

12. When I try to get rid of a bad habit, I first try to find out all the factors that maintain this habit.

-3	-2	-1	1	2	3

13. When an unpleasant thought is bothering me, I try to think about something pleasant.

-3	-2	-1	1	2	3

14. If I would smoke two packages of cigarettes a day, I probably would need outside help to stop smoking.*

-3	-2	-1	1	2	3

15. When I am in a low mood, I try to act cheerful so my mood will change.

-3	-2	-1	1	2	3

16. If I had the pills with me, I would take a tranquilizer whenever I felt tense and nervous.*

-3	-2	-1	1	2	3

17. When I am depressed, I try to keep myself busy with things that I like.

-3	-2	-1	1	2	3

18. I tend to postpone unpleasant duties even if I could perform them immediately.*

-3	-2	-1	1	2	3

19. I need outside help to get rid of some of my bad habits.*

-3	-2	-1	1	2	3

20. When I find it difficult to settle down and do a certain job, I look for ways to help me settle down.

-3	-2	-1	1	2	3

21. Although it makes me feel bad, I cannot avoid thinking about all kinds of possible catastrophes in the future.*

-3	-2	-1	1	2	3

22. First of all I prefer to finish a job that I have to do and then start doing the things I really like.

-3	-2	-1	1	2	3

23. When I feel pain in a certain part of my body, I try not to think about it.

-3	-2	-1	1	2	3

24. My self-esteem increases once I am able to overcome a bad habit.

-3	-2	-1	1	2	3

25. In order to overcome bad feelings that accompany failure, I often tell myself that it is not so catastrophic and that I can do something about it.

-3	-2	-1	1	2	3

26. When I feel that I am too impulsive, I tell myself "stop and think before you do anything."

-3	-2	-1	1	2	3

27. Even when I am terribly angry at somebody, I consider my actions very carefully.

-3	-2	-1	1	2	3

28. Facing the need to make a decision, I usually find out all the possible alternatives instead of deciding quickly and spontaneously.

-3	-2	-1	1	2	3

29. Usually I do first the things I really like to do even if there are more urgent things to do.*

-3	-2	-1	1	2	3

30. When I realize that I cannot help but be late for an important meeting, I tell myself to keep calm.

-3	-2	-1	1	2	3

31. When I feel pain in my body, I try to divert my thoughts from it.

-3	-2	-1	1	2	3

32. I usually plan my work when faced with a number of things to do.

-3	-2	-1	1	2	3

33. When I am short of money, I decide to record all my expenses in order to plan more carefully for the future.

-3	-2	-1	1	2	3

34. If I find it difficult to concentrate on a certain job, I divide the job into smaller segments.

-3	-2	-1	1	2	3

35. Quite often I cannot overcome unpleasant thoughts that bother me.*

-3	-2	-1	1	2	3

36. Once I am hungry and unable to eat, I try to divert my thoughts away from my stomach or try to imagine that I am satisfied.

-3	-2	-1	1	2	3

Not for publication

Appendix C: Experimental Instructions

C.1.: Original Instructions for the Public Good Game

Instrucciones

Usted va a participar en un experimento sobre toma de decisiones. A finales de este experimento, le pagará a usted una cantidad dependiendo de sus decisiones y las decisiones de otros. El pago se hará en efectivo al final del experimento.

A lo largo del experimento, todo tipo de comunicación queda completamente prohibida. Los participantes que se comuniquen quedarán excluidos del experimento y no recibirán ningún pago. Si tiene alguna pregunta, levante la mano para que algún monitor venga a responder la pregunta en privado.

Durante el experimento no hablaremos de Pesos sino de "fichas". Sus ganancias en el experimento serán en fichas. Al final del experimento, las fichas serán convertidas a pesos a una tasa de cambio de:

2 fichas = 1500 pesos

Independientemente de sus decisiones, recibirá 5000 pesos por participar en el experimento.

Además tendrá que responder algunos cuestionarios durante el experimento. Algunas preguntas pueden parecerle muy extrañas, sin embargo, le pedimos el favor que las responda seriamente. Todas sus respuestas serán **confidenciales y anónimas.** Para identificarlo usaremos el número de identificación que usted recibió al entrar en el salón. Este mismo número será usado para identificarlo cuando hagamos los correspondientes pagos después del experimento. Antes de que usted abandone el salón, debe entregar a uno de los monitores el número de identificación que recibió al entrar al salón. El monitor pone ese número dentro de un sobre, lo sella y se lo entrega a usted nuevamente. Cuando vaya a reclamar su pago, usted debe presentar el mismo sobre debidamente sellado y con su número de identificación adentro, tal y como el monitor del experimento se lo entregó antes de abandonar el salón.

A lo largo de estas instrucciones le presentaremos algunos ejemplos, por favor considere los números de los ejemplos como una ilustración. Los números que usted obtendrá en el experimento pueden ser diferentes.

Decisión Básica

El experimento va a funcionar así: Primero vamos a explicar la decisión básica, luego vamos a hacer unas preguntas de control que le ayudarán a entender esta decisión básica.

Usted pertenece a **un grupo de cuatro personas**. Nadie, excepto los monitores sabrán quién pertenece a qué grupo. Los grupos se formarán al azar. Al principio del experimento recibirá un papel donde dice la cantidad de fichas que tiene inicialmente para jugar. Esta será su dotación. Cada una de las cuatro personas del grupo tiene que decidir cómo distribuir su dotación de 20 fichas. Puede poner todas, una parte o ninguna ficha en una **cuenta de un proyecto**. Las fichas que no deposite en la cuenta del proyecto serán automáticamente transferidas a su **cuenta privada**.

Su ingreso de la cuenta privada:

Por cada ficha que deposite en la cuenta privada ganará exactamente una ficha. Por ejemplo, si tiene una dotación de 20 fichas y deposita cero fichas en la cuenta del proyecto (o sea que deposita 20 fichas en su cuenta privada), entonces recibe exactamente 20 fichas. Si en cambio deposita 14 fichas en la cuenta del proyecto (es decir 6 fichas en la cuenta privada) entonces su ingreso de la cuenta privada son 6 fichas. *Nadie, excepto usted mismo recibe fichas de la cuenta privada.*

Su ingreso de la cuenta del proyecto:

Todos reciben el mismo ingreso por las fichas que usted deposite en la cuenta del proyecto. Por supuesto, usted también obtendrá ingresos por las fichas que otras personas depositen en la cuenta del proyecto. Para cada persona el ingreso de la cuenta del proyecto se determina de la siguiente manera:

Ingreso de la cuenta del proyecto = *Suma de las contribuciones al proyecto x 0.4*

Por ejemplo, si la suma de las contribuciones a la cuenta del proyecto es 60 fichas, usted y los otras personas del grupo recibirán 60x0.4=24 fichas para cada uno. Si las cuatro personas del grupo depositan un total de 10 fichas en la cuenta del proyecto, entonces usted y todos los otros reciben 10x0.4=4 fichas por la cuenta del proyecto.

Ingreso total:

Su ingreso total es la suma del ingreso de su cuenta privada más el ingreso de la cuenta del proyecto.

Ingreso de la cuenta privada(= Numero de fichas de dotación inicial – su contribución a la cuenta del proyecto)
+ Ingreso de la cuenta del proyecto (= 0.4 × suma de todas las contribuciones a la cuenta del proyecto)
Ingreso total

Antes de terminar de leer las instrucciones le queremos pedir el favor que responda las siguientes preguntas de control que ayudarán a verificar si ha entendido todo correctamente. Si hay alguna pregunta, por favor levante la mano para que un monitor le responda en privado.

Preguntas de Control

Por favor responda las siguientes preguntas de control. El propósito de estas preguntas es familiarizarlo con los cálculos de los ingresos en fichas que resultan de las diferentes decisiones sobre cómo distribuir sus recursos disponibles. Por favor responda todas las preguntas y escriba los correspondientes cálculos.

1. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Supongamos también que nadie, ni usted, pone nada en la cuenta del proyecto.

   ¿Cuál es su ingreso total? ____

   ¿Cuál es el ingreso de las otras personas de su grupo? ____, ____, y ____

2. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Supongamos también que todos, incluido usted, ponen todos sus recursos en la cuenta del proyecto.

   ¿Cuál es su ingreso total? ____

   ¿Cuál es el ingreso de las otras tres personas de su grupo? ____, ____, y ____

3. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Supongamos también que las otras tres personas del grupo depositan 30 fichas en la cuenta del proyecto.

   a) ¿Cuál es su ingreso total si además de las 30 fichas de las otras tres personas de su grupo usted pone 0 fichas en la cuenta del proyecto?

   Su ingreso total: ____

b) ¿Cuál es su ingreso total si además de las 30 fichas de las otras tres personas de su grupo usted pone 8 fichas en la cuenta del proyecto?

Su ingreso total: ____

c) ¿Cuál es su ingreso total si además de las 30 fichas de las otras tres personas de su grupo usted pone 15 fichas en la cuenta del proyecto?

Su ingreso total: ____

4. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Usted pone 8 fichas en la cuenta del proyecto.

a) ¿Cuál es su ingreso total si además de sus 8 fichas las otras personas del grupo ponen en total otras 7 fichas en la cuenta del proyecto?

Su ingreso total: ____

b) ¿Cuál es su ingreso total si además de sus 8 fichas las otras tres personas del grupo ponen en total otras 12 fichas en la cuenta del proyecto?

Su ingreso total: ____

c) ¿Cuál es su ingreso total si además de sus 8 fichas las otras tres personas del grupo ponen otras 22 fichas en la cuenta del proyecto?

Su ingreso total: ____

Si ha terminado estas preguntas antes que los otros, le recomendamos que piense en otros ejemplos adicionales para que se familiarice con este tipo de situaciones.

Procedimiento

El experimento consiste en la decisión que acabamos de describir. En la siguiente parte explicaremos el procedimiento a usar en detalle.

Como usted sabe, usted tiene una dotación de 20 fichas. Puede poner esas fichas en la cuenta del proyecto y el resto de fichas automáticamente se depositan en la cuenta privada. Cada persona en el grupo tendrá la misma dotación.

Cada persona del grupo tiene que hacer dos tipos de decisiones, a las que nos referiremos de aquí en adelante como "**contribución incondicional"** y como "**tabla de contribuciones"**

- En la **contribución incondicional** usted debe decidir cuántas de las fichas que tiene disponibles deposita en la cuenta del proyecto. Escriba esta cantidad al lado de *"Su contribución incondicional a la cuenta del proyecto"* en la segunda página de su hoja de decisión. Usted debe escribir un **número entero** que **no puede ser menor a cero ni mayor a las 20 fichas que usted tiene en su dotación**. La diferencia entre su dotación de 20 fichas y las fichas que pone en la cuenta del proyecto es automáticamente depositada en su cuenta privada.

- Su segunda tarea es completar la **tabla de contribuciones** en la tercera página de la hoja de decisión. En la tabla de contribución usted debe indicar para cada posible contribución promedio de las otras tres personas del grupo (aproximado al siguiente entero; por ejemplo si el promedio es 17,5 piense en 18) el número de fichas que usted quiere poner en la cuenta del proyecto. Usted decidirá cuanto contribuir dependiendo de lo que las otras personas contribuyan. Esto será más claro cuando vea el siguiente ejemplo de una tabla de contribución.

Contribución promedio (aproximada) de las otras tres personas a la cuenta del proyecto	**Su contribución a la cuenta del proyecto es:**
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

Los números en la columna izquierda son los posibles valores promedio (aproximados) de las contribuciones de las otras personas del grupo. Supongamos en el ejemplo que las otras personas tienen 20, 20 y 20 fichas, entonces en promedio pueden contribuir máximo 20 fichas ((20+20+20)/3=20).

Usted simplemente tiene que escribir en la columna de la derecha cuántas fichas quiere contribuir en la cuenta del proyecto dado que los otros contribuyen en promedio (aproximado) la cantidad de fichas de la columna izquierda. Tiene que completar todas las casillas de la columna derecha. Por ejemplo, debe escribir cuántas fichas quiere contribuir a la cuenta del proyecto si los otros contribuyen en promedio 0 fichas a la cuenta del proyecto; cuánto contribuye si los otros contribuyen 1 o 2 o 3, fichas, etc. En cada casilla debe escribir un número entero no menor de cero y no mayor a su dotación de 20 fichas. Por supuesto que puede escribir el mismo número en diferentes casillas.

Después que todos los participantes en el experimento han hecho su contribución incondicional y han completado la tabla de contribuciones, una persona de cada grupo será seleccionada al azar. Para las personas seleccionadas aleatoriamente el ingreso se determinará de acuerdo con la tabla de contribuciones. Para las otras tres personas del grupo que no son seleccionadas aleatoriamente la contribución incondicional determinará el ingreso. Cuando usted está decidiendo la contribución incondicional y la tabla de contribuciones, usted no sabe si va a ser seleccionado aleatoriamente, así que piense cuidadosamente los dos tipos de decisiones porque cualquiera puede ser relevante para usted. Los siguientes dos ejemplos servirán para aclarar este punto:

Ejemplo 1: Suponga que después de que ha entregado sus decisiones usted es seleccionado al azar. Esto implica que la decisión relevante para sus ingresos es la tabla de contribuciones. Para las otras tres personas la decisión incondicional es la decisión relevante. Supongamos que ellos han hecho contribuciones incondicionales de 0, 2 y 4 fichas. La contribución promedio aproximada es entonces 2 (=(0+2+4)/3).

Si usted ha indicado en su tabla de contribuciones que usted contribuiría una ficha a la cuenta del proyecto si los otros contribuían 2 fichas en promedio, entonces la contribución a la cuenta del proyecto es 0+2+4+1=7. Entonces todas las personas del grupo ganan un ingreso de 0.4x7=2.8 de la cuenta del proyecto más el respectivo ingreso de la cuenta privada.

Si en cambio usted ha indicado que contribuiría 19 fichas a la cuenta del proyecto si los otros contribuían 2 en promedio, entonces la contribución total a la cuenta del proyecto es 0+2+4+19=25. Todas las personas del grupo ganan un ingreso de 0.4x25=10 fichas de la cuenta del proyecto más el respectivo ingreso de la cuenta privada.

Ejemplo 2: Ahora suponga que usted no es seleccionado aleatoriamente. Esto quiere decir que para usted y para dos otras persona del grupo la contribución incondicional es la decisión de ingreso relevante. Suponga además que su contribución incondicional al proyecto es de 16 y que la de las otras tres personas es 18 y 20. La contribución incondicional promedio del grupo es entonces 18 (=(16+18+20)/3).

Si la persona del grupo que ha sido seleccionada aleatoriamente indicó en la tabla de contribuciones que contribuiría una ficha a la cuenta del proyecto si las otras tres personas contribuían en promedio 18, entonces la contribución total a la cuenta del proyecto es 16+18+20+1=55 fichas. Por lo tanto, todas las personas del grupo ganarían 0.4x55=22 fichas de la cuenta del proyecto adicionalmente a sus respectivos ingresos de las cuentas privadas.

Si la persona del grupo seleccionada aleatoriamente indicó en la tabla de contribuciones que contribuiría 19 fichas a la cuenta del proyecto si las otras tres personas contribuían en promedio 18, entonces la contribución total a la cuenta del proyecto es 16+18+20+19=73 fichas. Por lo tanto, todas las personas del grupo ganarían 0.4x73=29.2 fichas de la cuenta del proyecto adicionalmente a sus respectivos ingresos de las cuentas privadas.

La selección aleatoria de los participantes se hará de la siguiente forma. A cada persona del grupo se le asigna un número entre 1 y 4 que puede ver en la última página de su hoja de decisión. Un participante seleccionará al azar una de las cuatro cartas **después** que todos los participantes han tomado su decisión incondicional y han completado la tabla de contribuciones y el cuestionario. Si la carta que es seleccionada corresponde al número en su hoja de decisión entonces la tabla de contribuciones condicionales en la tercera página es relevante para usted. Si no, la contribución incondicional en la primera página es la decisión relevante. Recuerde que usted sabe cuál de las dos decisiones es la decisión relevante para sus pagos sólo después que ha entregado sus decisiones, por lo tanto debe completar las dos páginas cuidadosamente.

La cantidad de fichas que usted gane se convertirá a pesos que serán pagados en efectivo. Tiene alguna pregunta? Por favor levante la mano y un monitor vendrá a responder la pregunta privadamente.

C.2.: Guessed Contributions for the Public Good Game

Identificación en el experimento:______

Cuánto cree usted que los otros integrantes contribuyeron incondicionalmente a la cuenta del proyecto?

Hace un rato usted escribió su contribución incondicional y llenó una tabla con un número de contribuciones condicionales. Ahora, usted nos dirá cuándo cree que los otros integrantes de su grupo han escrito como sus contribuciones **incondicionales**.

Por favor escriba el número de fichas que **usted cree que los otros tres jugadores** en su grupo han contribuido a la cuenta del proyecto. En otras palabras, qué número sospecha usted que ellos escribieron?

AHORA usted puede ganar más dinero **si adivina correctamente Contribución promedio (aproximada) de las otras tres personas a la cuenta del proyecto**. Usted puede ganar tres fichas adicionales si la verdadera contribución de los otros es igual a lo que usted adivinó, o si está una ficha por encima o por debajo. Asi por ejemplo, si la verdadera contribución de los otros tres es 7 y usted dice 6 o dice 8, usted gana 2 fichas más, pero si usted dice 9 no ganará.

Contribución promedio (aproximada) de las otras tres personas a la cuenta del proyecto	Marque con una x la casilla que usted cree que corresponde a los que los otros tres contribuyeron incondicionalmente a la cuenta del proyecto. ***SÓLO MARQUE UNA CASILLA***
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

C.3.: Original Instructions for the Dictator Game

Note: the last sentence of the instructions varied according to the location of the experiment

Instrucciones Generales

Usted va a participar en un experimento sobre toma de decisiones. Independientemente de sus decisiones, usted recibirá 5000 pesos colombianos (COP) sólo por participar en el experimento.

Cualquier tipo de comunicación queda completamente prohibida durante el experimento; los participantes que no cumplan esta regla y se comuniquen entre ellos quedarán excluidos del experimento y no recibirán ningún pago. Si tiene alguna pregunta, por favor levante la mano para que algún monitor venga a atenderle

Primero usted participará en una tarea de toma de decisiones, y después le pediremos responder algunos cuestionarios. Algunas preguntas pueden parecerle muy extrañas, sin embargo, le pedimos el favor que las responda seriamente. Todas sus respuestas serán **confidenciales y anónimas.** Para identificarlo durante y después del experimento usaremos el número que usted recibió al entrar en el salón; este mismo número será usado para identificarlo cuando hagamos los correspondientes pagos después del experimento.

IMPORTANTE: Por favor escriba su número de identificación en todas sus hojas de respuesta para que nosotros podamos pagarle.

Cuando se acabe el experimento, les pedimos salir del salón. Cuando usted salga, debe entregar a uno de los monitores el número de identificación que recibió al entrar. El monitor pondrá ese número dentro de un sobre, lo sellará y se lo entregará nuevamente para garantizar que sus respuestas quedan confidenciales y anónimas. Cuando vuelva a reclamar su pago (unos 20 minutos después de que termine el experimento), usted debe presentar el mismo sobre debidamente sellado y con su número de identificación adentro, tal y como el monitor del experimento se lo entregó antes de abandonar el salón. El pago se hará de manera individual y privada en otro sobre sellado para proteger la confidencialidad de sus ganancias.

Descripción del experimento.

Usted ha recibido una dotación de 15 000 COP; Su tarea en el experimento es decidir cómo repartir esta cantidad entre usted y la Cruz Roja Colombiana. En la “hoja de decisión” que le será entregada posteriormente, usted debe indicar qué cantidad de esos 15 000 COP quiere conservar para usted mismo y qué cantidad quiere donar. En este experimento cualquier decisión que usted tome es válida, es decir, usted puede donar todo, una parte o nada.

Al final del experimento, todas las donaciones de los participantes serán sumadas y mandadas a la Cruz Roja Colombiana y una copia del recibo de la donación será pegado en la cartelera del IDEA (/Departamento de Ingeniería Sanitaria) máximo 5 días después de que se acabe el experimento .

Appendix C.4.: Guessed Contributions for the Dictator Game

Identificación en el experimento:________________

Cuanto cree usted que los otros participantes de esta sesión donaron en promedio a la Cruz Roja Colombiana?

Usted recibirá 2000 COP adicionales si adivina correctamente o dentro un margen de 1000 COP por encima o por debajo del promedio correcto.

POR EJEMPLO: Si el promedio verdadero de las donaciones de los otros participantes es 10 000, usted recibirá 2000 COP adicionales si adivinó 9000, 10000 o 11000

Por favor, indique su estimación en la línea debajo:

____

Not for publication

Appendix D: Experimental Instructions (Translated from Spanish)

D.1.: Original Instructions for the Public Good Game

Instructions

You will be taking part in an experiment on decision-making. The experiment is designed so that your earnings will depend on both your own decisions and the decisions of others. Your earnings will be paid in cash at the end of the session.

Talking is not allowed throughout the entire session. Any violation of this rule will result in exclusion from the session and not receiving any payment. If you have any questions regarding these instructions, please raise your hand and a member of the experimenter team will attend to you.

Your earnings in this experiment will be in tokens. At the end of the experiment, the tokens will be converted into Colombian pesos (COP) at an exchange rate of:

2 tokens = 1500 COP.

Regardless of what decisions you make, you will receive a show-up fee of 5,000 COP.

During the experiment, you will have to answer a few questionnaires. Although some questions may appear strange to you, we ask you to still take them seriously. All your answers will be treated **confidentially and anonymously**. The identification number you received when entering the room will be used to identify you when paying you after the experiment. Before you leave the room, you should hand the identification number you received when entering the room to a member of the experimenter team. The experimenter will put this number in an envelope, seal it, and return it to you. When you go to collect your earnings, you should return the sealed envelope with your identification number still inside, the way it was handed to you before you left the room.

Along with these instructions, we will present you with a few examples. The numbers used are only for illustration purposes. The numbers you will encounter in the experiment could be different.

The basic decision

You will now learn how the experiment is conducted. First we will introduce the basic decision-making situation. Then we will ask you to answer control questions that will help you gain an understanding of the decision-making situation.

You will be a **member of a group of four people**. No one, except the experimenters, knows who belongs to what group. The groups are assembled randomly. At the beginning of the experiment, you will receive (on paper) **a number of tokens, called an "endowment."** Each of the four members of the group has to decide how to divide his or her endowment. You can put all, some, or none of your tokens into the project account. Each token you do not deposit in the project account will automatically be transferred to your **private account**.

Your income from the private account:

For each token you put into your private account, you will earn exactly one token. For example, if you have an endowment of 20 tokens and you put zero tokens into the project account (and therefore 20 tokens into the private account), then you will earn exactly 20 tokens **from the private account**. If instead you put 14 tokens into the project account (and therefore 6 tokens into the private account), then you will receive an income of 6 tokens from the private account. *Nobody except you earns tokens from your private account.*

<u>Your income from the project account:</u>

Everybody receives the same income from the project account, which is based on the total number of tokens the group puts into it. Your income from the project account will therefore be determined not only by the number of tokens you decide to put into the project account, but also by the number of tokens the other group members invest in it. For each group member, the income from the project account will be determined as follows:

Income from the project account = the sum of all contributions to the project account x 0.4

For example, if the sum of all contributions to the group account is 60 tokens, you and the other group members will earn 60x0.4=24 tokens from the project account. If the four group members deposit a total of 10 tokens into the project account, then you and the others will earn 10x0.4=4 tokens from the project account.

<u>Your total income:</u>

Your total income is the sum of the income from your private account and the income from the project account:

Income from your private account (=your endowment – your contribution to the project account)
+ Income from the project account (=0.4 x the sum of all contributions to the project account)
Total income

Before we finish reading the instructions, please answer the following control questions. This will help you make sure you have understood everything correctly. If you have any questions or problems, please raise your hand. A member of the experimenter team will attend to you and answer your question in private.

Control questions

Please answer the following control questions. Their purpose is to make you familiar with calculating the various incomes in tokens that you might earn depending on the decisions you will make about endowment allocation. Please answer all questions and write down all calculations.

1. Assume that you have an endowment of 20 tokens. Assume also that all group members (including yourself) put nothing into the project account.

    What is your total income? ______________

    What are the incomes of the three other group members?_____,____ and ____

2. Assume that you and the other team members each have an endowment of 20 tokens. the same as the other three group members. Assume also that all group members (including yourself) put their entire endowments into the project account.

    What is your total income? ______________

    What are the incomes of the three other group members?_____,____ and ____

3. Assume you have an endowment of 20 tokens. Assume also that the other group members collectively put a total of 30 tokens into the project account.

    a) What is your total income if you, in addition to the 30 tokens from the other three group members, put 0 tokens into the project account?

        i. Your total income is ___________.

b. What is your total income if you, in addition to the 30 tokens from the other three group members, put 8 tokens into the project account?

    i. Your total income is__________.

c. What is your total income if you, in addition to the 30 tokens from the other three group members, put 15 tokens into the project account?

    i. Your total income is__________.

4. Assume that you have an endowment of 20 tokens and that you put 8 tokens into the project account.

    a. What is your total income if the other three group members, in addition to your 8 tokens, put a total of 7 tokens into the project account?

        i. Your total income is __________.

    b. What is your total income if the other three group members, in addition to your 8 tokens, put a total of 12 tokens into the project account?

        i. Your total income is__________.

    c. What is your total income if the other three group members, in addition to your 8 tokens, put a total of 22 tokens into the project account?

        i. Your total income is__________.

If you finish these questions before the other participants, we advise you to think about additional examples to familiarize yourself further with these types of decision-making situations.

The Experimental Procedure

The experiment consists of decision-making situations similar to the one we just described. We will now explain the procedure in detail.

As you know, you have an endowment of 20 tokens. You can put these tokens into a project account. Any remaining tokens will automatically be deposited into your private account. Each person in the group will have the same endowment.

Each group member is asked to make two types of decisions. In the following instructions, we will refer to them as the **"unconditional contribution"** and the **"contribution table decision."**

- With the unconditional contribution, you decide how many tokens you want to put into the project account. Write this amount under *"Your unconditional contribution to the group account"* on the first page of your decision sheet. You must write down an **integer number that is neither smaller than zero nor larger than the total number of tokens you were given in your endowment (20).** The difference between your endowment of 20 tokens and the amount you put into the project account is automatically transferred to your private account.

- Your second task is to fill out the **contribution table** on page 3 of the decision sheet. In the contribution table, please indicate how many tokens you would like to put into the project account for each possible average contribution of the other three group members (rounded up or down to the nearest integer number; for example, if the average is 17.5, then write 18). What you actually contribute will depend on what the other group members actually contribute. This will become clear to you if you take a look at the following contribution table example:

(Rounded) Average contribution of the other group members to the project account.	Your contribution to the project account is:
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

The numbers in the left column are the possible (rounded) average contributions of the other three group members. Assume for this example that the other three group members can contribute a maximum of 20 tokens each ((20+20+20)/3=20).

Using the column on the right, simply write down how many tokens you would like to contribute to the project account for each possible average contribution of the others. You must make an entry in each field of the right column. For example, write down how many tokens you want to contribute to the group account if the others contribute an average of 0 tokens to the group account; how many you want to contribute if the others contribute an average of 0 tokens to the group account; how many you want to contribute if the others contribute an average 1 or 2 or 3 tokens, etc. In each field, you must write down an integer number that is neither smaller than zero nor larger than the total number of 20 tokens in your endowment. You can of course write down the same number in different fields.

After all participants have made their unconditional contribution decisions and have filled out their conditional contribution tables, one member of each group will be selected randomly. For the randomly selected group member, only the contribution table will be income relevant. For the three group members who are not selected, the unconditional contribution decision will be the income-relevant decision. When you make your unconditional contribution and when you fill out the contribution table, you do not know whether you will be selected randomly. You will therefore have to think carefully about both types of decisions since both could affect your earned amount. The following two examples should illustrate this:

Example 1. Assume that after you hand in your decisions, you are randomly selected. This implies that your income-relevant decision will be determined by your contribution table. For the other three group members, the unconditional contribution is the income-relevant decision. Assume they have made unconditional contributions of 0, 2, and 4 tokens. The rounded average contribution is therefore 2 ((0+2+4)/3=2).

If you have indicated in your contribution table that you will put 1 token into the project account if the others contribute 2 tokens on average, then the total contribution to the group account is 0+2+4+1=7. Thus, all group members earn an income of 0.4x7=2.8 from the project account plus the respective incomes from their private accounts.

If you have indicated instead that you will contribute 19 tokens to the project account if the others contribute 2 on average, then the total contribution to the project account is 0+2+4+19=25. All group members then earn an income of 0.4x25=10 tokens from the project account plus the respective incomes from their private accounts.

Example 2. Now assume that you are not selected randomly, which means that for you and two other group members, the unconditional contribution is the income-relevant decision. Assume further that your unconditional contribution to the project account is 16, and that those of the other two group members are 18 and 20. The average unconditional contribution is then 18 ((16+18+20)/3).

If the randomly selected group member indicated in the contribution table that he or she contributes 1 token to the group account when the other three group members contribute 18 on average, then the total contribution of the group to the group account is 16+18+20+1=55 tokens. All group members will therefore earn 0.4x55=22 tokens from the group account in addition to the respective incomes from their private accounts.

If the randomly selected group member instead indicated in the contribution table that he or she will contribute 19 tokens to the group account if the other three group members contribute 18 on average, then the total contribution of the group to the group account is 16+18+20+19=73 tokens. Each group member will therefore earn 0.4x73=29.2 tokens from the group account in addition to the income from his or her private accounts.

The random selection is arranged in the following manner. Every person in each group is assigned a number from 1 to 4. This number is found on the last page of your decision sheet. A participant will randomly pick one of four cards **after** all participants have made their unconditional contributions and have completed the contribution table and the questionnaire. If the card that is picked corresponds to the number on your decision sheet, then the contribution table on the third page becomes income-relevant for you. If not, then the unconditional contribution on the first page is your income-relevant decision. Remember that you do not know which of the two decisions will be relevant for your earnings until you have handed in all your decisions. You should therefore complete both pages carefully.

The amount of tokens you earn will be converted into pesos and then paid in cash. Do you have any questions? Please raise your hand and a member of the experimenter team will attend to you and answer your question in private.

D.2.: Guessed Contributions for the Public Good Game (Translated from Spanish)

Experimental ID number:______

How much do you think the other participants contributed unconditionally to the project account?

A moment ago, you wrote down your unconditional contribution and completed the conditional contribution table. Now, please tell us how much you think the other participants from your group contributed **unconditionally.**

Please write down the number of tokens **you believe the other three players** from your group have contributed on average to the project account. In other words, what number do you suspect they wrote down?

NOW you can earn more money if you **correctly guess the average contribution (rounded) of the other three players to the project account.** You can earn two additional tokens if the true average contribution of the others is equal to what you guessed, or if it is one token more or less. For example, if the true average contribution of the other three is 7 and you guess 6 or 8, you get two more tokens, while if you guess 9 you do not.

Average contribution (rounded) of the other three persons to the project account	Please mark an *x* in the box that corresponds to what you think the other three contributed unconditionally to the project account. ***ONLY MARK ONE BOX***
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

D.3.: Instructions for the Dictator Game (Translation)

Please note: variations are displayed within parentheses

General Instructions

You are about to participate in an experiment on decision-making. Regardless of what decision you make, you will receive 5,000 Colombian pesos (COP) for participating in the experiment.

Now that we have begun, all communication is strictly forbidden. Participants who communicate will be excluded from the experiment and will not receive payment. If you have any questions, please raise your hand and a member of the experimenter team will attend to you.

First you will take part in a decision-making task and then you will have to answer a few questionnaires. Although some questions may appear strange to you, we ask you to still take them seriously. All your answers will remain **confidential and anonymous**. To identify you during and after the experiment, we use only the numbers you received when you entered the room. These numbers will be used to identify you when paying you at the end of the experiment.

Please note: You must write your identification number on all your answer sheets in order for us to be able to pay you.

When the experiment is over, you will be asked to leave the room. As you leave the room, you should hand in your identification number to a member of the experimenter team. He or she will place the number in an envelope, seal it and hand it back to you to keep your decision anonymous and confidential. Then please walk to the room next door to claim you payment. To receive your payment (about 20 minutes after the end of the experiment), you will need to present the sealed envelope with your identification number still inside, just as it was handed to you before leaving the room. The payment will then be given to you in private in another sealed envelope to keep your earnings confidential.

Decision task: You have been given an endowment of 15,000 COP. Your task is to decide how to divide the 15,000 COP between Red Cross Colombia and yourself. Write down the amount you wish to donate to Red Cross Colombia and how much you would like to keep for yourself on your decision sheet, which will soon be handed to you. In this experiment, any decision is valid. This means that you can donate all, some, or nothing.

After the experiment is over, the experimenters will add all donations and send the total amount to Red Cross Colombia within 5 days. A receipt of the total donation will be posted on the notice board of the IDEA (departamento de Ingeniería Sanitaria).

D.4.: Guessed Contributions for the Dictator Game (Translation)

How much do you think other participates in this session have donated on average?

You will receive an extra 2,000 COP for a correct guess or for a guess that is within a margin of plus or minus one thousand.

EXAMPLE: If the average of the donations made by the other participants is 10,000, you will get an extra 2,000 COP if you guessed 9,000, 10,000, or 11,000.

Please indicate your guess on the line below:

____

Chapter 4

Conditional Cooperation and Self-Control

Peter Martinsson, University of Gothenburg[A]

Kristian Ove R. Myrseth, ESMT European School of Management and Technology[B]

Conny Wollbrant, University of Gothenburg[C]

Abstract

When facing the opportunity to act either in self-interest or in the interest of others, individuals may experience a self-control conflict between pro-social preferences and urges to act selfishly. We explore the domain of conditional contribution, and we test the hypothesis that an increase in an individual's belief about others' average contribution increases contributions more when her willpower is high than when it is low. We employ a subtle framing technique and the strategy method in a public goods experiment. Consistent with our hypothesis, we find that conditionally cooperative behavior is stronger when beliefs of high contributions are accompanied by high rather than low levels of self-control.

Keywords: Self-control, Pro-social behavior, Public good experiment, Conditional cooperation.

JEL Classification: D01, D03, D64, D70.

Acknowledgments: Financial support from the Swedish Research Council (*Vetenskapsrådet*) and the Swedish International Development Cooperation Agency (Sida) to the Environmental Economics Unit at the University of Gothenburg is gratefully acknowledged. We are grateful to Katarina Nordblom and Robert Östling for helpful comments. Horacio Antonio Villegas Rivera provided excellent logistical support and Angela Maria Gonzales provided excellent research assistance.

[A] Department of Economics, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 52 55; Fax +46 31 786 10 43; E-mail: peter.martinsson@economics.gu.se
[B] ESMT European School of Management and Technology, Schlossplatz 1, 10178 Berlin, Germany; Ph +49 (0) 30 212 31 1529; Fax +49 (0) 30 212 31 1281 E-mail: myrseth@esmt.org
[C] Department of Economics, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 26 15; Fax +46 31 786 10 43; E-mail: conny.wollbrant@economics.gu.se

1. Introduction

The problem of self-control persists in many domains of private goods consumption. Well-known examples include smoking and dieting (see e.g., Fredericks *et al*., 2002). Less is known, however, about the role of self-control in the domain of pro-social behavior, such as the voluntary provision of public goods.[1] The idea that problems of pro-social versus selfish behavior may be understood as problems of self-control has been explored only recently. Loewenstein (1996) suggests that selfish behavior may arise from visceral urges or drive-states, similar to those manifested in cravings for relief of hunger, pain, and sexual deprivation. O'Donaghue and Loewenstein (2007) argue that such selfish urges often conflict with the "colder," more abstract preferences for altruism, just as visceral urges for sweets could conflict with more abstract preferences for a fine figure or good health. This conceptualization of pro-social behavior mirrors the dual process modes of cognition from experimental psychology (e.g., Kahneman, 2003; Sloman, 1996). This perspective distinguishes between two modes of cognition. Roughly speaking, one is rule-based, effortful, "cold," and relies on serial processing (our "rational" self). The other is associative, parallel, "hot," and relies on parallel processing (our "emotional" self). The latter is known as "System 1" and the former "System 2."

Taken together, this implies that System 2 often is responsible for altruistic actions, while System 1 may be a driver of greed. The evidence to date, however, is relatively scarce, and much is indirect. For example, Pronin *et al*. (2008) show that decisions about others resemble decisions about "future selves," both classes of which contrast to decisions about less abstract "present selves." Moreover, Curry *et al.* (2008) find that contribution to a public good is negatively correlated with discount rates; the more "patient" individuals are, the more they contributed to the public good. More recently, Martinsson *et al*. (2010) find both in a dictator and in a one-shot public good game direct support for the hypothesis that individuals with higher self-control act more pro-socially when conflict is "identified."

[1] Pro-social preferences have attracted significant research interest (e.g., overview in Fehr and Schmidt, 2006). Much work explores why individuals voluntarily contribute to charity (e.g., Andreoni, 2006; List, 2008). Work on the underlying motivations for pro-social behavior is extensive. For example, Bénabou and Tirole (2006) classify the motivations into three broad categories: intrinsic (e.g., pure altruism), extrinsic (i.e., monetary rewards), and image motivation (e.g., reputation). A similar discussion is found in Ariely *et al.* (2009). For early work on motivation, see for example Deci (1975).

A prevalent form of pro-social behavior given much attention in the experimental economics literature is that of "conditional cooperation" (e.g., Fischbacher *et al*., 2001; Keser and van Winden, 2000). Conditional cooperation means that individuals are willing to contribute to the public good when others contribute; individuals' contributions to the public good rise when the expectation of others' contributions increases. Fischbacher *et al*. (2001) develop and apply the strategy method. In addition to reporting how much they would unconditionally contribute to the public good (as in a traditional public goods experiment), each subject reports how much she would contribute given all possible integer combinations of other subjects' average contributions. This latter contribution is denoted *conditional contribution* (not to be confused with the contributor type called *conditional contributor*). By investigating the relationship between own and others' contributions to the public good, tone may classify subjects into different contributor types. Previous experiments have found conditional cooperators to represent the majority, with a smaller fraction of free-riders (e.g., Fischbacher and Gächter, 2009; Herrmann and Thöni, 2009; Kocher *et al*., 2008).

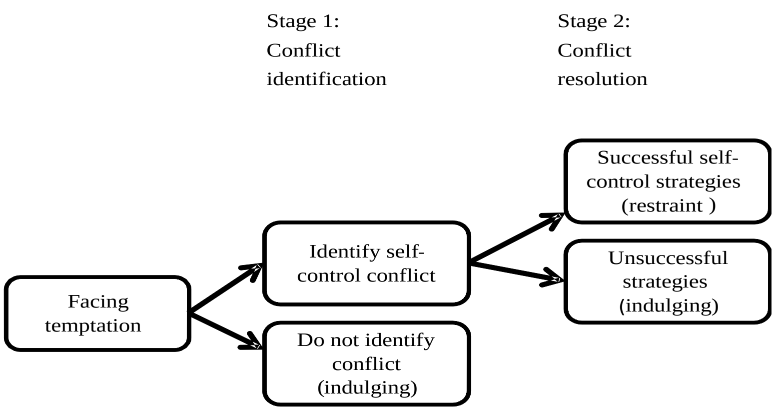
Nonetheless, one could imagine even that individuals with preferences for conditional cooperation may feel tempted to act selfishly and, therefore, not to completely match their expectations of others' behavior. Rather they may appear as "imperfect conditional cooperators" (e.g., Fischbacher and Gächter, 2009). That an individual's target contribution increases when the belief about *others'* contribution increases is relevant to self-control problems for two reasons. On one hand, for the pro-social self of the individual (System 2), the cost of not contributing a positive amount increases when others' contributions increase. Thus the pro-social self should be willing to exert more costly self-control effort in the pursuit of cooperative behavior. On the other hand, with the increase in belief about others' contributions, and the wish to match contributions, greed also becomes a relevant force. For the individual's greedy self (System 1), any contribution conflicts with self-interest. Because increasing beliefs dictate a larger contribution, the sensation of greed should increase as beliefs increase. Indeed, were the individual not to believe that anyone else contributes, or were the individual a free-rider (who holds no pro-social preferences), then greed would not conflict with any wish to contribute.

In this way, when System 2 of an individual is characterized by conditionally cooperative preferences, she wishes to increase her contribution when her belief about others' contribution increases. System 1, however, prefers not to contribute anything. In response to

System 2's belief about appropriate levels of contribution, System 1 exhibits greed, an impulse not to contribute. The more System 2 wants to contribute, the greater is System 1's impulse not to contribute. In response to System 1's impulse not to contribute, System 2 exercises self-control (or willpower) to resist the impulse of greed. The outcome of the conflict will depend on the relative strengths of the two modes of cognition (willpower vs. greed). Hence, this leads us to hypothesize that the degree of conditional cooperation may depend on willpower and could potentially help explain the often empirically observed phenomenon of "imperfect conditional cooperation."

Myrseth and Fishbach (2009) propose a two-stage model of self-control, according to which two conditions are necessary for successful restraint in the face of temptation. First, the individual identifies a conflict between indulging and pursuing a "higher-order goal," i.e., the most preferred alternative of System 2. Second, the individual employs self-control strategies, but only if conflict has been identified at the first stage (see Figure 1). Thus, self-control only is relevant when conflict has been identified. Such self-control strategies may take a variety of forms, and common examples include willpower (e.g., Baumeister *et al.*, 1994) and pre-commitment (Thaler and Shefrin, 1981; Schelling, 1984).

Figure 1. The two-stage model of self-control.



Source: Myrseth and Fishbach (2009).

Because self-control is relevant to the decision to indulge only when the individual has identified a self-control conflict, a possible strategy for testing whether the problem of conditional cooperation versus selfish behavior resembles one of self-control is to investigate whether self-control is positively correlated with conditional cooperation when individuals have identified self-control conflict, but to a lesser extent when individuals have not.

In some instances, the question of conflict identification in the face of temptation is trivial. For example, the diabetic dieter knows that having even a single, tempting chocolate may incur major costs in terms of ill-health, while for a non-diabetic dieter facing the same chocolate the question is more ambiguous. Having the chocolate only will incur substantial costs if consumed extensively. Similarly, the fair-minded individual might find that generally not contributing to his team of hardworking colleagues would threaten his self-image, but doing so on a few occasions is a more ambiguous matter.

Tempting opportunities that incur but trivial costs when consumed in small amounts, but potentially significant costs when consumed extensively, are termed epsilon cost temptations by Myrseth and Fishbach (2009). Individuals identify self-control conflict in the face of epsilon cost temptation, they argue, if and only if two conditions are met: (a) the focal consumption opportunity must be viewed in relation to multiple additional opportunities, and (b) the decision maker must assume that similar choices are made for each opportunity. That is, considering whether or not to consume a chocolate will evoke self-control conflict in the non-diabetic dieter if the consumption opportunity is viewed in relation to future opportunities for chocolate consumption, but not if the consumption opportunity is viewed in isolation, as a singular event. Similarly, for the fair-minded individual, the question of whether or not to contribute to the efforts of her hard-working colleagues may trigger self-control conflict if the choice of contributing is viewed in relation to future opportunities for cooperation, but not if viewed in isolation.

In the face of epsilon cost temptations, subtle framing manipulations may be sufficient to influence identification of self-control conflict (Myrseth and Fishbach, 2009). In their study, presenting to participants a calendar of the current month with a grid, which separated the dates of the month, increased participants' consumption of potato chips compared to presenting to participants a calendar with no such grid (and thus no visual separation of the dates). They argue that the gridded calendar activated an isolated (versus interrelated) frame of the choice opportunity, thereby raising to the likelihood that participants isolated the date in question. In other words, participants were less likely to view the decision task in relation to similar future opportunities. Consequently, the gridded calendar reduced the likelihood that participants identified a conflict. Consistent with this argument, participants who viewed the calendar with the grid reported that they experienced less conflict during their decision to have chips or not than did those who viewed a calendar without a grid. Participants' trait ability to implement self-control strategies, measured by Rosenbaum's (1980) psychometric scale, was positively correlated with chips consumption for those who viewed the calendar without the grid (and who were more likely to identify conflict), but not for those who viewed the calendar with a grid (and who were less likely to identify conflict).

The objective of this paper is to test the hypothesis that the degree of conditional cooperation, associated with an increase in the belief about others' contributions, increases as self-control increases. To test this hypothesis, we combine a strategy method version of the classic public goods game (Fischbacher *et al*., 2001), with the framing manipulation from Myrseth and Fishbach (2009). The Fischbacher *et al*. (2001) design, and its reliance on the strategy method, is suitable for the test since participants explicitly are asked to condition their decisions on their beliefs about others' behavior. To capture individual's self-control, we used the Rosenbaum Self-Control Schedule (henceforth, Rosenbaum score) (Rosenbaum, 1980).[2] If conditional cooperation versus selfish behavior in the experiment could represent a self-control conflict, one would expect that the interaction of participant's beliefs about other's contributions and their trait self-control, as measured by Rosenbaum's (1980) scale, is positively correlated with contributions for those participants who view a calendar of the current month without a grid, but less so for those participants who view the same calendar with the grid.

[2] The Rosenbaum score has been externally validated against several criteria (e.g., Rosenbaum and Rolnick, 1983) and measures individual's cognitive skills for exercising self-control in the face of temptation. Each subject is asked to respond to 36 statements using a 6-point Likert-scale (see appendix C). Each statement is graded from -3 to +3 (zero excluded). Thus, "perfect" self-control corresponds to +108 and "no self-control" to -108.

2. Experimental design and procedure

2.1. Experimental treatments

We employed three treatments – the isolated frame, the standard frame and the interrelated frame – using a between-subject design. The isolated versus interrelated frame was manipulated with the procedure from Myrseth and Fishbach (2009); participants viewed a calendar that displayed the present month, either containing a grid separating the dates or containing no such grid (see Appendix B). Furthermore, the date of the experiment was highlighted in grey in the calendar with a grid, but not in the calendar without.

Because we expected participants who viewed a calendar with the grid to adopt an isolated frame of their decisions (less likely to identify self-control conflict), we refer to this treatment as the *isolated treatment*. In contrast, we expected participants who viewed a calendar without the grid to adopt a less isolated frame. Because their decisions thus would be viewed relatively more related to similar future decisions, we refer to this treatment as the *interrelated treatment* (higher likelihood of self-control conflict identification). We expected the interaction of participant's beliefs about other's contributions and their trait self-control, as measured by Rosenbaum's (1980) scale, to be positively correlated with contributions in the experiment for participants who viewed a calendar of the current month without a grid, and less so for participants who viewed the same calendar with.

We denote the third treatment, without a calendar, as the *standard treatment*. Our main reason for including this treatment is related to the validity of the Rosenbaum score. One might claim that any positive correlation between the interaction of the Rosenbaum and beliefs about others' behavior with own contributions is due to the Rosenbaum score measuring pro-social preferences, and not due to the identification manipulation. If, however, we find in the standard frame a weaker correlation between the interaction and own contribution, we may rule out this possibility. Moreover, the standard frame allows us to detect irregularities within our sample compared to those in previous work, to ensure that treatment effects may not be attributed to sampling effects.

2.2. The public good game

We recruited student subjects from various undergraduate classes at Escuela de Ingenieria de Antiouqia (EIA), Medellín, Colombia, 2008. We held six sessions, two for each treatment, with 24-28 participants per session. Nobody participated in more than one experimental session, and nobody was a student of mathematics, psychology, or economics.

We employed a standard linear public goods experiment. Each member was endowed with 20 tokens, which they were to divide between a public and a private good. The payoff for member i, measured in units of tokens, was calculated according to the payoff function

$$\pi_i = 20 - c_i + 0.4\sum_{i=1}^{4} c_i, \qquad (1)$$

where c_i is individual $i's$ contribution to the public good. Contribution to the public good would yield to each member a marginal per capita return of 0.4 tokens. Each group consisted of four members. The choice of parameter values replicates the features of a public good since full contribution to the public good is Pareto optimal, while the dominant strategy is zero contribution to the public good – that is, to free-ride.

As in Fischbacher *et al.* (2001), participants were asked to make both unconditional and conditional contributions to the public good. In the case of unconditional contributions, they were asked simply how much they wished to contribute, as in a standard public goods game. However, in the case of conditional contributions, participants were asked how much they would like to contribute conditional on the average contribution of the other group members, ranging from 0 to 20, rounded to the nearest integer.

Participants were randomly assigned to groups of four (from the same session).[3] To make each decision incentive compatible, the unconditional contribution was the payoff-relevant decision for three randomly selected members. Using their average unconditional contribution, the contribution of the fourth member was given by her contribution table. Then, each member's monetary payoff would follow from equation (1). Each token in the experiment was exchanged for 750 Colombian pesos. The average earnings per participant

[3] The selection was anonymous and hence no participant knew to which group, she belonged.

were 25,000 Colombian pesos (including a 5,000 peso show-up fee).[4] The sessions lasted about 1.5 hours.

3. Results

We expected that conditional contributions would depend on the interaction between identification of self-control conflict, induced by treatments, and success at the conflict stage (see Figure 1), measured by the Rosenbaum Self-Control Schedule. We also expected trait self-control to exhibit a weaker correlation (if any) with contribution in the isolated treatment since participants in this treatment would adopt a more isolated view of their decisions. In contrast, we expected that the interrelated treatment would yield a higher likelihood of conflict identification. Thus, we expected self-control to exhibit a stronger positive correlation with contribution.

We hypothesized that success at the conflict stage would depend on the belief about the average contribution of other group members (henceforth *Others*) and that it would interact with willpower. Conditional contribution preferences dictate that one's contribution increases if one's expectation of *Others* contribution increases. An increased contribution, however, implies a larger material cost to the individual, and thus a stronger impulse to act selfishly. By applying self-control, the individual may resist the impulse to act selfishly. Therefore, we expected the interaction term between *Others* and the Rosenbaum score to exhibit a stronger positive correlation with contributions than would *Others* alone.

Regression analysis (reported in Table 1) supports our hypothesis. In each treatment, the variable *Others* is positive and significant at the one-percent level. Moreover, the interaction term between the Rosenbaum score and *Others* only is significant in the interrelated treatment, where we expected that identification of self-control conflict would be most likely. The effect is of economic significance because the marginal effect in the interrelated treatment of *Others* is: $0.215 + 0.006 \mathit{Rosenbaum\ score}$. The marginal effect evaluated at the mean of the Rosenbaum score (29.99) is approximately equal to 0.395. Increasing the Rosenbaum score by one standard deviation (19.94) implies a marginal effect of

[4] The exchange rate at the time of the experiment was approximately 1 USD = 1,762 Colombian Pesos.

approximately 0.120 higher than the marginal effect evaluated at the mean of the Rosenbaum score, totaling 0.515 and corresponding to a 30% increase.

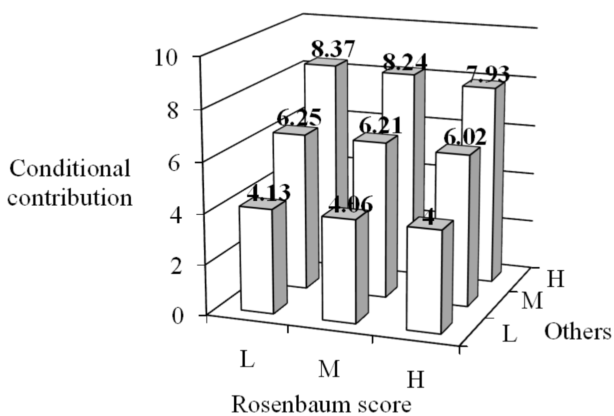
Table 1. Regression results.

Dependent variable: Conditional contribution	OLS
Isolated treatment	1.431
	(1.602)
Standard treatment	-0.705
	(-1.371)
Isolated treatment x Others	0.360***
	(3.350)
Standard treatment x Others	0.406***
	(4.699)
Interrelated treatment x Others	0.215***
	(2.710)
Isolated treatment x Rosenbaum score x Others	-0.001
	(-0.380)
Standard treatment x Rosenbaum score x Others	0.000
	(0.200)
Interrelated treatment x Rosenbaum score x Others	0.006***
	(2.810)
Constant	1.313***
	(3.768)
R-squared	0.177
Number of observations	3234

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Note: Standard errors clustered on individual obs.

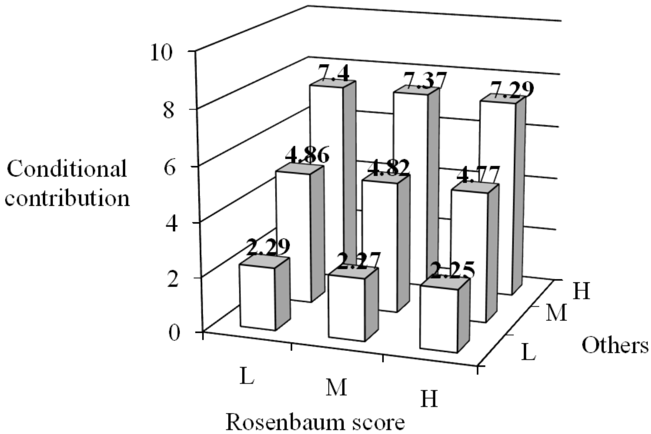
Using the regression results, we plot the data for three different values of each independent variable. These levels are, “Low”, “Mean” and “High.” Conditional contribution is reported on the vertical axis, as a function of Rosenbaum score and *Others* by treatment. The threshold for classification of the Rosenbaum score is set at one standard deviation (such that “Low” is the mean value of the variable less one standard deviation, and “High” is the mean plus one standard deviation) while for *Others* the corresponding levels for “Low,” "Medium" and "High" are set at 4, 10, and 16, respectively. This provides nine coordinates (3 *x* 3), for which we estimate the conditional contribution. We use the estimated model to predict values for each of the nine coordinates. The data plots are presented in Figures 2-4 for the isolated, standard and interrelated treatment, respectively. Figures 2 and 3 suggest that conditional contributions increase as *Others* increases, but they show no association between conditional contributions and the Rosenbaum score. This corresponds to the standard observation in public goods experiments. Similar sensitivity to beliefs about *Others* contributions also appears in figure 4. However, as hypothesized, it appears that the effect of increasing *Others* is stronger for higher levels of self-control. Indeed, the highest level of conditional cooperation is observed in the (High, High) cell in the Interrelated treatment, where we expected self-control to matter.

Figure 2. Conditional contribution by levels of self-control (Rosenbaum score) and belief about others' average contribution (Others) in the Isolated treatment.



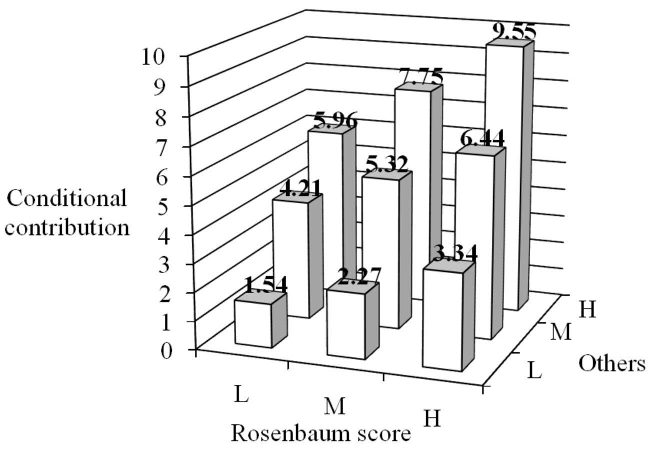
Note. L, M and H correspond to "Low," "Medium," and "High."

Figure 3. Conditional contribution by levels of self-control (Rosenbaum score) and belief about others' average contribution (Others) in the Standard treatment.



Note. L, M and H correspond to "Low," "Medium," and "High."

Figure 4. Conditional contribution by levels of self-control (Rosenbaum score) and belief about others' average contribution (Others) in the Interrelated treatment.



Note. L, M and H correspond to "Low," "Medium," and "High."

4. Discussion

This paper joins a larger line of research that attempts to understand how individuals act on basis of ostensibly conflicting preferences. While the economic literature on self-control has addressed this problem for quite a while (see Fredericks et al., 2002, for a review), the economic literature on pro-social behavior has not. This paper follows Martinsson et al (2010) in exploring the idea that the question of pro-social versus selfish behavior may represent one of self-control. As already established in the economic literature, cooperation behavior may be divided into two distinct domains: conditional versus unconditional cooperation (e.g., Fischbacher et al., 2001). While Martinsson et al (2010) approached the domain of unconditional cooperation, this paper takes aim at conditional cooperation.

As Loewenstein (1996) proposes, selfish impulses may resemble "hot" urges for food or water, whereas pro-social preferences may be of "colder," more abstract nature. Therefore, as when the urge for sweets conflicts with the preference to stay slim, self-control, or "willpower," may determine behavior when selfish urges conflict with pro-social preferences. Martinsson et al. (2010) found evidence that higher self-control was associated with higher unconditional contribution for subjects who had identified a self-control conflict. In line with these findings, this paper finds with the strategy method that higher self-control was associated with higher conditional contribution for subjects who had identified a self-control conflict. More specifically, we find that while the expectation of others' behavior indeed is a major predictor of contribution, it is a stronger predictor when individuals have high self-control. Furthermore, our results were of economic significance; the predicted conditional cooperation evaluated at the mean of the self-control score plus one standard deviation was 30% larger than the predicted contribution evaluated at the mean alone.

These findings are important not only because they shed light on the mechanisms behind an important aspect of pro-social behavior. They may also help explain the empirical regularity of "imperfect conditional cooperation," that individuals, in fact, contribute less than what they think others will do. If individuals believe that they ought to match others' contribution, but at the same time are tempted to keep their endowment for themselves, limited self-control may cause them to contribute less than what that they believe that they ought to contribute, or even to contribute nothing at all.

Our findings merit further investigation into possible field applications. The results suggest that there is quite a potential to boost pro-social behavior by helping individuals apply their own self-control resources. Such help could take the form of helping individuals boost their self-control, but it might also take the simple form of a simple reminder that their resources indeed are applicable and relevant to the task at hand.

References

Andreoni, J., (2006). Philanthropy. In: Kolm, S.C., Ythier, J.M. (Eds.). *Handbook of the Economics of Giving, Reciprocity, and Altruism*, vol. 2. North-Holland, Amsterdam.

Ariely, D., Bracha, A. and Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially, *The American Economic Review*, 99:1, 544-555.

Baumeister, R.F., Heatherton, T.F., & Tice, D.M. (1994). Losing control: How and why people fail at self-regulation. San Diego, CA: Academic Press.

Benabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review* 96:5, 1652-1678.

Burger, N., Charness, G., Lynham, J. (2009) Three field experiments on procrastination and willpower, Working paper, University of California Santa Barbara.

Curry, O.S., Price, M.E. and Price, J.G. (2008). Patience is a virtue: Cooperative people have lower discount rates. *Personality and individual differences* 44:3, 780-785.

Deci, E.L. (1975). *Intrinsic motivation*. New York: Plenum

Eckel, C.C. and Grossman, P.J. (1996). Altruism in anonymous dictator games, *Games and Economic Behavior* 16:2, 181–191.

Fehr, E., and K. M. Schmidt (2006). The economics of fairness, reciprocity and altruism: Experimental evidence, Chapter 8, in (eds.) Kolm, S. ans J.M. Ythier, *Handbook of the economics of giving, altruism and reciprocity, 1*, North-Holland.

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

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments, *American Economic Review* 100:1, 546-665.

Fredericks, S., Loewenstein, G. and O'Donoghue, T. (2002). Time discounting and time preference: A critical review, *Journal of Economic Literature* 15, 351-401.

Herrmann, B. and Thöni, C. (2009). Measuring conditional cooperation: A replication study in Russia, *Experimental Economics* 12, 87-92.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58, 697-720.

Keser, C. and van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods, *Scandinavian Journal of Economics* 102, 23-39.

List, J., (2008). Introduction to field experiments in economics with applications to the economics of charity, *Experimental Economics* 11, 203-212.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes* 65, 272–292.

Martinsson, P., Myrseth, K.O.R. and Wollbrant, C. (2009). Pro-social behavior: The role of self-control. Working Paper: The University of Chicago.

Myrseth, K.O.R. and Fishbach, A. (2009a). Self-control: A function of knowing when and how to exercise restraint. *Current Directions in Psychological Science*.

Myrseth, K.O.R. and Fishbach, A. (2009b). Activating self-control: Isolated vs. interrelated temptations. Working paper, the University of Chicago.

Myrseth, K.O.R., Fishbach, A., and Trope, Y. (2009). Counteractive self-control: When making temptation available makes temptation less tempting. *Psychological Science 20*, 159 - 163.

Myrseth, K.O.R. and Wollbrant, C. (2009). A theory of self-control conflict: The pyrrhic motions of reason and passion. Working Paper: The University of Chicago.

Plato (1986). Protagoras. In B.A.F. Hubbard and E.S. Karnofsky (Trans.), *The Dialogues of Plato.* New York: Bantam Books. (Original work published 380 BC).

Pronin, E., Olivola, C. Y., and Kennedy, K. A. (2008). Doing unto future selves as you would do unto *Others*: Psychological distance and decision making. *Personality and Social Psychology Bulletin* 34, 224 – 236.

Rosenbaum, M. (1980). A schedule for assessing self-control behaviors: Preliminary findings. *Behavior Therapy 11,* 109–121.

Rosenbaum, M. and Rolnick, A. (1983). Self-control behaviors and coping with seasickness. *Cognitive Therapy and Research 7*, 93 – 97.

Appendix A: Calendars

A.1 The isolated frame (the highlighted date is the same as today's date).

Before we continue with the experiment, please take a moment to consider this month's calendar:

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

*What is today's date?*__________

A.2 The interrelated frame.

Before we continue with the experiment, please take a moment to consider this month's calendar:

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

*What is today's date?*__________

Not for Publication

Appendix B: The Rosenbaum Self-Control Schedule

Note: * = item is reverse scored.

Directions - Indicate how characteristic or descriptive each of the following statements is of you by using the code given below

+3 very characteristic of me, extremely descriptive
+2 rather characteristic of me, quite descriptive
+1 somewhat characteristic of me, slightly descriptive
-1 somewhat uncharacteristic of me, slightly undescriptive
-2 rather uncharacteristic of me, quite undescriptive
-3 very uncharacteristic of me, extremely nondescriptive

l. When I do a boring job, I think about the less boring parts of the job and the reward that I will receive once I am finished.

-3	-2	-1	1	2	3

2. When I have to do something that is anxiety arousing for me, I try to visualize how I will overcome my anxieties while doing it.

-3	-2	-1	1	2	3

3. Often by changing my way of thinking I am able to change my feelings about almost everything.

-3	-2	-1	1	2	3

4. I often find it difficult to overcome my feelings of nervousness and tension without any outside help.*

-3	-2	-1	1	2	3

5. When I am feeling depressed I try to think about pleasant events.

-3	-2	-1	1	2	3

6. I cannot avoid thinking about mistakes I have made in the past.*

-3	-2	-1	1	2	3

7. When I am faced with a difficult problem, I try to approach its solution in a systematic way.

-3	-2	-1	1	2	3

8. I usually do my duties quicker when somebody is pressuring me.*

-3	-2	-1	1	2	3

9. When I am faced with a difficult decision, I prefer to postpone making a decision even if all the facts are at my disposal.*

-3	-2	-1	1	2	3

10. When I find that I have difficulties in concentrating on my reading, I look for ways to increase my concentration.

-3	-2	-1	1	2	3

11. When I plan to work, I remove all the things that are not relevant to my work.

-3	-2	-1	1	2	3

12. When I try to get rid of a bad habit, I first try to find out all the factors that maintain this habit.

-3	-2	-1	1	2	3

13. When an unpleasant thought is bothering me, I try to think about something pleasant.

-3	-2	-1	1	2	3

14. If I would smoke two packages of cigarettes a day, I probably would need outside help to stop smoking.*

-3	-2	-1	1	2	3

15. When I am in a low mood, I try to act cheerful so my mood will change.

-3	-2	-1	1	2	3

16. If I had the pills with me, I would take a tranquilizer whenever I felt tense and nervous.*

-3	-2	-1	1	2	3

17. When I am depressed, I try to keep myself busy with things that I like.

-3	-2	-1	1	2	3

18. I tend to postpone unpleasant duties even if I could perform them immediately.*

-3	-2	-1	1	2	3

19. I need outside help to get rid of some of my bad habits.*

-3	-2	-1	1	2	3

20. When I find it difficult to settle down and do a certain job, I look for ways to help me settle down.

-3	-2	-1	1	2	3

21. Although it makes me feel bad, I cannot avoid thinking about all kinds of possible catastrophes in the future.*

-3	-2	-1	1	2	3

22. First of all I prefer to finish a job that I have to do and then start doing the things I really like.

-3	-2	-1	1	2	3

23. When I feel pain in a certain part of my body, I try not to think about it.

-3	-2	-1	1	2	3

24. My self-esteem increases once I am able to overcome a bad habit.

-3	-2	-1	1	2	3

25. In order to overcome bad feelings that accompany failure, I often tell myself that it is not so catastrophic and that I can do something about it.

-3	-2	-1	1	2	3

26. When I feel that I am too impulsive, I tell myself "stop and think before you do anything."

-3	-2	-1	1	2	3

27. Even when I am terribly angry at somebody, I consider my actions very carefully.

-3	-2	-1	1	2	3

28. Facing the need to make a decision, I usually find out all the possible alternatives instead of deciding quickly and spontaneously.

-3	-2	-1	1	2	3

29. Usually I do first the things I really like to do even if there are more urgent things to do.*

-3	-2	-1	1	2	3

30. When I realize that I cannot help but be late for an important meeting, I tell myself to keep calm.

-3	-2	-1	1	2	3

31. When I feel pain in my body, I try to divert my thoughts from it.

-3	-2	-1	1	2	3

32. I usually plan my work when faced with a number of things to do.

-3	-2	-1	1	2	3

33. When I am short of money, I decide to record all my expenses in order to plan more carefully for the future.

-3	-2	-1	1	2	3

34. If I find it difficult to concentrate on a certain job, I divide the job into smaller segments.

-3	-2	-1	1	2	3

35. Quite often I cannot overcome unpleasant thoughts that bother me.*

-3	-2	-1	1	2	3

36. Once I am hungry and unable to eat, I try to divert my thoughts away from my stomach or try to imagine that I am satisfied.

-3	-2	-1	1	2	3

Not for publication

Appendix C: Experimental Instructions

C.1.: Original Instructions for the Public Good Game

Instrucciones

Usted va a participar en un experimento sobre toma de decisiones. A finales de este experimento, le pagará a usted una cantidad dependiendo de sus decisiones y las decisiones de otros. El pago se hará en efectivo al final del experimento.

A lo largo del experimento, todo tipo de comunicación queda completamente prohibida. Los participantes que se comuniquen quedarán excluidos del experimento y no recibirán ningún pago. Si tiene alguna pregunta, levante la mano para que algún monitor venga a responder la pregunta en privado.

Durante el experimento no hablaremos de Pesos sino de "fichas". Sus ganancias en el experimento serán en fichas. Al final del experimento, las fichas serán convertidas a pesos a una tasa de cambio de:

2 fichas = 1500 pesos

Independientemente de sus decisiones, recibirá 5000 pesos por participar en el experimento.

Además tendrá que responder algunos cuestionarios durante el experimento. Algunas preguntas pueden parecerle muy extrañas, sin embargo, le pedimos el favor que las responda seriamente. Todas sus respuestas serán **confidenciales y anónimas.** Para identificarlo usaremos el número de identificación que usted recibió al entrar en el salón. Este mismo número será usado para identificarlo cuando hagamos los correspondientes pagos después del experimento. Antes de que usted abandone el salón, debe entregar a uno de los monitores el número de identificación que recibió al entrar al salón. El monitor pone ese número dentro de un sobre, lo sella y se lo entrega a usted nuevamente. Cuando vaya a reclamar su pago, usted debe presentar el mismo sobre debidamente sellado y con su número de identificación adentro, tal y como el monitor del experimento se lo entregó antes de abandonar el salón.

A lo largo de estas instrucciones le presentaremos algunos ejemplos, por favor considere los números de los ejemplos como una ilustración. Los números que usted obtendrá en el experimento pueden ser diferentes.

Decisión Básica

El experimento va a funcionar así: Primero vamos a explicar la decisión básica, luego vamos a hacer unas preguntas de control que le ayudarán a entender esta decisión básica.

Usted pertenece a **un grupo de cuatro personas**. Nadie, excepto los monitores sabrán quién pertenece a qué grupo. Los grupos se formarán al azar. Al principio del experimento recibirá un papel donde dice la cantidad de fichas que tiene inicialmente para jugar. Esta será su dotación. Cada una de las cuatro personas del grupo tiene que decidir cómo distribuir su dotación de 20 fichas. Puede poner todas, una parte o ninguna ficha en una **cuenta de un proyecto**. Las fichas que no deposite en la cuenta del proyecto serán automáticamente transferidas a su **cuenta privada**.

Su ingreso de la cuenta privada:

Por cada ficha que deposite en la cuenta privada ganará exactamente una ficha. Por ejemplo, si tiene una dotación de 20 fichas y deposita cero fichas en la cuenta del proyecto (o sea que deposita 20 fichas en su cuenta privada), entonces recibe exactamente 20 fichas. Si en cambio deposita 14 fichas en la cuenta del proyecto (es decir 6 fichas en la cuenta privada) entonces su ingreso de la cuenta privada son 6 fichas. *Nadie, excepto usted mismo recibe fichas de la cuenta privada.*

Su ingreso de la cuenta del proyecto:

Todos reciben el mismo ingreso por las fichas que usted deposite en la cuenta del proyecto. Por supuesto, usted también obtendrá ingresos por las fichas que otras personas depositen en la cuenta del proyecto. Para cada persona el ingreso de la cuenta del proyecto se determina de la siguiente manera:

Ingreso de la cuenta del proyecto =

Suma de las contribuciones al proyecto x 0.4

Por ejemplo, si la suma de las contribuciones a la cuenta del proyecto es 60 fichas, usted y los otras personas del grupo recibirán 60x0.4=24 fichas para cada uno. Si las cuatro personas del grupo depositan un total de 10 fichas en la cuenta del proyecto, entonces usted y todos los otros reciben 10x0.4=4 fichas por la cuenta del proyecto.

Ingreso total:

Su ingreso total es la suma del ingreso de su cuenta privada más el ingreso de la cuenta del proyecto.

Ingreso de la cuenta privada(= Numero de fichas de dotación inicial – su contribución a la cuenta del proyecto)

+ Ingreso de la cuenta del proyecto (= 0.4 $\times$ suma de todas las contribuciones a la cuenta del proyecto)

Ingreso total

Antes de terminar de leer las instrucciones le queremos pedir el favor que responda las siguientes preguntas de control que ayudarán a verificar si ha entendido todo correctamente. Si hay alguna pregunta, por favor levante la mano para que un monitor le responda en privado.

Preguntas de Control

Por favor responda las siguientes preguntas de control. El propósito de estas preguntas es familiarizarlo con los cálculos de los ingresos en fichas que resultan de las diferentes decisiones sobre cómo distribuir sus recursos disponibles. Por favor responda todas las preguntas y escriba los correspondientes cálculos.

1. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Supongamos también que nadie, ni usted, pone nada en la cuenta del proyecto.

   ¿Cuál es su ingreso total? ____

   ¿Cuál es el ingreso de las otras personas de su grupo? ____, ____, y ____

2. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Supongamos también que todos, incluido usted, ponen todos sus recursos en la cuenta del proyecto.

   ¿Cuál es su ingreso total? ____

   ¿Cuál es el ingreso de las otras tres personas de su grupo? ____, ____, y ____

3. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Supongamos también que las otras tres personas del grupo depositan 30 fichas en la cuenta del proyecto.

   a) ¿Cuál es su ingreso total si además de las 30 fichas de las otras tres personas de su grupo usted pone 0 fichas en la cuenta del proyecto?

   Su ingreso total: ____

b) ¿Cuál es su ingreso total si además de las 30 fichas de las otras tres personas de su grupo usted pone 8 fichas en la cuenta del proyecto?

Su ingreso total: ____

c) ¿Cuál es su ingreso total si además de las 30 fichas de las otras tres personas de su grupo usted pone 15 fichas en la cuenta del proyecto?

d)

Su ingreso total: ____

4. Cada persona del grupo tiene una dotación de 20 fichas disponibles. Usted pone 8 fichas en la cuenta del proyecto.

a) ¿Cuál es su ingreso total si además de sus 8 fichas las otras personas del grupo ponen en total otras 7 fichas en la cuenta del proyecto?

Su ingreso total: ____

b) ¿Cuál es su ingreso total si además de sus 8 fichas las otras tres personas del grupo ponen en total otras 12 fichas en la cuenta del proyecto?

Su ingreso total: ____

c) ¿Cuál es su ingreso total si además de sus 8 fichas las otras tres personas del grupo ponen otras 22 fichas en la cuenta del proyecto?

Su ingreso total: ____

Si ha terminado estas preguntas antes que los otros, le recomendamos que piense en otros ejemplos adicionales para que se familiarice con este tipo de situaciones.

Procedimiento

El experimento consiste en la decisión que acabamos de describir. En la siguiente parte explicaremos el procedimiento a usar en detalle.

Como usted sabe, usted tiene una dotación de 20 fichas. Puede poner esas fichas en la cuenta del proyecto y el resto de fichas automáticamente se depositan en la cuenta privada. Cada persona en el grupo tendrá la misma dotación.

Cada persona del grupo tiene que hacer dos tipos de decisiones, a las que nos referiremos de aquí en adelante como "**contribución incondicional"** y como "**tabla de contribuciones"**

- En la **contribución incondicional** usted debe decidir cuántas de las fichas que tiene disponibles deposita en la cuenta del proyecto. Escriba esta cantidad al lado de *"Su contribución incondicional a la cuenta del proyecto"* en la segunda página de su hoja de decisión. Usted debe escribir un **número entero** que **no puede ser menor a cero ni mayor a las 20 fichas que usted tiene en su dotación**. La diferencia entre su dotación de 20 fichas y las fichas que pone en la cuenta del proyecto es automáticamente depositada en su cuenta privada.

- Su segunda tarea es completar la **tabla de contribuciones** en la tercera página de la hoja de decisión. En la tabla de contribución usted debe indicar para cada posible contribución promedio de las otras tres personas del grupo (aproximado al siguiente entero; por ejemplo si el promedio es 17,5 piense en 18) el número de fichas que usted quiere poner en la cuenta del proyecto. Usted decidirá cuanto contribuir dependiendo de lo que las otras personas contribuyan. Esto será más claro cuando vea el siguiente ejemplo de una tabla de contribución.

Contribución promedio (aproximada) de las otras tres personas a la cuenta del proyecto	**Su contribución a la cuenta del proyecto es:**
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

Los números en la columna izquierda son los posibles valores promedio (aproximados) de las contribuciones de las otras personas del grupo. Supongamos en el ejemplo que las otras personas tienen 20, 20 y 20 fichas, entonces en promedio pueden contribuir máximo 20 fichas ((20+20+20)/3=20).

Usted simplemente tiene que escribir en la columna de la derecha cuántas fichas quiere contribuir en la cuenta del proyecto dado que los otros contribuyen en promedio (aproximado) la cantidad de fichas de la columna izquierda. Tiene que completar todas las casillas de la columna derecha. Por ejemplo, debe escribir cuántas fichas quiere contribuir a la cuenta del proyecto si los otros contribuyen en promedio 0 fichas a la cuenta del proyecto; cuánto contribuye si los otros contribuyen 1 o 2 o 3, fichas, etc. En cada casilla debe escribir un número entero no menor de cero y no mayor a su dotación de 20 fichas. Por supuesto que puede escribir el mismo número en diferentes casillas.

Después que todos los participantes en el experimento han hecho su contribución incondicional y han completado la tabla de contribuciones, una persona de cada grupo será seleccionada al azar. Para las personas seleccionadas aleatoriamente el ingreso se determinará de acuerdo con la tabla de contribuciones. Para las otras tres personas del grupo que no son seleccionadas aleatoriamente la contribución incondicional determinará el ingreso. Cuando usted está decidiendo la contribución incondicional y la tabla de contribuciones, usted no sabe si va a ser seleccionado aleatoriamente, así que piense cuidadosamente los dos tipos de decisiones porque cualquiera puede ser relevante para usted. Los siguientes dos ejemplos servirán para aclarar este punto:

Ejemplo 1: Suponga que después de que ha entregado sus decisiones usted es seleccionado al azar. Esto implica que la decisión relevante para sus ingresos es la tabla de contribuciones. Para las otras tres personas la decisión incondicional es la decisión relevante. Supongamos que ellos han hecho contribuciones incondicionales de 0, 2 y 4 fichas. La contribución promedio aproximada es entonces 2 (=(0+2+4)/3).

Si usted ha indicado en su tabla de contribuciones que usted contribuiría una ficha a la cuenta del proyecto si los otros contribuían 2 fichas en promedio, entonces la contribución a la cuenta del proyecto es 0+2+4+1=7. Entonces todas las personas del grupo ganan un ingreso de 0.4x7=2.8 de la cuenta del proyecto más el respectivo ingreso de la cuenta privada.

Si en cambio usted ha indicado que contribuiría 19 fichas a la cuenta del proyecto si los otros contribuían 2 en promedio, entonces la contribución total a la cuenta del proyecto es 0+2+4+19=25. Todas las personas del grupo ganan un ingreso de 0.4x25=10 fichas de la cuenta del proyecto más el respectivo ingreso de la cuenta privada.

Ejemplo 2: Ahora suponga que usted no es seleccionado aleatoriamente. Esto quiere decir que para usted y para dos otras persona del grupo la contribución incondicional es la decisión de ingreso relevante. Suponga además que su contribución incondicional al proyecto es de 16 y que la de las otras tres personas es 18 y 20. La contribución incondicional promedio del grupo es entonces 18 (=(16+18+20)/3).

Si la persona del grupo que ha sido seleccionada aleatoriamente indicó en la tabla de contribuciones que contribuiría una ficha a la cuenta del proyecto si las otras tres personas contribuían en promedio 18, entonces la contribución total a la cuenta del proyecto es 16+18+20+1=55 fichas. Por lo tanto, todas las personas del grupo ganarían 0.4x55=22 fichas de la cuenta del proyecto adicionalmente a sus respectivos ingresos de las cuentas privadas.

Si la persona del grupo seleccionada aleatoriamente indicó en la tabla de contribuciones que contribuiría 19 fichas a la cuenta del proyecto si las otras tres personas contribuían en promedio 18, entonces la contribución total a la cuenta del proyecto es 16+18+20+19=73 fichas. Por lo tanto, todas las personas del grupo ganarían 0.4x73=29.2 fichas de la cuenta del proyecto adicionalmente a sus respectivos ingresos de las cuentas privadas.

La selección aleatoria de los participantes se hará de la siguiente forma. A cada persona del grupo se le asigna un número entre 1 y 4 que puede ver en la última página de su hoja de decisión. Un participante seleccionará al azar una de las cuatro cartas **después** que todos los participantes han tomado su decisión incondicional y han completado la tabla de contribuciones y el cuestionario. Si la carta que es seleccionada corresponde al número en su hoja de decisión entonces la tabla de contribuciones condicionales en la tercera página es relevante para usted. Si no, la contribución incondicional en la primera página es la decisión relevante. Recuerde que usted sabe cuál de las dos decisiones es la decisión relevante para sus pagos sólo después que ha entregado sus decisiones, por lo tanto debe completar las dos páginas cuidadosamente.

La cantidad de fichas que usted gane se convertirá a pesos que serán pagados en efectivo. Tiene alguna pregunta? Por favor levante la mano y un monitor vendrá a responder la pregunta privadamente.

C.2.: Guessed Contributions for the Public Good Game

Identificación en el experimento:______

Cuánto cree usted que los otros integrantes contribuyeron incondicionalmente a la cuenta del proyecto?

Hace un rato usted escribió su contribución incondicional y llenó una tabla con un número de contribuciones condicionales. Ahora, usted nos dirá cuándo cree que los otros integrantes de su grupo han escrito como sus contribuciones **incondicionales**.

Por favor escriba el número de fichas que **usted cree que los otros tres jugadores** en su grupo han contribuido a la cuenta del proyecto. En otras palabras, qué número sospecha usted que ellos escribieron?

AHORA usted puede ganar más dinero **si adivina correctamente Contribución promedio (aproximada) de las otras tres personas a la cuenta del proyecto**. Usted puede ganar tres fichas adicionales si la verdadera contribución de los otros es igual a lo que usted adivinó, o si está una ficha por encima o por debajo. Asi por ejemplo, si la verdadera contribución de los otros tres es 7 y usted dice 6 o dice 8, usted gana 2 fichas más, pero si usted dice 9 no ganará.

Contribución promedio (aproximada) de las otras tres personas a la cuenta del proyecto	Marque con una x la casilla que usted cree que corresponde a los que los otros tres contribuyeron incondicionalmente a la cuenta del proyecto. ***SÓLO MARQUE UNA CASILLA***
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

Not for publication

Appendix D: Experimental Instructions (Translated from Spanish)

D.1.: Original Instructions for the Public Good Game

Instructions

You will be taking part in an experiment on decision-making. The experiment is designed so that your earnings will depend on both your own decisions and the decisions of others. Your earnings will be paid in cash at the end of the session.

Talking is not allowed throughout the entire session. Any violation of this rule will result in exclusion from the session and not receiving any payment. If you have any questions regarding these instructions, please raise your hand and a member of the experimenter team will attend to you.

Your earnings in this experiment will be in tokens. At the end of the experiment, the tokens will be converted into Colombian pesos (COP) at an exchange rate of:

2 tokens = 1500 COP.

Regardless of what decisions you make, you will receive a show-up fee of 5,000 COP.

During the experiment, you will have to answer a few questionnaires. Although some questions may appear strange to you, we ask you to still take them seriously. All your answers will be treated **confidentially and anonymously**. The identification number you received when entering the room will be used to identify you when paying you after the experiment. Before you leave the room, you should hand the identification number you received when entering the room to a member of the experimenter team. The experimenter will put this number in an envelope, seal it, and return it to you. When you go to collect your earnings, you should return the sealed envelope with your identification number still inside, the way it was handed to you before you left the room.

Along with these instructions, we will present you with a few examples. The numbers used are only for illustration purposes. The numbers you will encounter in the experiment could be different.

The basic decision

You will now learn how the experiment is conducted. First we will introduce the basic decision-making situation. Then we will ask you to answer control questions that will help you gain an understanding of the decision-making situation.

You will be a **member of a group of four people**. No one, except the experimenters, knows who belongs to what group. The groups are assembled randomly. At the beginning of the experiment, you will receive (on paper) **a number of tokens, called an "endowment."** Each of the four members of the group has to decide how to divide his or her endowment. You can put all, some, or none of your tokens into the project account. Each token you do not deposit in the project account will automatically be transferred to your **private account**.

<u>Your income from the private account:</u>

For each token you put into your private account, you will earn exactly one token. For example, if you have an endowment of 20 tokens and you put zero tokens into the project account (and therefore 20 tokens into the private account), then you will earn exactly 20 tokens **from the private account**. If instead you put 14 tokens into the project account (and therefore 6 tokens into the private account), then you will receive an income of 6 tokens from the private account. *Nobody except you earns tokens from your private account.*

Your income from the project account:

Everybody receives the same income from the project account, which is based on the total number of tokens the group puts into it. Your income from the project account will therefore be determined not only by the number of tokens you decide to put into the project account, but also by the number of tokens the other group members invest in it. For each group member, the income from the project account will be determined as follows:

Income from the project account = the sum of all contributions to the project account x 0.4

For example, if the sum of all contributions to the group account is 60 tokens, you and the other group members will earn 60x0.4=24 tokens from the project account. If the four group members deposit a total of 10 tokens into the project account, then you and the others will earn 10x0.4=4 tokens from the project account.

Your total income:

Your total income is the sum of the income from your private account and the income from the project account:

Income from your private account (=your endowment – your contribution to the project account) *+ Income from the project account (=0.4 x the sum of all contributions to the project account)* ______________________________

Before we finish reading the instructions, please answer the following control questions. This will help you make sure you have understood everything correctly. If you have any questions or problems, please raise your hand. A member of the experimenter team will attend to you and answer your question in private.

Control questions

Please answer the following control questions. Their purpose is to make you familiar with calculating the various incomes in tokens that you might earn depending on the decisions you will make about endowment allocation. Please answer all questions and write down all calculations.

1. Assume that you have an endowment of 20 tokens. Assume also that all group members (including yourself) put nothing into the project account.

   What is your total income? ______________

   What are the incomes of the three other group members?_____,____ and ____

2. Assume that you and the other team members each have an endowment of 20 tokens. the same as the other three group members. Assume also that all group members (including yourself) put their entire endowments into the project account.

   What is your total income? ______________

   What are the incomes of the three other group members?_____,____ and ____

3. Assume you have an endowment of 20 tokens. Assume also that the other group members collectively put a total of 30 tokens into the project account.

   a) What is your total income if you, in addition to the 30 tokens from the other three group members, put 0 tokens into the project account?

      i. Your total income is ___________.

   b. What is your total income if you, in addition to the 30 tokens from the other three group members, put 8 tokens into the project account?

      i. Your total income is___________.

   c. What is your total income if you, in addition to the 30 tokens from the other three group members, put 15 tokens into the project account?

      i. Your total income is___________.

4. Assume that you have an endowment of 20 tokens and that you put 8 tokens into the project account.

    a. What is your total income if the other three group members, in addition to your 8 tokens, put a total of 7 tokens into the project account?

        i. Your total income is __________.

    b. What is your total income if the other three group members, in addition to your 8 tokens, put a total of 12 tokens into the project account?

        i. Your total income is__________.

    c. What is your total income if the other three group members, in addition to your 8 tokens, put a total of 22 tokens into the project account?

        i. Your total income is__________.

If you finish these questions before the other participants, we advise you to think about additional examples to familiarize yourself further with these types of decision-making situations.

The Experimental Procedure

The experiment consists of decision-making situations similar to the one we just described. We will now explain the procedure in detail.

As you know, you have an endowment of 20 tokens. You can put these tokens into a project account. Any remaining tokens will automatically be deposited into your private account. Each person in the group will have the same endowment.

Each group member is asked to make two types of decisions. In the following instructions, we will refer to them as the **"unconditional contribution"** and the **"contribution table decision."**

- With the unconditional contribution, you decide how many tokens you want to put into the project account. Write this amount under *"Your unconditional contribution to the group account"* on the first page of your decision sheet. You must write down an **integer number that is neither smaller than zero nor larger than the total number of tokens you were given in your endowment (20).** The difference between your endowment of 20 tokens and the amount you put into the project account is automatically transferred to your private account.

- Your second task is to fill out the **contribution table** on page 3 of the decision sheet. In the contribution table, please indicate how many tokens you would like to put into the project account for each possible average contribution of the other three group members (rounded up or down to the nearest integer number; for example, if the average is 17.5, then write 18). What you actually contribute will depend on what the other group members actually contribute. This will become clear to you if you take a look at the following contribution table example:

(Rounded) Average contribution of the other group members to the project account.	Your contribution to the project account is:
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

The numbers in the left column are the possible (rounded) average contributions of the other three group members. Assume for this example that the other three group members can contribute a maximum of 20 tokens each ((20+20+20)/3=20).

Using the column on the right, simply write down how many tokens you would like to contribute to the project account for each possible average contribution of the others. You must make an entry in each field of the right column. For example, write down how many tokens you want to contribute to the group account if the others contribute an average of 0 tokens to the group account; how many you want to contribute if the others contribute an average of 0 tokens to the group account; how many you want to contribute if the others contribute an average 1 or 2 or 3 tokens, etc. In each field, you must write down an integer number that is neither smaller than zero nor larger than the total number of 20 tokens in your endowment. You can of course write down the same number in different fields.

After all participants have made their unconditional contribution decisions and have filled out their conditional contribution tables, one member of each group will be selected randomly. For the randomly selected group member, only the contribution table will be income relevant. For the three group members who are not selected, the unconditional contribution decision will be the income-relevant decision. When you make your unconditional contribution and when you fill out the contribution table, you do not know whether you will be selected randomly. You will therefore have to think carefully about both types of decisions since both could affect your earned amount. The following two examples should illustrate this:

Example 1. Assume that after you hand in your decisions, you are randomly selected. This implies that your income-relevant decision will be determined by your contribution table. For the other three group members, the unconditional contribution is the income-relevant decision. Assume they have made unconditional contributions of 0, 2, and 4 tokens. The rounded average contribution is therefore 2 ((0+2+4)/3=2).

If you have indicated in your contribution table that you will put 1 token into the project account if the others contribute 2 tokens on average, then the total contribution to the group account is 0+2+4+1=7. Thus, all group members earn an income of 0.4x7=2.8 from the project account plus the respective incomes from their private accounts.

If you have indicated instead that you will contribute 19 tokens to the project account if the others contribute 2 on average, then the total contribution to the project account is 0+2+4+19=25. All group members then earn an income of 0.4x25=10 tokens from the project account plus the respective incomes from their private accounts.

Example 2. Now assume that you are not selected randomly, which means that for you and two other group members, the unconditional contribution is the income-relevant decision. Assume further that your unconditional contribution to the project account is 16, and that those of the other two group members are 18 and 20. The average unconditional contribution is then 18 ((16+18+20)/3).

If the randomly selected group member indicated in the contribution table that he or she contributes 1 token to the group account when the other three group members contribute 18 on average, then the total contribution of the group to the group account is 16+18+20+1=55 tokens. All group members will therefore earn 0.4x55=22 tokens from the group account in addition to the respective incomes from their private accounts.

If the randomly selected group member instead indicated in the contribution table that he or she will contribute 19 tokens to the group account if the other three group members contribute 18 on average, then the total contribution of the group to the group account is 16+18+20+19=73 tokens. Each group member will therefore earn 0.4x73=29.2 tokens from the group account in addition to the income from his or her private accounts.

The random selection is arranged in the following manner. Every person in each group is assigned a number from 1 to 4. This number is found on the last page of your decision sheet. A participant will randomly pick one of four cards **after** all participants have made their unconditional contributions and have completed the contribution table and the questionnaire. If the card that is picked corresponds to the number on your decision sheet, then the contribution table on the third page becomes income-relevant for you. If not, then the unconditional contribution on the first page is your income-relevant decision. Remember that you do not know which of the two decisions will be relevant for your earnings until you have handed in all your decisions. You should therefore complete both pages carefully.

The amount of tokens you earn will be converted into pesos and then paid in cash. Do you have any questions? Please raise your hand and a member of the experimenter team will attend to you and answer your question in private.

D.2.: Guessed Contributions for the Public Good Game (Translated from Spanish)

Experimental ID number:______

How much do you think the other participants contributed unconditionally to the project account?

A moment ago, you wrote down your unconditional contribution and completed the conditional contribution table. Now, please tell us how much you think the other participants from your group contributed **unconditionally.**

Please write down the number of tokens **you believe the other three players** from your group have contributed on average to the project account. In other words, what number do you suspect they wrote down?

NOW you can earn more money if you **correctly guess the average contribution (rounded) of the other three players to the project account.** You can earn two additional tokens if the true average contribution of the others is equal to what you guessed, or if it is one token more or less. For example, if the true average contribution of the other three is 7 and you guess 6 or 8, you get two more tokens, while if you guess 9 you do not.

Average contribution (rounded) of the other three persons to the project account	Please mark an *x* in the box that corresponds to what you think the other three contributed unconditionally to the project account. ***ONLY MARK ONE BOX***
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

Chapter 5

Conditional Cooperation and Social Group

- *Experimental Results from Colombia*[1]

Peter Martinsson,[a] Clara Villegas-Palacio,[b,c] and Conny Wollbrant[d]

Abstract

In contrast to previous studies on cross-group comparisons of conditional cooperation, this study keeps cross- and within-country dimensions constant. The results reveal significantly different cooperation behavior between social groups in the same location.

Keywords: Conditional cooperation, experiment, public goods, social group

JEL Classification: C91, H41.

---

[1]The authors gratefully acknowledge the financial support from the Swedish Research Council (Vetenskapsrådet) and the Swedish International Development Cooperation Agency (Sida) to the Environmental Economics Unit at the University of Gothenburg. We are also grateful to the School of Engineering (Escuela de Ingeniería de Antioquia), Medellín, Colombia, and to Universidad Nacional de Colombia, Sede Medellín, for their support for our experiment. Special thanks go to Antonio Villegas-Rivera for excellent logistic support, and to Martin Kocher, Marta Matute, Katarina Nordblom, Patrik Söderholm, and Alba Upegui for their helpful comments.

[a] Corresponding author: Department of Economics, School of Business, Economics and Law, University of Gothenburg, P.O. Box 640, 405 30 Gothenburg, Sweden, (email) Peter.Martinsson@economics.gu.se, (ph) + 46 (0) 31 786 5255, (fax) + 46 (0) 31 786 1043.

[b]Department of Economics, School of Business, Economics and Law, University of Gothenburg, (email) Clara.Villegas@economics.gu.se.

[c]Facultad de Minas, Universidad Nacional de Colombia, Sede Medellín, Colombia, (email) civillegas@unal.edu.co.

[d] Department of Economics, School of Business, Economics and Law, University of Gothenburg, P.O. Box 640, 405 30 Gothenburg, Sweden, (email) Conny.Wollbrant@economics.gu.se.

1. Introduction

Voluntary contribution to public goods is frequently found both in the field and in the laboratory (e.g., see Gächter, 2007). Fischbacher et al. (2001) developed a one-shot public goods experiment in which subjects are asked for 1) an unconditional contribution to a public good, as in standard public goods experiments; and 2) a conditional contribution to the public good, given all possible average contributions (rounded to the nearest integer) of other group members. By investigating the profile of conditional contributions, subjects can be grouped into contributor types, such as free riders and conditional cooperators. (In other words, their degree of cooperation is conditional on their beliefs about others' cooperation.) Early evidence from experiments using the type classification following Fischbacher et al.'s approach used university students in Western countries as subjects (see, e.g., Gächter, 2006, for an overview). Generally, conditional cooperators are the dominating type (Fischbacher et al., 2001); however, most conditional cooperators are not perfect conditional contributors, but contribute slightly less than others. Kocher et al. (2008) replicated the experiment by Fischbacher et al. (2001) in three different countries and found differences in both the distribution of types and the share of conditional cooperation. Herrmann and Thöni (2009) conducted the same experiment in two rural and two urban locations in Russia and found that their fractions of conditional cooperators varied 48–60 percent within location, but that the differences between the locations were insignificant. The evidence from studies testing the effect of cultural background on behavior, using a standard multi-period public goods game, has been mixed as well (e.g., Brandts et al., 2004; Burlando and Hey, 1997; Herrmann et al., 2008).

When comparing experimental findings between locations, we identified three dimensions along which different locations may differ: 1) cross-country differences (e.g., religion and social norms), 2) within-country differences (e.g., rural versus urban areas), and 3) social group differences (e.g., age, trust, and income). Given these differences, it is not surprising that different locations yield different behavior. In this vein, Heinrich et al. (2005) found that those who see greater payoffs for cooperation in everyday life exhibit greater levels of prosociality in experimental games. La Ferrara (2002) found that relatively wealthy individuals are less likely to be a part of any group because benefits from cooperation do not outweigh the cost of membership for them. This opens up the question of whether preferences for cooperation vary across social groups.

The objective of the present paper is to investigate cooperative behavior in different social groups by keeping cross- and within-country differences constant. We used university students recruited from two universities in Medellin, Colombia, who differed in social-class: 1) socio-economic strata 2 and 3 (i.e., the “medium-low” group), and 2) socio-economic strata 4, 5, and 6 (i.e., the ”high” group).[2] We used the design of Fischbacher et al. (2001) to measure cooperative behavior in a public goods context.

2. Experimental Design and Procedure

We conducted a standard linear public goods experiment, following the same format as Fischbacher et al. (2001), where subject *i*'s payoff in tokens is given by:

$$\pi_i = 20 - c_i + 0.4\sum_{i=1}^{4} c_i \quad , \tag{1}$$

where 20 is the endowment and c the amount invested in the public good. Each group consisted of four randomly matched members. The marginal return from the public good was set to 0.4, ensuring a conflict between the dominant strategy to contribute zero, i.e., to free ride, and the full contribution Pareto optimum solution.

We asked our subjects to indicate how much they would like to contribute, both unconditionally and conditionally, to the public good. In the case of conditional contributions, subjects were asked how much they would like to contribute, conditional on the average contribution of the other members of the group, which included all integers numbers from 0 to 20 (i.e., the strategy method). To ensure incentive

[2]There are six social strata in Colombia: 1 (low-low), 2 (low), 3 (medium-low), 4 (medium), 5 (medium-high), and 6 (high). Strata 1–3 receive domestic public service subsidies, such as provision of water, electricity, and gas; 5–6 pay additional contributions toward the cost of public services. Stratum 4 receives no subsidies, but this group does not contribute either. The strata are indicators of people's socio-economic conditions.

compatibility for all decisions, the payoff relevant decision for three randomly selected members was the unconditional contribution. By using their average unconditional contribution, the contribution of the fourth member was given by his/her conditional contribution for that specific average contribution. Then, each member's monetary payoff could be calculated using equation (1). After the experiment, subjects were asked to guess the total contribution of the other three group members, and accuracy of guesses was monetarily rewarded.

The experiments were conducted at one socio-economic "medium-low" university (Universidad Nacional de Colombia) and one "high" university (Escuela de Ingeniería de Antioquia), both in Medellín, Colombia.[3] At both places, we ran two sessions with 24 subjects each; students of mathematics, psychology, and economics were excluded. The procedure of the experiment was the same at both places. Examples and individual exercises were used to ensure that subjects understood the experiment. Each session lasted approximately 90 minutes and the payoffs were calibrated to reflect opportunity costs. For the medium-low group, each token equaled COP 750, while the corresponding figure was COP 1,000 for the high group.[4] Average earnings were COP 25,000 for the high group and COP 23,000 for the medium-low group. (Both figures include a show-up fee of COP 5,000.)

3. Results

We followed the standard approach when defining the four contributor types (see Fischbacher et al., 2001). Conditional contributors submitted a contribution table showing

[3] At Universidad Nacional de Colombia (the medium-low group), approximately 80% of the student population belongs to strata 2 and 3, 11% to stratum 4, and only 5% to strata 5 and 6 (see Rico 2005). This is a public university where the cost of a six-month term is about the minimum monthly salary for students of stratum 3. At Escuela de Ingeniería de Antioquia, a private university, students mainly belong to strata 4, 5, or 6, and the cost is 10 times higher.

[4] In cases with samples with different opportunity costs, either the absolute amount in the experiment or the opportunity cost can be kept constant. We decided to keep the opportunity cost constant; it should be noted that Kocher et al. (2008) did not find a significant stake effect in one-shot public goods game. COP = Columbian Pesos; the exchange rate at the time of the experiment was US$ 1 = approximately COP 2,000. A lunch in the medium low–social class university costs approximately 75% of a lunch at the high social-class university.

a monotonically increasing own contribution for an increasing average contribution of the other members.[5] Free riders were characterized by a zero contribution for every possible average of the other members. Unconditional contributors submitted the same positive contribution independent of others' average contribution. Hump-shape contributors (also known as triangle contributors*)* showed monotonically increasing contributions up to a given average level of others' contributions, after which their contributions decreased. The category referred to as "Others" constituted the remaining participants.

Table 1 displays the distributions of types by social group. The dominating type is conditional cooperators, comprising 51 percent and 62 percent of the high group and the medium-low group, respectively. This is very close to the figures reported by, e.g., Fischbacher et al. (2001) and Fischbacher and Gächter (2006). Interestingly, 25 percent of the subjects in the high group were classified as free riders, compared to 4 percent in the medium-low group.

We rejected the null hypothesis of no differences in distribution of types between groups at the 5-percent significance level (p = 0.03; Chi2-test).[6] This is explained by a rejection of the hypothesis of no differences in share of free riders between the two groups at the 1 percent significance level (p = 0.004; Chi2-test). Table 1 also presents the average unconditional contribution for each type; the difference between the groups is statistically insignificant at conventional levels.

The relationship between the subjects' own conditional contribution and the average contribution of other group members is shown in figure 1. When the average contribution of others was zero, subjects in the medium-low group contributed more than those in the high group. Also, the difference in slope between the perfect conditional cooperation line and the plotted line, which represents degree of self-serving bias, was significantly larger in the high group. The regression results confirm the results shown in figure 1.

[5] We also included those without a monotonically increasing contribution, but with a highly significant (at 1%) positive Spearman rank correlation coefficient between own and others' contributions (see Fischbacher et al., 2001; Fischbacher and Gächter, 2006).
[6] This result is robust to systematic exclusion of types, e.g., excluding "others" (p = 0.026, Chi2-test).

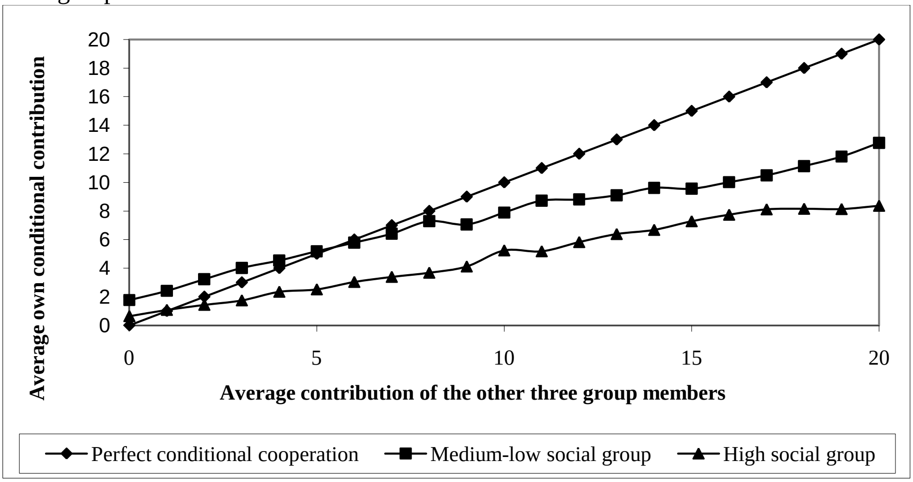
Table 1. Distribution of player types, average unconditional contribution, and guessed contribution.

	High socio-economic group			Medium-low socio-economic group		
	Distribution	Avg.uncond. contrib.	Avg. guessed contribution	Distri-bution	Avg. uncond. contrib.	Avg. guessed contribution
Unconditional cooperators	0.00%	0.00 (0.00)	0.00 (0.00)	4.17%	0.50 (0.71)	0.00 (0.00)
Conditional cooperators	54.17%	9.64 (4.68)	9.88 (4.78)	62.50%	9.33 (5.12)	9.50 (4.93)
Hump-shape contributors	8.33%	6.50 (7.85)	11.00 (7.53)	8.33%	8.75 (7.46)	8.25 (6.40)
Free-riders	25.00%	3.83 (7.02)	6.50 (7.43)	4.17%	0.50 (0.71)	2.00 (0.00)
Others	12.50%	8.00 (4.47)	7.67 (4.23)	20.83%	6.60 (3.95)	7.30 (4.16)

Note: Avg. uncond. contrib = average unconditional contributions; standard errors in parentheses.

Note: Avg. guessed. contrib = average guessed contributions; standard errors in parentheses.

Figure 1. Average own conditional contribution vs. average contribution of the other three group members.



Using two-sided Mann-Whitney U-tests, we found no significant difference in mean unconditional contribution between groups; it was 7.98 tokens in the medium-low group and 7.68 in the high group ($p = 0.75$). These levels of unconditional contributions, around 40 percent of the endowment, are in line with earlier findings (e.g., Kocher el at., 2008).

We elicited beliefs about others' contribution in the unconditional case, and found no significant differences in beliefs between the high group (8.83) and the medium-low group (8.23, where $p = 0.71$). Furthermore, regression results revealed that both groups can be classified as imperfect conditional cooperators (table 2). In addition, the high group displayed a significantly higher level of self-serving bias, which is similar to findings from the analysis of the conditional contribution tables.

Table 2. Regression results.

Dep. var: unconditional contribution in tokens	Tobit
	Coef.
Guessed contribution	0.948**
	(0.102)
Guessed contribution x High socio-economic group	-0.312*
	(0.140)
High socio-economic group	1.778
	(1.387)
Constant	0.181
	(0.984)
Sigma	4.079
	(0.341)
Number of observations	94
R-squared	0.58

Note: *** denotes significance at the 1% level, ** at the 5% significance level, * at 10% significance level. and t-statistics in parenthesis.

4. Conclusion

There is a growing interest in understanding whether behavior is the same across locations. By holding cross- and within-country dimensions constant, we investigated cooperative behavior between social groups in the same location. Our results suggest that different social groups exhibit differences both in terms of composition of types and extent of conditional cooperation.

As shown by Fischbacher and Gächter (2009), the decline in cooperation over time is caused by imperfect conditional cooperation. Thus, even if the unconditional contributions are similar across locations, the degree of imperfect conditional cooperation and the fraction of free riders are important factors determining the long-term differences in contributions to public goods. As a consequence, policymakers may need to consider different policy schemes. Following Gächter (2006), a social group where most individuals are conditional cooperators needs policies that sustain beliefs for cooperation of its integrants. In contrast, in situations where free riding dominates, policies involving monitoring and penalties may be required to enhance cooperation. Because a substantial part of public goods is local (e.g., teamwork and local environmental public goods governed by common property regimes such as lakes, pastures and irrigation systems), it is important to understand local preference heterogeneity.

References

Brandts, J., T. Saijo and A.J.H.C. Schram, 2004. A Four-Country Comparison of Spite and Cooperation in Public Goods Games, *Public Choice* 119, 381–424.

Burlando, R.M. and J.D. Hey, 1997. Do Anglo-Saxons Free Ride More?, *Journal of Public Economics* 64, 41-60.

Fischbacher, U.and S. Gächter, 2009. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments, *American Economic Review*, forthcoming.

Fischbacher, U., S. Gächter and E. Fehr, 2001. Are people conditionally cooperative?, Evidence from a public goods experiment, *Economic Letters* 71, 397-404.

Gächter, S., 2007. Conditional cooperation: Behavioural regularities from the lab and the field and their policy implications. In: Psychology and economics: A promising new cross-disciplinary field (Cesinfo seminar Series), Eds. B.S. Frey and A. Stuzter, Cambridge, MIT Press, 19-50.

Gächter, S. and B. Herrmann, 2006. The limits of self-governance in the presence of spite: Experimental evidence from urban and rural Russia, Working Paper 2006-13, University of Nottingham.

Heinrich, J., 2005. Economic man in cross-cultural perspective: behavioral experiments in 15 small-scale societies, *Behavioral and Brain Sciences* 28, 795-855.

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

Herrmann, B. and C. Thöni, 2009. Measuring conditional cooperation: A replication study in Russia, *Experimental Economics* 12, 87-92.

Kocher, M., P. Martinsson and M. Visser, 2008. Does stake size matter for cooperation and punishment?, *Economics Letters* 99, 508-511.

Kocher, M., T. Cherry, S. Kroll, R. Netzer and M. Sutter, 2008. Conditional cooperation on three continents, *Economics Letters* 101, 175-178.

La Ferrara, E., 2002. Inequality and group participation: theory and evidence from rural Tanzania, *Journal of Public Economics* 85, 235-273.

Rico, D., 2005. Evaluación del costo de las matrículas en la Universidad Nacional de Colombia, Sede Medellín, Oficina de Planeación. Universidad Nacional De Colombia.

Chapter 6

The role of beliefs, trust, and risk preferences in contributions to a public good*

Martin Kocher, University of Munich, Germany[A]

Peter Martinsson, University of Gothenburg, Sweden[B]

Dominik Matzat, University of Munich, Germany[C]

Conny Wollbrant, University of Gothenburg, Sweden[D]

Abstract

This paper experimentally investigates the role of beliefs, trust, and risk preferences in shaping cooperative behavior. By using a linear public goods game and the strategy method for revealing conditional contribution schedules, we categorize subjects into different types of contributors. Our results support the notion that beliefs about others behavior and trust are positively associated with cooperation while risk preferences do not seem to matter.

JEL Classification: *D01; D03; D64; D70.*

Keywords: *Experiment; Pro-social behavior; Public goods; Risk; Trust.*

---

* **Acknowledgments:** Financial support from the Swedish Research Council (*Vetenskapsrådet*) and the *Ideenfonds* of the University of Munich is gratefully acknowledged. The help of Nadja Furtner in conducting the experiments is gratefully acknowledged.

[A] Department of Economics, University of Munich, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany; Ph +49 89 2180 9726; Fax +49 89 2180 99 9726; E-mail martin.kocher@lrz.uni-muenchen.de.

[B] Department of Economics, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 52 55; Fax +46 31 786 10 43; E-mail peter.martinsson@economics.gu.se.

[C] Department of Economics, University of Munich, Geschwister-Scholl-Platz 1, D-80539 Munich, Germany; Ph +49 89 2180 9786; Fax +49 89 2180 99 9786; E-mail dominik.matzat@lrz.uni-muenchen.de.

[D] Department of Economics, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden; Ph +46 31 786 26 15; Fax +46 31 786 10 43; E-mail conny.wollbrant@economics.gu.se.

1. Introduction

Many situations in our daily lives possess the properties of a public good. Examples range from widely disparate domains from teamwork to recycling. Being non-rival and non-excludable, public goods are plagued by free-riding problems. Many people, however, do not free-ride in these situations although it is a dominant strategy for a rational and selfish individual, but we observe significant heterogeneity with regards to cooperative behavior across people. More precisely, recent evidence from public good experiments documents the prevalence of two main behavioral types when faced with a public good situation: free-riders and conditional cooperators. Conditional cooperators are subjects whose contributions are positively correlated with the (expected) contribution of others (e.g. Fischbacher *et al.*, 2001; Fischbacher and Gächter, 2010).[1] Conditional cooperation has also been investigated using e.g. lagged contributions by others to explain the contribution in current period (e.g. Keser and van Winden, 1998).

At present, little is known about what shapes cooperation and especially the connection between behavioral concepts such as trusting behavior and cooperation. The objective of this paper is to empirically investigate the links between beliefs about others' contributions, trust, and risk preferences on one hand and voluntary cooperation as well as contributor types in public goods game on the other hand. While each of these links has been studied separately and subsets of them in combination (e.g., Thöni et al., 2009, who investigated both self-reported trust and beliefs), we are the first to address these issues in a comprehensive manner using an incentivized experimental approach rather than relying on self-reported measures of cooperation, beliefs, trust, and risk.

[1] For field experiments investigating the relationship between own contributions and others' contributions, see e.g. Alpizar *et al.*, 2008, Frey and Meier, 2004, and Shang and Croson, 2006.

In our experiment, cooperative behavior is elicited by using a one-shot public goods experiment, which has the advantage of ruling out strategic motives. We apply a variant of the experimental design introduced by Fischbacher et al. (2001) to be able to classify cooperation types. Moreover, our subjects participate in a trust game similar to the design used by Berg et al. (1995), and a risk experiment using the same design as in Holt and Laury (2002) to elicit their attitudes towards risk.

The relationship between trust[2] and cooperation has been discussed in other disciplines for decades (e.g. Deutsch, 1958; Dawes, 1980), while more recently, it has been investigated by economists. Gächter *et al.* (2004) conducted a traditional public goods experiment and found positive and significant effect of self-reported trust questions related to beliefs about people's fairness, helpfulness and trust in strangers on contributions. However, they found no effect of a stated trust question related to trust of people in general.[3] Anderson *et al.* (2004) provided mixed evidence regarding the correlation between cooperation and self-reported trust in different domains. In an experiment among the Danish population, Thöni *et al.* (2009), however, found that self-reported trust explains cooperative behavior to a significant extent. The main innovation in our study compared to the previous ones is that we use an incentivized trust game to analyze the association between trusting behavior and cooperation. Given the debate regarding the predictive power of stated trust questionnaires, this is an essential robustness check.

---

[2] A related branch of literature has established the importance of trust especially for economic growth (e.g. Knack and Keefer, 1997; Keefer, 2002). While the literature includes many definitions of social capital (see e.g. the overview in Durlauf and Fafchamps, 2004), several emphasize trust as a key component (e.g. Bowles and Gintis, 2002; Putnam, 1993).

[3] A debated issue is if trust elicited in a trust experiment correlates with trust reported in surveys. Glaeser et al. (2000) compared results from trust experiments and stated trust and found poor correlations between the amounts sent in the trust experiment and stated trust. They concluded "that most work using these survey questions needs to be somewhat reinterpreted" (p. 814). On the other hand, Fehr et al. (2002) and Bellemare and Kröger (2003) find a positive relationship.

Another issue that may influence contributions to a public good is risk preferences. Individuals with higher risk aversion may choose to contribute less to the public good in order to compensate for the risk of others not contributing. Since individuals cannot be sure whether others contribute or not, contributing is a risky choice, although it involves social risk rather than natural risk (for an excellent experiment that demonstrates this difference, see Bohnet *et al.*, 2008). In line with this notion, Charness and Villeval (2009) find that subjects who invested more in a risky asset also contributed more to a public good. A similar result has been reported by Sabater-Grande and Georgantzis (2002) based on a multi-period prisoner's dilemma game. However, risk may also indirectly influence contributions as indicated by a few recent experiments that have focused on whether trust itself is determined by risk preferences.[4] The experimental results on the association between trust and risk are mixed. Whereas Schechter (2007) found a correlation of individual behavior in a trust game and a risk experiment in rural Paraguay, Bahry and Wilson (2005) and Eckel and Wilson (2004) did not find any relationship between risk attitudes and the amount sent in a trust game.

2. Experimental design

Our experimental design consists of three different parts conducted in the following order: (i) a one-shot linear public goods experiment with the strategy vector method as well as an elicitation of belief on others' contributions, (ii) a risk attitude elicitation experiment, and (iii) a trust experiment. The decisions in all experiments were monetarily rewarded, and it was clearly stated that the experiments were independent of each other.

[4] Others have focused on additional potential explanations of behavior in trust games and used experiments to tease out these effects, e.g. the impact of altruism. Cox (2004) with his triadic game design is an excellent example.

2.1. One-shot public goods game

We used the one-shot public goods experiment based on the strategy method as developed by Fischbacher *et al.* (2001). It builds on the following linear payoff function for subject *i*

$$\pi_i = 20 - c_i + 0.4\sum_{i=1}^{4} c_i \text{ ,} \quad (1)$$

where c_i denotes the contribution of subject *i* to the public good. Each group consists of four randomly matched subjects, and each subject receives an endowment of 20 tokens. In this part of the experiment each token was exchanged for 0.33 euro. The marginal per capita return (MPCR) from investing in the public good is 0.4. Assuming that participants are rational and selfish, it is obvious that any MPCR < 1 yields a dominant strategy for every group member to free-ride, i.e.,, to contribute nothing to the public good. From a social perspective, it is optimal to contribute the whole endowment because MPCR•*n* > 1. The details of the preference elicitation and the incentive mechanism in our experiment follow Fischbacher *et al.* (2001).

Subjects are asked to make two decisions: first an unconditional contribution to the public good, and thereafter a conditional contribution (a contribution schedule because we use the strategy method here). The unconditional contribution is a single integer number that satisfies $0 \leq c_i \leq 20$. For the conditional contributions, subjects have to indicate how much they would contribute to the public good for any possible average contribution of the three other players within their group (rounded to integers). For each of the 21 possible averages from 0 to 20, subjects must decide on a contribution between and including 0 and 20.

In order to ensure incentive compatibility, both the unconditional as well as the conditional contribution are potentially payoff relevant. For one group member, who is randomly determined by the throw of a four-sided dice, the conditional contribution is relevant, whereas the unconditional contributions are relevant for the other three group members. More specifically, the three unconditional contributions within a group and the corresponding conditional contribution (for the specific average of the three unconditional contributions) determine the sum of money contributed to the public good. Individual earnings can then be calculated according to equation (1).

Furthermore, subjects were asked to guess the average unconditional contribution of the other three group members (rounded to integers). The guessing stage is implemented after the conditional contribution stage. As in Gächter and Renner (2006), subjects were monetarily rewarded depending on the accuracy of their guesses. However, we use a slightly different and stronger incentive mechanism. If a subject's guess equals exactly the average unconditional contribution of the other three group members, the subject earns 9 tokens from the guess; if there is a difference of one between the guess and the average, 6 tokens are earned; and a difference of two still results in 3 tokens earned. Larger differences are neither rewarded nor punished.

2.2. Elicitation of risk attitudes

In the second part of the experiment, we used the design by Holt and Laury (2002) to measure individual risk attitudes. Each subject makes ten risky decisions. In each decision they choose between Option A or Option B, where both options include a lottery with the same probabilities but different payoffs. Option A is the relatively safer option because both possible lottery outcomes are between the outcomes of option B. Throughout the decisions, the payoffs are fixed, but the probability of receiving the higher payoff increases by 10 percentage points from 10% in decision 1 to 100% in decision 10 in both options. Depending on the subject's risk attitude, the subject should, moving down the decisions, switch at some point from Option A to Option B (or in the unlikely case of extreme risk-loving always choose Option B). Switching from B to A or choosing A always is incompatible with consistent behavior. The point at which subjects switch from Option A to Option B can then be used to calculate the degree of risk aversion. One of the ten lotteries was randomly selected and played for real. Subjects can earn up to 3.85 euros in this part.

2.3. The trust game

The trust experiment followed the classical design by Berg *et al*. (1995), but each subject plays both the role of sender and receiver (as for instance in Burks *et al.*, 2003). In the experiment the sender is given an endowment of 20 tokens, and he or she decides how much of the endowment in integers to send to the receiver. The amount sent by the sender is tripled before it reaches the receiver. The final stage of the game is when the receiver decides on how much to return to the sender (the returned amount is not tripled). A rational and selfish individual would send nothing to the receiver, because backward induction implies that the receiver has no incentive to send anything back. There is a possibility for a Pareto improvement, however, if the receiver returns at least one-third of the tripled amount received.

The amount sent by the sender is typically seen as an indication of trust, while the amount returned by the receiver is a measure of the level of trustworthiness. Since we wanted to obtain trust measures for all subjects, all of them had to make decisions in both roles without knowing which role they would finally be playing. In the role of receiver, we used the strategy method like in the public goods experiment above, i.e., subjects were asked to indicate how much they would send back for all the 21 possible amounts that they could receive. For monetary payoff, we randomly matched the subjects into pairs with roles of sender and receiver. The monetary payoff was then determined by their actions, i.e., the amount sent by the sender and the amount indicated to send back by the receiver conditional on the amount sent. Each token in the experiment was exchanged for 0.33 euro like in the public goods experiment.

2.4. Procedure and questionnaires

The computer-based experiments were conducted at the experimental laboratory MELESSA of the University of Munich in October 2009 and March 2010, using the experimental software z-Tree (Fischbacher, 2007) and the organizational software Orsee (Greiner, 2004). 144 undergraduate students from all disciplines except economics participated in 6 sessions with 24 subjects each. Sessions lasted up to 1½ hours, and the average payoff was 16.98 euro, including a show-up fee of 4 euro.

The experiment started with instructions for the public goods game. At that time, subjects received instructions only for the public good stage, but they knew that there would be two more parts in the experiment and that these further parts were unrelated. Subjects received written instructions, which were read aloud, and had the opportunity to ask questions in private. The experiment only began when all subjects correctly understood the procedures and after all subjects had passed through some computerized exercises, where they had to compute profits for different contribution levels in the public goods game. Upon completion, subjects received instructions for the second part, the risk attitude elicitation part, and finally for the trust part. We took care that matching of groups in the public goods game and the trust game was different, and this was clearly stated to the subjects. Decisions and results of the different parts were only revealed at the end of the entire experiment in order to avoid any effects from earnings in one part on behavior in subsequent parts. Before revelation of the results subjects had to fill in a short survey questionnaire eliciting a few personal characteristics (such as gender, age and academic major). Finally, subjects were paid privately and in cash and, then, were free to leave.

3. Results

In table 1, we present the descriptive results from our experiment. In the analyses below, we have excluded 12 subjects who did not answer consistently in the risk experiment (i.e., those who did not switch back from option B to A which is incompatible with consistent behavior). The average unconditional contribution is 6.83 tokens (34.2% of the endowment) and the corresponding guessed contribution by others is 7.32 tokens (36.6%). These levels correspond well to previous findings in German speaking countries (e.g. Fischbacher et al., 2001; Kocher et al., 2008). In the trust game, 7.59 tokens are on average sent by the sender, and the resulting level of 38.0% of endowment as transfers also corresponds to what has been previously found (e.g., Cardenas and Carpenter, 2008). In the risk experiment, Option A is chosen, on average 6.22 times. A risk-neutral subject would choose Option A four times, and thus our data indicates that subjects are on average risk averse. The results are very similar to the results in Holt and Laury (2002).

Using the design by Fischbacher et al. (2001), we can categorize subjects into different types of contributors based on the conditional contribution schedule. If a subject's own conditional contribution increases monotonically with the average contribution of the other members, the subject is classified as a c*onditional cooperator*. Moreover, a subject is also classified as conditional contributor if the relationship between own and others' average contributions is positive and significant at the 1% significance level based on the Spearman rank correlation coefficient (see Fischbacher et al., 2001; Fischbacher and Gächter, 2010). *Hump-shaped contributors* are subjects who show monotonically increasing contributions up to a given level of others' contributions or fulfilling a significant Spearman rank correlation coefficient at 1% significance level; above that level, their conditional contributions decrease monotonically. A *free rider* is a subject who has a conditional contribution of zero for all levels of the other members' contributions. Finally, those who cannot be categorized are referred to as *others*.

We find that 19.7% are classified as free-riders, 58.3% as conditional cooperators, 11.3% as hump-shaped and 10.6% as others, which again is very similar to the proportions reported in, e.g., Fischbacher et al. (2001) and Kocher et al. (2008). In the next four rows of table 1, we show descriptive statistics on the behavioral variables that we discussed for the whole sample above, but now separately for each type of contributors. As expected, the unconditional contribution differs significantly at the 1% level between the four types of contributors based on a Kruskal-Wallis test. Conditional cooperators on average contribute 8.18 tokens unconditionally, while free-riders only contribute 1.12 tokens. The average unconditional contributions for the hump-shaped and other contributors are 8.00 and 9.07 tokens, respectively. In our analysis, we focus on conditional cooperators and free-riders for two reasons. First, they exhibit clear and consistent patterns of behavior, and, second, they comprise the majority (78.0%) of types in our sample. Not surprisingly, the unconditional contribution differs significantly between free-riders and conditional cooperators according to a Mann-Whitney test ($p < 0.01$).

We find similar differences between the types when we investigate guessed contributions by others. The free-riders on average guessed that others would contribute 4.33 tokens compared to conditional contributors who guessed 7.88 tokens. We can reject the hypothesis of equality in guessed contributions both for all four types of contributors as well as for free-riders and conditional contributors at 0.1% significance levels.

Interestingly, the pattern between types is almost the same for the amount sent in the trust game. Free riders sent on average 2.58 tokens, compared to conditional cooperators who sent 9.06 tokens. Again, statistical tests reject equality both of all four types of contributors, and of free-riders and conditional contributors at 0.1% significance levels.

However, when it comes to risk preferences, there are neither statistical differences between the four types of contributors at the 5% significance level (p = 0.83), nor for the pairwise comparison of free-riders and conditional cooperators (p = 0.93).

Table 1. Descriptive statistics of the experiments (n=132).

Type of subject	*Proportion of subjects*	*Uncond-itional contribution*	*Guessed contribution on by others*	*Amount sent in trust game*	*Risk*
Free-rider	19.7%	1.12	4.31	2.58	6.27
Conditional cooperator	58.3%	8.18	7.88	9.06	6.38
Hump-shaped	11.3%	6.80	8.00	8.80	5.73
Others	10.6%	10.07	9.07	7.50	6.29
H0: No difference between types (Kruskal-Wallis test (p-value))		<0.01	<0.01	<0.01	0.83
H0: No difference between free-riders and conditional cooperators (Mann-Whitney test (p-value))		<0.01	<0.01	<0.01	0.93
All types	100%	6.83	7.32	7.59	6.22

Next we investigate what explains unconditional contributions. Previous research has tested the hypothesis that risk affects trust, and the results have been mixed. We cannot reject the hypothesis of no effect of risk on trust in a regression framework (p = 0.32) in table 2. As discussed in Thöni et al. (2009), there is a correlation between trust and the stated belief regarding others' contribution. We follow their approach by estimating models that include only beliefs or trust and models that include both. In model 1 in table 2, where we included the stated belief together with risk, only the belief is significant ($p < 0.01$). In model 2, we included trust instead of the belief, and only trust is significant in the regression ($p < 0.01$). In the third regression, where both belief and trust are included, we find that only the belief is significant at the 5% level ($p < 0.01$). Clearly, trust and the stated belief on others' contributions are associated.

Table 2. Estimation results from OLS model – Unconditional contributions (standard error in brackets).

	Model 1	**Model 2**	**Model 3**
	Coeff.	**Coeff.**	**Coeff.**
Belief about others' contribution	1.105**	-	1,056**
	(0.067)	-	(0.084)
Trust	-	0.349**	0.086
	-	(0.091)	(0.069)
Risk	0.094	0.178	0.120
	(0.225)	(0.335)	(0.225)
Constant	-1.839	3.073	-2.295
	(1.379)	(2.209)	(1.139)
Number of observations	132	132	132

Note: **p<0.01, * p<0.05. Robust standard errors. The results are similar if we use a tobit regression model.

In the following, we investigate the determinants of contributor types using a multinomial logit model. In the analyses, we merge the *hump-shaped* and *others* to one category denoted "others". We analyze the factors that influence the classification of free-riders, conditional cooperators and other type of contributors. The reference group is conditional cooperators, and thus the coefficients show how the different variables increase or decrease the probability of being classified as a free rider or as other compared to the case of conditional cooperators. We run three models since we included the belief and trust both separately as well as together. In all three models, trust and beliefs are significant at the 1% significance level when included. In other words, both lower levels of trust as well as lower belief in others' contributions explain free-riders. Again, risk does not significantly affect the probability of being classified into a certain type.

Table 3. Estimation results from multinomial logit model – contributor type (standard error in brackets)

	Model 1		**Model 2**		**Model 3**	
	Free-riders	**Others**	**Free-riders**	**Others**	**Free-riders**	**Others**
Belief about others' contribution	-0.282**	0.035	-	-	-0.215*	0.052
	(0.078)	(0.052)	-	-	(0.085)	(0.056)
Trust	-	-	-0.244**	-0.026	-0.203**	-0.038
	-	-	(0.060)	(0.035)	(0.061)	(0.038)
Risk	0.055	-0.121	-0.032	-0.132	0.059	-0.136
	(0.156)	(0.143)	(0.165)	(0.141)	(0.172)	(0.144)
Constant	0.234	-0.518	0.406	0.062	0.886	-0.246
	(1.023)	(1.012)	(1.090)	(0.963)	(1.121)	(1.046)
Number of observations	132		132		132	

Note: **p<0.01, * p<0.05. Robust standard errors. The contributor type other includes hump-shaped contributors. The reference group is conditional contributors.

4. Conclusions

By using a laboratory experiment, we have isolated how beliefs about others' contributions, trust, and risk preferences play a role in shaping contributions in a public goods experiment. According to Fischbacher *et al.* (2001), we classify subjects into contribution types. Previous findings document that conditional cooperation is a widespread type are supported by our experimental results. We further find that beliefs about others' contributions and trust elicited by a trust game are significantly associated with public good contributions, while risk preferences do neither affect contributions nor trust behavior in our experiment. Our findings regarding the correlation between trust and cooperation are similar to those in Thöni *et al.* (2009) despite the fact that we use an incentivized game, while Thöni et al. (2009) employ self-reported measures.

The fact that trust and cooperation are highly correlated is not surprising. It is intuitively clear that voluntary contribution to a public good involves a certain level of trust in the contribution of others. The association between trust and cooperation can be seen in actual behavior and/or in stated beliefs. Interestingly, free riders do not only contribute and trust less, but also have less optimistic expectations about other' contributions, in line with the false consensus effect. It is surprising that risk does not seem to play a role at all, neither in shaping trust, nor in explaining cooperation in our experiments. It seems that social risk is indeed something different to natural risk, as has already been indicated by Bohnet *et al.* (2008). Of course, we cannot exclude that our risk measure does not measure risk attitudes properly. What we know is that we have been using a widely accepted and often used method for eliciting risk preferences that has been validated a lot. The literature on trust and the literature on cooperation in economics, and specifically in experimental economics, have been distinct to a certain extent. Our results are one more piece of evidence showing that one should see them as much related concepts and that it would make sense to improve knowledge of the interactions between beliefs, trust, and cooperation.

For policy makers our results highlight the importance of high levels of trust as a prerequisite for achieving high degrees of voluntary cooperation. Thus, this indicates that trust building is an important alternative policy aimed to increase the number of conditional cooperators, who by their behavior will both contribute more to public goods as well as reducing the speed of decay to public goods over time. Hence, trust building is an alternative to previously tested institutions in public goods games involving the member in the group with the objective to increase contributions. Both monetary punishment (e.g. Bochet *et al.*, 2006; Fehr and Gächter, 2000; Ostrom *et al.*, 1992) and exclusion by voting (e.g. Cinyabuguma *et al.* (2005) have increased contributions substantially. In case of the monetary punishment, the overall effect on efficiency, i.e., when considering the negative effect on punishment, has shown to be negative in the short-run while in the long-run as degree of punishment decreases over time the effect is positive (Gächter et al., 2008). Trust building is also a costly activity. However, the effect of trust is more long-term compared to the sharp reduction to public goods when monetary punishment possibility is taken away (e.g. Fehr and Gächter, 2000), and similar findings from using non-monetary punishment in Masclet et al. (2003) where the effect from approval only had effect on contributions over initials periods. It is not difficult to predict that future research in economics will strengthen its focus on trust building and its institutional requirements, and to investigate the results of these activities on contributions to public goods.

References

Alpizar, F., Carlsson, F., and Johansson-Stenman., O. 2008. Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica. *Journal of Public Economics 92*, 1047-1060.

Anderson, L.R., Mellor, J.M., and Milyo, J. 2004. Social Capital and Contributions in a Public-Goods Experiment, *American Economic Review, Papers and Proceedings 94*, 373–376.

Bahry, D., and Wilson, R.K., 2005. Trust in Transitional Societies: Experimental results from Russia, paper presented at the American Political Science Association conference in 2004.

Bellemare, C. and Kröger, S., 2007. On Representative Social Capital, *European Economic Review 51*, 183-202.

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

Bochet, O., T. Page and L. Putterman (2006) Communication and Punishment in Voluntary Contribution Experiments, *Journal of Economic Behavior and Organization 60*, 11-26.

Bohnet, I., Greig, F., Herrmann, B., Zeckhauser, R. 2008. Betrayal Aversion, *American Economic Review 98*, 294-310.

Bowles, S. and H. Gintis, 2002. Social Capital and Community Governance, *Economic Journal 112*, 419-436.

Burks, S.V., Carpenter, J.P. and Verhoogen, E., 2003., Playing Both Roles in the Trust Game, *Journal of Economic Behavior and Organization 51*, 195-216.

Cardenas, J.C. and Carpenter, J., 2008. Behavioral Development Economics: Lessons from Field Labs in the Developing World. *Journal of Development Studies 44*, 337–64.

Charness, G. and Villeval, M.C., 2009. Cooperation and Competition in Intergenerational Experiments in the Field and the Laboratory, *American Economic Review 99*, 956-978.

Cinyabuguma, M., T. Page and L. Putterman (2005) Cooperation under the Threat of Expulsion in a Public Goods Experiment, *Journal of Public Economics 89,* 1421-1435.

Cox, J., 2004. How to Identify Trust and Reciprocity, *Games and Economic Behavior 46*, 260-281.

Dawes, R., 1980. Social Dilemmas. *Annual Review of Psychology 31*, 169–193.

Deutsch, M., 1958. Trust and Suspicion. *Journal of Conflict Resolution 2*, 265–279.

Durlauf, S. and Fafchamps, M., 2004., Social Capital. In *Handbook of Economic Growth*, S. Durlauf and Aghion, P. (eds.).

Eckel, C. C. and Wilson, R.K., 2004. Is Trust a Risky Decision? *Journal of Economic Behavior and Organization 55*, 447–465.

Fehr, E. and S. Gächter (2000) Cooperation and Punishment in Public Goods Experiments, *American Economic Review 90*, 980-994.

Fehr, E., Fischbacher, U., Rosenbladt, B., Schupp, J., and Wagner, G.G., 2002. A Nation-wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys, *Schmollers Jahrbuch 122*, 519-542.

Fischbacher, U. and Gächter., S. 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments. *American Economic Review 100*, 541-556

Fischbacher, U., 2007, z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics 10*, 171-178.

Fischbacher, U., S. Gächter and E. Fehr, E. 2001. Are People Conditionally Cooperative? Evidence from a Public Goods Experiment, *Economic Letters 71*, 397-404.

Frey, B. and Meier, S. 2004. Social Comparison and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment, *American Economic Review 94*, 1717-1722.

Gächter, S., Herrmann, B., and Thöni, C. 2004. Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence, *Journal of Economic Behavior and Organization 55*, 505–531.

Gächter, S., and Renner, E., 2006. The Effects of (Incentivized) Belief Elicitation in Public Good Experiments. *Discussion paper 2006-16*. Nottingham University.

Gächter, S., Renner, E., and Sefton, M. 2008. The Long-Run Benefits of Punishment. *Science 322*, 1510.

Glaeser, E. L., Laibson, D. I., Scheinkman, J.A., Soutter, C.L., 2000. Measuring Trust, *Quarterly Journal of Economics 115*, 811-846.

Greiner, B., 2004. An Online Recruitment System for Economic Experiments. In Kurt Kremer, V. M. (ed.) Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, Göttingen 79-93.

Holt, C. A., Laury, S. K., 2002. Risk Aversion and Incentive Effects. *American Economic Review 92*, 1644-1655.

Keser, C. and van Winden, F. 2000. Conditional Cooperation and Voluntary Contributions to Public Goods, *Scandinavian Journal of Economics 102*, 23-39.

Knack, S., 2002. Social Capital and the Quality of Government: Evidence from the United States, *American Journal of Political Science 46*, 772–785.

Knack, S., Keefer, P., 1997. Does Social Capital have an Economic Payoff? A Cross-Country Investigation, *Quarterly Journal of Economics 112*, 1251-1288.

Kocher, M., T. Cherry, T., Kroll, S., Netzer, R., and Sutter, M. 2008. Conditional Cooperation on three Continents, *Economics Letters 101*, 175-178.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. 2003. Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism, *American Economic Review 93*, 366-380.

Ostrom, E., Walker, J., and Gardner, R. 1992. Covenants With and Without a Sword: Self-Governance is Possible, *American Political Science Journal 86*, 404-417.

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

Sabater-Grande, G. and Georgantzis, N., 2002. Accounting for Risk Aversion in Repeated Prisoners' Dilemma Games: An Experimental Test, *Journal of Economic Behavior and Organization 48,* 37-50.

Schechter, L., 2007. Traditional Trust Measurement and the Risk of Confound: An Experiment in Rural Paraguay, *Journal of Economic Behavior and Organization 62*, 72-92

Shang, J. and R. Croson, 2006. The Impact of Social Comparisons on Nonprofit Fundraising, *Research in Experimental Economics 11*, 143-156.

Thöni, C., Tyran, J-R., and Wengström, E. 2009. Microfoundations of Social Capital, *Disucussion Paper 09-24,* Department of Economics, University of Copenhagen.

Previous doctoral theses in the Department of Economics, Gothenburg

Avhandlingar publicerade innan serien Ekonomiska Studier startades
(Theses published before the series Ekonomiska Studier was started):

Östman, Hugo (1911), Norrlands ekonomiska utveckling
Moritz, Marcus (1911), Den svenska tobaksindustrien
Sundbom, I. (1933), Prisbildning och ändamålsenlighet
Gerhard, I. (1948), Problem rörande Sveriges utrikeshandel 1936/38
Hegeland, Hugo (1951), The Quantity Theory of Money
Mattsson, Bengt (1970), Cost-Benefit analys
Rosengren, Björn (1975), Valutareglering och nationell ekonomisk politik
Hjalmarsson, Lennart (1975), Studies in a Dynamic Theory of Production and its Applications
Örtendahl, Per-Anders (1975), Substitutionsaspekter på produktionsprocessen vid massaframställning
Anderson, Arne M. (1976), Produktion, kapacitet och kostnader vid ett helautomatiskt emballageglasbruk
Ohlsson, Olle (1976), Substitution och odelbarheter i produktionsprocessen vid massaframställning
Gunnarsson, Jan (1976), Produktionssystem och tätortshierarki – om sambandet mellan rumslig och ekonomisk struktur
Köstner, Evert (1976), Optimal allokering av tid mellan utbildning och arbete
Wigren, Rune (1976), Analys av regionala effektivitetsskillnader inom industribranscher
Wästlund, Jan (1976), Skattning och analys av regionala effektivitetsskillnader inom industribranscher
Flöjstad, Gunnar (1976), Studies in Distortions, Trade and Allocation Problems
Sandelin, Bo (1977), Prisutveckling och kapitalvinster på bostadsfastigheter
Dahlberg, Lars (1977), Empirical Studies in Public Planning
Lönnroth, Johan (1977), Marxism som matematisk ekonomi
Johansson, Börje (1978), Contributions to Sequential Analysis of Oligopolistic Competition

<u>Ekonomiska Studier</u>, utgivna av Nationalekonomiska institutionen vid Göteborgs Universitet. Nr 1 och 4 var inte doktorsavhandlingar. (The contributions to the department series 'Ekonomiska Studier' where no. 1 and 4 were no doctoral theses):

2. **Ambjörn, Erik** (1959), Svenskt importberoende 1926-1956: en ekonomisk-statistisk kartläggning med kommentarer
3. **Landgren, K-G.** (1960), Den "Nya ekonomien" i Sverige: J.M. Keynes, E. Wigfors och utecklingen 1927-39
5. **Bigsten, Arne** (1979), Regional Inequality and Development: A Case Study of Kenya
6. **Andersson, Lars** (1979), Statens styrning av de kommunala budgetarnas struktur (Central Government Influence on the Structure of the Municipal Budget)
7. **Gustafsson, Björn** (1979), Inkomst- och uppväxtförhållanden (Income and Family Background)

8. **Granholm, Arne** (1981), Interregional Planning Models for the Allocation of Private and Public Investments
9. **Lundborg, Per** (1982), Trade Policy and Development: Income Distributional Effects in the Less Developed Countries of the US and EEC Policies for Agricultural Commodities
10. **Juås, Birgitta** (1982), Värdering av risken för personskador. En jämförande studie av implicita och explicita värden. (Valuation of Personal Injuries. A comparison of Explicit and Implicit Values)
11. **Bergendahl, Per-Anders** (1982), Energi och ekonomi - tillämpningar av input-output analys (Energy and the Economy - Applications of Input-Output Analysis)
12. **Blomström, Magnus** (1983), Foreign Investment, Technical Efficiency and Structural Change - Evidence from the Mexican Manufacturing Industry
13. **Larsson, Lars-Göran** (1983), Comparative Statics on the Basis of Optimization Methods
14. **Persson, Håkan** (1983), Theory and Applications of Multisectoral Growth Models
15. **Sterner, Thomas** (1986), Energy Use in Mexican Industry.
16. **Flood, Lennart** (1986), On the Application of Time Use and Expenditure Allocation Models.
17. **Schuller, Bernd-Joachim** (1986), Ekonomi och kriminalitet - en empirisk undersökning av brottsligheten i Sverige (Economics of crime - an empirical analysis of crime in Sweden)
18. **Walfridson, Bo** (1987), Dynamic Models of Factor Demand. An Application to Swedish Industry.
19. **Stålhammar, Nils-Olov** (1987), Strukturomvandling, företagsbeteende och förväntningsbildning inom den svenska tillverkningsindustrin (Structural Change, Firm Behaviour and Expectation Formation in Swedish Manufactury)
20. **Anxo, Dominique** (1988), Sysselsättningseffekter av en allmän arbetstidsför-kortning (Employment effects of a general shortage of the working time)
21. **Mbelle, Ammon** (1988), Foreign Exchange and Industrial Development: A Study of Tanzania.
22. **Ongaro, Wilfred** (1988), Adoption of New Farming Technology: A Case Study of Maize Production in Western Kenya.
23. **Zejan, Mario** (1988), Studies in the Behavior of Swedish Multinationals.
24. **Görling, Anders** (1988), Ekonomisk tillväxt och miljö. Förorenings-struktur och ekonomiska effekter av olika miljövårdsprogram. (Economic Growth and Environment. Pollution Structure and Economic Effects of Some Environmental Programs).
25. **Aguilar, Renato** (1988), Efficiency in Production: Theory and an Application on Kenyan Smallholders.
26. **Kayizzi-Mugerwa, Steve** (1988), External Shocks and Adjustment in Zambia.

27. **Bornmalm-Jardelöw, Gunilla** (1988), Högre utbildning och arbetsmarknad (Higher Education and the Labour Market)
28. **Tansini, Ruben** (1989), Technology Transfer: Dairy Industries in Sweden and Uruguay.
29. **Andersson, Irene** (1989), Familjebeskattning, konsumtion och arbetsutbud - En ekonometrisk analys av löne- och inkomstelasticiteter samt policysimuleringar för svenska hushåll (Family Taxation, Consumption and Labour Supply - An Econometric Analysis of Wage and Income Elasticities and Policy Simulations for Swedish Households)
30. **Henrekson, Magnus** (1990), An Economic Analysis of Swedish Government Expenditure
31. **Sjöö, Boo** (1990), Monetary Policy in a Continuous Time Dynamic Model for Sweden
32. **Rosén, Åsa** (1991), Contributions to the Theory of Labour Contracts.
33. **Loureiro, Joao M. de Matos** (1992), Foreign Exchange Intervention, Sterilization and Credibility in the EMS: An Empirical Study
34. **Irandoust, Manuchehr** (1993), Essays on the Behavior and Performance of the Car Industry
35. **Tasiran, Ali Cevat** (1993), Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States
36. **Milopoulos, Christos** (1993), Investment Behaviour under Uncertainty: An Econometric Analysis of Swedish Panel Data
37. **Andersson, Per-Åke** (1993), Labour Market Structure in a Controlled Economy: The Case of Zambia
38. **Storrie, Donald W.** (1993), The Anatomy of a Large Swedish Plant Closure
39. **Semboja, Haji Hatibu Haji** (1993), Energy and Development in Kenya
40. **Makonnen, Negatu** (1993), Labor Supply and the Distribution of Economic Well-Being: A Case Study of Lesotho
41. **Julin, Eva** (1993), Structural Change in Rural Kenya
42. **Durevall, Dick** (1993), Essays on Chronic Inflation: The Brazilian Experience
43. **Veiderpass, Ann** (1993), Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change
44. **Odeck, James** (1993), Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector
45. **Mwenda, Abraham** (1993), Credit Rationing and Investment Behaviour under Market Imperfections: Evidence from Commercial Agriculture in Zambia
46. **Mlambo, Kupukile** (1993), Total Factor Productivity Growth: An Empirical Analysis of Zimbabwe's Manufacturing Sector Based on Factor Demand Modelling
47. **Ndung'u, Njuguna** (1993), Dynamics of the Inflationary Process in Kenya
48. **Modén, Karl-Markus** (1993), Tax Incentives of Corporate Mergers and Foreign Direct Investments
49. **Franzén, Mikael** (1994), Gasoline Demand - A Comparison of Models
50. **Heshmati, Almas** (1994), Estimating Technical Efficiency, Productivity Growth And Selectivity Bias Using Rotating Panel Data: An Application to Swedish Agriculture

51. **Salas, Osvaldo** (1994), Efficiency and Productivity Change: A Micro Data Case Study of the Colombian Cement Industry
52. **Bjurek, Hans** (1994), Essays on Efficiency and Productivity Change with Applications to Public Service Production
53. **Cabezas Vega, Luis** (1994), Factor Substitution, Capacity Utilization and Total Factor Productivity Growth in the Peruvian Manufacturing Industry
54. **Katz, Katarina** (1994), Gender Differentiation and Discrimination. A Study of Soviet Wages
55. **Asal, Maher** (1995), Real Exchange Rate Determination and the Adjustment Process: An Empirical Study in the Cases of Sweden and Egypt
56. **Kjulin, Urban** (1995), Economic Perspectives on Child Care
57. **Andersson, Göran** (1995), Volatility Forecasting and Efficiency of the Swedish Call Options Market
58. **Forteza, Alvaro** (1996), Credibility, Inflation and Incentive Distortions in the Welfare State
59. **Locking, Håkan** (1996), Essays on Swedish Wage Formation
60. **Välilä, Timo** (1996), Essays on the Credibility of Central Bank Independence
61. **Yilma, Mulugeta** (1996), Measuring Smallholder Efficiency: Ugandan Coffee and Food-Crop Production
62. **Mabugu, Ramos E.** (1996), Tax Policy Analysis in Zimbabwe Applying General Equilibrium Models
63. **Johansson, Olof** (1996), Welfare, Externalities, and Taxation; Theory and Some Road Transport Applications.
64. **Chitiga, Margaret** (1996), Computable General Equilibrium Analysis of Income Distribution Policies in Zimbabwe
65. **Leander, Per** (1996), Foreign Exchange Market Behavior Expectations and Chaos
66. **Hansen, Jörgen** (1997), Essays on Earnings and Labor Supply
67. **Cotfas, Mihai** (1997), Essays on Productivity and Efficiency in the Romanian Cement Industry
68. **Horgby, Per-Johan** (1997), Essays on Sharing, Management and Evaluation of Health Risks
69. **Nafar, Nosratollah** (1997), Efficiency and Productivity in Iranian Manufacturing Industries
70. **Zheng, Jinghai** (1997), Essays on Industrial Structure, Technical Change, Employment Adjustment, and Technical Efficiency
71. **Isaksson, Anders** (1997), Essays on Financial Liberalisation in Developing Countries: Capital mobility, price stability, and savings
72. **Gerdin, Anders** (1997), On Productivity and Growth in Kenya, 1964-94
73. **Sharifi, Alimorad** (1998), The Electricity Supply Industry in Iran: Organization, performance and future development
74. **Zamanian, Max** (1997), Methods for Mutual Fund Portfolio Evaluation: An application to the Swedish market
75. **Manda, Damiano Kulundu** (1997), Labour Supply, Returns to Education, and the Effect of Firm Size on Wages: The case of Kenya
76. **Holmén, Martin** (1998), Essays on Corporate Acquisitions and Stock Market Introductions

77. **Pan, Kelvin** (1998), Essays on Enforcement in Money and Banking
78. **Rogat, Jorge** (1998), The Value of Improved Air Quality in Santiago de Chile
79. **Peterson, Stefan** (1998), Essays on Large Shareholders and Corporate Control
80. **Belhaj, Mohammed** (1998), Energy, Transportation and Urban Environment in Africa: The Case of Rabat-Salé, Morocco
81. **Mekonnen, Alemu** (1998), Rural Energy and Afforestation: Case Studies from Ethiopia
82. **Johansson, Anders** (1998), Empirical Essays on Financial and Real Investment Behavior
83. **Köhlin, Gunnar** (1998), The Value of Social Forestry in Orissa, India
84. **Levin, Jörgen** (1998), Structural Adjustment and Poverty: The Case of Kenya
85. **Ncube, Mkhululi** (1998), Analysis of Employment Behaviour in Zimbabwe
86. **Mwansa, Ladslous** (1998), Determinants of Inflation in Zambia
87. **Agnarsson, Sveinn** (1998), Of Men and Machines: Essays in Applied Labour and Production Economics
88. **Kadenge, Phineas** (1998), Essays on Macroeconomic Adjustment in Zimbabwe: Inflation, Money Demand, and the Real Exchange Rate
89. **Nyman, Håkan** (1998), An Economic Analysis of Lone Motherhood in Sweden
90. **Carlsson, Fredrik** (1999), Essays on Externalities and Transport
91. **Johansson, Mats** (1999), Empirical Studies of Income Distribution
92. **Alemu, Tekie** (1999), Land Tenure and Soil Conservation: Evidence from Ethiopia
93. **Lundvall, Karl** (1999), Essays on Manufacturing Production in a Developing Economy: Kenya 1992-94
94. **Zhang, Jianhua** (1999), Essays on Emerging Market Finance
95. **Mlima, Aziz Ponary** (1999), Four Essays on Efficiency and Productivity in Swedish Banking
96. **Davidsen, Björn-Ivar** (2000), Bidrag til den økonomisk-metodologiske tenkningen (Contributions to the Economic Methodological Thinking)
97. **Ericson, Peter** (2000), Essays on Labor Supply
98. **Söderbom, Måns** (2000), Investment in African Manufacturing: A Microeconomic Analysis
99. **Höglund, Lena** (2000), Essays on Environmental Regulation with Applications to Sweden
100. **Olsson, Ola** (2000), Perspectives on Knowledge and Growth
101. **Meuller, Lars** (2000), Essays on Money and Credit
102. **Österberg, Torun** (2000), Economic Perspectives on Immigrants and Intergenerational Transmissions
103. **Kalinda Mkenda, Beatrice** (2001), Essays on Purchasing Power Parity, RealExchange Rate, and Optimum Currency Areas
104. **Nerhagen, Lena** (2001), Travel Demand and Value of Time - Towards an Understanding of Individuals Choice Behavior

105. **Mkenda, Adolf** (2001), Fishery Resources and Welfare in Rural Zanzibar
106. **Eggert, Håkan** (2001), Essays on Fisheries Economics
107. **Andrén, Daniela** (2001), Work, Sickness, Earnings, and Early Exits from the Labor Market. An Empirical Analysis Using Swedish Longitudinal Data
108. **Nivorozhkin, Eugene** (2001), Essays on Capital Structure
109. **Hammar, Henrik** (2001), Essays on Policy Instruments: Applications to Smoking and the Environment
110. **Nannyonjo, Justine** (2002), Financial Sector Reforms in Uganda (1990-2000): Interest Rate Spreads, Market Structure, Bank Performance and Monetary Policy
111. **Wu, Hong** (2002), Essays on Insurance Economics
112. **Linde-Rahr, Martin** (2002), Household Economics of Agriculture and Forestry in Rural Vienam
113. **Maneschiöld, Per-Ola** (2002), Essays on Exchange Rates and Central Bank Credibility
114. **Andrén, Thomas** (2002), Essays on Training, Welfare and Labor Supply
115. **Granér, Mats** (2002), Essays on Trade and Productivity: Case Studies of Manufacturing in Chile and Kenya
116. **Jaldell, Henrik** (2002), Essays on the Performance of Fire and Rescue Services
117. **Alpizar, Francisco, R.** (2002), Essays on Environmental Policy-Making in Developing Countries: Applications to Costa Rica
118. **Wahlberg, Roger** (2002), Essays on Discrimination, Welfare and Labor Supply
119. **Piculescu, Violeta** (2002), Studies on the Post-Communist Transition
120. **Pylkkänen, Elina** (2003), Studies on Household Labor Supply and Home Production
121. **Löfgren, Åsa** (2003), Environmental Taxation – Empirical and Theoretical Applications
122. **Ivaschenko, Oleksiy** (2003), Essays on Poverty, Income Inequality and Health in Transition Economies
123. **Lundström, Susanna** (2003), On Institutions, Economic Growth and the Environment
124. **Wambugu, Anthony** (2003), Essays on Earnings and Human Capital in Kenya
125. **Adler, Johan** (2003), Aspects of Macroeconomic Saving
126. **Erlandsson, Mattias** (2003), On Monetary Integration and Macroeconomic Policy
127. **Brink, Anna** (2003), On the Political Economy of Municipality Break-Ups
128. **Ljungwall, Christer** (2003), Essays on China's Economic Performance During the Reform Period
129. **Chifamba, Ronald** (2003), Analysis of Mining Investments in Zimbabwe
130. **Muchapondwa, Edwin** (2003), The Economics of Community-Based Wildlife Conservation in Zimbabwe
131. **Hammes, Klaus** (2003), Essays on Capital Structure and Trade Financing
132. **Abou-Ali, Hala** (2003), Water and Health in Egypt: An Empirical Analysis
133. **Simatele, Munacinga** (2004), Financial Sector Reforms and Monetary Policy in Zambia
134. **Tezic, Kerem** (2004), Essays on Immigrants' Economic Integration
135. INSTÄLLD

136. **Gjirja, Matilda** (2004), Efficiency and Productivity in Swedish Banking
137. **Andersson, Jessica** (2004), Welfare Environment and Tourism in Developing Countries
138. **Chen, Yinghong** (2004), Essays on Voting Power, Corporate Governance and Capital Structure
139. **Yesuf, Mahmud** (2004), Risk, Time and Land Management under Market Imperfections: Applications to Ethiopia
140. **Kateregga, Eseza** (2005), Essays on the Infestation of Lake Victoria by the Water Hyacinth
141. **Edvardsen, Dag Fjeld** (2004), Four Essays on the Measurement of Productive Efficiency
142. **Lidén, Erik** (2005), Essays on Information and Conflicts of Interest in Stock Recommendations
143. **Dieden, Sten** (2005), Income Generation in the African and Coloured Population – Three Essays on the Origins of Household Incomes in South Africa
144. **Eliasson, Marcus** (2005), Individual and Family Consequences of Involuntary Job Loss
145. **Mahmud, Minhaj** (2005), Measuring Trust and the Value of Statistical Lives: Evidence from Bangladesh
146. **Lokina, Razack Bakari** (2005), Efficiency, Risk and Regulation Compliance: Applications to Lake Victoria Fisheries in Tanzania
147. **Jussila Hammes, Johanna** (2005), Essays on the Political Economy of Land Use Change
148. **Nyangena, Wilfred** (2006), Essays on Soil Conservation, Social Capital and Technology Adoption
149. **Nivorozhkin, Anton** (2006), Essays on Unemployment Duration and Programme Evaluation
150. **Sandén, Klas** (2006), Essays on the Skill Premium
151. **Deng, Daniel** (2006), Three Essays on Electricity Spot and Financial Derivative Prices at the Nordic Power Exchange
152. **Gebreeyesus, Mulu** (2006), Essays on Firm Turnover, Growth, and Investment Behavior in Ethiopian Manufacturing
153. **Islam, Nizamul Md.** (2006), Essays on Labor Supply and Poverty: A Microeconometric Application
154. **Kjaer, Mats** (2006), Pricing of Some Path-Dependent Options on Equities and Commodities
155. **Shimeles, Abebe** (2006), Essays on Poverty, Risk and Consumption Dynamics in Ethiopia
156. **Larsson, Jan** (2006), Four Essays on Technology, Productivity and Environment
157. **Congdon Fors, Heather** (2006), Essays in Institutional and Development Economics
158. **Akpalu, Wisdom** (2006), Essays on Economics of Natural Resource Management and Experiments
159. **Daruvala, Dinky** (2006), Experimental Studies on Risk, Inequality and Relative Standing
160. **García, Jorge** (2007), Essays on Asymmetric Information and Environmental Regulation through Disclosure

161. **Bezabih, Mintewab** (2007), Essays on Land Lease Markets, Productivity, Biodiversity, and Environmental Variability
162. **Visser, Martine** (2007), Fairness, Reciprocity and Inequality: Experimental Evidence from South Africa
163. **Holm, Louise** (2007), A Non-Stationary Perspective on the European and Swedish Business Cycle
164. **Herbertsson, Alexander** (2007), Pricing Portfolio Credit Derivatives
165. **Johansson, Anders C.** (2007), Essays in Empirical Finance: Volatility, Interdependencies, and Risk in Emerging Markets
166. **Ibáñez Díaz, Marcela** (2007), Social Dilemmas: The Role of Incentives, Norms and Institutions
167. **Ekbom, Anders** (2007), Economic Analysis of Soil Capital, Land Use and Agricultural Production in Kenya
168. **Sjöberg, Pål** (2007), Essays on Performance and Growth in Swedish Banking
169. **Palma Aguirre, Grisha Alexis** (2008), Explaining Earnings and Income Inequality in Chile
170. **Akay, Alpaslan** (2008), Essays on Microeconometrics and Immigrant Assimilation
171. **Carlsson, Evert** (2008), After Work – Investing for Retirement
172. **Munshi, Farzana** (2008), Essays on Globalization and Occupational Wages
173. **Tsakas, Elias** (2008), Essays on Epistemology and Evolutionary Game Theory
174. **Erlandzon, Karl** (2008), Retirement Planning: Portfolio Choice for Long-Term Investors
175. **Lampi, Elina** (2008), Individual Preferences, Choices, and Risk Perceptions – Survey Based Evidence
176. **Mitrut, Andreea** (2008), Four Essays on Interhousehold Transfers and Institutions in Post-Communist Romania
177. **Hansson, Gustav** (2008), Essays on Social Distance, Institutions, and Economic Growth
178. **Zikhali, Precious** (2008), Land Reform, Trust and Natural Resource Management in Africa
179. **Tengstam, Sven** (2008), Essays on Smallholder Diversification, Industry Location, Debt Relief, and Disability and Utility
180. **Boman, Anders** (2009), Geographic Labour Mobility – Causes and Consequences
181. **Qin, Ping** (2009), Risk, Relative Standing and Property Rights: Rural Household Decision-Making in China
182. **Wei, Jiegen** (2009), Essays in Climate Change and Forest Management
183. **Belu, Constantin** (2009), Essays on Efficiency Measurement and Corporate Social Responsibility
184. **Ahlerup, Pelle** (2009), Essays on Conflict, Institutions, and Ethnic Diversity
185. **Quiroga, Miguel** (2009), Microeconomic Policy for Development: Essays on Trade and Environment, Poverty and Education
186. **Zerfu, Daniel** (2010), Essays on Institutions and Economic Outcomes
187. **Wollbrant, Conny** (2010), Self-Control and Altruism